

An Exploratory LASSO–Cox Model for Mortality Risk in Hospitalized Patients With Idiopathic Pulmonary Fibrosis and Type 2 Diabetes: A Single-Center Retrospective Study

Zhuo-Jing Wu

The Second Affiliated Hospital of Zhengzhou University

Wei-Wei Gao

13598832393@163.com

The Second Affiliated Hospital of Zhengzhou University

Research Article

Keywords: Idiopathic pulmonary fibrosis, Type 2 diabetes mellitus, Prognosis, LASSO–Cox regression, Survival analysis, Nomogram, Risk prediction, Inflammation, Internal validation

Posted Date: October 5th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10852999/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

**An Exploratory LASSO–Cox Model for Mortality Risk in Hospitalized Patients With
Idiopathic Pulmonary Fibrosis and Type 2 Diabetes: A Single-Center Retrospective Study**

Zhuo-Jing Wu¹, Wei-Wei Gao^{2*}

*Corresponding author:

Address:

Wei-Wei Gao

The Second Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, CT: 450014

13598832393@163.com

Abstract

Background: Idiopathic pulmonary fibrosis (IPF) is a progressive fibrotic interstitial lung disease characterized by poor prognosis and substantial clinical heterogeneity. Type 2 diabetes mellitus (T2DM) is a common comorbidity in IPF and may exacerbate outcomes through metabolic dysfunction, chronic low-grade inflammation, and microvascular injury. However, prognostic tools specifically designed for patients with both IPF and T2DM remain limited.

Methods: We conducted a single-center retrospective cohort study of consecutive adult inpatients with physician-diagnosed IPF and T2DM at the Second Affiliated Hospital of Zhengzhou University between December 20, 2017 and January 20, 2023. The primary outcome was overall survival (OS), defined as the time from the baseline inpatient assessment (when predictors were collected) to death or last follow-up (January 20, 2026) . Candidate predictors included demographics, vital signs, pulmonary function, HRCT fibrosis grading, inflammatory markers, metabolic indices, and routine

laboratory variables. Variables with more than 20% missingness were excluded, and missing data in the remaining variables were handled by multiple imputation. Predictor selection was performed via LASSO–Cox regression with 10-fold cross-validation, followed by multivariable Cox modeling. Internal validation was performed using 1000 bootstrap resamples. Model performance was assessed using Harrell's C-index, time-dependent area under the curve (AUC), calibration plots, and decision-curve analysis (DCA). Reporting followed TRIPOD.

Results: Of 196 screened patients, 102 were included in the final analysis, of whom 74 died during follow-up. The LASSO procedure identified a parsimonious five-predictor model comprising age, pulmonary function grading, HRCT fibrosis grading, C-reactive protein (CRP), and interleukin-6 (IL-6). In the multivariable Cox model, age (HR 1.05, 95% CI 1.02–1.08; $P<0.001$), pulmonary function grading (HR 1.72, 95% CI 1.42–2.08; $P<0.001$), and HRCT fibrosis grading (HR 1.99, 95% CI 1.28–3.09; $P=0.002$) were independently associated with mortality. Although CRP and IL-6 were retained by the penalized selection process, they did not reach statistical significance in the multivariable model. The apparent C-index was 0.822 (SE 0.022), and the optimism-corrected C-index after 1000 bootstraps was 0.809 (95% CI 0.790–0.828). Time-dependent AUCs were 0.882 (95% CI 0.814–0.951) at 1 year and 0.856 (95% CI 0.778–0.923) at 3 years. Calibration plots demonstrated reasonable agreement between predicted and observed survival probabilities. DCA suggested potential net benefit across a range of threshold probabilities within this internally validated dataset.

Conclusions: In hospitalized adults with coexisting IPF and T2DM, older age, worse pulmonary function, and greater HRCT fibrosis burden were independently associated with higher mortality. An exploratory LASSO–Cox model based on routine clinical variables showed good internal

performance, however it should be regarded as hypothesis-generating, and external validation is required before clinical application.

Keywords: Idiopathic pulmonary fibrosis; Type 2 diabetes mellitus; Prognosis; LASSO–Cox regression; Survival analysis; Nomogram; Risk prediction; Inflammation; Internal validation

Introduction

IPF is a progressive fibrotic interstitial lung disease characterized by a poor prognosis and highly heterogeneous clinical trajectories. Despite the availability of contemporary antifibrotic therapy, median survival remains limited for many patients, and the clinical course varies widely from rapid progression to relative stability over several years [1, 2].

T2DM is a common comorbidity in IPF cohorts, with reported prevalence ranging from 16% to 40% in various registry and observational studies [3, 4]. Beyond its role as a concurrent condition, T2DM may actively influence the course of IPF through multiple interconnected pathways. Chronic low-grade inflammation, oxidative stress, metabolic reprogramming, and microvascular injury—hallmarks of T2DM—may aggravate alveolar epithelial injury, promote fibroblast activation, and accelerate fibrogenesis [5]. Recent evidence further suggests that diabetes-related microangiopathy can affect pulmonary capillary networks, potentially impairing gas exchange and contributing to disease progression [6].

Although comorbidity burden is known to influence outcomes in IPF [7, 8], most existing prognostic studies and prediction models have been developed in unselected IPF populations and have treated diabetes as a single covariate rather than as a defining characteristic of a distinct subgroup. T2DM may influence prognosis through pathways—including systemic inflammation, insulin resistance,

and metabolic derangement—that are not fully captured by conventional IPF prognostic indices. Whether and how routinely available clinical variables can stratify mortality risk specifically among patients with coexisting IPF and T2DM has not been well explored. Moreover, existing IPF prediction models were developed in populations with relatively low diabetes prevalence and may not perform optimally in this metabolically distinct subgroup. We therefore conducted a single-center, retrospective study in hospitalized adults with IPF and T2DM. Our aims were to: (1) evaluate the associations between clinical, functional, radiological, and inflammatory variables and overall survival; and (2) develop an exploratory LASSO–Cox-based prognostic model tailored to this specific population.

Methods

Study design and setting

This single-center retrospective cohort study was conducted at the Department of Respiratory and Critical Care Medicine, the Second Affiliated Hospital of Zhengzhou University. Consecutive hospitalized adults with documented diagnoses of both IPF and T2DM between December 20, 2017, and January 20, 2023, were screened for eligibility. Patients were follow-up until January 20, 2026. The study was approved by the institutional ethics committee (KY2025495). Owing to the retrospective nature of the analysis using de-identified data, the requirement for individual informed consent was waived in accordance with institutional policy.

Participants

Eligible patients were adults aged ≥ 18 years with a clinical and radiological diagnosis of IPF according to the contemporary ATS/ERS/JRS/ALAT guideline (2022) [1] and a diagnosis of

T2DM based on the Chinese national diabetes guideline [9]. All included patients were hospitalized in the Department of Respiratory and Critical Care Medicine and had available baseline clinical data from the index hospitalization, as well as documented follow-up for survival status.

We excluded patients with secondary interstitial lung diseases, including connective tissue disease-associated ILD, hypersensitivity pneumonitis, pneumoconioses, and drug-induced lung injury.

Additionally, patients with major co-existing pulmonary disorders—such as chronic obstructive pulmonary disease or acute pulmonary embolism—were excluded to avoid confounding of lung function assessments and disease progression. Furthermore, to reduce confounding from non-IPF-related mortality, we excluded individuals with active malignancies or severe organ failure (e.g., decompensated cirrhosis or dialysis-dependent renal failure). Finally, patients with missing survival outcomes or >20% missingness in key predictor variables were excluded. Antifibrotic therapy was not systematically captured in this retrospective dataset, and its potential impact on survival is acknowledged as a limitation. Of the 196 screened patients, 102 met all eligibility criteria and were included in the final analysis.

Outcome and index date

The primary outcome was overall survival (OS). To align predictor measurement with time zero, the index date for survival analysis was defined as the date of the baseline inpatient assessment at which predictor variables were collected. For patients in this cohort, the baseline inpatient assessment typically occurred at or shortly after the time of combined diagnosis; thus, this index date aligns closely with the clinical diagnosis date used in the original study design. OS was calculated from this index date to death from any cause or to censoring at the last follow-up. Vital status was ascertained from hospital records and structured telephone follow-up. Given the objective

nature of this outcome, blinding was not applicable

Predictors

All predictors were obtained from the baseline inpatient assessment performed during the index hospitalization. We collected demographic and lifestyle variables, including age, sex, and smoking history, along with admission vital signs, including systolic and diastolic blood pressure, heart rate, and respiratory rate.

Pulmonary function variables included forced vital capacity as percent predicted (FVC%pred) and diffusing capacity for carbon monoxide as percent predicted (DLCO%pred). To reduce collinearity and improve clinical interpretability, these measures were summarized as a prespecified ordinal pulmonary function grading (mild, moderate, or severe) based on established thresholds reflecting both restrictive impairment and gas-exchange limitation [1].

High-resolution computed tomography (HRCT) features of fibrosis were evaluated based on the ATS/ERS/JRS/ALAT clinical practice guidelines (2022). The core features of the usual interstitial pneumonia (UIP) pattern, including reticulation, traction bronchiectasis, and honeycombing (the most specific diagnostic hallmark of UIP), were systematically assessed. The extent of fibrotic lung involvement was visually estimated and dichotomized utilizing a threshold of $< 20\%$ versus $\geq 20\%$, adapted from the Goh method [10]. To ensure diagnostic accuracy and reproducibility, all HRCT scans were independently reviewed by two senior thoracic radiologists (each with an associate chief title) blinded to the clinical outcomes, with discrepancies resolved through consensus. Based on the core fibrotic features and the volumetric extent of involvement, a three-level ordinal grading system was established as follows: Mild (presence of reticulation or traction bronchiectasis, without honeycombing); Moderate (localized honeycombing, affecting $< 20\%$ of the lung); or Severe

(extensive honeycombing, with a total fibrotic extent of $\geq 20\%$). Baseline Laboratory Measurements All laboratory variables were consistently obtained from the index admission during which patients were diagnosed with both IPF and T2DM, reflecting their baseline status prior to any therapeutic intervention. The inflammatory markers panel included CRP, IL-6, and erythrocyte sedimentation rate (ESR). Metabolic parameters included fasting or random plasma glucose, triglycerides (TG), total cholesterol (TC), and high-density lipoprotein cholesterol (HDL-C). Standard hematologic and biochemical profiles included white blood cell count (WBC), red blood cell (RBC), hemoglobin (Hb), serum creatinine (Cr), sodium (Na^+), potassium (K^+), and alanine aminotransferase (ALT). Additionally, semiquantitative urinalysis data for glucose and protein were recorded.

Data management and missing data

Variables with more than 20% missing values were excluded a priori. For the remaining variables, missing data were handled using multiple imputation with chained equations in SPSS version 26.0, generating five imputed datasets under an assumed missing-at-random mechanism. The imputed data were subsequently used for model-development analyses. Because the *glmnet* package does not natively support Rubin's pooling rules for penalized regression, a stacked imputation approach was employed: the five imputed datasets were combined vertically with case weights of 1/5, and LASSO-Cox variable selection was performed on this stacked dataset. The selected variables were then fitted using standard Cox regression on each imputed dataset separately, and the resulting hazard ratios and standard errors were pooled using Rubin's rules. Continuous variables were inspected for distributional properties, and categorical variables were coded as binary or ordinal variables as appropriate. Collinearity was assessed using variance inflation factors (VIFs), with VIF

< 5 considered acceptable. Aspartate aminotransferase (AST) was excluded because its VIF (5.002) slightly exceeded the prespecified threshold.

Sample size considerations

A total of 74 deaths occurred among the 102 included patients. Given that the number of initial candidate predictors was relatively large compared with the available number of events, standard maximum likelihood estimation would carry a high risk of overfitting and selection bias. To strictly mitigate this risk, we utilized the least absolute shrinkage and selection operator (LASSO) for variable selection, which actively shrinks coefficients and minimizes overfitting during the exploratory phase [11, 12]. To ensure a parsimonious and clinically interpretable final multivariable model, the number of sustained degrees of freedom was strictly constrained. The final model retained approximately 5 predictors, yielding a post-selection events-per-variable (EPV) ratio of approximately 15 (74 events / 5 retained predictors). The initial candidate predictor set comprised approximately 20 variables, corresponding to a candidate-stage EPV of approximately 3.7; LASSO penalization was employed specifically to address this low candidate-stage EPV. This strategy aligns with methodological recommendations for sustaining robust statistical power and model stability in smaller event-count cohorts [13].

Statistical analysis

Baseline characteristics were summarized as mean $\pm$ standard deviation (SD) for approximately normally distributed continuous variables, median with interquartile range (IQR) for skewed continuous variables, and frequency (percentage) for categorical variables. Comparisons between survivors and non-survivors were performed using the independent-samples T-test or Wilcoxon rank-sum test for continuous variables and the chi-square test or Fisher's exact test for categorical

variables, as appropriate.

Univariate Cox proportional hazards models were fitted to describe crude associations between candidate predictors and OS. These analyses were used to characterize the data but were not used as the sole criterion for final predictor selection.

Predictor selection was performed using LASSO Cox regression in R version 4.5.2 with the glmnet package. The full prespecified candidate predictor set was entered into the penalized model, and the tuning parameter (λ) was examined via 10-fold cross-validation. The λ value minimizing the cross-validated partial likelihood deviance retained 15 predictors. Because this number was considered too large relative to the available number of events and the intended clinical interpretability of the model, we selected a more strongly penalized and more parsimonious solution from the cross-validation path that retained five predictors (post hoc criterion: the smallest lambda along the cross-validation path yielding a parsimonious model with EPV ≥ 10): age, pulmonary function grading, HRCT fibrosis grading, CRP, and IL-6. For transparency, the lambda.1se solution was also evaluated, and its results are reported in the sensitivity analysis[14]. These variables were then entered into a multivariable Cox proportional hazards model, and results were reported as hazard ratios (HRs) with 95% confidence intervals (CIs). The proportional hazards assumption was assessed via Schoenfeld residuals for each covariate and for the global model[15].

Internal validation and model performance

Internal validation was performed using 1000 bootstrap resamples with replacement to estimate optimism-corrected model performance[16]. Discrimination was assessed using Harrell's C-index, reported as both apparent and optimism-corrected values. Time-dependent receiver operating characteristic (ROC) analysis was used to estimate AUC values at 1 and 3 years, with 95% bootstrap

199 CIs. Calibration was evaluated visually using bootstrap-corrected calibration plots at 1 and 3 years.
200 DCA was used as an exploratory assessment of potential clinical usefulness within the development
201 dataset. A nomogram was constructed to provide a graphical representation of the model and is
202 presented for illustration rather than direct clinical deployment. Reporting followed the TRIPOD
203 statement [17]; the completed checklist is available as supplementary material.

204 **Sensitivity analysis**

205 Given that CRP and IL-6 were not statistically significant in the multivariable Cox model, we
206 performed a post hoc sensitivity analysis using a reduced three-predictor model including only age,
207 pulmonary function grading, and HRCT fibrosis grading. Discrimination and DCA results were
208 compared qualitatively between the full five-predictor model and the reduced model to assess
209 whether the inflammatory markers provided incremental prognostic information.

211 **Results**

212 **Study population**

213 Of the 196 screened patients with documented IPF and T2DM, 102 met the inclusion criteria and
214 were included in the final analysis (Figure 1). During follow-up, 74 patients died and 28 survived .
215 The mean age of the cohort was 67.3 years (SD 10.9), and 65 patients (63.7%) were male. Median
216 follow-up duration was 15.2 months (IQR 6.8–36.5). As expected for an inpatient cohort, the event
217 rate was high (72.5%), indicating a relatively severe case mix.

218 **Baseline characteristics**

219 Baseline characteristics stratified by survival status are summarized in Table 1. Compared with
220 survivors, non-survivors were older (69.97 vs 60.25 years; $P<0.001$) and had higher CRP (median

13.04 vs 1.92 mg/L; $P<0.001$), higher IL-6 (median 16.54 vs 5.64 pg/mL; $P=0.005$), and higher ESR (median 45.50 vs 16.00 mm/h ; $P<0.001$). Non-survivors also had lower RBC counts (4.19 vs $4.52 \times 10^{12}/L$; $P=0.003$), lower hemoglobin levels (123.77 vs 136.89 g/L; $P=0.003$), lower serum sodium concentrations (137.90 vs 140.36 mmol/L $P=0.028$), and lower triglyceride levels (1.26 vs 2.12 mmol/L $P=0.001$). In addition, non-survivors had more severe impairment in FVC%pred and DLCO%pred (both $P<0.001$). Several other routine clinical and laboratory variables did not differ significantly between the two groups.

Univariate associations with mortality

Univariate Cox regression analyses identified 13 variables associated with mortality (Table 2). These included age (HR 1.06, 95% CI 1.03–1.09 ; $P<0.001$), HRCT fibrosis grading (moderate vs mild: HR 1.87, 95% CI 1.16–3.02 ; $P=0.011$; severe vs mild: HR 5.02, 95% CI 1.91–13.21; $P=0.001$), pulmonary function grading (moderate vs mild: HR 2.79, 95% CI 1.46–5.32; $P=0.002$; severe vs mild: HR 7.72, 95% CI 3.92–15.21; $P<0.001$), CRP (HR 1.01, 95% CI 1.01–1.02; $P<0.001$), IL-6 (HR 1.02, 95% CI 1.01–1.02; $P<0.001$), ESR (HR 1.01, 95% CI 1.01–1.02; $P<0.001$), glucose (HR 1.16, 95% CI 1.00–1.34; $P=0.045$), WBC (HR 1.10, 95% CI 1.03–1.17; $P=0.004$), RBC (HR 0.57, 95% CI 0.40–0.81; $P=0.002$), hemoglobin (HR 0.98, 95% CI 0.97–0.99; $P=0.001$), creatinine (HR 1.01, 95% CI 1.00–1.02; $P=0.023$), serum sodium (HR 0.88, 95% CI 0.83–0.93; $P<0.001$), and triglycerides (HR 0.68, 95% CI 0.51–0.91; $P=0.010$).

Predictor selection and multivariable model

LASSO–Cox regression with 10-fold cross-validation identified a parsimonious five-predictor solution comprising age, pulmonary function grading, HRCT fibrosis grading, CRP, and IL-6 (Figure 2). Collinearity among these final predictors was low, with all VIF values <2 .

In the multivariable Cox model (Table 3), age (HR 1.05, 95% CI 1.02–1.08; $P<0.001$), pulmonary function grading (HR 1.72, 95% CI 1.42–2.08; $P<0.001$), and HRCT fibrosis grading (HR 1.99, 95% CI 1.28–3.09; $P=0.002$) were independently associated with higher mortality. CRP ($P=0.056$) and IL-6 ($P=0.199$) did not retain conventional statistical significance after multivariable adjustment, but retained in the final model because they were selected by penalized regression and appeared to provide modest incremental prognostic information in sensitivity analyses. Schoenfeld residual testing did not suggest violation of the proportional hazards assumption for the individual covariates (all $P>0.05$) or for the global model (global $\chi^2=7.95$, $df=5$, $P=0.133$).

Model performance and internal validation

The apparent Harrell's C-index of the final model was 0.822 (SE 0.022). After 1000 bootstrap resamples, the optimism-corrected C-index was 0.809 (95% CI 0.790–0.828), indicating good internal discrimination.

Time-dependent ROC analysis showed AUC values of 0.882 (95% CI 0.814–0.951) at 1 year and 0.856 (95% CI 0.778–0.923) at 3 years. The time-dependent AUCs at 1 and 3 years were higher than the overall C-index, which is expected in this high-event-rates cohort: the C-index measures concordance across all time points, whereas time-dependent AUCs reflect discrimination within specific time windows where events are concentrated. (Figure 3). Bootstrap-corrected calibration plots demonstrated reasonable agreement between predicted and observed survival probabilities at both 1 and 3 years (Figure 4).

Decision-curve analysis suggested potential net benefit of the model compared with treat-all or treat-none strategies across approximately 5%–70% threshold probabilities at 1 year and 10%–80% threshold probabilities at 3 years within this internally validated dataset (Figure 5). A nomogram

representing the five-predictor model is shown in Supplementary Figure S1 for illustrative purposes; the nomogram facilitates visual appreciation of the relative contribution of each predictor to the overall risk estimate.

Sensitivity analysis

To evaluate the robustness of the final prediction model, we performed internal validation using 1000 bootstrap resamples. The final LASSO-selected Cox model comprised five predictors: age, pulmonary function grading, HRCT fibrosis grading, CRP, and IL-6. Although CRP and IL-6 did not reach statistical significance in the multivariable Cox model, they were retained because they were selected by the LASSO procedure and are biologically plausible inflammatory markers in both IPF and T2DM. The apparent Harrell's C-index of the five-predictor model was 0.822 (SE 0.022), and the optimism-corrected C-index after bootstrap validation was 0.809 (95% CI 0.790–0.828), indicating good discrimination. The time-dependent AUCs were 0.882 (95% CI 0.814–0.951) at 1 year and 0.856 (95% CI 0.778–0.923) at 3 years. Calibration curves showed good agreement between predicted and observed survival probabilities at both time points. Decision-curve analysis also suggested positive net benefit across clinically relevant threshold probability ranges. These findings support the internal stability and potential clinical utility of the five-predictor model; however, given the retrospective single-center design and limited sample size, external validation is required before clinical application.

Discussion

In this single-center cohort of hospitalized adults with IPF and T2DM, older age, worse pulmonary function, and more extensive HRCT fibrosis were independently associated with mortality. A five-

287 predictor model incorporating these variables along with CRP and IL-6 showed good internal
288 discrimination and calibration. However, given the retrospective design, modest sample size, and
289 absence of external validation, these findings should be interpreted as hypothesis-generating.

290 Our findings reinforce the concept that, even in a metabolically distinct subgroup of patients with
291 IPF, the major drivers of prognosis remain the severity of physiologic impairment and the extent of
292 structural fibrotic lung damage. Age, pulmonary function, and HRCT fibrosis burden are well-
293 established prognostic markers in unselected IPF populations [18–20]. The present study extends
294 those observations to patients with concomitant T2DM and suggests that diabetes does not displace
295 the central prognostic importance of baseline lung disease severity. In practical terms, these results
296 support continued emphasis on pulmonary function testing and careful radiologic assessment when
297 evaluating risk in this comorbid population.

298 The role of inflammatory markers in the final prognostic model warrants nuanced interpretation.
299 Although CRP and IL-6 were selected by the penalized regression procedure and modestly enhanced
300 overall model performance, neither reached conventional statistical significance within the final
301 multivariable Cox model. This phenomenon likely indicates that circulating inflammatory markers
302 capture shared variance and prognostic information already deeply accounted for by pulmonary
303 function impairment and radiologic fibrotic severity. Nonetheless, their clinical consideration
304 remains biologically compelling. IPF and T2DM share common pathophysiological trajectories
305 governed by chronic low-grade inflammation, oxidative stress, accelerated epithelial injury, and
306 aberrant tissue repair [21–24]. Recent Mendelian randomization evidence suggests that genetically
307 predicted higher CRP levels are associated with increased IPF susceptibility [25], whereas elevated
308 IL-6 drives functional decline across various fibrotic interstitial lung diseases [26]. Within this

cohort, CRP and IL-6 are best viewed as adjunctive biological indicators that enrich the clinical picture, rather than as stand-alone, independent prognostic determinants. Interpretation of univariate predictors and variable selection logic: several additional variables exhibited significant associations with mortality in univariate analyses but were not retained in the final prognostic tool. Specifically, lower triglyceride levels, anemia-related indices, hyponatremia, and elevated baseline creatinine were significant during univariable screening. Rather than representing isolated comorbidities, these alterations likely capture the systemic manifestations of advanced IPF, wherein chronic hypoxia and sustained metabolic expenditure culminate in generalized cachexia, catabolism, and subclinical multi-organ strain. For instance, the inverse association between triglycerides and mortality should not be misconstrued as a protective effect of hyperlipidemia. Instead, it underscores a classic reverse-causation paradigm, where declining lipid levels serve as a surrogate for advanced metabolic wasting. From a methodological standpoint, while these parameters achieved univariate significance, they were ultimately excluded during the LASSO-Cox procedure based on the optimized tuning parameter λ . This data-driven regularization implies that once primary predictors (such as lung function and imaging features) were adjusted for, the conditional prognostic contribution of these systemic markers became redundant due to collinearity, justifying their omission to ensure maximum model parsimony and robustness.

Few studies have specifically focused on prognostic model development in patients with coexisting IPF and T2DM using routinely available clinical variables. This clinical subgroup is of interest because T2DM may influence outcomes through mechanisms not fully captured in conventional IPF prediction tools, including systemic inflammation, insulin resistance, endothelial dysfunction, and microvascular injury [5–8]. At the same time, the present findings suggest that any future

disease-specific risk model for IPF with T2DM should still be anchored in core markers of pulmonary disease severity. If confirmed in larger cohorts, inflammatory or metabolic markers may provide complementary rather than dominant prognostic information.

The potential clinical implications of this work should be framed cautiously. The current model is not intended for bedside deployment or treatment decision-making. Rather, it highlights a small set of routinely collected variables that may be useful for exploratory risk stratification in hospitalized patients with IPF and T2DM. If externally validated, such a model might eventually support risk communication, follow-up planning, or enrichment of higher-risk patients in observational studies. Future studies should incorporate variables that were unavailable or incompletely captured in the present cohort, including HbA1c, duration of diabetes, diabetes-related complications, glucose-lowering medications, antifibrotic treatment, acute exacerbation status, and longitudinal changes in lung function.

Limitations

This study has several limitations. First, it was a retrospective analysis from a single center, and all included patients were hospitalized. Accordingly, the cohort likely represents a higher-risk and more clinically severe subgroup than the broader outpatient IPF population, which limits generalizability.

The high mortality rate observed in our cohort should therefore be interpreted in the context of inpatient case selection rather than as representative of all patients with IPF and T2DM.

Second, although the final model contained only five predictors, the initial number of candidate variables was relatively large for the 74 observed death events. Overfitting may therefore have occurred during the variable-selection stage despite the use of penalized regression and bootstrap correction. The reported performance measures should be regarded as internally validated estimates

within the development dataset, not as proof of transportable model performance.

Third, some predictors were based on clinically pragmatic categorization rather than fully standardized quantitative assessment. In particular, pulmonary function grading and HRCT fibrosis grading may be influenced by measurement variability and reader interpretation. Formal inter-reader agreement was not assessed, and future studies would benefit from centralized review or quantitative imaging methods.

Fourth, several clinically relevant diabetes-specific and IPF-specific variables were unavailable or incompletely captured, including HbA1c, duration of diabetes, microvascular complications, glucose-lowering medication class, antifibrotic therapy, acute exacerbation status, and serial pulmonary function changes. Residual confounding is therefore likely. Regarding antifibrotic therapy, which was not systematically captured, the likely direction of bias is as follows: if patients receiving antifibrotic treatment (who may have had less severe disease or better access to care) were unevenly distributed across predictor strata, the HRs for pulmonary function and HRCT fibrosis grading may be overestimated, as the protective effect of treatment could be partially attributed to these baseline severity markers. Conversely, if antifibrotic use was evenly distributed, confounding would be minimal.

Fifth, missing data were addressed using multiple imputation under an assumed missing-at-random mechanism. Although this approach is preferable to complete-case analysis when used appropriately, it cannot fully eliminate the possibility of bias, especially in a relatively small dataset.

Finally, no external validation was performed. For this reason, the nomogram and decision-curve findings should be interpreted as exploratory only. The model should be considered hypothesis-generating and requires testing in larger, independent, multicenter cohorts before any clinical

application can be recommended.

Conclusions

Among hospitalized adults with IPF and T2DM, older age, worse pulmonary function, and more extensive HRCT fibrosis were independently associated with mortality. An exploratory LASSO–Cox model combining age, pulmonary function grading, HRCT fibrosis grading, CRP, and IL-6 demonstrated good internal performance but requires external validation before clinical application. These findings should be regarded as hypothesis-generating and a foundation for multicenter studies. Future work should incorporate detailed glycemic control metrics, diabetes treatment data, and longitudinal follow-up to refine risk stratification in this high-risk comorbid population.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of the Second Affiliated Hospital of Zhengzhou University (approval number: KY2025495). Because this was a retrospective analysis of de-identified data, the requirement for individual informed consent was waived in accordance with institutional policy and relevant regulations. All methods were performed in accordance with the relevant guidelines and regulations (Declaration of Helsinki).

Consent for publication

Not applicable. This manuscript does not contain any individual person's data in any form.

Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available due to institutional restrictions and patient privacy protection, but may be available from the corresponding author on reasonable request, subject to institutional review and data protection regulations.

Competing interests

The authors declare that they have no competing interests.

Funding

This study received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' contributions

ZJ W: conceptualization, data curation, formal analysis, methodology, writing—original draft. W- W G: conceptualization, data curation, investigation, supervision, writing—review and editing. Both authors read and approved the final manuscript. We are grateful to HC R for assistance with manuscript preparation and language polishing.

Acknowledgements

We thank the Department of Respiratory and Critical Care Medicine at the Second Affiliated Hospital of Zhengzhou University for support with data collection.

Figures and Tables (Placeholders)

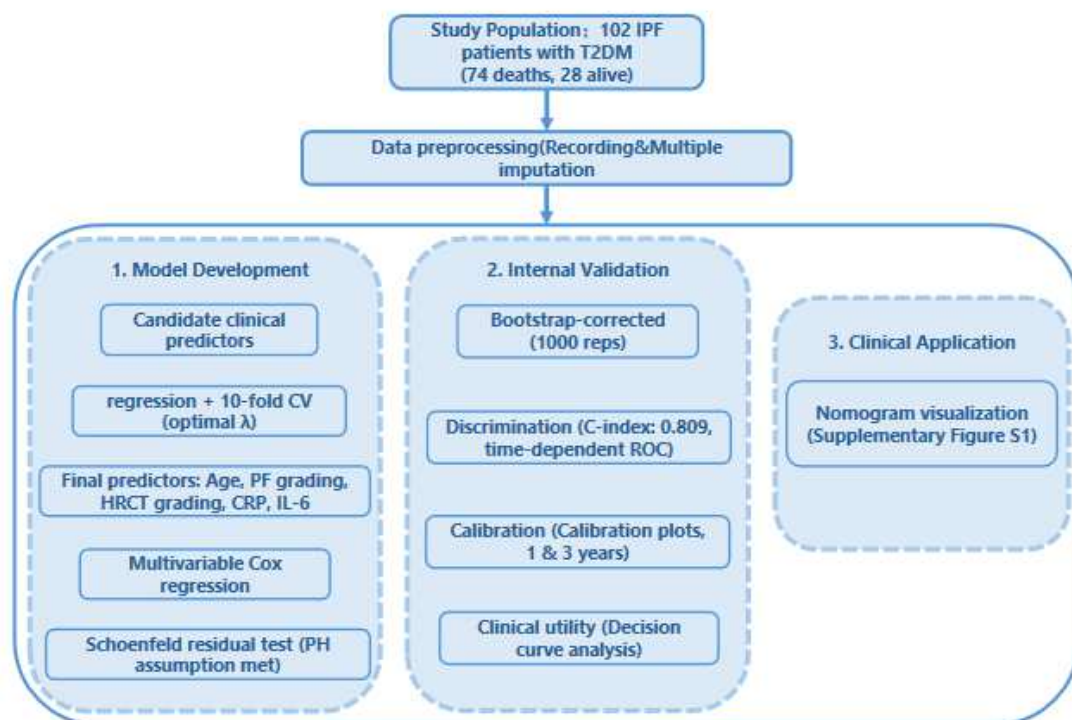


Figure 1. Study flow diagram.

Flow diagram showing patient screening, exclusion criteria, and the final analytical cohort. Of 196 screened hospitalized patients with idiopathic pulmonary fibrosis (IPF) and type 2 diabetes mellitus (T2DM), 102 were included in the final analysis.

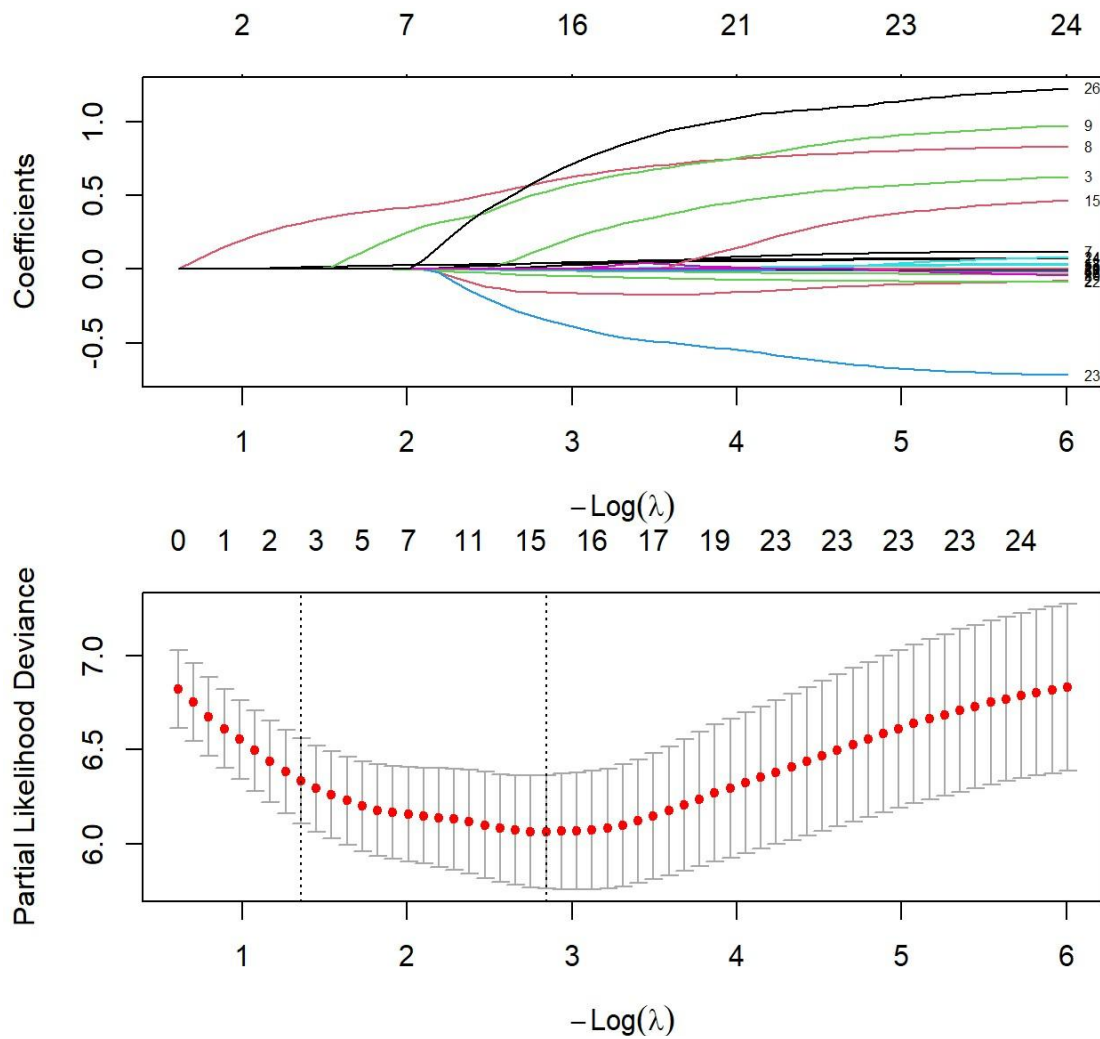


Figure 2. LASSO-Cox regression for predictor selection.

(A) Coefficient profiles of candidate predictors plotted against $\log(\lambda)$ in the LASSO-Cox model.

(B) Ten-fold cross-validation curve showing the cross-validated partial likelihood deviance across values of $\log(\lambda)$. A more parsimonious penalization level was selected to obtain a five-predictor exploratory model comprising age, pulmonary function grading, HRCT fibrosis grading, C-reactive protein (CRP), and interleukin-6 (IL-6).

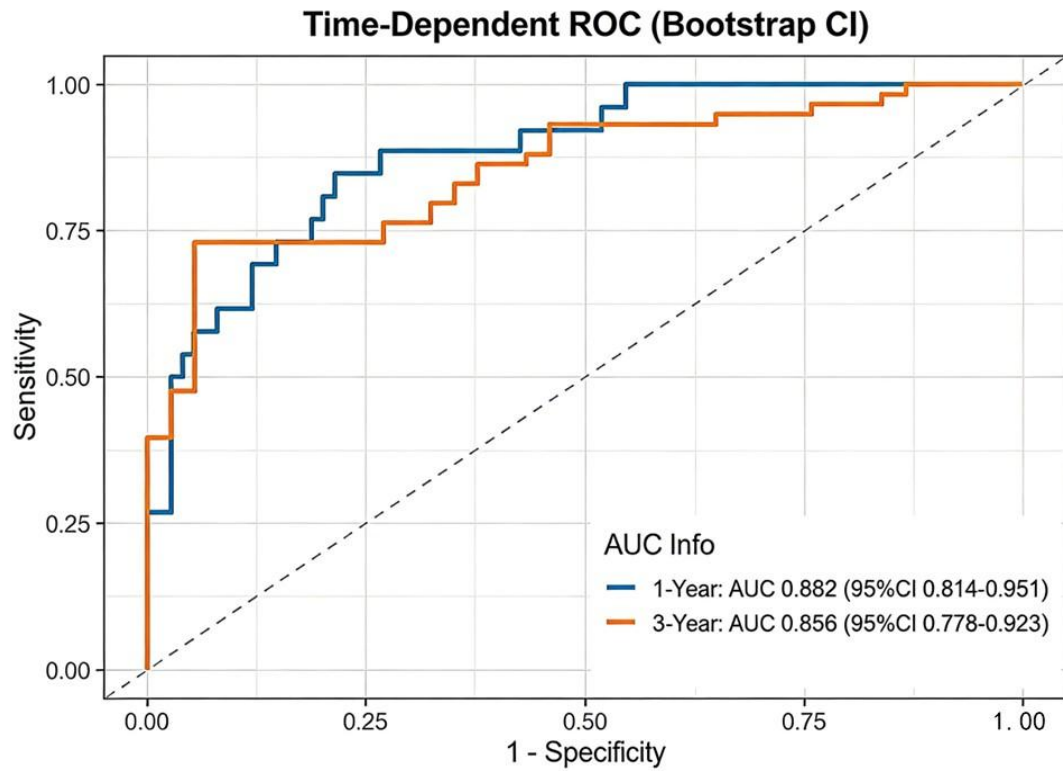


Figure 3. Time-dependent ROC curves for the five-predictor model.

Time-dependent receiver operating characteristic (ROC) curves of the final model at 1 year and 3 years. The area under the curve (AUC) was 0.882 (95% CI 0.814–0.951) at 1 year and 0.856 (95% CI 0.778–0.923) at 3 years.

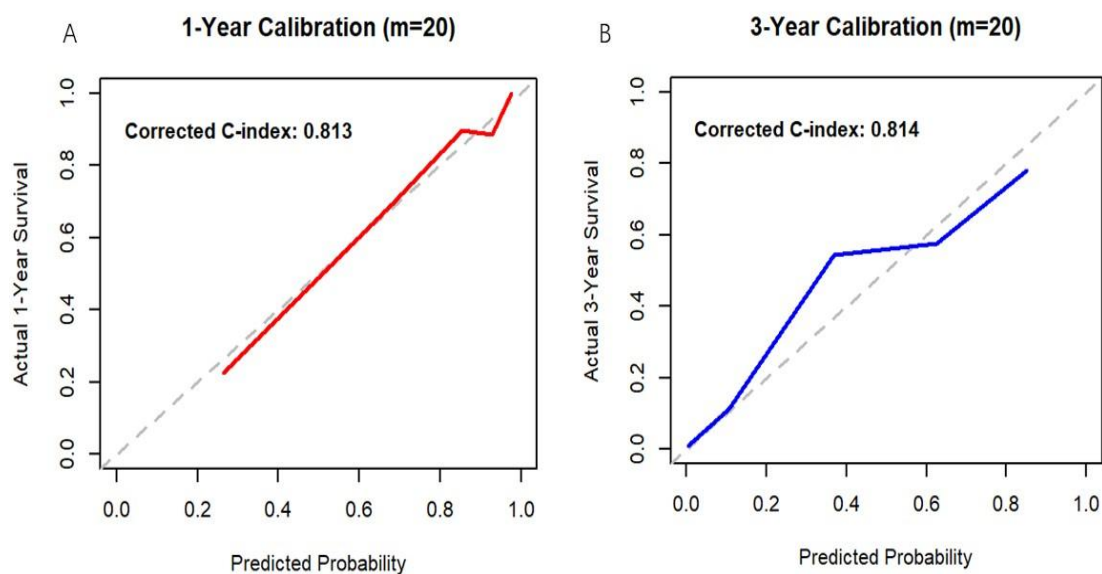


Figure 4. Bootstrap-corrected calibration plots of the five-predictor model.

Calibration plots at 1 year and 3 years comparing predicted survival probabilities with observed survival probabilities. The plots suggest reasonable agreement between predicted and observed outcomes within the development dataset after bootstrap correction.

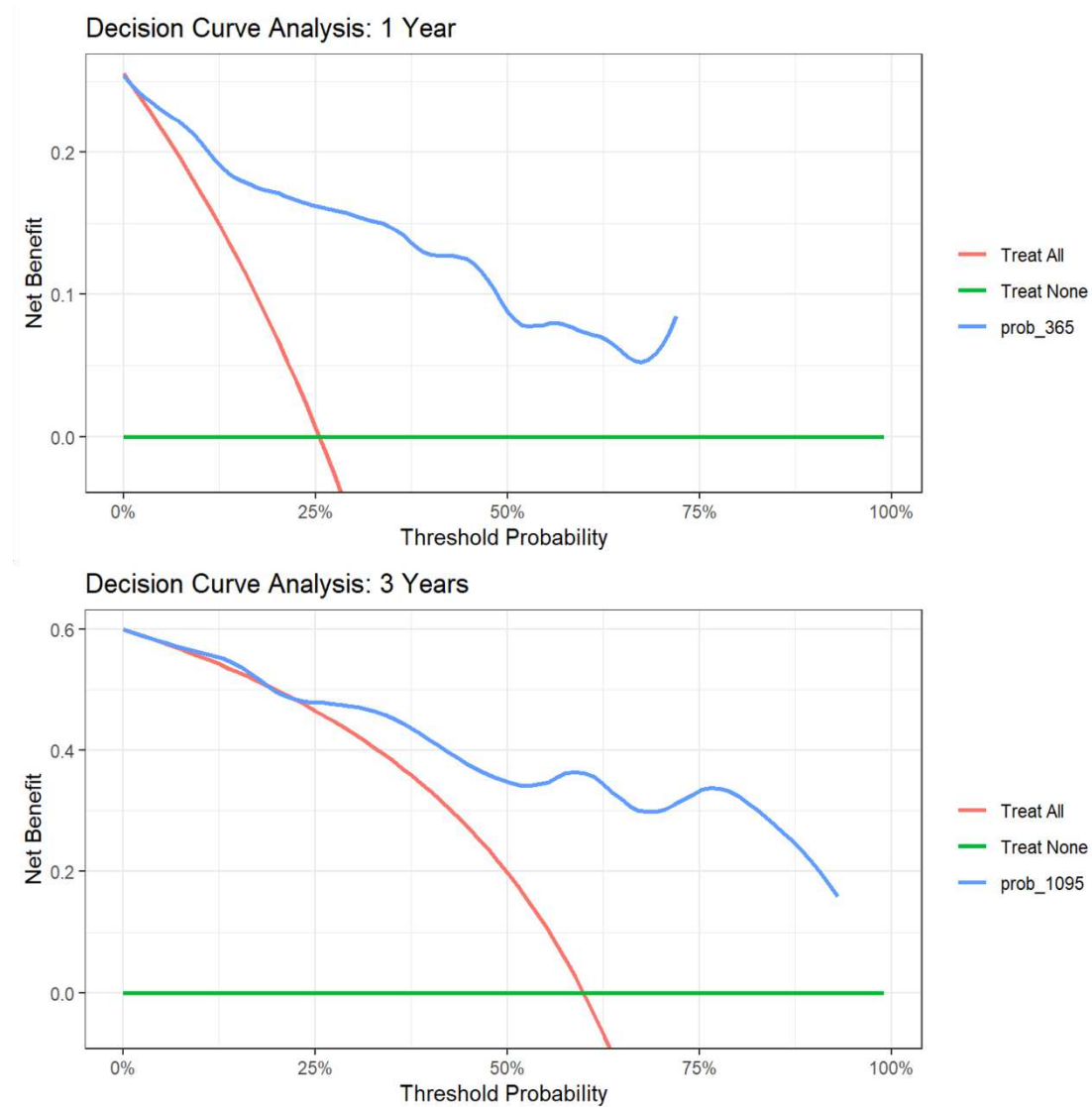


Figure 5. Decision-curve analysis of the five-predictor model.

DCA at 1 year and 3 years showing the net benefit of the model across a range of threshold probabilities, compared with treat-all and treat-none strategies. These results represent exploratory evidence of potential clinical usefulness within the internally validated development dataset.

List of Abbreviations.

Abbreviation	Definition
ALT	Alanine aminotransferase

Abbreviation	Definition
AUC	Area under the curve
CI	Confidence interval
Cr	Creatinine
CRP	C-reactive protein
DCA	Decision-curve analysis
DLCO%pred	Diffusing capacity for carbon monoxide as percent predicted
EPV	Events-per-variable
ESR	Erythrocyte sedimentation rate
FVC%pred	Forced vital capacity as percent predicted
Hb	Hemoglobin
HDL-C	High-density lipoprotein cholesterol
HR	Hazard ratio
HRCT	High-resolution computed tomography
IL-6	Interleukin-6
IPF	Idiopathic pulmonary fibrosis
IQR	Interquartile range
K ⁺	Potassium
LASSO	Least absolute shrinkage and selection operator
Na ⁺	Sodium
OS	Overall survival
RBC	Red blood cell
ROC	Receiver operating characteristic
SD	Standard deviation
TC	Total cholesterol

Abbreviation	Definition
TG	Triglycerides
UIP	Usual interstitial pneumonia
VIF	Variance inflation factor
WBC	White blood cell count

453

454 **Table 1.** Baseline characteristics of hospitalized patients with idiopathic pulmonary fibrosis and
455 type 2 diabetes mellitus, stratified by survival status.

Variable	Survivors (n=28)	Non-survivors (n=74)	P
Demographics and follow-up			
Age, years	60.25 ± 12.16	69.97 ± 10.05	<0.001
Sex, male/female	13/15	52/22	0.025
Smoking history, no/yes	21/7	46/28	0.223
Follow-up time, days	1868 [1297, 2318]	508 [172.5, 906.5]	<0.001
Vital signs			
SBP, mmHg	131.86 ± 13.26	128.99 ± 19.28	0.470
DBP, mmHg	84.43 ± 10.74	79.84 ± 12.72	0.093
Heart rate, beats/min	84.21 ± 10.94	86.65 ± 14.65	0.427
Respiratory rate, breaths/min	20.50 ± 2.22	21.04 ± 3.05	0.395
Laboratory markers			
CRP, mg/L	1.92 [1.33, 4.18]	13.04 [3.60, 75.23]	<0.001
IL-6, pg/mL	5.64 [3.06,	16.54 [6.36,	0.005

Variable	Survivors (n=28)	Non-survivors (n=74)	P
	14.46]	41.30]	
ESR, mm/h	16.00 [5.75, 24.25]	45.50 [20.50, 71.00]	<0.001
Glucose, mmol/L	5.13 [5.00, 6.27]	5.76 [4.83, 6.73]	0.556
WBC, $\times 10^9/L$	7.17 $\pm$ 2.01	8.42 $\pm$ 4.05	0.120
RBC, $\times 10^{12}/L$	4.52 $\pm$ 0.49	4.19 $\pm$ 0.69	0.003
Hemoglobin, g/L	136.89 $\pm$ 19.35	123.77 $\pm$ 19.79	0.003
ALT, U/L	18.00 [14.75, 26.00]	16.00 [12.00, 25.50]	0.201
Creatinine, $\mu\text{mol}/L$	62.00 [49.50, 65.50]	62.50 [49.25, 78.00]	0.574
Sodium, mmol/L	140.36 $\pm$ 2.10	137.90 $\pm$ 5.69	0.028
Potassium, mmol/L	3.97 $\pm$ 0.36	4.04 $\pm$ 0.57	0.543
Triglycerides, mmol/L	2.12 $\pm$ 1.80	1.26 $\pm$ 0.70	0.001
Total cholesterol, mmol/L	4.59 $\pm$ 1.06	4.31 $\pm$ 1.07	0.247
HDL-C, mmol/L	1.23 $\pm$ 0.28	1.21 $\pm$ 0.30	0.698
Pulmonary function and imaging			
FVC%pred (no impairment/mild/moderate/severe)	7/18/3/6	0/32/28/14	<0.001
DLCO%pred (no impairment/mild/moderate/severe)	5/16/6/1	2/29/27/16	<0.001
HRCT fibrosis (mild/moderate/severe)	19/9/0	33/36/5	0.071

456 Values are presented as mean $\pm$ SD, median [IQR], or n. Abbreviations: ALT, alanine
457 aminotransferase; CRP, C-reactive protein; DBP, diastolic blood pressure; DLCO%pred, diffusing
458 capacity for carbon monoxide as percent predicted; ESR, erythrocyte sedimentation rate;

FVC%pred, forced vital capacity as percent predicted; HDL-C, high-density lipoprotein cholesterol; HRCT, high-resolution computed tomography; IL-6, interleukin-6; RBC, erythrocyte count; SBP, systolic blood pressure; WBC, leukocyte count.

Table 2. Univariate Cox proportional hazards analyses for overall survival.

Variable	HR (95% CI)	P
Demographics		
Age	1.06 (1.03–1.09)	<0.001
Sex (female vs male)	1.54 (0.93–2.54)	0.092
Smoking history (yes vs no)	0.80 (0.50–1.28)	0.353
Imaging		
HRCT grading (reference: mild)		0.001
Moderate vs mild	1.87 (1.16–3.02)	0.011
Severe vs mild	5.02 (1.91–13.21)	0.001
Pulmonary function		
Pulmonary function grading (reference: mild)		<0.001
Moderate vs mild	2.79 (1.46–5.32)	0.002
Severe vs mild	7.72 (3.92–15.21)	<0.001
Vital signs		
Heart rate	1.01 (0.99–1.02)	0.362
Respiratory rate	1.08 (0.99–1.17)	0.082
SBP	1.00 (0.98–1.01)	0.490
DBP	0.98 (0.97–1.00)	0.091
Laboratory markers		

Variable	HR (95% CI)	P
Demographics		
CRP	1.01 (1.01–1.02)	<0.001
IL-6	1.02 (1.01–1.02)	<0.001
ESR	1.01 (1.01–1.02)	<0.001
Glucose	1.16 (1.00–1.34)	0.045
WBC	1.10 (1.03–1.17)	0.004
RBC	0.57 (0.40–0.81)	0.002
Hemoglobin	0.98 (0.97–0.99)	0.001
Creatinine	1.01 (1.00–1.02)	0.023
ALT	1.00 (0.98–1.01)	0.648
Sodium	0.88 (0.83–0.93)	<0.001
Triglycerides	0.68 (0.51–0.91)	0.010
Total cholesterol	0.86 (0.68–1.08)	0.186
HDL-C	1.03 (0.46–2.30)	0.950

Abbreviations: ALT, alanine aminotransferase; CI, confidence interval; CRP, C-reactive protein; DBP, diastolic blood pressure; ESR, erythrocyte sedimentation rate; HDL-C, high-density lipoprotein cholesterol; HR, hazard ratio; HRCT, high-resolution computed tomography; IL-6, interleukin-6; RBC, erythrocyte count; SBP, systolic blood pressure; WBC, leukocyte count.

Table 3. Multivariable Cox proportional hazards model of the five-predictor model for overall survival.

Variable	β	HR	95% CI	P
Age (per 1-year increase)	0.046	1.05	1.02–1.08	<0.001

Variable	β	HR	95% CI	P
Pulmonary function grading (per grade)	0.543	1.72	1.42–2.08	<0.001
HRCT fibrosis grading (per grade)	0.688	1.99	1.28–3.09	0.002
CRP (per mg/L)	0.006	1.01	0.999–1.01	0.056
IL-6 (per pg/mL)	0.004	1.00	0.998–1.01	0.199

Apparent C-index = 0.822 (SE 0.022). Abbreviations: CI, confidence interval; CRP, C-reactive protein; HR, hazard ratio; HRCT, high-resolution computed tomography; IL-6, interleukin-6.

References

1. Raghu G, Remy-Jardin M, Richeldi L, et al. Idiopathic pulmonary fibrosis (an update) and progressive pulmonary fibrosis in adults: an official ATS/ERS/JRS/ALAT clinical practice guideline. *Am J Respir Crit Care Med*. 2022;205(9):18-47.
2. Tondo P, Scioscia G, De Pace CC, Murgolo F, Maci F, Stella GM, et al. Gender Differences Are a Leading Factor in 5-Year Survival of Patients with Idiopathic Pulmonary Fibrosis over Antifibrotic Therapy Reduction. *Life (Basel)*. 2025;15(1):106.
3. Alhamad EH, Cal JG, Alrajhi NN, Aharbi WM, AlRikabi AC, AlBoukai AA. Clinical characteristics, comorbidities, and outcomes in patients with idiopathic pulmonary fibrosis. *Ann Thorac Med*. 2020;15(4):208-14.
4. Sauleda J, Núñez B, Sala E, Soriano JB. Idiopathic Pulmonary Fibrosis: Epidemiology, Natural History, Phenotypes. *Med Sci (Basel)*. 2018;6(4):110.
5. Mzimela N, Dimba N, Sosibo A, Khathi A. Evaluating the impact of type 2 diabetes mellitus on pulmonary vascular function and the development of pulmonary fibrosis. *Front Endocrinol (Lausanne)*. 2024;15:1431405.

- 489 6. Chance WW, Rhee C, Yilmaz C, Dane DM, Pruneda ML, Raskin P, Hsia CC. Diminished
490 alveolar microvascular reserves in type 2 diabetes reflect systemic microangiopathy. *Diabetes*
491 *Care*. 2008;31(8):1596-601.
- 492 7. Qinxue Shen, Shiting Zhou, Xiaoli Ouyang, Yuexin Tan, Sijin Chen, Jingcheng Shi, Hong Peng,
493 et al. Comorbidity burden and survival in patients with idiopathic pulmonary fibrosis: a real-
494 world cohort study. *Heart Lung*. 2026;79:102812.
- 495 8. Sia LC, Minnis P, Bloye G, Morris H, Naqvi M, Hodgkinson C, et al. Effect of comorbidities
496 on mortality in IPF: Analysis from the British Thoracic Society (BTS) UK interstitial lung
497 disease (ILD) registry. *Eur Respir J*. 2025;66(suppl 69):PA5091.
- 498 9. Liu HX. [Chinese Diabetes Prevention and Treatment Guideline (2024 Edition) released].
499 *Chinese Medical Information Report*. 2025;40(3):14. Chinese.
- 500 10. Goh NS, Desai SR, Veeraraghavan S, Hansell DM, Copley SJ, Maher TM, et al. Interstitial
501 lung disease in systemic sclerosis: a simple staging system. *Am J Respir Crit Care Med*.
502 2008;177(11):1248-1254.
- 503 11. Lemercier P, Colin-Chevalier R, Teruel E, Winter A, Roca L, Sophie G. P14 - Benchmark of
504 variable selection methods for survival analysis in clinical trials, a simulation study. *J*
505 *Epidemiol Popul Health*. 2025;73(suppl 2):203044.
- 506 12. Bosson-Amedenu S, Ayitey E, Ayiah-Mensah F, et al. Evaluating key predictors of breast
507 cancer through survival: a comparison of AFT frailty models with LASSO, ridge, and elastic
508 net regularization. *BMC Cancer*. 2025;25:665.
- 509 13. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number
510 of events per variable in logistic regression analysis. *J Clin Epidemiol*. 1996;49(12):1373-9.

- 511 14. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Series B Stat*
512 *Methodol.* 1996;58(1):267-88.
- 513 15. Xue X, Xie X, Gunter M, et al. Testing the proportional hazards assumption in case-cohort
514 analysis. *BMC Med Res Methodol.* 2013;13:88.
- 515 16. Austin PC. Bootstrap vs asymptotic variance estimation when using propensity score
516 weighting with continuous and binary outcomes. *Stat Med.* 2022;41(22):4426-43
- 517 17. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable
518 prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD statement.
519 *Ann Intern Med.* 2015;162(1):55-63.
- 520 18. Nishikiori H, Chiba H. Prognosis and prognostic factors for chronic fibrosing idiopathic
521 interstitial pneumonias. *Respir Investig.* 2025;63(5):762-70.
- 522 19. Sun X, Lei S, Zhao H, Guo L, Wang Y. Mortality-related risk factors of idiopathic pulmonary
523 fibrosis: a systematic review and meta-analysis. *J Thorac Dis.* 2024;16(12):8338-49.
- 524 20. Martka M, Miłkowska-Dymanowska J, Piotrowski WJ, Majewski S. Short-Term Longitudinal
525 Assessment of Automated Lung Parenchyma Analysis in IPF. *Eur Respir J.* 2025;66(suppl
526 69):PA2969.
- 527 21. Rennerova M, Zaloudikova M, Haluzik M, Stemberkova Hubackova S. Breaking barriers in
528 pulmonary health of patients with metabolic diseases: senolytics and beyond. *Front Aging.*
529 2026;7:1812724.
- 530 22. Kopf S, Kumar V, Kender Z, Han Z, Fleming T, Herzig S, et al. Diabetic Pneumopathy—A New
531 Diabetes-Associated Complication: Mechanisms, Consequences and Treatment
532 Considerations. *Front Endocrinol.* 2021;12:765201.

- 533 23. Mari YM, Fraix MP, Agrawal DK. Pulmonary Fibrosis and Diabetes Mellitus: Two coins with
534 the same face. *Arch Intern Med Res.* 2024;7(1):53-70.
- 535 24. Wang D, Ma Y, Tong X, Zhang Y, Fan H. Diabetes Mellitus Contributes to Idiopathic
536 Pulmonary Fibrosis: A Review From Clinical Appearance to Possible Pathogenesis. *Front*
537 *Public Health.* 2020;8:196.
- 538 25. Zhang K, Li A, Zhou J, Zhang C, Chen M. Genetic association of circulating C-reactive protein
539 levels with idiopathic pulmonary fibrosis: a two-sample Mendelian randomization study.
540 *Respir Res.* 2023;24(1):7.
- 541 26. De Lauretis A, Sestini P, Pantelidis P, Hoyles R, Hansell DM, Goh NS, et al. Serum interleukin
542 6 is predictive of early functional decline and mortality in interstitial lung disease associated
543 with systemic sclerosis. *J Rheumatol.* 2013;40(4):435-46.