



HHS Public Access

Author manuscript

J Thorac Oncol. Author manuscript; available in PMC 2016 December 01.

Published in final edited form as:

J Thorac Oncol. 2015 December ; 10(12): 1694–1702. doi:10.1097/JTO.0000000000000691.

A Prognostic Model to Predict Mortality among Non-Small Cell Lung Cancer Patients in the U.S. Military Health System

Jie Lin, PhD, MPH¹, Corey A. Carter, MD¹, Katherine A. McGlynn, PhD², Shelia H. Zahm, ScD², Joel A. Nations, MD¹, William F. Anderson, MD, MPH², Craig D. Shriver, MD^{1,3}, and Kangmin Zhu, MD, PhD^{1,3}

¹John P. Murtha Cancer Center, Walter Reed National Military Medical Center, Bethesda, MD

²Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD

³Uniformed Services University, Bethesda, MD

Abstract

Introduction—Accurate prognosis assessment after non-small cell lung cancer (NSCLC) diagnosis is an essential step for making effective clinical decisions. This study is aimed to develop a prediction model with routinely available variables to assess prognosis in patients with NSCLC in the U.S. Military Health System.

Methods—We used the linked database from the Department of Defense’s Central Cancer Registry (CCR) and the Military Health System Data Repository (MDR). The dataset was randomly and equally split into a training set to guide model development and a testing set to validate the model prediction. Stepwise Cox regression was used to identify predictors of survival. Model performance was assessed by calculating area under the receiver operating curves (AUC) and construction of calibration plots. A simple risk scoring system was developed to aid quick risk score calculation and risk estimation for NSCLC clinical management.

Results—The study subjects were 5,054 patients diagnosed with NSCLC between 1998 and 2007. Age, sex, tobacco use, tumor stage, histology, surgery, chemotherapy, peripheral vascular disease, cerebrovascular disease and diabetes mellitus were identified as significant predictors of survival. Calibration showed high agreement between predicted and observed event rates. The AUC reached 0.841, 0.849, 0.848, and 0.838 over one, two, three and five years, respectively.

Conclusions—This is the first NSCLC prognosis model for quick risk assessment within the MHS. After external validation, the model can be translated into clinical use both as a web-based tool and through mobile applications easily accessible to physicians, patients and researchers.

Correspondence to: Kangmin Zhu, Division of Military Epidemiology and Population Sciences, John P. Murtha Cancer Center, Walter Reed National Military Medical Center, 11300 Rockville Pike, Suite 1120, Rockville, MD 20852 Phone: (301) 816-4786; kzhu@murthacancercenter.org.

Disclaimers: The views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the Departments of the Navy and Army, the Uniformed Services University of the Health Sciences, the Department of Defense, National Cancer Institute, or the U.S. Government. Nothing in the presentation implies any Federal/DoD endorsement.

The authors declare no conflicts of interest or financial disclosures.

Keywords

Model; military health system; mortality; non-small cell lung cancer; risk prediction

Introduction

Non-small cell lung cancer (NSCLC) comprises over 85% of lung cancers¹. The five-year survival rates for NSCLC range from 58.2% for early stage disease to a dismal 4.5% for advanced disease². Prognosis assessment upon NSCLC diagnosis is the first essential step towards making informed medical care decisions. Currently, cancer stage remains the most widely used prognostic factor in risk assessment for NSCLC³. However, the heterogeneity of the disease coupled with comorbidities results in substantial variability in survival among patients diagnosed at the same stage⁴. A more accurate risk stratification tool will likely aid in shared clinical-decision making, designs of clinical trials, and a better allocation of health care resources⁵.

To date, most models are derived from patient populations of clinical trials with small numbers of patients, confinement to specific tumor stages, and homogeneous patient characteristics^{6–12}. These models are often aimed at patients with advanced stage NSCLC and lack applicability to nonclinical trial patients^{13–16}. Some models have variables that are not readily available in routine clinical practice. In regard to population-based models, Blanchon et al. has developed one using medical records and questionnaire data from study participants diagnosed with NSCLC in French general hospitals¹³. This model demonstrated good discrimination accuracy and calibration by internal validation. However, the application of the model to U.S. populations has not been conducted with an external validation. A recent U.S. based model¹⁴ derived from the Surveillance, Epidemiology and End Results (SEER) database identified age, sex, tumor grade, tumor stage, and race as prognostic factors. However, the clinical application of this model is limited by the lack of chemotherapy data. An updated version was based on SEER-Medicare population and incorporated chronic obstructive pulmonary disease (COPD) as an additional predictor¹⁵. However, the model applies only to patients 65 of age or older.

The U.S. Military Health System (MHS) is an equal access healthcare system that provides universal health care to its beneficiaries including military service members, retirees, and their dependents. The Department of Defense (DoD) has a Central Cancer Registry (CCR) that collects detailed diagnosis, treatment and follow-up information for patients diagnosed with cancer. The DoD also maintains a Military Data Repository (MDR) that contains administrative and medical care information for MHS beneficiaries. The linked CCR and MDR database contains comprehensive data on demographics, tumor characteristics, medical history, and treatment information for MHS beneficiaries^{17–21}, which offers a unique resource to comprehensively study cancer prognosis. So far, there is no NSCLC prognosis prediction tool for MHS beneficiaries and their physicians. The major independent risk factors for predicting survival among NSCLC patients receiving care from the MHS have not been identified. It is not clear whether risk factors identified from the general population apply to patients in the MHS system. Therefore, this study aimed to

develop a prognostic assessment tool, which can be applied upon the diagnosis of NSCLC to the MHS beneficiaries, using the data in the MHS system.

Materials and Methods

Sources of data

Linked data from the DoD's Central Cancer Registry (CCR) and the MHS Data Repository (MDR) were used in this study, as previously described^{20, 21}. Currently, the linked database contains the data with cancer diagnosis from 1998 to 2007. The CCR contains information for cancer patients diagnosed or treated at military treatment facilities (MTFs), including active duty military personnel, retirees and their dependents. The CCR Data included demographic variables, tumor characteristics, cancer diagnosis, treatment, recurrence and vital status. The registry staff conduct lifetime follow-up on patients. Quality assurance was conducted following the guidelines established by the North America Association of Central Cancer Registries. The MDR contains administrative and medical care information that includes both inpatient and outpatient care provided at MTFs and civilian facilities paid for by the DoD. The MDR database includes information on clinical diagnoses of all medical conditions, which are coded using the diagnostic and treatment procedures or Current Procedural Terminology of the International Classification of Disease, 9th Revision (ICD-9). The Institutional Review Boards of the Walter Reed National Military Medical Center, TRICARE Management Activity, and the National Institutes of Health Office of Human Subjects Research approved the data linkage project.

Study subjects and variables

A total of 5,054 patients diagnosed with histologically confirmed primary NSCLC between 1998 and 2007 were identified from the linked database. Cancer site and histology were classified using the topography (C34.0 to C34.3, C34.8, C34.9) and morphology codes (8050–8078, 8083, 8084, 8250–8260, 8480–8490, 8570–8574, 8140, 8211, 8230, 8231, 8323, 8550, 8551, 8576, 8010–8012, 8014–8031, 8035, 8310, and any NSCLC codes between 8010 to 8576) of the International Classification of Diseases for Oncology, third edition (ICD-O-3)²².

Demographic variables, tobacco use history and tumor characteristics were obtained from the CCR. Demographic variables included age, sex, race, marital status, active duty status and military service branch of patient or sponsor at the time of diagnosis. Tumor characteristics included tumor stage, histology and tumor recurrence. Comorbidity data were obtained from the MDR. Comorbidities were considered as present if a diagnosis was recorded in at least one inpatient record or three or more outpatient records. Comorbidities were included if diagnosed at or before the diagnosis of NSCLC. Vital status and date of death were obtained from CCR. Both CCR and MDR data were used to determine the receipt of surgery, chemotherapy and radiation therapy. Missing values in a variable were coded as a separate missing /unknown category.

Statistical Analyses

The survival time was calculated as the difference between date of diagnosis and date of death, or censored at the date of last contact or the end of the study, December 31, 2009. The dataset was randomly and equally split into a training set (50% of the data) to guide the building of the risk model, and a testing set (the remaining 50% of the data) to validate the model prediction. Model development was performed in both the training and further repeated using the full dataset. As the results were similar, only results from the full dataset are presented in the final model. The assessment of model discriminatory accuracy and calibration was performed in training, testing and the full datasets.

We first performed univariate Cox regression to assess the association between individual variables and death. Variables with statistical significance ($P < 0.05$) and clinical relevance were considered as candidates for stepwise Cox regression analysis. Stepwise Cox regression was performed to choose the final subset of predictors. The model's discriminatory accuracy for predicting mortality was assessed by constructing the time-dependent Receiver Operating Characteristic (ROC) curves for censored survival data²³ and calculated area under curve (AUC). We assessed model calibration capability by assessing the agreement between predicted and observed death rates²⁴.

To facilitate the utility of the models in the clinical setting, we derived risk scores based on regression coefficients in the Cox proportional hazards model following standard procedures²⁵. The risk score was calculated by dividing each regression coefficient by the smallest coefficient significantly different from zero, and then rounding that number to the nearest integer. The lowest category of each risk factor was assigned a score of zero. Total risk score was calculated for each patient by summing the scores from all risk factors.

We calculated the individualized risk of death from baseline probability (probability of death at the reference level of all risk factors) and relative risk estimated from the Cox regression model. The predicted risks of death were estimated using the following equation²⁶:

$$F(t, X) = 1 - S(t, X) = 1 - [S_0(t)]^{\exp(\sum_{j=1}^p b_j X_j - \sum_{j=1}^p b_j M_j)}$$

Where $F(t)$ denotes the probability of death in t years given covariates X (x_1 to x_p). M_j denotes the mean level of X_j . $S(t)$ denotes the probability of alive until t ; $S_0(t)$ denotes the baseline survival function; and b_j denotes the regression coefficients.

Statistical analyses were conducted using the SAS software version 9.3.0 (SAS Institute, Inc.) and the R software. All reported P values are two sided, with the significance level set at $P < 0.05$.

Results

Among the 5,054 patients, 3,504 died during the follow-up period. The distributions of patient characteristics by vital status are shown in Supplementary Table 1. After stepwise selection, as shown in Table 1, the final multivariate model shows a significant increase in

mortality associated with older age (age 70 to 74 vs. age <50, HR=1.209, 95% CI = 1.013 to 1.443; age 80 and older vs. age < 50, HR=1.278, 95% CI = 1.056 to 1.545), male gender (male vs. female, HR = 1.177, 95% CI = 1.088 to 1.272), tobacco use (previous use vs. never use, HR = 1.230, 95% CI= 1.055 to 1.435; current use vs. never use, HR = 1.371, 95% CI = 1.173 to 1.602), late tumor stage (IB vs. IA, HR=1.410, 95% CI = 1.208 to 1.647; IIA vs. IA, HR = 1.710, 95% CI = 1.316 to 2.222; IIB vs. IA, HR = 2.141, 95% CI = 1.782 to 2.571; IIIA vs. IA, HR = 2.670, 95% CI = 2.284 to 3.121; IIIB vs. IA, HR = 3.265, 95% CI = 2.799 to 3.808; IV vs. IA, HR = 5.247, 95% CI = 4.560 to 6.038), large cell histology (large cell carcinoma vs. squamous cell carcinoma, HR = 1.290, 95% CI = 1.136 to 1.465), no surgery (no vs. yes, HR = 2.746, 95% CI = 2.493 to 3.024), no chemotherapy (no vs. yes, HR = 1.775, 95% CI = 1.635 to 1.926), peripheral vascular disease (yes vs. no, HR = 1.165, 95% CI = 1.039 to 1.306), cerebrovascular disease (yes vs. no, HR = 1.185, 95% CI = 1.043 to 1.346) and diabetes mellitus (yes vs. no, HR = 1.124, 95% CI = 1.024 to 1.234).

The calibration plot showed that the observed probabilities of survival were all within the 95% CI of the predicted probabilities of survival at two, three and five years after diagnosis, respectively (Figure 1, Panels b, c and d, respectively). The model slightly underestimated survival among individuals having survival probability greater than 0.8 in the first year after diagnosis (Figure 1, Panel a).

Figure 2 shows that the AUC reached 0.841 (Panel a), 0.849 (Panel b), 0.848 (Panel c), 0.838 (Panel d) for one-, two-, three- and five year-prediction, respectively. The AUCs were similar in the full, training and testing data sets (Table 2). For example, the AUCs for predicting survival in two years were 0.849 (95% CI = 0.829 to 0.864), 0.848 (95% CI = 0.830 to 0.864), and 0.849 (95% CI = 0.835 to 0.862) for the full, training and testing datasets, respectively.

Risk scores were assigned to each significant risk factor identified in the model (Supplementary Table 2). Probability of survival decreased in all risk groups over time after diagnosis. Given a time point, patients with a higher risk score exhibited lower survival probability than patients with a lower risk score (Figure 3).

We next applied the model to predict probability of survival using five hypothetical examples (Table 3). Example 1 is a 71-year-old man with stage IB squamous cell carcinoma, a current smoker with a history of cerebrovascular disease and diabetes. The patient has refused to receive surgery. The total risk score for this patient is then 23 and the predicted probability of survival in one year is 0.458 (95% CI=0.386 to 0.543). If the same patient is treated with surgery (example 2), then his total risk score decreases to 15 and one year survival probability is significantly increased to 0.752 (95% CI=0.709 to 0.799). The risk score decreases to 13 and the one year survival probability is further increased to 0.826 (95% CI=0.800 to 0.852), as shown in example 3, provided that the patient had stopped smoking and become a former smoker (e.g. had stopped smoking for at least a year or more) and had not had comorbidities. In example 4, the patient is a 66-yr-old woman with adenocarcinoma diagnosed at stage IIIB, a never smoker with a history of peripheral vascular disease. The total risk score is 25 and the predicted probability of survival in one year is 0.377 (95% CI=0.302 to 0.471). If she does not have a history of peripheral vascular

disease and is willing to undergo chemotherapy, as shown in example 5, then the risk score decreases to 16 and her survival probability increases to 0.624 (95% CI=0.571 to 0.682). The predictions in two, three and five years under these hypothetical scenarios are also shown in Table 3.

Discussion

Accurate prognostic assessment tool is important to guide shared treatment decision making and disease management. In building risk prediction tools, one important consideration is the tradeoff between statistical accuracy and clinical utility. The more predictors a model includes, the more accurate the prediction could be. However, the unavailability of the many predictors in routine clinical settings could dampen the model's clinical utility. While a simple model with a few predictors could be statistically less accurate than a model that exhausts all possible predictors, it has higher clinical utility with readily available variables from a clinical setting. It is ideal to have a simple model based on routinely available predictors while achieving high prediction accuracy.

Our model was based on variables from routine medical care settings and generated high prediction accuracy and calibration with a discrimination accuracy of 84–85%. In clinical settings, besides tumor stage and histology information usually obtained from routine diagnosis workout, a particular patient will only need to provide simple information on demographics (age, sex), tobacco use history (never, previous or current), diagnosis of a few comorbidities (yes or no) and prior cancer treatment (yes or no) to have risk score calculated and probability of survival accurately estimated.

Unlike the few previous published tools that used clinical trial participants^{6–12, 27, 28}, our utility comes from taking real-world known outcomes and developing a tool within the closed military health system. Patients enrolled in specific clinical trials are mostly a homogeneous group with defined characteristics (e.g. certain stage and age groups) to satisfy eligibility criteria of the trials. These models are therefore not generalizable to other populations. Moreover, the homogeneity of treatment regimens and agents in certain clinical trials makes these models trial-specific and less suitable for general clinical use. Some of these models also require data from additional procedures that are not readily available during the initial clinic visit (7–10). Although external validation of these models have been performed, the number of patients, both in the development and validation cohorts, are still relatively small (7, 8, 10). Our model was derived from larger patient population with the rich epidemiologic and medical care data and can be applied to a wider range of patient population.

Currently, there are few population-based models with a wide range of patient population. A French study¹³ developed a prognostic model that identified age, sex, performance status at diagnosis, histological type, tumor stage as independent predictors of mortality with good calibration and a high discrimination accuracy. However, therapeutic treatments were not considered as candidate variables in model development and tobacco use history was not identified to affect survival in the French population. Among studies from U.S. populations, a study based on the National Cancer Institute's SEER data¹⁴ identified age, sex, tumor

grade, tumor stage, and race as prognostic factors. This model had a much lower discriminatory accuracy of 0.73¹⁴. Due to the lack of chemotherapy and comorbidity data in the SEER database, the prognostic significance of treatment and comorbidity could not be assessed. An update of this model was derived from the SEER-Medicare data¹⁵. However, because SEER-Medicare data contain only patients sixty-five years or older, the model cannot apply to younger patients.

Unlike the SEER and the SEER-Medicare based model, our model covers all ages and has the capacity to assess treatment effects and thus is able to aid both clinical decision-making and shared decision making. It is noteworthy that race was not identified as a significant predictor of survival in our population, while both the SEER and SEER-Medicare models^{14, 15} identified black race to be associated with an increased risk of mortality. This is most probably a result of all our patients being treated in an equal health care system regardless of race²⁹.

Although our model was developed within the MHS system, the factors identified are consistent with those identified in the general population. All variables included in our model have high clinical relevance to survival in NSCLC patients. Specifically, advanced tumor stage, older age, male sex and tobacco use are well-established risk factors associated with poor survival in NSCLC patients^(32, 33). Large cell histology has also been reported to be associated with worse survival compared to other histologic types (12, 14). The associations between type II diabetes mellitus and high risk of mortality in lung cancer patients have been reported in previous studies³⁰⁻³³ although the association was not consistently observed in other studies³⁴. Peripheral vascular and cerebrovascular diseases are among the most commonly found comorbidities among lung cancer patients that contributed to decreased survival in NSCLC patients^{35, 36}. Our results showed that the associations of the comorbidities with survival remain significant in the presence of strong predictors such as tumor stage and treatments. Although these comorbid conditions may not be completely curable, the results suggest the importance of better management and treatment of the comorbid conditions to improve survivorship while receiving cancer treatments. Our results also suggest the potential role of tobacco cessation for improving survival.

This study can be further improved. First, progress in molecular markers has made it possible to integrate molecular profiles^{4, 37-39} convenient in clinical use, such as the epidermal growth factor receptor (EGFR), a well-established marker influencing treatment response and outcome in NSCLC patients⁴⁰. The integration of clinical molecular markers into our model is warranted and is planned as they become available. Second, performance status, which was not available in our study, may be a potential prognostic factor to be tested and integrated into the model if it helps improve model prediction. Third, there have been advances in lung cancer diagnosis, staging and treatment during the time covered in this study and since. The introduction of PET (positron emission tomography) and EBUS (endobronchial ultrasound) aid diagnosis and staging. The ability to provide targeted therapy has improved the treatment armamentarium. Thus, this model could be further tested and improved with the inclusion of current standards of care. Finally, external validation of our model in the general population would be desirable. Although our results are similar to some

previous studies in the general population, the model may not be directly generalized to the general population without external validation.

In conclusion, our model is the first NSCLC prognosis model for the DoD's MHS system that could help physicians and patients perform quick prognosis assessment. The model prediction has high statistical accuracy and the variables are readily obtainable in routine clinical setting. The risk scores are simple to calculate and allow for ease in communication. After external validation, the model can be translated into clinical use as a web-based tool or through portable mobile devices easily accessible to physicians, patients and researchers^{16,41}.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Sources of Support: This project was supported by John P. Murtha Cancer Center, Walter Reed National Military Medical Center via the Uniformed Services University of the Health Sciences under the auspices of the Henry M. Jackson Foundation for the Advancement of Military Medicine and by the intramural research program of the National Cancer Institute. The original data linkage was supported by the United States Military Cancer Institute and Division of Cancer Epidemiology and Genetics, National Cancer Institute.

References

1. American Cancer Society. <http://www.cancer.org/cancer/lungcancer-non-smallcell/detailedguide/non-small-cell-lung-cancer-what-is-non-small-cell-lung-cancer>
2. Howlader, NNA.; Krapcho, M.; Garshell, J.; Miller, D.; Altekruse, SF.; Kosary, CL.; Yu, M.; Ruhl, J.; Tatalovich, Z.; Mariotto, A.; Lewis, DR.; Chen, HS.; Feuer, EJ.; Cronin, KA., editors. SEER Cancer Statistics Review, 1975–2011. National Cancer Institute; Bethesda, MD: Apr. 2014 http://seer.cancer.gov/csr/1975_2011/, based on November 2013 SEER data submission, posted to the SEER web site
3. Osarogiagbon RU. Predicting survival of patients with resectable non-small cell lung cancer: Beyond TNM. *J Thorac Dis.* 2012; 4:214–216. [PubMed: 22833830]
4. Kratz JR, He J, Van Den Eeden SK, et al. A practical molecular assay to predict survival in resected non-squamous, non-small-cell lung cancer: development and international validation studies. *Lancet.* 2012; 379:823–832. [PubMed: 22285053]
5. Onn A, Dickey BF. A better crystal ball to predict lung-cancer survival? *Lancet Oncol.* 2006; 7:789–790. [PubMed: 17012037]
6. Park MJ, Lee J, Hong JY, et al. Prognostic model to predict outcomes in nonsmall cell lung cancer patients treated with gefitinib as a salvage treatment. *Cancer.* 2009; 115:1518–1530. [PubMed: 19177485]
7. Dehing-Oberije C, Aerts H, Yu S, et al. Development and validation of a prognostic model using blood biomarker information for prediction of survival of non-small-cell lung cancer patients treated with combined chemotherapy and radiation or radiotherapy alone (NCT00181519, NCT00573040, and NCT00572325). *Int J Radiat Oncol Biol Phys.* 2011; 81:360–368. [PubMed: 20888135]
8. Dehing-Oberije C, De Ruyscher D, van der Weide H, et al. 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. *Int J Radiat Oncol Biol Phys.* 2008; 70:1039–1044. [PubMed: 17889446]
9. Dehing-Oberije C, Yu S, De Ruyscher D, et al. Development and external validation of prognostic model for 2-year survival of non-small-cell lung cancer patients treated with chemoradiotherapy. *Int J Radiat Oncol Biol Phys.* 2009; 74:355–362. [PubMed: 19095367]

10. Mandrekar SJ, Schild SE, Hillman SL, et al. A prognostic model for advanced stage nonsmall cell lung cancer. Pooled analysis of North Central Cancer Treatment Group trials. *Cancer*. 2006; 107:781–792. [PubMed: 16847887]
11. 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; 7:1361–1368. [PubMed: 22843087]
12. Hoang T, Xu R, Schiller JH, et al. Clinical model to predict survival in chemo-naïve patients with advanced non-small-cell lung cancer treated with third-generation chemotherapy regimens based on eastern cooperative oncology group data. *J Clin Oncol*. 2005; 23:175–183. [PubMed: 15625371]
13. Blanchon F, Grivaux M, Asselain B, et al. 4-year mortality in patients with non-small-cell lung cancer: development and validation of a prognostic index. *Lancet Oncol*. 2006; 7:829–836. [PubMed: 17012045]
14. Putila J, Remick SC, Guo NL. Combining clinical, pathological, and demographic factors refines prognosis of lung cancer: a population-based study. *PLoS One*. 2011; 6:e17493. [PubMed: 21364765]
15. Putila J, Guo NL. Combining COPD with clinical, pathological and demographic information refines prognosis and treatment response prediction of non-small cell lung cancer. *PLoS One*. 2014; 9:e100994. [PubMed: 24967586]
16. Zhang M, Liu Y, Jiang Y, et al. Model based user interface design for predicting lung cancer treatment outcomes. *Conf Proc IEEE Eng Med Biol Soc*. 2011; 2011:75–78. [PubMed: 22254254]
17. Gill AA, Enewold L, Zahm SH, et al. Colon cancer treatment: are there racial disparities in an equal-access healthcare system? *Dis Colon Rectum*. 2014; 57:1059–1065. [PubMed: 25101601]
18. Enewold L, McGlynn KA, Zahm SH, et al. Surveillance mammography among female Department of Defense beneficiaries: a study by race and ethnicity. *Cancer*. 2013; 119:3531–3538. [PubMed: 23913448]
19. Zhou J, Enewold L, Zahm SH, et al. Breast conserving surgery versus mastectomy: the influence of comorbidities on choice of surgical operation in the Department of Defense health care system. *Am J Surg*. 2013; 206:393–399. [PubMed: 23866763]
20. Andaya AA, Enewold L, Zahm SH, et al. Race and colon cancer survival in an equal-access health care system. *Cancer Epidemiol Biomarkers Prev*. 2013; 22:1030–1036. [PubMed: 23576691]
21. Enewold L, Zhou J, McGlynn KA, et al. Racial variation in breast cancer treatment among Department of Defense beneficiaries. *Cancer*. 2012; 118:812–820. [PubMed: 21766298]
22. Fritz, A.; Percy, C.; Jack, A. *International Classification of Diseases for Oncology*. 3. World Health Organization; 2000.
23. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*. 2000; 56:337–344. [PubMed: 10877287]
24. D’Agostino, RD.; Nam, BH. Evaluation of performance of survival analysis models: discrimination and calibration measures. In: Balakrishnan, N.; Rao, CR., editors. *Handbook of Statistics*. Vol. 23. Amsterdam: Elsevier; 2004. p. 1-25.
25. Sullivan LM, Massaro JM, D’Agostino RB Sr. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. *Stat Med*. 2004; 23:1631–1660. [PubMed: 15122742]
26. Hosmer, DW.; Lemeshow, S. *Applied Survival Analysis: Regression Modeling of Time to Event Data*. New York: John Wiley & Sons, Inc; 1999.
27. Williams BA, Sugimura H, Endo C, et al. Predicting postrecurrence survival among completely resected nonsmall-cell lung cancer patients. *Ann Thorac Surg*. 2006; 81:1021–1027. [PubMed: 16488713]
28. Sawyer TE, Bonner JA, Gould PM, et al. Patients with stage I non-small cell lung carcinoma at postoperative risk for local recurrence, distant metastasis, and death: implications related to the design of clinical trials. *Int J Radiat Oncol Biol Phys*. 1999; 45:315–321. [PubMed: 10487551]
29. Zheng L, Enewold L, Zahm SH, et al. Lung cancer survival among black and white patients in an equal access health system. *Cancer Epidemiol Biomarkers Prev*. 2012; 21:1841–1847. [PubMed: 22899731]

30. Tseng CH. Higher risk of mortality from lung cancer in Taiwanese people with diabetes. *Diabetes Res Clin Pract.* 2013; 102:193–201. [PubMed: 24262943]
31. Inal A, Kaplan MA, Kucukoner M, et al. Is diabetes mellitus a negative prognostic factor for the treatment of advanced non-small-cell lung cancer? *Rev Port Pneumol.* 2014; 20:62–68. [PubMed: 24210228]
32. Chiou WK, Hwang JS, Hsu KH, et al. Diabetes mellitus increased mortality rates more in gender-specific than in nongender-specific cancer patients: a retrospective study of 149,491 patients. *Exp Diabetes Res.* 2012; 2012:701643. [PubMed: 22701471]
33. Shieh SH, Probst JC, Sung FC, et al. Decreased survival among lung cancer patients with comorbid tuberculosis and diabetes. *BMC Cancer.* 2012; 12:174. [PubMed: 22578056]
34. Ioacara S, Guja C, Ionescu-Tirgoviste C, et al. Cancer specific mortality in insulin-treated type 2 diabetes patients. *PLoS One.* 2014; 9:e93132. [PubMed: 24667573]
35. Battafarano RJ, Piccirillo JF, Meyers BF, et al. Impact of comorbidity on survival after surgical resection in patients with stage I non-small cell lung cancer. *J Thorac Cardiovasc Surg.* 2002; 123:280–287. [PubMed: 11828287]
36. Ahn DH, Mehta N, Yorio JT, et al. Influence of medical comorbidities on the presentation and outcomes of stage I-III non-small-cell lung cancer. *Clin Lung Cancer.* 2013; 14:644–650. [PubMed: 23886797]
37. Xie Y, Minna JD. Predicting the future for people with lung cancer. *Nat Med.* 2008; 14:812–813. [PubMed: 18685594]
38. Xie Y, Minna JD. A lung cancer molecular prognostic test ready for prime time. *Lancet.* 2012; 379:785–787. [PubMed: 22386017]
39. Shedden K, Taylor JM, Enkemann SA, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med.* 2008; 14:822–827. [PubMed: 18641660]
40. Lee CK, Brown C, Gralla RJ, et al. Impact of EGFR inhibitor in non-small cell lung cancer on progression-free and overall survival: a meta-analysis. *J Natl Cancer Inst.* 2013; 105:595–605. [PubMed: 23594426]
41. Gegg-Harrison T, Zhang M, Meng N, et al. Porting a cancer treatment prediction to a mobile device. *Conf Proc IEEE Eng Med Biol Soc.* 2009; 2009:6218–6221. [PubMed: 19965083]

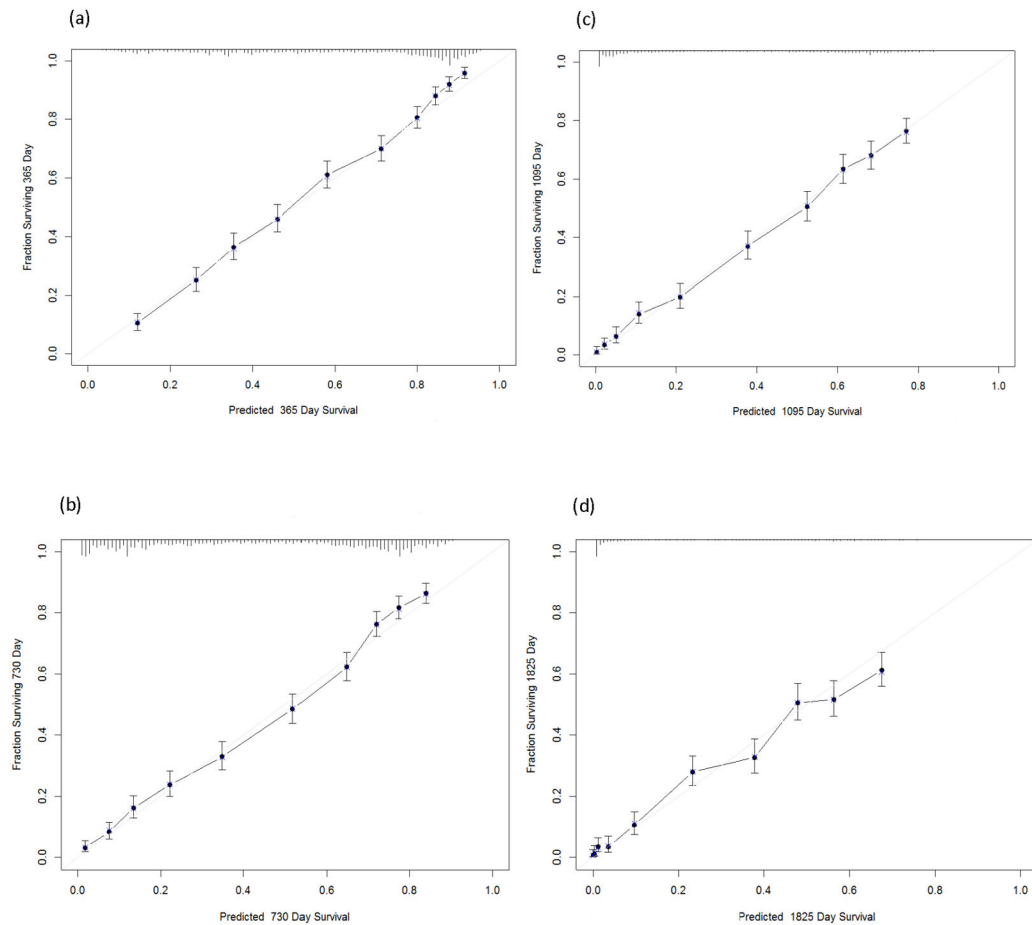


Figure 1. Calibration plots (observed probability vs. predicted probability) for different time periods: (a) one year; (b) two years; (c) three years; (d) five years. Y axis represents observed probability. X axis represents predicted probability. The predicted and observed probabilities of survival are graphed on the horizontal and vertical axes, respectively. The grey line indicates the reference line, on which an ideal model would lie. Solid dots mark the predictions; X's mark the cross-validated predictions. Vertical bars indicate 95% confidence intervals around the prediction.

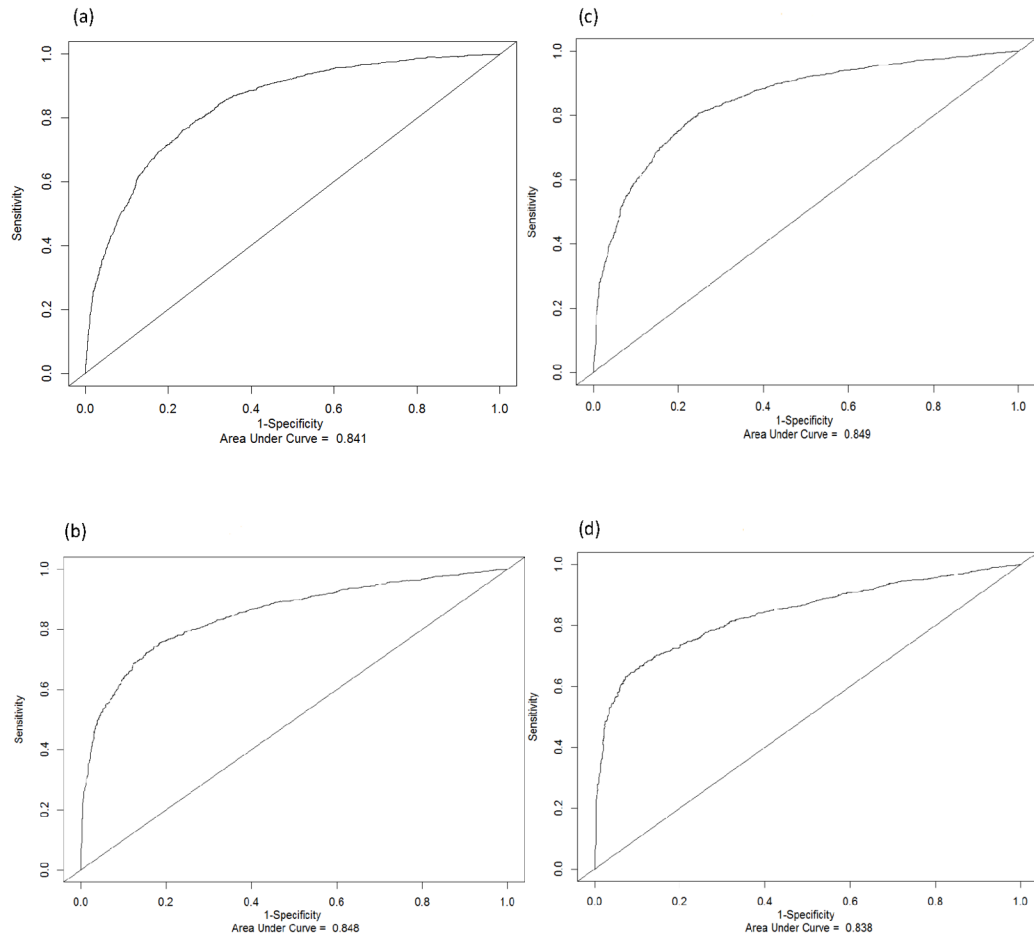


Figure 2. Time-dependent Receiver Operating Characteristic (ROC) curves and area under curve (AUC) for different time periods: (a) one year; (b) two years; (c) three years; (d) five years

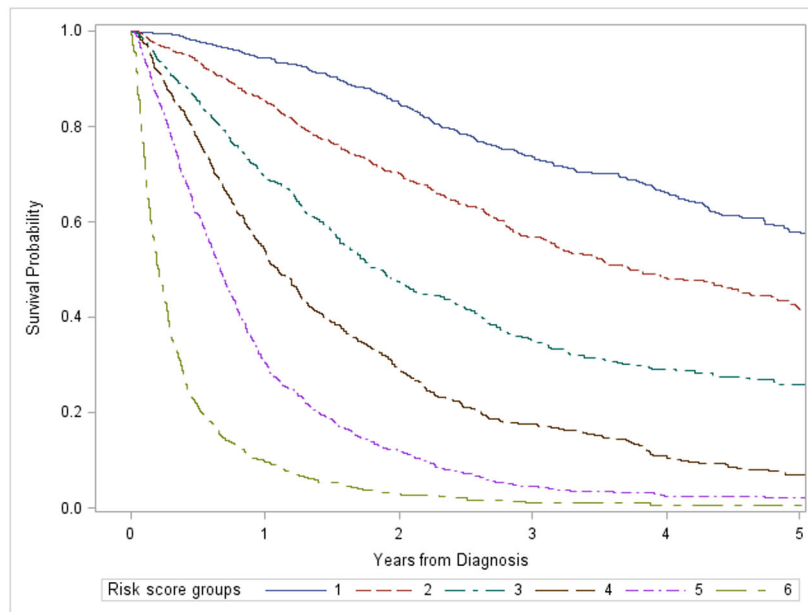


Figure 3. Survival probability by risk score groups. Risk scores were assigned to each risk factor by dividing each regression coefficient by the smallest coefficient significantly different from zero, and then rounded to the nearest integer. A risk score was assigned to each patient by summing the points for each risk factor present. Risk score groups are defined as following: group 1: risk score=0 to 9; group 2: risk score=10 to 14; group 3: risk score=15 to 19; group 4: risk score=20 to 24; group 5: risk score=25 to 29; group 6: risk score 30.

Table 1 Predictors of mortality among non-small cell lung cancer patients diagnosed between 1998 and 2007 in the Military Health System

Risk Factors	Regression Coefficient (β)	Standard Error (SE)	Hazard Ratio (HR)	95% Confidence Interval (CI)	P-value
Age					
<50	0.000	0.000	Ref.	Ref.	
50–59	0.037	0.082	1.037	0.883 to 1.218	0.6543
60–69	0.149	0.077	1.161	0.998 to 1.350	0.053
70–74	0.170	0.085	1.185	1.003 to 1.399	0.0454
75–79	0.190	0.090	1.209	1.013 to 1.443	0.0353
80 and older	0.245	0.097	1.278	1.056 to 1.545	0.0116
Sex					
Female	0.000	0.000	Ref.	Ref.	
Male	0.163	0.040	1.177	1.088 to 1.272	<0.0001
Tobacco use history					
Never used	0.000	0.000	Ref.	Ref.	
Previous use	0.207	0.078	1.230	1.055 to 1.435	0.0083
Current use	0.315	0.080	1.371	1.173 to 1.602	<0.0001
Stage					
IA	0.000	0.000	Ref.	Ref.	
IB	0.344	0.079	1.410	1.208 to 1.647	<0.0001
IIA	0.537	0.134	1.710	1.316 to 2.222	<0.0001
IIB	0.761	0.094	2.141	1.782 to 2.571	<0.0001
IIIA	0.982	0.080	2.670	2.284 to 3.121	<0.0001
IIIB	1.183	0.079	3.265	2.799 to 3.808	<0.0001
IV	1.658	0.072	5.247	4.560 to 6.038	<0.0001
Histology					
Squamous cell carcinoma	0.000	0.000	Ref.	Ref.	
Adenocarcinoma	0.015	0.047	1.016	0.927 to 1.113	0.7408
Large cell carcinoma	0.255	0.065	1.290	1.136 to 1.465	<0.0001
Other NSCLC, NOS *	0.028	0.055	1.028	0.924 to 1.144	0.6128
Surgery					

Risk Factors	Regression Coefficient (β)	Standard Error (SE)	Hazard Ratio (HR)	95% Confidence Interval (CI)	P-value
Yes	0.000	0.000	Ref.	Ref.	
No	1.010	0.049	2.746	2.493 to 3.024	<0.0001
Chemotherapy					
Yes	0.000	0.000	Ref.	Ref.	
No	0.574	0.042	1.775	1.635 to 1.926	<0.0001
Peripheral vascular disease					
No	0.000	0.000	Ref.	Ref.	
Yes	0.153	0.058	1.165	1.039 to 1.306	0.0091
Cerebrovascular disease					
No	0.000	0.000	Ref.	Ref.	
Yes	0.170	0.065	1.185	1.043 to 1.346	0.0092
Type II diabetes mellitus					
No	0.000	0.000	Ref.	Ref.	
Yes	0.117	0.047	1.124	1.024 to 1.234	0.014

* NSCLC=non-small cell lung cancer; NOS=Not otherwise specified

Table 2

Time-dependent area under curve (AUC) of the prediction model in training, testing and full datasets for non-small cell lung cancer patients diagnosed between 1998 and 2007 in Military Health System

	AUC	95% CI
In 1 year		
Training set	0.840	0.823 to 0.857
Testing set	0.842	0.826 to 0.860
Full dataset	0.841	0.829 to 0.853
In 2 years		
Training set	0.849	0.829 to 0.864
Testing set	0.848	0.830 to 0.864
Full dataset	0.849	0.835 to 0.859
In 3 years		
Training set	0.849	0.832 to 0.869
Testing set	0.851	0.831 to 0.868
Full dataset	0.848	0.835 to 0.862
In 5 years		
Training set	0.843	0.812 to 0.862
Testing set	0.840	0.822 to 0.865
Full dataset	0.838	0.822 to 0.854

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Application of the prediction model to hypothetical individuals with different risk profiles

	Example 1	Example 2	Example 3	Example 4	Example 5
Age	71	71	71	66	66
Sex	Male	Male	Male	Female	Female
Tobacco use	Current	Current	Previous	Never	Never
Stage	IB	IB	IB	IIIB	IIIB
Histology	Squamous cell carcinoma	Squamous cell carcinoma	Squamous cell carcinoma	Adenocarcinoma	Adenocarcinoma
Peripheral vascular disease	No	No	No	Yes	No
Cerebrovascular disease	Yes	Yes	No	No	No
Type II diabetes mellitus	Yes	Yes	No	No	No
Surgery	No	Yes	Yes	No	No
Chemotherapy	No	No	No	No	Yes
Total risk score	23	15	13	25	16
Predicted probability of survival in 1 yr (95% CI)	0.458 (0.386 to 0.543)	0.752 (0.709–0.799)	0.826 (0.800–0.852)	0.377 (0.302–0.471)	0.624 (0.571–0.682)
Predicted probability of survival in 2 yr (95% CI)	0.219 (0.158 to 0.305)	0.576 (0.513–0.645)	0.689 (0.649–0.731)	0.150 (0.097–0.232)	0.400 (0.337–0.475)
Predicted probability of survival in 3 yr (95% CI)	0.105 (0.064 to 0.171)	0.439 (0.371–0.521)	0.575 (0.536–0.627)	0.060 (0.031–0.114)	0.256 (0.198–0.330)
Predicted probability of survival in 5 yr (95% CI)	0.033 (0.016 to 0.071)	0.289 (0.223–0.375)	0.434 (0.380–0.495)	0.014 (0.005–0.038)	0.128 (0.086–0.189)