Development of a local nomogram-based scoring system for predicting overall survival in idiopathic pulmonary fibrosis: A rural appalachian experience

Rowida Mohamed[1], **Rahul G. Sangani**[2], **Khalid M. Kamal**[3], **Traci J. LeMaster**[3], **Toni Marie Rudisill**[4], **Virginia G. Scott**[3], **George A. Kelley**[4,5], **Sijin Wen**[4]

[1]Biological Sciences Division, University of Chicago, Chicago, Illinois, USA

[2]Pulmonary and Critical Care Medicine, Northeast Georgia Health System, Gainesville, Georgia, USA

[3]Department of Pharmaceutical Systems and Policy, School of Pharmacy, West Virginia University, Morgantown, West Virginia, USA

[4]Department of Epidemiology and Biostatistics, School of Public Health, West Virginia University, Morgantown, West Virginia, USA

[5]School of Public and Population Health and Department of Kinesiology, Boise State University, Boise, Idaho, USA

Abstract

Background: Accurate staging systems are essential for assessing the severity of idiopathic pulmonary fibrosis (IPF) and guiding clinical management. This study aimed to evaluate the prognostic value of pulmonary comorbidities and body mass index (BMI) in IPF, develop a nomogram predicting overall survival (OS), and create a nomogram-based survival prediction model.

Methods: Patients with IPF were identified from electronic medical records of the West Virginia hospital system. Least Absolute Shrinkage and Selection Operator (LASSO) Cox regression analysis was used for variable selection, and a nomogram was constructed. Risk groups were defined based on the nomogram's probability tertiles. The performance of the nomogram-based model was evaluated using Harrell's concordance index (C-index) and the Hosmer–Lemeshow test.

Results: The study included 152 patients with IPF. The majority of the patients were elderly, male, and had a BMI above 24 kg/m^2. The median survival duration was 7.6 years. The survival rates were 91% at 1 year, 78% at 3 years, and 68% at 5 years. LASSO regression selected carbon monoxide lung diffusion capacity percentage predicted (DLco%), BMI, pulmonary hypertension, pulmonary embolism, and sleep apnea as independent predictive variables. The nomogram demonstrated good discrimination (C-index = 0.71) and calibration.

Conclusions: Pulmonary comorbidities and BMI have significant prognostic value in IPF, emphasizing the necessity for consistent screening, assessment, and management of these factors in IPF care. Furthermore, the nomogram-based staging system showed promising performance in predicting OS and represents an actionable staging system that could potentially improve clinical management in IPF. Further validation of the nomogram is warranted to confirm its utility in clinical practice.

Keywords

body mass index; hypertension; idiopathic pulmonary fibrosis; nomogram; staging system; survival

1 | INTRODUCTION

The progression of idiopathic pulmonary fibrosis (IPF) is influenced by the disease severity at presentation. Therefore, accurate assessment of IPF disease severity using valid staging systems is integral to the clinical management of patients with IPF. Unfortunately, there is no widely-recognized staging system for IPF [1]. Currently, the most commonly used prognostic index for IPF is the Gender-Age-Physiological (GAP) index [2]. The GAP index classifies patients into three categories based on sex, age, and physiological variables: forced vital capacity percentage predicted (FVC%) and carbon monoxide lung diffusion capacity percentage predicted (DLco%). However, age and sex are already incorporated in the percentage predicted values of lung function estimates [1]. Nevertheless, over the last decade, the knowledge of IPF prognosis has evolved; patients with IPF prognosis depend on their concomitant comorbidities.

As IPF occurs mainly in elderly patients, it is often associated with comorbidities that can significantly impact the health outcomes of these patients [3]. Thus, managing comorbid conditions is crucial in the comprehensive care of patients with IPF [4]. Pulmonary comorbidities, such as pulmonary hypertension, pulmonary embolism, sleep apnea, emphysema, lung cancer, and malnutrition, have been found to be consistently associated with shorter survival in IPF [4]. However, there is still a dearth of research on the value of comorbidities for improving the prediction of survival in IPF. Torrisi et al. found

that comorbidities (gastroesophageal reflux, pulmonary hypertension, atrial arrhythmias, lung cancer, and valvular heart disease) significantly improved the prediction of risk of death when added to GAP [5]. Also, body mass index (BMI) is one of the most accepted methods to assess malnutrition, which is common among patients with IPF. Previous work found that the inclusion of BMI reinforced the GAP performance in patients with IPF [6, 7]. Thus, including comorbidities in the staging system could potentially improve the real-life clinical approach to the management of IPF.

Tools that accurately enable the clinician to evaluate a patient's situation, also known as decision aids, are crucial for making the final decision on clinical management [8]. Among the available decision aids, nomograms represent the most accurate tools for predicting outcomes; the superior performance of nomograms compared to risk grouping has been well-documented [9]. Risk stratification is a standard procedure in clinical practice to categorize patients into risk groups. However, when continuous variables are used to stratify patients into risk groups, group assignment becomes less certain for patients with values close to cutoff points. In this situation, a nomogram is more desirable as it allows continuous variables to be integrated into risk prediction [10]. Nomograms provide a reliable visualization of multivariable prognostic regression models that quantify the risk of a clinical event. Regression coefficients, representing the weightings of different variables, determine the length and scale of corresponding variable axes in the nomogram construction [10]. In other words, nomograms take the weights of parameters into account, whereas conventional staging systems do not. Given the above-mentioned considerations, this study aimed to (1) investigate the prognostic value of pulmonary comorbidities and BMI (as a proxy for malnutrition) along with pulmonary function tests (PFTs), (2) establish a nomogram predicting the 5-year survival probability in IPF, and (3) develop a nomogram-based survival prediction model.

2 | METHODS

2.1 | Patient participants

Patients with IPF were identified using the International Classification of Diseases (ICD) coding (ICD 9-CM: 515–516.9; ICD-10-CM J84–84.9) from the electronic medical records of the West Virginia University Medicine hospitals system from January 1, 2015 to December 31, 2019. The West Virginia University Medicine system includes both tertiary centers and peripheral hospitals. The study inclusion criteria were based on the latest American Thoracic Society/European Respiratory Society/Japanese Respiratory Society/ Latin American Thoracic Association (ATS/ERS/JRS/ALAT) diagnostic guidelines [11]. Details regarding study patient selection have been published elsewhere [12, 13].

2.2 | Data collection

Demographic and clinical data were extracted manually from electronic medical records at the time of the first available PFTs, which were the closest to the recorded ICD code. Extracted data included age (60/61–65/>65 years), gender (males/females), BMI (>24/ 24 kg/m^2), lung transplant (yes/no), pulmonary comorbidities, including pulmonary hypertension (absent/present), lung cancer (absent/present), emphysema (absent/

present), pulmonary embolism (absent/present), and sleep apnea (present/absent), PFTs, including FVC% (>75%/50%–75%/<50%/missing/unknown), DLco% (>55%/36%–55%/ 35%/missing/unknown), and anti-fibrotic use (yes/no). In order to ensure the accuracy of our analysis, we decided not to include smoking status as a variable in the analysis. Respiratory research has suggested a potential bias known as the "healthy smoker effect," where current smokers may have a better prognosis than never or former smokers, even after adjusting for individual and socioeconomic variables [14, 15]. Emphysema diagnosis was confirmed by checking computed tomography scans. The cutoff points for BMI and PFTs were parallel to those used in the GAP and GAP-Plus-BMI staging systems [2, 7]. Except for PFTs. We did not encounter any missing data for demographic and clinical data. In order to assess the pattern of missing data for PFTs, we compared the overall survival (OS) between two groups: patients with available PFT data and patients with missing PFT values, using Kaplan–Meier curves (Figure S1). This analysis revealed a significant difference in OS between patients with available PFT data and those with missing values. Specifically, the group with missing PFT values showed significantly worse OS compared to the group with available PFT data. These results strongly suggest that the missingness of PFT data was not random. Since the assumption of missing at random (MAR) was not satisfied to permit multiple imputation or complete case analysis, we adopted a coding approach where missing PFT values were designated as "unknown" [16–18] This approach aligns with the GAP index methodology and enhances the nomogram's practicality and applicability to real-world clinical scenarios, where high levels of missingness in PFTs occur due to challenges in adhering to guidelines and patients' deteriorating ability to perform these tests [18–20]. This coding strategy allowed us to explicitly acknowledge and account for the missing data in our analysis. Data were aggregated using REDCap, a HIPAA-compliant data aggregation tool [21].

2.3 | Follow-up and outcome assessment

The primary endpoint was OS. The cohort was followed up until June 2022.

2.4 | Statistical analysis

This study aimed to assess the association between independent predictors and OS. Variables were described using numbers and percentages. For data dimension reduction and variable selection, the Least Absolute Shrinkage and Selection Operator (LASSO) Cox regression analysis with cross-validation was used to identify the best subset of variables as the predictive variables in the regression models [22]. Lambda (often denoted as λ), which refers to the regularization parameter or penalty term applied to the model, was selected by tenfold cross-validation and was used to determine the best subset of the variables. A smaller lambda allows more flexibility in the model, allowing coefficients to take on larger values. On the other hand, a larger lambda increases the amount of regularization, resulting in more coefficients being set to zero and, subsequently, to sparsity in the model. The significance of each variable in the best subset was evaluated by Cox regression analysis. The proportional hazard assumption was confirmed by checking the scaled Schoenfeld residuals [23]. All variables were considered in multivariable Cox proportional hazard regression models. The nomogram was constructed based on results from the multivariable Cox proportional hazard model. Values for each of the model covariates were mapped on a scale ranging from 0

to 100 points, with total points obtained for each covariate mapped to a 5-year survival probability associated with that combination of covariate values. In addition, calibration curves were plotted to assess the predictive accuracy of the nomogram. For clinical use of the model, a risk classification system was established based on the survival probability of each patient. Patients were categorized into three prognostic groups according to the following nomogram probability tertiles: patients with >66.6% 5-year survival probability (low-risk), 33.3%–66.6% 5-year survival probability (moderate risk), and <33.3% 5-year survival probability (high-risk). In our data set, censoring times (the outcome variable) varied among subjects, with entry times differing randomly among subjects due to the retrospective design of the study. These variations in censoring may introduce potential biases in the data analysis. However, the Kaplan–Meier and Cox regression analysis, which are based on maximum likelihood estimation, are commonly used to handle time-to-event outcomes and provide unbiased estimates under random censoring conditions [24]. Kaplan–Meier curves were constructed to analyze the difference in OS between risk groups, and hazard ratios were estimated using Cox proportional hazard models. Harrell's concordance index (C-index) was used to examine the discrimination of the fitted model. A C-index score of 0.5 suggests random chance, while a score of 1.0 implies perfect discrimination [25]. The Hosmer–Lemeshow test was performed where $p > 0.05$ indicates a good fit of the predictive model (well calibrated). Statistical analyses were performed using STATA (version 16) and R (version 3.4.4). The "glmnet," "Rms," and "survival" packages were used for variable selection, survival analysis, and building the nomogram. The R codes used for variable selection and nomogram development are available in Table S1. All tests were 2-sided, and a p-value of <0.05 was considered statistically significant. The reporting of this prognostic model study followed the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement [26].

3 | RESULTS

3.1 | Characteristics of patients

A total of 152 patients with IPF were included in the study. The majority of patients were elderly, males (59%), had a BMI above 24 kg/m^2 (78%), and were on supplemental oxygen (78%). Around one-third of the study population received anti-fibrotics (37%). The most common pulmonary comorbidities were emphysema (40%), pulmonary hypertension (36%), and sleep apnea (35%). Coronary artery disease (78%), hyperlipidemia (73%), and gastroesophageal reflux disease (70%) were the most common non-pulmonary comorbidities. Seven patients (46%) underwent lung transplantation. The patient characteristics are shown in Table 1.

3.2 | Prognostic data

At the end of the study follow-up, one-third (36.8%) of patients had died. The median survival was 5.8 years. The survival rates were 91% at 1 year, 78% at 3 years, and 68% at 5 years. LASSO regression was used for variable selection. The 10-fold cross-validation method was applied, and a model with excellent performance but a minimum number of variables was obtained when λ was 0.019 (Figure S2). The selected variables included DLco%, BMI, pulmonary hypertension, pulmonary embolism, and sleep apnea. The Cox

regression model was established based on parameters selected by LASSO regression (Table 2). The C-index was 0.71 (95% CI: 0.63–0.77) for OS.

3.3 | Nomogram development

We constructed the nomogram based on five independent variables: DLco%, BMI, pulmonary hypertension, pulmonary embolism, and sleep apnea (Figure 1). Each variable was mapped on a scale from 0 to 100, and the range of the total points was 0–350. Finally, the total points on the risk axis represent the 5-year survival probability. The calibration curve indicated that the predicted survival rates were consistent with the actual survival rates (Figure 2).

Risk scores were calculated based on the 5-year survival probability tertiles displayed by the nomogram. The median follow-up duration for the low-risk, intermediate-risk, and high-risk groups was 6.72, 6.11, and 5.56 years, respectively. Univariable Cox hazard regression revealed significant differences between risk groups OS (Figure 3). The C-index of the nomogram-based risk score was 0.71 (95% CI: 0.65–0.78) for OS, indicating stable and favorable performance of the model. Moreover, the Hosmer–Lemeshow test showed good concordance between predicted and actual observations ($p > 0.05$).

4 | DISCUSSION

With accurate risk estimation, patient management could be carried out more conveniently. Unfortunately, because IPF is remarkably heterogeneous regarding prognosis, the prediction of survival using current staging systems is imprecise [27]. In this study, we developed a nomogram and derived a risk-staging system using local inpatient data from the West Virginia University Medicine hospital systems. We identified DLco%, BMI (as a proxy for malnutrition), pulmonary hypertension, pulmonary embolism, and sleep apnea as independent prognostic factors for OS. These findings align with earlier research on risk factors for mortality among patients with IPF. Previous studies represented efforts to characterize the rural Appalachian IPF patient population, which is also uniquely challenged with the highest cigarette smoking rates nationally. While Sangani et al. identified BMI and pulmonary hypertension as significant prognostic predictors for the IPF cohort, the current study explored additional prognostic factors and examined their relative importance to each other using a nomogram [12, 13].

Several studies have suggested that a reduced DLco% is associated with higher death rates and worse outcomes in patients with IPF [28]. However, FVC% was not determined as an independent prognostic factor for inclusion in the nomogram. One plausible explanation could be its high collinearity with DLco%; LASSO regression selects one representative predictor from a set of highly correlated variables while shrinking the coefficients of the others. Another explanation could be the high representation of patients with emphysema in our sample (40%), which may alter the physiological characteristics of IPF because it mitigates the impact of fibrosis on ventilatory physiology and delays lung function decline [29]. Thus, FVC% may have limited predictive ability in the presence of emphysema, consistent with the results reported in the literature. For example, Kim et al. showed that low DLco% was independently associated with higher death rates in patients with combined

pulmonary fibrosis and emphysema, but FVC% was not [30]. In a previous study, DLco% changes over time were shown to be a better predictor of mortality than FVC% changes [31]. Remarkably, to the best of the authors' knowledge, there are no clinical trials in IPF where DLco% has been used as a marker of pulmonary function monitoring. Therefore, our findings support the inclusion of DLco% as an endpoint in clinical trials, which can potentially improve the precision of results.

The prognostic value of BMI in IPF has been investigated in several studies. Consistent with our findings, multiple studies have shown that a lower BMI is associated with worse outcomes and increased mortality in IPF [13, 32, 33]. This phenomenon is often referred to as the "obesity paradox" [34]. Low BMI can indicate an overall poor nutritional status [35]. Malnourished individuals have weaker immune systems, decreased overall functional capacity, and shorter survival times [36]. Both IPF itself and gastrointestinal-related adverse events associated with anti-fibrotic therapy can affect patients' nutritional status [37, 38]. Therefore, addressing malnutrition and optimizing nutritional status are important considerations in IPF management. Nutritional support is crucial in the comprehensive care of patients with lung cancer [39]. Given the resemblance with IPF, dietary counseling, and adequate nutritional support that could help improve BMI, muscle mass, functional capacity, quality-of-life, and potentially prolong survival appears warranted.

Among pulmonary comorbidities, pulmonary hypertension, pulmonary embolism, and sleep apnea were identified as independent prognostic factors for mortality, whereas emphysema and lung cancer were not. These findings align with previous studies. The impact of pulmonary hypertension on survival in patients with IPF is well characterized. Pulmonary hypertension complicates IPF [40], is associated with a worse prognosis [41], and is an independent predictor of mortality [42]. In fact, high pulmonary arterial pressure values at the initial evaluation of patients with IPF are an independent predictor of survival [43]. The prognostic value of pulmonary hypertension has been evaluated and was one of the comorbidities that improved GAP's all-cause mortality prediction in IPF [5]. Unfortunately, until recently, patients with IPF did not routinely undergo regular screening for pulmonary hypertension, partly because there was no approved pulmonary hypertension-targeted therapy for the IPF population [44]. However, with the approval of inhaled Treprostinil for pulmonary hypertension associated with intestinal lung disease in 2021 [45], routine screening for this comorbidity is a potential cornerstone in pulmonary fibrosis patient care.

Emphysema is one of the most common pulmonary comorbidities of IPF, presenting in 40% of our sample. Past analyses have shown conflicting results regarding the impact of emphysema on mortality. A previous single-center study demonstrated that emphysema in patients with IPF did not independently predict mortality once age, gender, smoking status, baseline severity using DLco%, and the extent of interstitial lung disease and emphysema had been considered [46]. Another registry-based study supported these findings [47]. However, some authors describe a negative impact on prognosis when IPF is associated with emphysema [48, 49]. In our analysis, we did not find a significant difference in OS in Patients with IPF with or without emphysema after adjusting for other pulmonary comorbidities. Also, consistent with our findings, previous studies have suggested that mortality in patients with concomitant lung cancer and IPF was no greater than in those

with IPF without lung cancer [50]. The lack of a prognostic impact of lung cancer in these patients confirms that IPF has a worse prognosis than most cancers. In fact, IPF has a worse prognosis than many forms of cancer [51]. However, more recent studies have argued that the mortality among patients with co-existing IPF and lung cancer is higher than that among patients with IPF alone [52–54]. In fact, the survival rates vary depending on the stage and type of lung cancer, which we did not account for in our study. However, there are currently no guidelines available for clinicians on how to manage lung cancer in IPF.

There are several strengths to our study. First, we accounted for survivor bias by creating a distinct group of patients with missing PFTs. As IPF progresses, missed spirometry visits promote survivor bias by raising the mean PFTs because missing values are associated with exacerbation of the condition or mortality among patients [55]. Second, we utilized LASSO regression for variable selection. LASSO regression is instrumental when there is a high degree of multicollinearity among predictors, such as DLco% and FVC %, as it tends to select one representative predictor from a set of highly correlated variables while shrinking the coefficients of the others. Also, as a regularization technique, it effectively helps prevent overfitting, improve interpretability, and simplify the model by eliminating irrelevant predictors. Third, the study has at least 10 events (death) per variable modeled, commonly recommended in survival analysis to maintain statistical power and reliability. This helps to ensure that the model is not overfitting the data and that the estimated hazard ratios are more likely to reflect true associations [56]. Finally, we evaluated model performance by assessing model discrimination and calibration, the two cardinal aspects of model fit [57].

We recognize the retrospective and the limited sample size as important limitations to our study. Also, our study focused on a cohort of locally based patients with IPF with distinct characteristics within this group [12, 13]. These included older age, high burden of comorbidities, diagnostic delay, limited access to specialized interstitial lung disease centers, and restricted utilization of anti-fibrotic therapy and lung transplantation [12, 13]. Also, the presence of rural health disparities and socioeconomic challenges impacts the eligibility of our patient population for advanced therapeutic options [12, 13]. It is important to note that these distinct characteristics may limit the generalizability of our findings to broader populations, but they hold significant value for understanding and addressing the specific needs of the Appalachian population. Additionally, it is important to highlight that the Appalachian population in our study has one of the highest smoking rates in the United States. Smoking is a well-known common risk factor for both IPF and emphysema. This high prevalence of smoking within the population likely contributes to the observed distinct characteristics and further emphasizes the importance of addressing these specific health challenges in the Appalachian region [12, 13]. Also, we did not precisely evaluate the extent and severity of the comorbidities or define the stage and the type of lung cancer. Additionally, we used BMI as a proxy for malnutrition, which could have underestimated malnutrition among patients with IPF [58]. Furthermore, this nomogram has not been tested in an independent population, and thus, its external validity remains to be established [59]. As a result, our findings should be considered as suggestive rather than definitive. Finally, some important prognostic factors, including non-pulmonary comorbidities and D-dimer levels, were not included due to the study's limited sample size. Further efforts on

prospective data collection, wider geographic recruitment, and incorporation of the extent and severity of comorbidity are encouraged to improve this model. More reliable methods of systematically screening patients with IPF for malnutrition are also warranted.

Another potential limitation is that we did not directly compare the performance of the developed nomogram-based scoring system with an established staging system, such as the GAP index. However, in comparison to the GAP index, which is widely used for staging and prognosis in similar patient populations, our developed model incorporates additional clinical predictors, such as pulmonary comorbidities, and utilizes a different modeling approach. Additionally, while the GAP index is developed to predict 3-year survival, our model focuses on OS. Also, the discriminative ability of our model as measured by the c-statistic is comparable to values reported in the literature, indicating similar predictive performance [2, 60, 61]. It is also important to acknowledge that a substantial proportion of the data for FVC and DLco, key predictors in our model, were missing. Although we have included all available data in this analysis by categorizing individuals with missing data as "unknown," enabling the model to borrow information from the complete observed data in the multivariable analysis, a nomogram based on a complete dataset may provide a more definitive model. However, real-world data often show high levels of missingness in PFTs due to difficulties in adhering to guidelines and the patients' ability to perform these tests as their condition worsens. Including patients with missing data on PFTS makes the nomogram more pragmatic and reflective of actual clinical scenarios. The former notwithstanding, future research could consider utilizing advanced imputation methods to address missing data more effectively, for example, multiple imputation or complete case analysis, provided the missing data follows a missing at random pattern. However, in our dataset, the data were missing not at random, and thus restricts the use of these approaches. Such methods, if applicable, would allow for more refined handling of incomplete datasets, preserving sample size, and potentially improving the model's robustness and predictive accuracy.

Our findings have significant implications for research and patient care, as physicians and patients could perform an individualized survival prediction through this easy-to-use scoring method. Identifying subgroups of patients at different risks of poor survival might impact treatment or care options. We believe that the nomogram represents a promising prognostic model as it is medically actionable, including manageable parameters that can be directly influenced or controlled through deliberate interventions or decisions to achieve desired outcomes.

5 | CONCLUSIONS

The present study confirmed the prognostic value of BMI and pulmonary comorbidities in IPF. Thus, routine screening and assessment of such parameters is fundamental in IPF care. The promising performance of the nomogram signifies its potential value in clinical practice and offers a practical approach to enhance the clinical management of IPF. However, further validation of the nomogram is essential to confirm its effectiveness in real-world clinical practice. Furthermore, recognizing the complexities of rural healthcare structures and the distinct health characteristics of West Virginia's population, we propose the adoption of a nomogram-based risk-staging system leveraging local patient cohort data. This tailored

approach may present a promising solution to address the inherent inaccuracies in survival prognostication for IPF.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

DATA AVAILABILITY STATEMENT

The data supporting the findings of this study are not publicly available due to restrictions on data sharing. For any questions or inquiries regarding the data, please contact the co-senior author Dr. Rahul G. Sangani.

Abbreviations:

BMI	body mass index
C-index	Harrell's concordance index
DLco%	carbon monoxide lung diffusion capacity percentage predicted
FVC%	forced vital capacity percentage predicted
GAP index	the gender-age-physiological index
HR	hazard ratio
ICD	International Classification of Diseases
IPF	idiopathic pulmonary fibrosis
LASSO	Least Absolute Shrinkage and Selection Operator
OS	overall survival
PFTs	pulmonary function tests
Ref	reference group

REFERENCES

[1]. O'Connell OJ, Egan JJ. The burden of disease and the need for a simple staging system in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2014;189(7):765–7. 10.1164/rccm.201402-0306ED [PubMed: 24684356]

[2]. Ley B, Ryerson CJ, Vittinghoff E, Ryu JH, Tomassetti S, Lee JS, et al. A multidimensional index and staging system for idiopathic pulmonary fibrosis. Ann Intern Med. 2012;156(10):684–91. 10.7326/0003-4819-156-10-201205150-00004 [PubMed: 22586007]

[3]. Caminati A, Lonati C, Cassandro R, Elia D, Pelosi G, Torre O, et al. Comorbidities in idiopathic pulmonary fibrosis: an underestimated issue. Eur Respir Rev. 2019;28(153):190044. 10.1183/16000617.0044-2019 [PubMed: 31578211]

[4]. Raghu G, Amatto VC, Behr J, Stowasser S. Comorbidities in idiopathic pulmonary fibrosis patients: a systematic literature review. Eur Respir J. 2015;46(4):1113–30. 10.1183/13993003.02316-2014 [PubMed: 26424523]

[5]. Torrisi SE, Ley B, Kreuter M, Wijsenbeek M, Vittinghoff E, Collard HR, et al. The added value of comorbidities in predicting survival in idiopathic pulmonary fibrosis: a multicentre observational study. Eur Respir J. 2019;53(3):1801587. 10.1183/13993003.01587-2018 [PubMed: 30578385]

[6]. Lacedonia D, De Pace CC, Rea G, Capitelli L, Gallo C, Scioscia G, et al. Machine learning and BMI improve the prognostic value of GAP index in treated IPF patients. Bioengineering. 2023;10(2):251. 10.3390/bioengineering10020251 [PubMed: 36829744]

[7]. Suzuki Y, Mori K, Aono Y, Kono M, Hasegawa H, Yokomura K, et al. Combined assessment of the GAP index and body mass index at antifibrotic therapy initiation for prognosis of idiopathic pulmonary fibrosis. Sci Rep. 2021;11(1):18579. 10.1038/s41598-021-98161-y [PubMed: 34535738]

[8]. Chun FKH, Briganti A, Karakiewicz PI, Graefen M. Should we use nomograms to predict outcome? Eur Urol Suppl. 2008;7(5):396–9. 10.1016/j.eursup.2008.01.011

[9]. Shariat SF, Karakiewicz PI, Godoy G, Lerner SP. Use of nomograms for predictions of outcome in patients with advanced bladder cancer. Ther Adv Urol. 2009;1(1):13–26. 10.1177/1756287209103923 [PubMed: 21789050]

[10]. Lee W, Lam SK, Zhang Y, Yang R, Cai J. Review of methodological workflow, interpretation and limitations of nomogram application in cancer study. Radiat Med Prot. 2022;3(4):200–7. 10.1016/j.radmp.2022.08.004

[11]. Raghu G, Remy-Jardin M, Myers JL, Richeldi L, Ryerson CJ, Lederer DJ, et al. Diagnosis of idiopathic pulmonary fibrosis. An official ATS/ERS/JRS/ALAT clinical practice guideline. Am J Respir Crit Care Med. 2018;198(5):e44–68. 10.1164/rccm.201807-1255st [PubMed: 30168753]

[12]. Sangani R, Ghio A, Culp S, Patel Z, Sharma S. Combined pulmonary fibrosis emphysema: role of cigarette smoking and pulmonary hypertension in a rural cohort. Int J Chronic Obstr Pulm Dis. 2021;16:1873–85. 10.2147/COPD.S307192

[13]. Sangani RG, Ghio AJ, Mujahid H, Patel Z, Catherman K, Wen S, et al. Outcomes of idiopathic pulmonary fibrosis improve with obesity: a rural Appalachian experience. South Med J. 2021;114(7):424–31. 10.14423/SMJ.0000000000001275 [PubMed: 34215896]

[14]. Yoon HY, Kim H, Bae Y, Song JW. Smoking status and clinical outcome in idiopathic pulmonary fibrosis: a nationwide study. Respir Res. 2024;25(1):191. 10.1186/s12931-024-02819-w [PubMed: 38685071]

[15]. Becklake MR, Lalloo U. The 'healthy smoker': a phenomenon of health selection? Respiration. 1990;57(3):137–44. 10.1159/000195837 [PubMed: 2274712]

[16]. van Walraven C, McCudden C, Austin PC. Imputing missing laboratory results may return erroneous values because they are not missing at random. J Clin Epidemiol. 2023;154:65–74. 10.1016/j.jclinepi.2022.12.011 [PubMed: 36528233]

[17]. Ross RK, Breskin A, Westreich D. When is a complete-case approach to missing data valid? The importance of effect-measure modification. Am J Epidemiol. 2020;189(12):1583–9. 10.1093/aje/kwaa124 [PubMed: 32601706]

[18]. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009;338(jun29 1):b2393. 10.1136/bmj.b2393

[19]. Behr J, Prasse A, Wirtz H, Koschel D, Pittrow D, Held M, et al. Survival and course of lung function in the presence or absence of antifibrotic treatment in patients with idiopathic pulmonary fibrosis: long-term results of the INSIGHTS-IPF registry. Eur Respir J. 2020;56(2):1902279. 10.1183/13993003.02279-2019 [PubMed: 32381492]

[20]. Fainberg HP, Oldham JM, Molyneau PL, Allen RJ, Kraven LM, Fahy WA, et al. Forced vital capacity trajectories in patients with idiopathic pulmonary fibrosis: a secondary analysis of a multicentre, prospective, observational cohort. Lancet Digit Health. 2022;4(12):e862–72. 10.1016/S2589-7500(22)00173-X [PubMed: 36333179]

[21]. Patridge EF, Bardyn TP. Research electronic data capture (REDCap). J Med Libr Assoc. 2018;106(1):142–4. 10.5195/JMLA.2018.319

[22]. Tibshirani R. Regression shrinkage and selection via the Lasso. J R Stat Soc Series B Stat Methodol. 1996;58(1):267–88. 10.1111/j.2517-6161.1996.tb02080.x

[23]. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika. 1994;81(3):515–26. 10.1093/biomet/81.3.515

[24]. Turkson AJ, Ayiah-Mensah F, Nimoh V. Handling censoring and censored data in survival analysis: a standalone systematic literature review. Int J Math Math Sci. 2021;2021:9307475. 10.1155/2021/9307475

[25]. Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. JAMA. 1982;247(18):2543–6. 10.1001/JAMA.1982.03320430047030 [PubMed: 7069920]

[26]. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Mac-Askill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162(1):W1–73. 10.7326/M14-0698 [PubMed: 25560730]

[27]. Ley B, Collard HR, King TE. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2011;183(4):431–40. 10.1164/rccm.201006-0894ci [PubMed: 20935110]

[28]. Hyldgaard C, Møller J, Bendstrup E. Changes in management of idiopathic pulmonary fibrosis: impact on disease severity and mortality. Eur Clin Respir J. 2020;7(1):1807682. 10.1080/20018525.2020.1807682 [PubMed: 32944203]

[29]. Yoon HY, Kim TH, Seo JB, Lee SM, Lim S, Lee HN, et al. Effects of emphysema on physiological and prognostic characteristics of lung function in idiopathic pulmonary fibrosis. Respirology. 2019;24(1):55–62. 10.1111/resp.13387 [PubMed: 30136753]

[30]. Kim YS, Jin GY, Chae KJ, Han YM, Chon SB, Lee YS, et al. Visually stratified CT honeycombing as a survival predictor in combined pulmonary fibrosis and emphysema. Br J Radiol. 2015;88(1055):20150545. 10.1259/bjr.20150545 [PubMed: 26388110]

[31]. Doubková M, Švancara J, Svoboda M, Šterclová M, Bartoš V, Pla ková M, et al. EMPIRE Registry, Czech Part: impact of demographics, pulmonary function and HRCT on survival and clinical course in idiopathic pulmonary fibrosis. Clin Res J. 2018;12(4):1526–35. 10.1111/crj.12700

[32]. Zinellu A, Carru C, Pirina P, Fois AG, Mangoni AA. A systematic review of the prognostic significance of the body mass index in idiopathic pulmonary fibrosis. J Clin Med. 2023;12(2):498. 10.3390/jcm12020498 [PubMed: 36675428]

[33]. Nishiyama O, Yamazaki R, Sano H, Iwanaga T, Higashimoto Y, Kume H, et al. Fat-free mass index predicts survival in patients with idiopathic pulmonary fibrosis. Respirology. 2017;22(3):480–5. 10.1111/resp.12941 [PubMed: 27868303]

[34]. Prado CM, Gonzalez MC, Heymsfield SB. Body composition phenotypes and obesity paradox. Curr Opin Clin Nutr Metab Care. 2015;18(6):535–51. 10.1097/MCO.0000000000000216 [PubMed: 26335310]

[35]. Jensen GL, Cederholm T, Correia MITD, Gonzalez MC, Fukushima R, Higashiguchi T, et al. GLIM criteria for the diagnosis of malnutrition: a consensus report from the global clinical nutrition community. JPEN - J Parenter Enter Nutr. 2019;43(1):32–40. 10.1002/jpen.1440

[36]. Shakersain B, Santoni G, Faxén-Irving G, Rizzuto D, Fratiglioni L, Xu W. Nutritional status and survival among old adults: an 11-year population-based longitudinal study. Eur J Clin Nutr. 2016;70(3):320–5. 10.1038/ejcn.2015.109 [PubMed: 26153193]

[37]. Richeldi L, du Bois RM, Raghu G, Azuma A, Brown KK, Costabel U, et al. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. N Engl J Med. 2014;370(22):2071–82. 10.1056/NEJMoa1402584 [PubMed: 24836310]

[38]. Jouneau S, Kerjouan M, Rousseau C, Lederlin M, Llamas-Guttierez F, De Latour B, et al. What are the best indicators to assess malnutrition in idiopathic pulmonary fibrosis patients? A cross-sectional study in a referral center. Nutrition. 2019;62:115–21. 10.1016/j.nut.2018.12.008 [PubMed: 30878815]

[39]. Kasprzyk A, Bilmin K, Chmielewska-Ignatowicz T, Pawlikowski J, Religioni U, Merks P. The role of nutritional support in malnourished patients with lung cancer. In Vivo. 2021;35(1):53–60. 10.21873/invivo.12231 [PubMed: 33402449]

[40]. Waxman AB, Elia D, Adir Y, Humbert M, Harari S. Recent advances in the management of pulmonary hypertension with interstitial lung disease. Eur Respir Rev. 2022;31(165):210220. 10.1183/16000617.0220-2021 [PubMed: 35831007]

[41]. Nikkho SM, Richter MJ, Shen E, Abman SH, Antoniou K, Chung J, et al. Clinical significance of pulmonary hypertension in interstitial lung disease: a consensus statement from the Pulmonary Vascular Research Institute's innovative drug development initiative-Group 3 pulmonary hypertension. Pulm Circ. 2022;12(3):e12127. 10.1002/pul2.12127 [PubMed: 36016668]

[42]. Patel NM, Lederer DJ, Borczuk AC, Kawut SM. Pulmonary hypertension in idiopathic pulmonary fibrosis. Chest. 2007;132(3):998–1006. 10.1378/chest.06-3087 [PubMed: 17873194]

[43]. Kimura M, Taniguchi H, Kondoh Y, Kimura T, Kataoka K, Nishiyama O, et al. Pulmonary hypertension as a prognostic indicator at the initial evaluation in idiopathic pulmonary fibrosis. Respiration. 2013;85(6):456–63. 10.1159/000345221 [PubMed: 23257350]

[44]. Olschewski H, Behr J, Bremer H, Claussen M, Douschan P, Halank M, et al. Pulmonary hypertension due to lung diseases: updated recommendations from the Cologne Consensus Conference 2018. Int J Cardiol. 2018;272S:63–8. 10.1016/j.ijcard.2018.08.043 [PubMed: 30131231]

[45]. Waxman A, Restrepo-Jaramillo R, Thenappan T, Ravichandran A, Engel P, Bajwa A, et al. Inhaled treprostinil in pulmonary hypertension due to interstitial lung disease. N Engl J Med. 2021;384(4):325–34. 10.1056/NEJMoa2008470 [PubMed: 33440084]

[46]. Jacob J, Bartholmai BJ, Rajagopalan S, Kokosi M, Maher TM, Nair A, et al. Functional and prognostic effects when emphysema complicates idiopathic pulmonary fibrosis. Eur Respir J. 2017;50(1):1700379. 10.1183/13993003.00379-2017 [PubMed: 28679612]

[47]. Kim HJ, Snyder LD, Neely ML, Hellkamp AS, Hotchkin DL, Morrison LD, et al. Clinical outcomes of patients with combined idiopathic pulmonary fibrosis and emphysema in the IPF-PRO Registry. Lung. 2022;200(1):21–9. 10.1007/s00408-021-00506-x [PubMed: 34997268]

[48]. Sugino K, Ishida F, Kikuchi N, Hirota N, Sano G, Sato K, et al. Comparison of clinical characteristics and prognostic factors of combined pulmonary fibrosis and emphysema versus idiopathic pulmonary fibrosis alone. Respirology. 2014;19(2):239–45. 10.1111/resp.12207 [PubMed: 25198924]

[49]. Zhang L, Zhang C, Dong F, Song Q, Chi F, Liu L, et al. Combined pulmonary fibrosis and emphysema: a retrospective analysis of clinical characteristics, treatment and prognosis. BMC Pulm Med. 2016;16(1):137. 10.1186/s12890-016-0300-7 [PubMed: 27809901]

[50]. Ozawa Y, Suda T, Naito T, Enomoto N, Hashimoto D, Fujisawa T, et al. Cumulative incidence of and predictive factors for lung cancer in IPF. Respirology. 2009;14(5):723–8. 10.1111/j.1440-1843.2009.01547.x [PubMed: 19659650]

[51]. Vancheri C, Failla M, Crimi N, Raghu G. Idiopathic pulmonary fibrosis: a disease with similarities and links to cancer biology. Eur Respir J. 2010;35(3):496–504. 10.1183/09031936.00077309 [PubMed: 20190329]

[52]. Karampitsakos T, Spagnolo P, Mogulkoc N, Wuyts WA, Tomassetti S, Bendstrup E, et al. Lung cancer in patients with idiopathic pulmonary fibrosis: a retrospective multicentre study in Europe. Respirology. 2023;28(1):56–65. 10.1111/resp.14363 [PubMed: 36117239]

[53]. Tomassetti S, Gurioli C, Ryu JH, Decker PA, Ravaglia C, Tantalocco P, et al. The impact of lung cancer on survival of idiopathic pulmonary fibrosis. Chest. 2015;147(1):157–64. 10.1378/chest.14-0359 [PubMed: 25166895]

[54]. Kim H, Yoo H, Pyo H, Ahn YC, Noh JM, Ju SG, et al. Impact of underlying pulmonary diseases on treatment outcomes in early-stage non-small cell lung cancer treated with definitive radiotherapy. Int J Chronic Obstr Pulm Dis. 2019;14:2273–81. 10.2147/COPD.S210759

[55]. Vasakova M, Sterclova M, Mo ulkoç N, Lewandowska K, Müller V, Hajkova M, et al. Healthy survivor bias in patients with idiopathic pulmonary fibrosis in clinical registries. Eur Respir J. 2021;58(Suppl 65):PA477. 10.1183/13993003.congress-2021.pa477

[56]. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. Am J Epidemiol. 2007;165(6):710–8. 10.1093/aje/kwk052 [PubMed: 17182981]

[57]. D'Agostino RB, Nam BH. Evaluation of the performance of survival analysis models: discrimination and calibration measures. In: Handbook of statistics. Amsterdam: Elsevier; 2003. p. 1–25. 10.1016/s0169-7161(03)23001-7

[58]. Topoluk N, Patel JJ. The implications of using BMI alone in outcomes research. Chest. 2022;162(4):e204. 10.1016/j.chest.2022.06.043

[59]. Bleeker SE, Moll HA, Steyerberg EW, Donders ART, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research: a clinical example. J Clin Epidemiol. 2003;56(9):826–32. 10.1016/s0895-4356(03)00207-5 [PubMed: 14505766]

[60]. Zhang X, Ren Y, Xie B, Wang S, Geng J, He X, et al. External validation of the GAP model in Chinese patients with idiopathic pulmonary fibrosis. Clin Res J. 2023;17(9):831–40. 10.1111/crj.13564

[61]. Lee SH, Kim SY, Kim DS, Kim YW, Chung MP, Uh ST, et al. . Predicting survival of patients with idiopathic pulmonary fibrosis using GAP score: a nationwide cohort study. Respir Res. 2016;17(1):131. 10.1186/s12931-016-0454-0 [PubMed: 27756398]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript
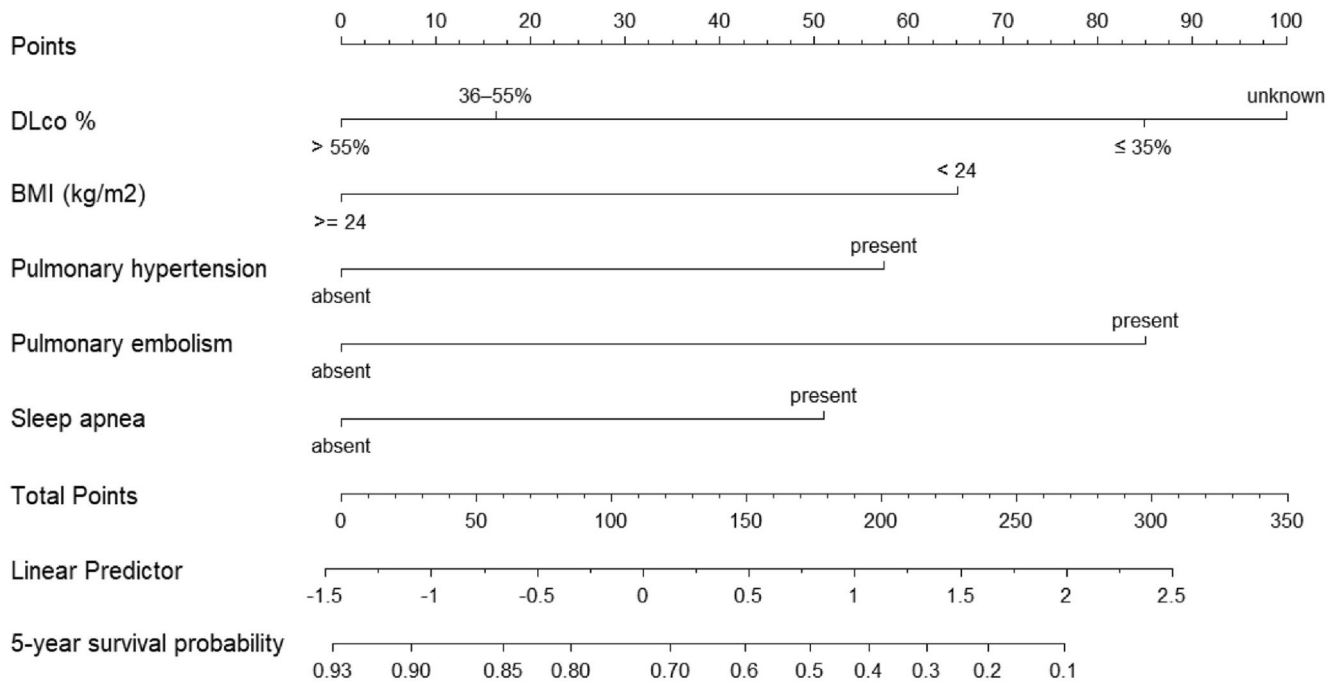
FIGURE 1.
Nomogram used to predict 5-year survival probability in patients with IPF. The usage of the nomogram is illustrated in a hypothetical patient with DLco% of 49%, BMI of 26, and pulmonary embolism. According to the nomogram, points for DLco%, BMI, and pulmonary embolism were 17, 0, and 87, respectively. The total points added up to 104 for this patient, representing approximately 74% 5-year survival probability. BMI, body mass index; DLco%, carbon monoxide lung diffusion capacity percentage predicted; IPF, idiopathic pulmonary fibrosis.
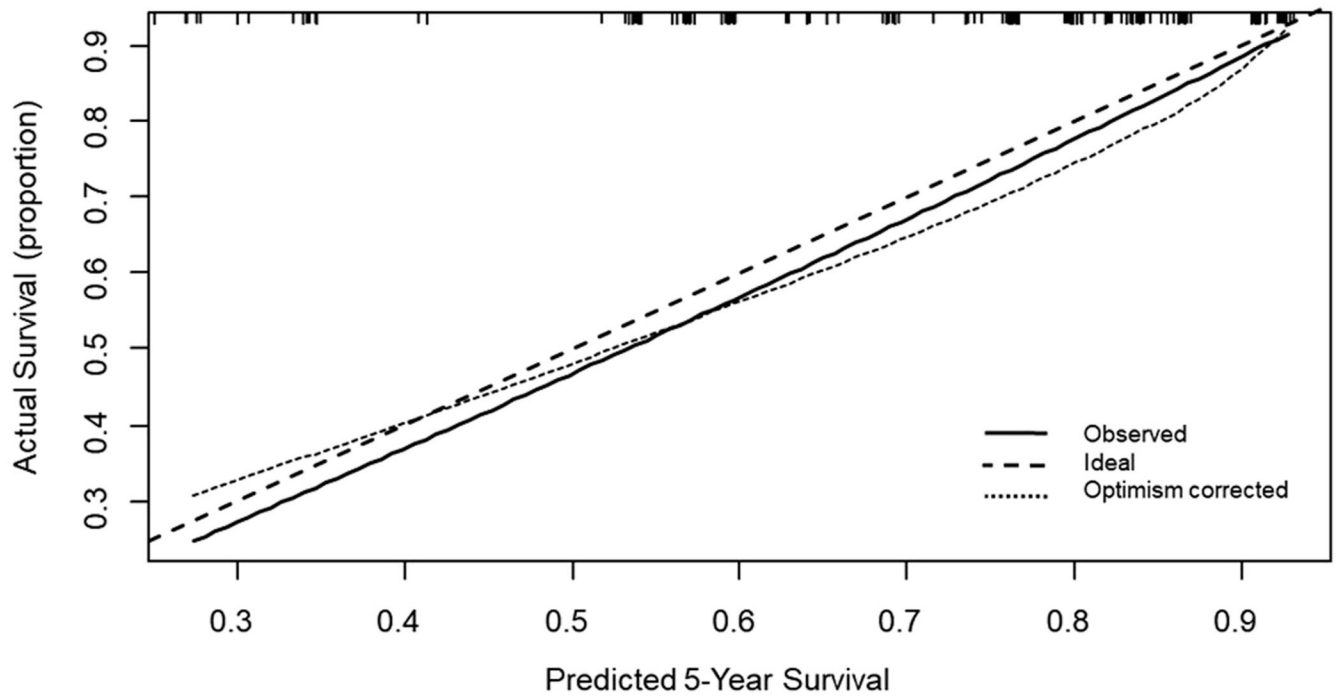
FIGURE 2.
Calibration curve of overall survival at 5 years. The nomogram-predicted probability of survival is plotted on the *x*-axis, and the actual survival is plotted on the *y*-axis. Dashed lines along the 45-degree line through the point of origin represent the perfect calibration models where the predicted probabilities are identical to the actual probabilities.
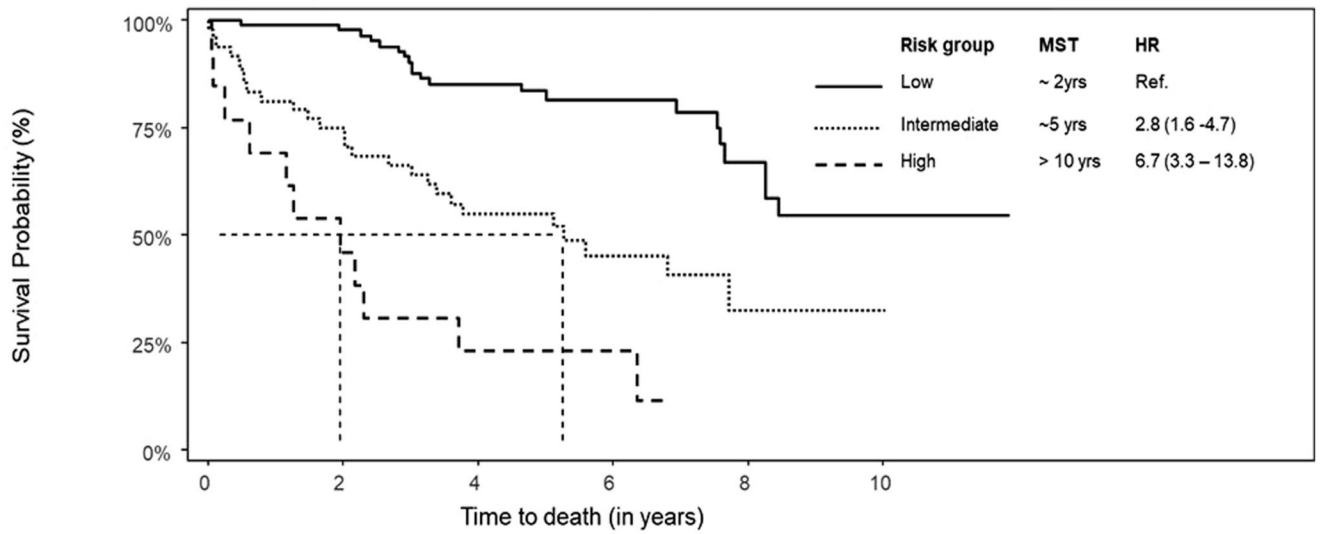
FIGURE 3.

Kaplan–Meier survival curves of a nomogram-based scoring system. One hundred and fifty-two patients were categorized into three prognostic groups according to the nomogram probability tertiles: patients with >66.6% 5-year survival probability as "low-risk group," 33.3%–66.6% 5-year survival probability as "moderate risk group" and <33.3% 5-year survival probability as "high-risk group." HR, hazard ratio; MST, median survival time; Ref, reference group.

TABLE 1

Demographic characteristics.

	Total (N = 152)	Alive (N = 87)	Deceased (N = 65)
Gender			
Female	63 (41.4%)	36 (41.4%)	27 (41.5%)
Male	89 (58.6%)	51 (58.6%)	38 (58.5%)
Age			
Mean (SD)	75.7 (9.97)	75.1 (9.53)	76.5 (10.5)
Median [min, max]	77.0 [49.0, 99.0]	76.0 [51.0, 95.0]	78.0 [49.0, 99.0]
BMI			
24	34 (22.4%)	15 (17.2%)	19 (29.2%)
>24	118 (77.6%)	72 (82.8%)	46 (70.8%)
FVC%			
>75%	48 (31.6%)	32 (36.8%)	16 (24.6%)
50%–75%	53 (34.9%)	30 (34.5%)	23 (35.4%)
<50%	12 (7.9%)	4 (4.6%)	8 (12.3%)
Unknown/missing	39 (25.7%)	21 (24.1%)	18 (27.7%)
DLco%			
>55%	30 (19.7%)	21 (24.1%)	9 (13.8%)
36%–55%	46 (30.3%)	30 (34.5%)	16 (24.6%)
35%	28 (18.4%)	12 (13.8%)	16 (24.6%)
Unknown/missing	48 (31.6%)	24 (27.6%)	24 (36.9%)
Pulmonary comorbidities			
Pulmonary hypertension	54 (35.5%)	20 (23.0%)	34 (52.3%)
Pulmonary embolism	12 (7.9%)	3 (3.4%)	9 (13.8%)
Lung cancer	7 (4.6%)	4 (4.6%)	3 (4.6%)
Sleep apnea	53 (34.9%)	27 (31.0%)	26 (40.0%)
Emphysema	60 (39.5%)	32 (36.8%)	28 (43.1%)

	Total (N = 152)	Alive (N = 87)	Deceased (N = 65)
Non-pulmonary comorbidities			
GERD	107 (70.4%)	61 (70.1%)	46 (70.8%)
Diabetes	52 (34.2%)	30 (34.5%)	22 (33.8%)
Hyperlipidemia	111 (73.0%)	63 (72.4%)	48 (73.8%)
CAD	82 (53.9%)	44 (50.6%)	38 (58.5%)
Systemic hypertension	118 (77.6%)	64 (73.6%)	54 (83.1%)
CKD	37 (24.3%)	11 (12.6%)	26 (40.0%)
Smoking status			
Never	46 (30.3%)	27 (31.0%)	19 (29.2%)
Former	94 (61.8%)	50 (57.5%)	44 (67.7%)
Current	12 (7.9%)	10 (11.5%)	2 (3.1%)
Antifibrotic use	57 (37.5%)	34 (39.1%)	23 (35.4%)
Lung transplant	7 (4.6%)	7 (8.0%)	0 (0%)
Oxygen use	119 (78.3%)	62 (71.3%)	57 (87.7%)

Abbreviations: BMI, body mass index; CAD, coronary artery disease; CKD, chronic kidney disease; DLco%, carbon monoxide lung diffusion capacity percentage predicted; FVC%, forced vital capacity percentage predicted; GERD, gastroesophageal reflux disease.

TABLE 2

Cox proportional hazard regression for predicting OS among patients with IPF.

Variables	Unadjusted		Adjusted	
	HR (95% CI)	*p*-value	HR (95% CI)	*p*-value
DLco%				
>55%	Ref		Ref	
36%–55%	1.18 (0.46–2.99)	0.73	1.23 (0.48–3.15)	0.66
35%	3.08 (0.25–7.58)	0.014	2.95 (1.17–7.44)	0.022
Unknown	3.25 (1.39–7.62)	0.007	3.58 (1.48–8.66)	0.005
BMI				
>24	Ref		Ref	
24	2.06 (1.19–3.56)	0.01	2.29 (1.27–4.17)	0.006
Pulmonary hypertension				
Absent	Ref		Ref	
Present	2.39 (1.43–4.02)	0.001	2.08 (1.22–3.55)	0.007
Pulmonary embolism				
Absent				
Present	2.46 (1.16–5.21)	0.019	2.96 (1.32–6.59)	0.008
Sleep apnea				
Absent	Ref		Ref	
Present	1.19 (0.71–2.03)	0.49	1.92 (1.09–3.39)	0.025

Abbreviations: BMI, body mass index; DLco, carbon monoxide lung diffusion capacity; HR, hazard ratio; IPF, idiopathic pulmonary fibrosis; OS, overall survival.