U.PORTO

POSTGRADUATE DEGREE

Mortality Risk Prediction in Idiopathic Pulmonary Fibrosis: A Systematic Review and Meta-Analysis of Prediction Model Performance

Alexandra Sofia Morgado Neves

MEDS
13ªEDICTION

ADVISORS

Prof. Dr. Cláudia Camila Dias

Faculdade de Medicina da Universidade do Porto

Prof Dr. Alda Marques

Universidade de Aveiro

Prof. Dr. Elisabetta Renzoni

Royal Brompton Hospital

September 2026

FACULDADE DE **MEDICINA**

Acknowledgements

I am truly grateful to my supervisor, Professor Camila Dias, and co-supervisors Professor Alda Marques and Professor Elisabetta Renzoni, for their support, encouragement, expertise, and helpful feedback during the development of this thesis.

I am especially thankful to Professor Camila Dias for always being available and willing to help, no matter how simple my questions were.

I am also very grateful to Professor Alda Marques for her guidance, for helping me set clear goals and milestones, and for her valuable feedback throughout the project. I appreciate her patience and understanding throughout this process.

I would also like to thank Dr. Guilherme Rodrigues for generously sharing his knowledge, support, and time during this work. I am especially grateful to both Professor Alda Marques and Dr. Guilherme Rodrigues for letting me continue and further develop this project, which they had started and registered in PROSPERO.

I also want to thank Professor Matilde Soares for inspiring my interest in systematic reviews and for reminding me how important it was to continue this work during a difficult time in my life. Her encouragement, along with Prof. Camila's, helped me stay committed to this project throughout a long and challenging process.

I am also very thankful to my colleagues at work for their understanding and flexibility, which made it possible for me to balance my job with the demands of this degree.

Finally, I want to thank my family for their patience, understanding, and ongoing support throughout this journey. Thank you for accepting the many hours when, even though I was there, my mind was often focused on this work, and for supporting me through the final stages of this project.

1 Abstract

Background

Idiopathic Pulmonary Fibrosis (IPF) is a progressive fibrotic interstitial lung disease with an unpredictable and heterogeneous course. Accurate prognostic tools are critical for effective management, including decisions about lung transplantation, clinical trial inclusion, and patient counseling. Over the past three decades, numerous mortality prediction models have been developed for IPF. However, evidence regarding their validity and applicability across diverse populations remains limited.

Methods

This systematic review and meta-analysis aimed to identify and critically appraise mortality prediction models for IPF. A comprehensive literature search was conducted across PubMed, Web of Science, and Scopus. Studies included those that developed, validated, or updated mortality prediction models in adults with IPF. Data extraction followed the CHARMS checklist, and quality was assessed using the PROBAST tool. Model performance measures, such as discrimination and calibration, were synthesized, and meta-analyses were performed when feasible.

Results

A total of 4,751 records were screened, and 99 studies met the eligibility criteria, describing 59 original prognostic models, 26 updated models, and 80 external validation studies. Cox proportional hazards regression was the predominant modeling approach. Calibration was rarely reported, and most studies showed a high risk of bias, particularly in the analysis domain. The GAP model was the most extensively validated, with a

pooled Harrell's C-index of 0.686 (95% CI 0.659–0.713). Other models showed higher pooled discrimination, such as GAP-CR (Harrell's C-index 0.735, 95% CI 0.697–0.771), but this finding was based on only two performance estimates from a single study and should be interpreted cautiously. Overall, newer or updated models received substantially less independent validation than GAP.

Conclusion

Although numerous mortality prediction models have been developed for IPF, few have undergone rigorous, independent external validation. Overall, the models demonstrated moderate discrimination, but methodological limitations limit confidence in their clinical implementation. The GAP model remains the best-validated tool, while newer models require further independent validation.

Keywords

Idiopathic Pulmonary Fibrosis, Prognostic Models, Mortality Prediction, Systematic Review, Meta-Analysis, Model Validation

Table of Contents

Acknowledgements 3

1 Abstract 4

Background 4
Methods 4
Results 4
Conclusion 5
Keywords 5

3 Acronyms and abbreviations 9

4 Figure index 11

5 Table index 12

6 Thesis structure 13

Motivation 13
Research question and aims 13
State of the art 13
Materials and Methods 13
Results 13
Discussion 14
Conclusions and recommendations 14

7 Scientific and financial results 15

8 Motivation 16

9 Research question and aims....17

10 State of the art....18

10.1 Interstitial Lung Diseases....18

10.2 Idiopathic Pulmonary Fibrosis....19

11 Material and Methods....30

11.1 Type of study....30

11.2 Study eligibility criteria and information sources....30

11.3 Search strategy....31

11.4 Study selection process and Data collection process....32

11.5 Risk of bias and applicability assessment....33

11.6 Synthesis methods....34

12 Results....35

12.1 Number of studies identified through the search and number of studies included....35

12.2 Original models identified in the included studies....62

12.3 Models with external validation and identification of models that were updated.....66

12.4 Risk of bias assessment....72

12.5 Meta-analysis of prognostic mortality models....79

13 Discussion....84

14 Conclusions:....89

15 References....90

3 Acronyms and abbreviations

AIP - acute interstitial pneumonia

AE - Acute exacerbation

AUC - Area under the curve

BMI- Body mass index

CHARMS - CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies

CCI - Charlson Comorbidity Index

COP - Cryptogenic organizing pneumonia

CRP-Clinical Radiology Physiology Score

CTD - Connective Tissue Disease

DAD – Diffuse Alveolar Damage

DALY- Disability-Adjusted Life Years.

DC- Derivation Cohort

DIP - Desquamative Interstitial Pneumonia

DLCO - Diffusion Capacity for Carbon Monoxide

ERS – European Respiratory Society

FVC- Forced Vital Capacity

GAP – Gender Age Physiology

GEO - Gene Expression Omnibus

HL test -- Hosmer-Lemeshow test

HRQoL - Health-Related Quality of Life

HRCT - High-Resolution Computed Tomography

i.e. - id est

ILA - Interstitial Lung Abnormalities

ILD - Interstitial Lung Diseases

IPF - Idiopathic Pulmonary Fibrosis

LAS - Lung Allocation Score

LIPI- Lung Immune Prognostic Index

NSIP - non-specific interstitial pneumonia

N-number

NA – Not applicable

NI – No information; VC-Validation cohort

OS – overall survival

PET- CT – Positron Emission Tomography- Computed Tomography

PROBAST - Prediction model Risk Of Bias ASsessment Tool

RISE - Risk Stratification Score

ROC- Receiver operating characteristic

RB-ILD- respiratory bronchiolitis- interstitial lung disease

6-MWT – six-Minute Walk Test

SD – Standard deviation

SPP- Portuguese Pulmonology Society

TRIPOD-SRMA - Transparent Reporting of multivariable prediction models for Individual Prognosis Or Diagnosis-checklist for Systematic Reviews and Meta-Analyses.

UIP – Usual Interstitial Pneumonia

VC-Validation cohort

WASOG- World Association of Sarcoidosis and Other Granulomatous Disorders

4 Figure index

Figure 1 - Literature screening flow chart - PRISMA 2020 35
Figure 2 - Included predictors within the original models 64
Figure 3 – External validation studies TRIPOD 3/4 67
Figure 4 - External validation studies TRIPOD type 3 67
Figure 5 - New Predictors included in the Updated Prognostic Models 69
Figure 6- Risk of Bias for Development Models 72
Figure 7 - Applicability for Development models 73
Figure 8 - Overall Development models 73
Figure 9- Risk of Bias Updated Models 75
Figure 10 - Applicability-updated models 75
Figure 11 -Overall updated models 76
Figure 12 - Risk of Bias for External Validation studies 77
Figure 13 - Applicability for External Validation studies 77
Figure 14 - Overall- External Validation Studies 78
Figure 15 - Forest plot of pooled Harrell's C-index estimates for prognostic mortality models in IPF 80
Figure 16 - Forest plot of pooled Harrell's C-index estimates for the original GAP model, external validation studies, updated GAP-derived models 82

5 Table index

Table 1 - Idiopathic Pulmonary Fibrosis diagnosis based on HRCT and histopathology pattern ..... **Erro! Marcador não definido.**
Table 2- Original Models 38
Table 3 - External validation studies and updated models 48
Table 4 - Methodological characteristics of the original models..... 63
Table 5 - Methodological characteristics of the updated models 70

6 Thesis structure

Motivation

In this section, a brief overview of Idiopathic Pulmonary Fibrosis and an explanation about the unpredictable and heterogeneous course of this disease are presented. This section also explains the relevance of the present work.

Research question and aims

The research question, as well as the aim, are presented in this section.

State of the art

This section is subdivided into topics that provide context for the work developed.

Materials and Methods

This section is subdivided into subsections. It begins by describing the type of research study, followed by the study's eligibility criteria and information sources. After that, the search strategy, study selection process, and data collection methods are presented. Then, data items and methods for assessing risk of bias and applicability concerns are described. Finally, the synthesis methods are reported.

Results

This section presents the detailed results obtained after applying the statistical methods to achieve the proposed aims.

Discussion

In the discussion section, the results are reflected on and contextualized in relation to previous work. Additionally, the limitations of the present study are explored here.

Conclusions and recommendations

This section summarizes the main results of this work and answers the research question.

7 Scientific and financial results

The primary aim of this study is to produce an original research manuscript for submission to a peer-reviewed international journal in the field of pulmonology. No financial support was received for the conduct of this study.

8 Motivation

Diffuse parenchymal lung diseases, also known as interstitial lung diseases (ILD), comprise a large number of disorders with different aetiologies, clinical manifestations, imaging and pathological characteristics, and disease courses (1).

Idiopathic pulmonary fibrosis (IPF) is the most common fibrotic ILD and is characterized by a progressive decline in lung function, ultimately leading to death from respiratory failure. Without treatment, median survival after diagnosis is typically 3-5 years (2). This decline in lung function is reduced in IPF patients treated with the antifibrotic agents pirfenidone, nintedanib, or nerandomilast (3-5). Nevertheless, rapid or accelerated progression of the disease may occur both in patients on antifibrotic treatment and in those who have been stable for several years without pharmacological treatment. As the disease course and rate of progression are often unpredictable and heterogeneous (6), a simple staging system or prediction model that accurately informs prognosis is crucial for identifying those at higher risk. Such models could guide management decisions, including the timing of lung transplantation, life-planning decisions, enrolment in randomized clinical trials, personalized medicine, optimizing the use of effective interventions, and reducing waste of healthcare resources (7).

Several models for predicting mortality in IPF have been developed and published over the past three decades(8). Still, there is a lack of knowledge about the performance of the different models and their applicability to populations other than those included in the development datasets.

The availability of several models with the same purpose motivates this systematic review and meta-analysis, which aims to evaluate and summarize the available evidence on the performance of mortality prediction models in IPF.

9 Research question and aims.

The dissertation research question was:

What mortality prediction models are available for people with idiopathic pulmonary fibrosis, and what is their predictive performance?

P - Adults (≥18 years of age) diagnosed with Idiopathic Pulmonary Fibrosis
I - N/A
C - N/A
O - Mortality

Aim:

To systematically identify and critically appraise mortality prediction models for people with IPF and, where appropriate, synthesise and compare their predictive performance through meta-analysis.

10 State of the art

10.1 Interstitial Lung Diseases

Diffuse parenchymal lung diseases or interstitial lung diseases (ILD) are a group of disorders that embrace more than 150 distinct entities, which mostly affect the pulmonary interstitium (1). They are characterized by varying degrees of inflammation and fibrosis in the interstitial space, which impair the alveolar-capillary barrier, leading to dyspnoea and eventually respiratory failure and death.

ILD can be categorized as: idiopathic interstitial pneumonias, autoimmune-related ILD, exposure-related (including iatrogenic) ILD, interstitial lung diseases with cysts or airspace filling, sarcoidosis, and orphan diseases (9-11).

The 2025 ERS/ATS classification expands the concept of interstitial pneumonias to encompass both idiopathic and secondary forms and organizes them by their predominant morphologic pattern. It distinguishes between interstitial and alveolar filling disorders and recognizes combined patterns and unclassifiable ILD ((12)). Regarding idiopathic forms, it divides them into:

Interstitial disorders

- Idiopathic pulmonary fibrosis (idiopathic usual interstitial pneumonia)
- Idiopathic nonspecific interstitial pneumonia (iNSIP)
- Idiopathic Bronchiolocentric interstitial pneumonia (iBIP) (provisional diagnosis)
- Idiopathic diffuse alveolar damage
- Idiopathic pleuroparenchymal fibroelastosis (iPPFE)
- Idiopathic lymphoid interstitial pneumonia (iLIP)

Alveolar filling disorders

- Cryptogenic organizing pneumonia (COP)
- Respiratory bronchiolitis-associated interstitial lung disease (RB-ILD)
- Alveolar macrophage pneumonia (AMP; replacing the term DIP)

- Rare alveolar filling disorders (e.g., eosinophilic pneumonia, pulmonary alveolar proteinosis, lipoid pneumonia)

10.2 Idiopathic Pulmonary Fibrosis

IPF is a specific type of chronic, progressive fibrotic interstitial pneumonia of unknown cause that is associated with radiological and histologic features of usual interstitial pneumonia (UIP) (13, 14). This disorder was previously known as cryptogenic fibrosing alveolitis, idiopathic fibrosing alveolitis, or common interstitial pneumonia; in 2000, the concept of IPF was elucidated and validated with diagnostic criteria, establishing it as a distinct clinical entity (13).

I.Diagnosis

In the last 3 decades, several statements have reviewed the diagnostic criteria for IPF (13-16). As stated by the guidelines published in 2018 and reviewed in 2022, the diagnosis of IPF requires: 1) the exclusion of other known causes of ILD (e.g., domestic, and occupational environmental exposures, connective tissue disease (CTD), drug toxicity), and 2) the presence of a high-resolution computed tomography (HRCT) pattern of UIP or specific combinations of HRCT patterns and histopathology patterns in patients subjected to lung tissue sampling (14). (Table 1)

Table 1-Idiopathic Pulmonary Fibrosis diagnosis based on HRCT and histopathology pattern

IPF suspected	**Histopathology pattern**				
HRCT pattern		**UIP**	**Probable UIP**	**Indeterminate or biopsy not performed**	**Alternative diagnosis**
	UIP	IPF	IPF	IPF	Non-IPF diagnosis
	Probable UIP	IPF	IPF	IPF (likely)	Non-IPF diagnosis
	Indeterminate	IPF	IPF (likely)	indeterminate	Non-IPF diagnosis
	Alternative Diagnosis	IPF (likely)	Indeterminate	Non-IPF diagnosis	Non-IPF diagnosis

Legend: HRCT – High-Resolution Computed Tomography; IPF-Idiopathic Pulmonary Fibrosis; UIP – Usual Interstitial Pneumonia. Adapted from Raghu et al (16).

From the point of view of HRCT, imaging is categorized into four groups: a UIP pattern, a probable UIP pattern, an indeterminate pattern for UIP, and a pattern suggestive of an alternative diagnosis. The UIP pattern in HRCT is characterized by the presence of honeycombing with or without traction bronchiectasis/ bronchiolectasis; the presence of irregular thickening of interlobular septa, usually superimposed with a reticular pattern and/or mild ground-glass opacity; and pulmonary ossification in some instances. When the UIP radiological pattern is present, the confidence in a histopathological UIP diagnosis is superior to 90%.

The histopathological pattern is also categorized in 4 groups: UIP, probable UIP, indeterminate for UIP, and an alternative diagnosis.

The histopathological UIP pattern is characterized by patchy dense fibrosis with architectural distortion (i.e., destructive scarring and/or honeycombing); a predilection for subpleural and paraseptal lung parenchyma; fibroblastic foci; and the absence of features that suggest an alternative diagnosis. When all these features are present, a UIP pattern can be established with confidence.

IPF should be considered in all adult patients with unexplained chronic exertional dyspnoea, and commonly presenting cough, bibasilar inspiratory crackles, and finger clubbing. The incidence of the disease increases with older age, typically occurring after the age of 60, and has been more commonly reported in males, with a history of cigarette exposure. Patients with IPF younger than 50 years old are rare; in such patients, an underlying CTD or familial pulmonary fibrosis should be suspected (15). Clinical features determine the likelihood of IPF and should inform clinical decision-making and the diagnostic workup (17).

The ATS/ERS statements published in 2011 and 2018 on the diagnosis and management of IPF highlight the importance of an experienced multidisciplinary team in diagnosing IPF (14, 15). Multidisciplinary discussion that integrates clinical, radiological, and, when available, histopathological findings has been shown to improve diagnostic confidence in a substantial number of patients (14). Multidisciplinary discussion should also be considered before invasive procedures, and cases should be reviewed again after invasive testing is performed. This discussion is particularly relevant in patients with a radiological pattern that is indeterminate for UIP or suggestive of an alternative diagnosis. At present, multidisciplinary discussion is the gold standard for diagnosing IPF and the remaining ILD subtypes.

II. Epidemiology and burden of disease

Epidemiological data for IPF can vary widely and depend on the country/geographic region, case definition, and population demographics. Incidence and prevalence increase with age.

In 2019, the global incidence and mortality of ILD and pulmonary sarcoidosis were 24.2 million and 169 833 cases, respectively. From 1990 to 2019, the global incidence, deaths and disability adjusted life years (DALYs) due to ILD and pulmonary sarcoidosis increased by 118.6%, 166.63% and 122.87%, respectively (18).

A systematic review published in 2023 concerning the incidence and prevalence of ILD reported a prevalence of IPF ranging from 7 to 1650 per 100 000 persons (19).

The PERSEIDS researchers reported a retrospective two-phase study that assessed incidence, prevalence, and rate of progression of ILDs across several European countries, including Portugal. The analysis allowed for an estimate for 2018 of an incidence and prevalence of IPF ranging from 0.4 to 10.3 and from 2.8 to 31 per 100000 person-year, respectively (20).

III.Mortality

Zeng et al., in a 2022 systematic review of mortality and survival in IPF, reported that age-standardized mortality rates for IPF ranged from 0.5 to 12 per 100 000 population annually after 2000. The overall 3-year and 5-year cumulative survival rates (CSRs) were 61.8% (95% CI 58.7-64.9; I2 = 97.1%) and 45.6% (95% CI 41.5-49.7; I2 = 97.7%), respectively. Prior to 2010, the pooled 3-year CSR was 59.9% (95% CI 55.8-64.1; I2 = 95.8%), and it increased to 66.2% (95% CI 62.9-69.5; I2 = 92.6%) in the 2010s, although this change was not significant. After excluding three post-2010 studies in which no patients received antifibrotics, the pooled 3-year CSR in the 2010s increased significantly to 67.4% (95% CI 63.9-70.9; I2 = 93.1%) (21).

IV.Quality of life and costs

Diamantopoulos et al., in a comprehensive review of the literature on health-related quality of life (HRQoL) and costs, included 60 studies on HRQoL and 28 on costs or healthcare resource use (HCRU). The analysis reported that despite regional differences, there is some agreement between the absolute and relative level of HRQoL for patients with IPF compared with the general population. HRQoL was found to be lower for patients with IPF when compared with overall population. Regarding costs, after adjustments for the cost years and currency, the suggested annual per capita cost of patients with IPF in North America was estimated to be around US$20 000, 2.5–3.5 times higher than the national healthcare expenditure. Furthermore, research comparing patients with IPF with a matched control cohort suggested a significant increase in resource use and cost.(21)

OASIS study, a prospective, observational, multicentric study of patients with confirmed IPF in Spain, evaluated annual IPF-related costs per patient, categorized for Forced Vital Capacity (FVC). The mean (standard deviation) total annual IPF related cost was €26,997 (17,555), with statistically significant differences (p = 0.0002) between groups: €44,412 (33,389) for the FVC < 50%, €25,803 (14,688) for the FVC

50–80% and €23,242 (13,642) for the FVC > 80%. Annual direct health costs had the greatest weight and included pharmacological treatments [€22,324 (13,773)] and hospitalization days [€1659 (7362)]. Mean total cost of an acute exacerbation of IPF was €10,372. According to the multivariate analysis, an impaired lung function (FVC < 50%) and use of antifibrotic treatment were determinants of cost ($p < 0.0001$ both) (22).

Current data suggest that the incidence of IPF is similar to conditions such as oesophageal, liver, pancreatic, and cervical cancers (23), with considerable impact on patients and healthcare providers (21, 22).

V. Disease Course

As mentioned before, IPF is a chronic, progressive disease characterized by a gradual decline in lung function, respiratory failure, and death. Across clinical trials, the decline in FVC in placebo-treated patients with IPF and mild or moderate impairment in lung function at baseline was approximately 150–200 mL/ year (23). In an analysis of pooled data from placebo-treated patients in the CAPACITY and ASCEND trials, change in FVC was highly variable and could not be predicted based on change in the previous 6 months (4, 24, 25). Interestingly, patients with FVC>90% of predicted at baseline experienced a decline in FVC over 52 weeks similar to that of those with baseline FVC ⩽90% of predicted, suggesting that patients with preserved FVC should not be regarded as being at low risk of disease progression (26).

In the years before IPF diagnosis, which usually occurs in the sixth decade or later, patients may be asymptomatic or have mild symptoms and are often seen and managed by a general medical practitioner. Given that many symptoms are nonspecific, a substantial delay between the onset of symptoms and diagnosis, with a median time of 1-2 years, often occurs (27, 28).

Despite the progressive nature of the disease over time, the patient course and the rate of disease progression are unpredictable and heterogeneous. Patients generally follow one of three courses: slow physiologic deterioration since diagnosis with worsening dyspnea, rapid deterioration over months leading to death, or periods of relative stability interposed with periods of acute respiratory decline, also known as acute exacerbation, sometimes manifested by hospitalizations for respiratory failure (23, 29).

VI. Acute exacerbation

Acute exacerbation (AE) is defined as an acute, clinically significant respiratory deterioration characterized by evidence of new widespread alveolar abnormality, in patients with: 1) previous or concurrent diagnosis of IPF; 2) an acute worsening or development of dyspnoea typically with less than 1 month duration; 3) computed tomography with new bilateral ground-glass opacity and/or consolidation superimposed on a background pattern consistent with usual interstitial pneumonia pattern and 4) deterioration not fully explained by cardiac failure or fluid overload (29).

AE is a frightening event with profound prognostic implications. Available data suggest that up to 46% of deaths in IPF are preceded by an acute exacerbation, and respiratory failure from an acute exacerbation is associated with high in-hospital mortality (29).

The reported incidence of AE is variable, higher in cohort studies with an annual incidence range from 8.6 to 14.2%, per 100 IPF patient years (29, 30). Several triggers have been described such as infection, aspiration, drugs, and thoracic surgery(6, 29).

Risk factors for AE include reduced forced vital capacity and total lung capacity, poor baseline PaO2 and a higher alveolar-arterial oxygen difference. Mechanical procedures, higher serum KL-6 concentration and secondary pulmonary hypertension, might be risk factors for AE-IPF (29, 31).

VII. Antifibrotic Therapy

In 2014, the publication of the results of 3 randomized controlled trials: assessment of pirfenidone to confirm efficacy and safety in idiopathic pulmonary fibrosis (ASCEND) and efficacy and safety of nintedanib in the treatment of idiopathic pulmonary fibrosis (INPULSIS 1 and 2) was a historical mark in the management of IPF(3, 32).

Pirfenidone and nintedanib, have been shown to slow lung function decline in patients with IPF (3, 32, 33). In a systematic review and meta-analysis, that included 26 studies and 12956 patients, antifibrotic treatment was also associated to improvement of clinical outcomes such as mortality, AE and respiratory hospitalizations (34, 35).

Recently, nerandomilast has emerged as a new drug that reduces the decline in lung function in IPF patients (5).

Regardless of the undoubtful reduction on lung function decline with antifibrotic therapy, IPF patients can still witness a rapid progression of disease even those who have been stable for several years without pharmacological treatment (2, 36, 37).

VIII. Staging and Prognostication in IPF

Staging disease severity and accurately predicting survival in IPF have clear clinical value, namely: facilitating counseling of patients and families about end-of-life decisions, aiding communication among healthcare providers, guiding the optimal timing of lung transplantation, and improving clinical decisions such as entry into IPF trials (6, 12, 33, 43).

Multiple independent factors have been identified as prognostic indicators: at baseline - level of dyspnea, DLCO < 40% predicted, desaturation to values below 88% during the 6-minute walk test (6MWT), extent of honeycombing on HRCT, and pulmonary hypertension; at follow-up - increase in level of dyspnea, decrease in FVC > 10% in absolute value, decrease in diffusion capacity for carbon monoxide (DLCO) > 15% in absolute value, and worsening of fibrosis on HRCT (12, 44).

Several studies have proposed multidimensional indices to achieve a more accurate prediction of disease course. A variety of prognostic variables have been used as part of multidimensional scoring systems (clinical prediction models) including age, gender,

PFT measures (including longitudinal trends), extent of disease on HRCT, exercise tests (6-min walk test) and dyspnea scores and biomarkers.

IX. Predictive models

Numerous decisions in medicine are made based on an estimated probability that a specific event will occur in the future in an individual. Information from a single predictor is often insufficient to provide reliable estimates of diagnostic and prognostic probabilities (38). Multivariable prediction models are used in many medical fields to provide accurate prognostic information and disease staging (39).

Clinical prediction models typically use statistical or machine learning methods that combine clinical findings from history, physical examination, and/or test results to estimate the probability of an outcome, usually a diagnosis or prognosis. Other names for a prediction model include risk prediction model, predictive model, prognostic (or prediction) index or rule, and risk score (38).

After developing a model, it is important to evaluate its predictive performance, for example, in terms of calibration and discrimination, overall fit, and clinical utility.

Discrimination refers to the ability of a prediction model to distinguish between individuals who experience the outcome of interest and those who do not. Calibration refers to the agreement between the outcomes predicted by the model and the outcomes observed.

It is common to distinguish between internal and external validation. In internal validation, performance is evaluated within the model development dataset itself, potentially while estimating and adjusting for optimism due to overfitting. Conversely, external validation evaluates performance on data not used for model development, potentially from a different target population or setting (39-42).

Over the last 3 decades, many prediction models have been developed to predict mortality risk in patients with IPF (43).

Although a 2014 literature review identified several mortality prediction models for patients with idiopathic pulmonary fibrosis (IPF), the performance and generalisability of these models across different populations and clinical settings remained unclear(43).

More recently, Wu et al.(44) systematically evaluated the risk of bias and the applicability of existing IPF mortality risk models but excluded studies that focused solely on validating existing models. Khalid et al. (45) systematically reviewed the use of artificial intelligence (AI), machine learning, and radiomics approaches for the diagnosis and prognosis of idiopathic pulmonary fibrosis (IPF) based on computed tomography (CT) imaging. In contrast, the present systematic review adopts a broader scope, encompassing mortality prediction models that incorporate clinical, physiological, imaging, biomarker, genomic, and AI-derived predictors. Given the increasing number of prognostic models developed for IPF, a comprehensive synthesis of the available evidence is warranted to characterize the predictors and modeling approaches used and to evaluate the methodological quality, clinical validity, and applicability of these models.

11 Material and Methods

The study was registered in PROSPERO under registration number CRD42022366785.

11.1 Type of study

A systematic review of the performance of prediction models was conducted and reported according to the Transparent reporting of multivariable prediction models for individual prognosis or diagnosis-checklist for systematic reviews and meta-analyses (TRIPOD-SRMA) (39).

11.2 Study eligibility criteria and information sources

A systematic search with search strategies was conducted in PubMed, Web of Science and Scopus from inception to July 2025. EMBASE was not searched as initially planned and registered because the institution did not have a subscription. A manual search of the reference lists of the included studies was also performed. No language restrictions were implemented.

Inclusion criteria, following the PICOTS format were: 1) P - studies that assessed mortality predictive models in adult patients with Idiopathic Pulmonary Fibrosis; 2) Index models - studies with prediction models developed by regression or machine learning modelling, reporting derivation, validation or updating of the model; 3) Comparator-not applicable; 4) Outcome- Studies reporting predictive models, with mortality as a binary or time to event outcome with complete information about the modelling (i.e. all regression coefficients and model intercept or baseline survival at a given time point) and performance; or with information that allow to calculate

validation, calibration or association measures; 4) Timing - all prediction horizons; 5) Setting -in and outpatients

11.3 Search strategy

Studies were retrieved using the following queries:

1) PubMed

((("predict*"[Title/Abstract] OR "predictive value of tests"[MeSH Terms] OR "score"[Title/Abstract] OR "scores"[Title/Abstract] OR "scoring system"[Title/Abstract] OR "scoring systems"[Title/Abstract] OR "observ*"[Title/Abstract] OR "observer variation"[MeSH Terms]) AND ("idiopathic pulmonary fibrosis"[MeSH Terms] OR ("idiopathic"[All Fields] AND "pulmonary"[All Fields] AND "fibrosis"[All Fields]) OR "idiopathic pulmonary fibrosis"[All Fields]) AND ("mortality"[MeSH Terms] OR "mortality"[All Fields] OR "mortalities"[All Fields] OR "mortality"[MeSH Subheading])) NOT ("animals"[MeSH Terms] NOT "humans"[MeSH Terms])) NOT (("child"[MeSH Terms] OR "Infant"[MeSH Terms] OR "Adolescent"[MeSH Terms]) NOT "adult"[MeSH Terms]);

2) Web of Science

(predict* OR predictive value of test OR score OR scores OR scoring system OR scoring systems OR observ*OR observer variation) AND (idiopathic pulmonary fibrosis OR (idiopathic AND pulmonary AND fibrosis)) AND mortality

3) Scopus

((predict*) OR (predictive value of test) OR (score) OR (scores) OR (scoring system) OR (scoring systems) OR (observ*) OR (observer variation)) AND ((idiopathic pulmonary fibrosis) OR (idiopathic AND pulmonary AND fibrosis)) AND (mortality)

11.4 Study selection process and Data collection process

Search results were initially managed in EndNote (Clarivate, version 21). After duplicate records were removed, the remaining citations were exported to Rayyan. Two independent reviewers (SN, GR), blinded to each other's assessments, selected a random subset comprising 20% of the total retrieved records. Titles and abstracts were screened for potential eligibility, and the full texts of potentially eligible studies were subsequently assessed against the predefined criteria. Disagreements regarding study inclusion were resolved through consensus or, when necessary, by consultation with a third reviewer.

Data extraction was performed by one reviewer using a predefined Excel spreadsheet in accordance with the checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS)(40, 45). Data concerning the following items were included:

1. first author
2. year of publication
3. country
4. journal
5. study design (cohort, randomized trial participants, registry data)
6. participants (participant eligibility and recruitment method, participant description, and study dates)
7. outcome to be predicted (definition, assessment of outcome blindness to candidate predictors)
8. candidate predictors (number, type, timing of predictor measurement, handling of predictors in the modelling)
9. sample size (number of participants and outcomes, number of outcomes in relation to candidate predictors)
10. missing data (number of participants with missing data, number of participants with missing data for each predictor, handling of missing data)

11. type of model (development, validation or updating)
12. model development (modelling method, method for selection of predictors in multivariable modelling and during multivariable modelling and criteria used)
13. model performance (calibration, discrimination measures with confidence intervals, validate measures)
14. method used for testing model evaluation (development dataset only or separate external validation) and in case of poor validation if model was updated
15. results (final and other multivariable models presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures and if done alternative presentation of the prediction models as nomograms, sum scores with performance)

Eligible articles were categorized into three groups: model development studies, external validation studies, and model updating studies. Separate data extraction forms were used for each group. Once the articles were selected, data were extracted into a prespecified Excel spreadsheet in accordance with the checklist for critical appraisal and data extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS)(40, 45).

11.5 Risk of bias and applicability assessment

The risk of bias and applicability of prognostic prediction model studies were assessed by one reviewer using the Prediction Model Risk of Bias Assessment Tool (PROBAST) (46). The PROBAST consists of four domains assessing risk of Bias (participants, predictors, outcome, and analysis) and 3 domains assessing applicability concerns (participants, predictors and outcome). Each item was scored as yes, no, or unclear.

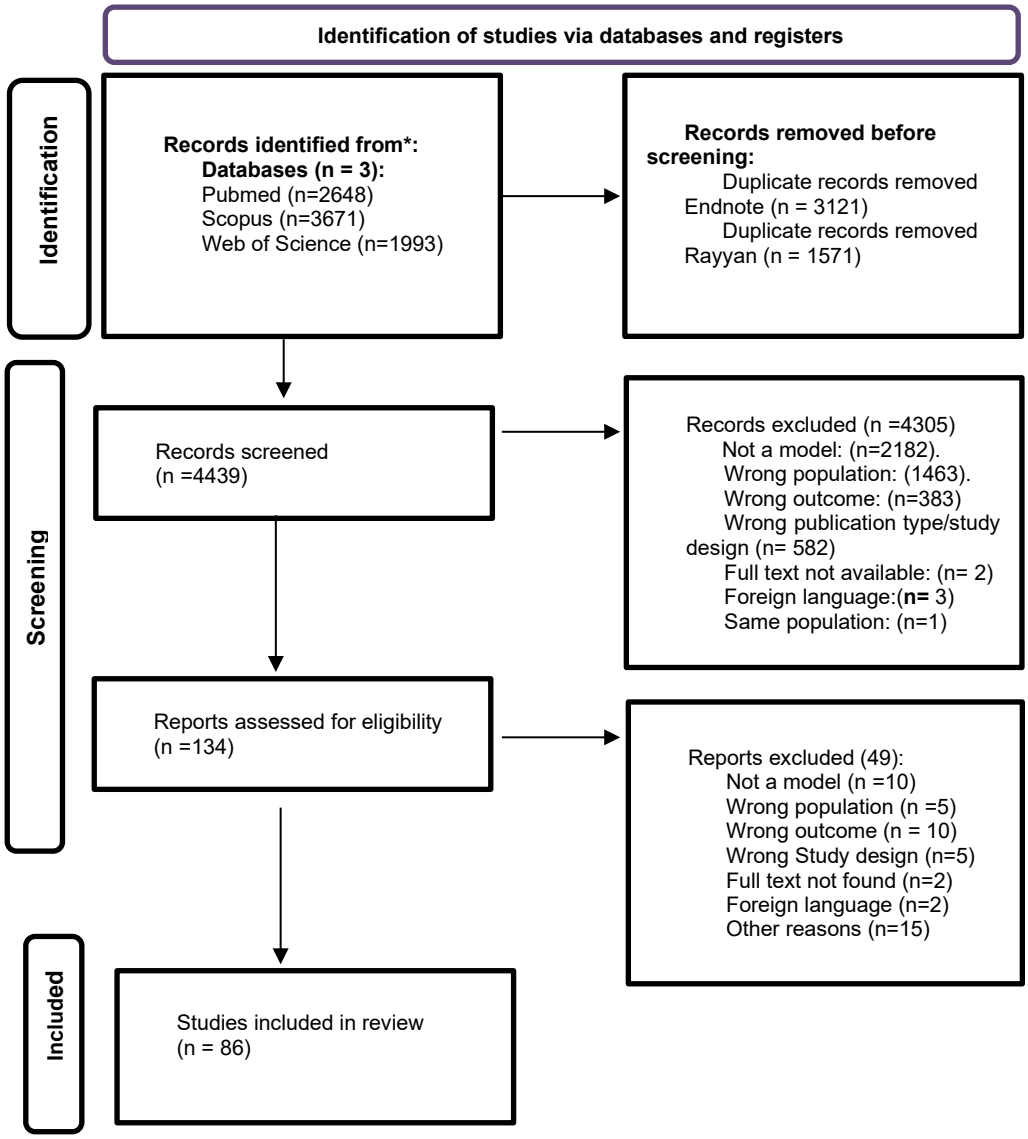
11.6 Synthesis methods

Measures of discrimination (C-statistic, D-statistic) and calibration (calibration plots, total number of observed [O] and expected [E] outcomes, and the O/E ratio) were extracted to enable the quantitative synthesis of model performance(42). A meta-analysis was performed using R (version 4.6.1; R Foundation for Statistical Computing, Vienna, Austria) when at least two independent performance estimates, together with corresponding confidence intervals or standard errors for the C-index and/or calibration measure, were available for the same model. A random-effects model was applied to summarize estimates of model discrimination and calibration. Heterogeneity was investigated by means of forest plot and I2 statistic.

12 Results

12.1 Number of studies identified through the search and number of studies included

Figure 1 - Literature screening flow chart - PRISMA 2020



Study selection and data extraction were performed by 2 independent reviewers (blinded to each other´s assessment). The second reviewer evaluated 20% of the records identified, after removing duplicates, chosen randomly. (Figure 1)

Cohen´s Kappa (interrater reliability) between reviewers was 0.70.

A literature screening update covering the period from 6/2024 to 3/7/2025 was conducted across 3 databases, identifying 535 records (PubMed: 165, Scopus: 257, Web of Science: 113). After removing duplicates with EndNote and Rayyan, 352 records were screened, and 13 studies were included

In summary, the initial bibliographic search retrieved 4,439 citations, and an updated search yielded an additional 312 papers, resulting in a total of **4,751** screened records. Of these, **99** articles fulfilled the eligibility criteria and were included in the final analysis.

These studies described the development of **59** original prognostic models for Idiopathic Pulmonary Fibrosis (IPF) and **26** updated models. Additionally, **80** external validations were performed; among these, **7** involved the application of prognostic models originally developed for health outcomes other than mortality or diseases other than IPF.

The data extracted from the included studies were organized into summary tables and classified according to the prediction model stage into development models and updated /external validation studies (Tables 2 and 3)

Original Models

Table 2- Original Models

Model	Author,Year (Sites, enrolment dates)	TRIPOD type model	Modelling method/algorithm/alternative presentation	Development cohort: Sample (N; Age (SD), % male)	Analysis sample	Events (n)	Validation cohort: Sample (n)/Events (n)	Outcome (binary or time-to-event)	Prediction horizon	Discrimination (C-statistic, 95% CI) VC	Discrimination (C-statistic, 95% CI) DC	Calibration
Bao Y, 2024 (47)	Bao Y, 2024	3	Cox regression/ Nomogram	75; 69 (8.1), 69	-	NI	45/NI	time-to - event	1,2,3 Years OS	Time-dependent ROC analysis (AUC 1y, 2y, 3y - 0.72; 0.76; 0.66)	time-dependent ROC curve analysis (AUC 1y ,2y ,3y -0,68;0,71; 0,68)	Calibration Plot
Cai H, 2022 (48)	Cai H, 2022	1a	Cox regression/ Score System	179; NI, 66.5	-	100	-	time-to-event	1 year OS		time-dependent ROC curve analysis (AUC 1 y - 0.789)	Calibration Plot
Chen H, 2023 (49)	Chen H,2023	1a	Cox regression/ nomogram)	175; NI	-	NI	-	Binary	2,3,5 years OS		time-dependent ROC curve analysis (AUC 0.708)	Absent
du Bois RM 2011, (50)	du Bois RM, 2011 (United States, Europe, Canada, and South Africa; 2000-2001)	1 a	Cox regression /(risk score/nomogram)	1099; 65 (8), 70.2	1854 patient visits	142	.	Time-to-event	1-year mortality		C-statistic NS (0.75; 0.71–0.79)	Observed vs predicted (quintiles); χ^2 test p = 0.316, HL test
Fan G, 2022(51)	Fan G, 2022	3	Cox regression/ Nomogram	75; NI, NI	-	NI	45/NI	time-to-event	1,2,3 years OS	time-dependent ROC analysis (AUC 1 y,2 y,3y - 0.742, 0.729, 0.634)	time-dependent ROC curve analysis (AUC 1 y,2y,3y - 0.670, 0.787, 0.86)	Calibration plot
Feng A, 2023 (52)	Feng A, 2023 (Genes data set) (52)	3	Cox regression/Risk score system	45; NI/84.4	-	13	VC1-51/31; VC2 76/51	time-to-event	1,2,3 years OS	time-dependent ROC analysis (VC1 AUC 1,2 y-0.709;0.715; VC2 AUC 1 y, 2y,3y - 0.638; 0.629;0.709)	time-dependent ROC curve analysis (AUC 1 y, 2y,3y -0.882; 0.814; 0.758)	Absent

Legend: Age (SD) – Media (standard deviation); AUC-area under the curve; DC-derivation cohort; GAP- Gender- Age- Physiology; CRP-Clinical Radiology Physiology Score; OS-overall survival; ROC - Receiver operating characteristic; HL Test - Hosmer-Lemeshow test; N-number; NA – Not applicable; NI – No information; NS-not specified; VC-Validation cohort

Original Models

Model	Author,Year (Sites, enrolment dates)	TRIPOD type model	Modelling method/algorithm/alternative presentation	Development cohort: Sample (N; Media (SD) or Median [IQR]), % male)	Analysis sample	Events (n)	Validation cohort: Sample (n)/Events (n)	Outcome (binary or time-to-event)	Prediction horizon	Discrimination (C-statistic, 95% CI) VC	Discrimination (C-statistic, 95% CI) DC	Calibration
CPI	Wells A, (Uk,2003)(53)	3	Cox regression	106;61.8 (10.9),79	-	NI	VC (split sample) - 106/NI; Ext VC-36/NI	Time to event	5 years mortality	NI	NI	Absent
DOS score (64)	Fukuda CY, 2020 (Brazil-multicentric: 1993-2016)(54)	1b/Internal validation-Bootstrap	Cox regression/Score system	173; 70, NI		154	-	time-to-event	1,2,3 years mortality	-	Uno C-index (70.0, 59.0 – 80.0)	Absent
Gudmundsson E, 2021 (55)	Gudmundsson E, 2021 (Turkey; 2008-2015)(55)	3	Cox regression/Sum Score	142; 66 [42-86], 80		70	145/81	time-to-event	2,3 years mortality	Harrell's C-index -0,72	Harrell's C-index -0,79	Absent
Hamai, 2016(66)	Hamai, 2016 (Hiroshima-japan, NI)(56)	1a	Cox regression	65; 69.3(8.5), 76.9		NI	-	time-to-event	5-year mortality		C-index NS-(0.816 ,0.707-0.923)	Absent
Functional Genomic Model (57)	Huang Y 2015 (USA; 2007)(57)	3	Cox regression	45; 67.1 (8.2), 90		17	VC1-21; VC2-75	time-to-event	1,3,5 years OS	NI	Time - dependent ROC curve analysis (AUC 1,3, 5 y - 0.889, 0.816, 0.725)	Absent
He Y,2022 (58)	He Y,2022 (Gene Expression Omnibus (GEO))(58)	3	Cox regression (Internal validation-Cross-validation)	62; 65.7 (10.1) /83.9		45	VC1 - 64/24; VC2 - 50/31	time-to-event	1,2,3,5 years OS	**VC1** (AUC 1,2,3,5,7, y- 0.813, 0.806, 0.835, 0.635, 0,632); **VC2** (AUC1,2,3 y- 0,815, 0,849; 0,75)	Time-dependent ROC curve analysis (AUC 1,2,3,5, y- 0.869, 0.845, 0.83, 0.936)	Absent

Legend: AUC-area under the curve; CPI -Composite physiologic index; CRP-Clinical Radiology Physiology Score; DC-derivation cohort; DOS - dyspnoea, oxygen, and spirometry; OS-overall survival; ROC -Receiver operating characteristic; NA – Not applicable; NI – No information; NS-not specified; VC-Validation cohort

Original Models

Model	Author,Year (Sites, enrolment dates)	TRIPOD type model	Modelling method/algorithm/alternative presentation	Development cohort: Sample (N; Age Media (SD) or Median [IQR], % male)	Analysis sample	Events (n)	Validation cohort: Sample (n)/Events (n)	Outcome (binary or time-to-event)	Prediction horizon	Discrimination (C-statistic, 95% CI) VC	Discrimination (C-statistic, 95% CI) DC	Calibration
Hou Z, 2022 (59)	Hou Z, 2022 (GEO) (59)	3/bootstrap	Cox regression/ Nomogram	112; Median 69.5 [62.0, 76.0], 83		72	64/41	time-to-event	OS	**C-statistic NS (0.812, 0.703-0.921)**	C-index NS (0.815, 0.769 – 0.861)	Calibration Plot
Huang G,2022 (60)	Huang G,2022 (GEO) (60)	2 a	Cox regression	56; 68(10.06), 83		37	56/39	time-to-event	1,3,5 y OS	Time-dependent ROC estimation (AUC 1,3,5 Y- 0.706, 0.818, and 0.819)	Time-dependent ROC curve analysis (AUC 1,3,5 Y- 0.889, 0.816, 0.725)	Absent
Jacob, 2017 (61)	Jacob, 2017 (UK, USA; 2007-2011) (61)	1b/Bootstrap	Cox regression/ Score system	283; Median 67, 77		210	-	time-to-event	2–3 years Mortality		Harrell's C-index 0.67	Absent
Jacob, 2020 (62)	Jacob, 2020 (UK,Netherlands, Turkey; 2004-2015) (62)	1a	Cox regression	211	422	NI	-	time-to-event	OS		C-statistic NS 0.70	Absent
Kalluri M, 2024 (63)	Kalluri M,2024 (Canada; 2012 - 2018) (63)	1a	Logistic Regression	100; 73 (9), 65	-	28	-	Binary outcome	1-year mortality prediction		C-statistic NS 0.90	Absent
Karampitsakos T 2024 (64)	Karampitsakos T 2024 (GEO: GSE28221) (64)	3	Cox regression	75;69 (8.2), 69		NI	VC1-45/NI, VC2 176/NI; VC3 120/NI	time-to-event	3-4 years OS	NI	Time-dependent ROC curve analysis (NI)	Absent

Legend: AUC-area under the curve; DC-derivation cohort; GEO - Gene Expression Omnibus; GSE -Geo Series Experiment; OS-overall survival; ROC -Receiver operating Characteristic; NA – Not applicable; NI – No information; NS-not specified; VC-Validation cohort

Original Models

Model	Author,Year (Sites, enrolment dates)	TRIPOD type model	Modelling method/algorithm/ alternative presentation	Development cohort: Sample (N; Age (SD), % male)	Analysis sample	Events (n)	Validation cohort: Sample (n)/Events (n)	Outcome (binary or time-to-event)	Prediction horizon	Discrimination (C-statistic, 95% CI) VC	Discrimination (C-statistic, 95% CI) DC	Calibration
CRP score	King TE, 2001 (Denver; 1982-1996)(65)	1 b /Bootstrap	Cox regression/ sum score	238; 61.4 (10.5), 64		133		time-to-event	5-year survival		NI	Absent
BFR staging	Kishaba T, 2016 (Japan)(66)	1a	Cox regression/ score system	65; 71.9 (1.8), 63		39	-	time-to-event	NI		C-statistic NS (0.765)	Absent
Japanese Revised System	**Kondoh Y, 2017 (Japan - Multicentric; 2005-2007)(67)**	1 b/ Bootstrap	NI/stage system	215; 67.2 (8.2), 80		116	-	time-to-event			Harrell's C-index (0.677, NI)	Absent
Lee SM,2018 (68)	Lee SM,2018 (Korea; 2004-2010)(68)	1a	Cox regression	144; 63.0 (7.6), 79		106	-	time-to-event			Harrell's C-index (0.768; 0.707–0.829)	Absent
GAP	Ley B, 2012 (San Francisco, 2001-2010)(69)	3	Cox regression/score and stage system	228; 69.7 (8.7), 72.8		89	330/186	time-to-event	1,2,3Y mortality	Harrell's C-index (69.1, 64.3-73.5)	Harrell's C-index (70.8, 63.7-75.0)	Table O/E
Kam, 2019 (70)	Kam MLW 2019 (Singapore; 2012-2017) (70)	1b	Cox regression/normogram	181; 66 (11.3), 60		48		time-to-event	1,2,3 Y mortality		Harrell's C-index 0,817	Calibration Plot

Legend: AUC-area under the curve; BFR- Body Mass Index, forced vital capacity, respiratory hospitalizations; CRP-Clinical Radiology Physiology Score; DC-derivation cohort; GAP - Gender Age Physiology; OS-overall survival; ROC -Receiver operating Characteristic; NA – Not applicable; NI – No information; NS-not specified; VC-Validation cohort

Original Models

Model	Author,Year (Sites, enrolment dates)	TRIPOD type model	Modelling method/algorithm/ alternative presentation	Development cohort: Sample (N; Age (SD), % male)	Analysis sample	Events (n)	Validation cohort: Sample (n)/Events (n)	Outcome (binary or time-to-event)	Prediction horizon	Discrimination (C-statistic, 95% CI) VC	Discrimination (C-statistic, 95% CI) DC	Calibration
Liao 2024 (71)	Liao YW 2024 (Tawain; 2018-2019)(71)	1a	Cox regression	104; 69 [62–74], 73		12		time-to-event			None	Absent
Inflammation and Hypoxia-Related Genes model	Liu J 2022 (GEO - GSE70866) (72)	3	Cox regression/Nomogram	112; NI, 83		76	64/24	time-to-event	1,3,5 y OS	time-dependent ROC curve analysis (AUC 1,2,3,5 Y 0.78, 0.79, 0.80, 0.91)	time-dependent ROC curve analysis (AUC 1,2,3,5 Y- 0.83, 0.82, 0.80, 0.81,0.94)	Calibration Plot
Ferroptosis-Related Genes model	Li M 2021 (GEO-GSE70866) (73)	2a	Cox regression/Nomogram	124; NI, 82		NI	52/NI	time-to-event	1,2,3 years OS	time-dependent ROC curve analysis (AUC 1-, 2-, 3-y: 0.891, 0.870, 0.678)	time-dependent ROC curve analysis (AUC 1-, 2-, 3-y: 0.784, 0.801, 0.805	Calibration Plot
Plasma cells-related gene signature ST6GAL1 model	Lin F 2025 (GEO: GSE70866) (74)	3	Machine learning techniques	112;67.97 (10.07), 83		76	64/24	time-to-event	1,2,3,4,5-year OS	Harrell's C-index 0:711	Harrell's C-index 0:76	Calibration Plot
Li X, 2021(75)	Li X, 2021 (GEO-GSE70866) (75)	3	Cox regression/ Nomogram	176; NI, 66.5		NI	VC1 57/NI; VC2121/NI; VC3-179/NI	time-to-event	1-, 2-, 3, 4,5 -year OS	NI	Time dependent ROC curve analysis AUC 1,2,3 Y: 0.789, 0.768, 0754, 0798, 0.913	Absent
Li Y2023 (76)	Li Y 2023 (GEO – GSE 28042) (76)	3	Cox regression/Nomogram	75; 69 (8.16), NI		NI	VC1 -75/NI; VC2-45/NI; VC3 -176/NI	Time-to-event	1, 2 and 3year OS	Time dependent ROC curve analysis VC 2 AUC 1,2,3 Y :0.600, 0.671, 0.683	Time dependent ROC curve analysis AUC 1,2,3 Y :0.666, 0.651, 0.717	Absent

Legend: AUC-area under the curve; DC-derivation cohort; GEO - Gene Expression Omnibus; GSE -Geo Series Experiment; OS-overall survival; ROC -Receiver operating Characteristic; NA – Not applicable; NI – No information; VC-Validation cohort

Original Models

Model	Author,Year (Sites, enrolment dates)	TRIPOD type model	Modelling method/algorithm/ alternative presentation	Development cohort: Sample (N; Age (SD), % male)	Analysis sample	Events (n)	Validation cohort: Sample (n)/Events (n)	Outcome (binary or time-to-event)	Prediction horizon	Discrimination (C-statistic, 95% CI) VC	Discrimination (C-statistic, 95% CI) DC	Calibration
Prognostic Index Based on Inflammation-Related Gene Analysis	Lu Y 2021(GEO; GSE28042) (77)	3	Cox regression/Nomogram	75; 75, NI		NI	45/NI	time-to-event	1-, 2-, and 3-year OS	Time dependent ROC curve analysis AUC 1,2,3 Y:0.678, 0.695«6, 0.671	Time dependent ROC curve analysis AUC 1,2,3 Y:0.611, 0.695, 0.681	Absent
Extracellular matrix-related genes model	Luo H 2023 (GEO; GSE70866) (78)	3	Cox regression/Nomogram	112; Median 69.5 [62, 76], 52.8		76	64/24	time-to-event	1-, 2-, and 3-year OS (nomogram)	C-statistic NS 0.727. 0.619–0.834)	C-statistic NS 0.707 (0.649 –0.76)	Calibration Plot
Cold exposure-related genes (CERGs) model	Luo P 2025 (GEO)(79)	3	Cox regression/Nomogram	62; NI		NI	VC1-50/NI; VC2-64/NI	time-to-event	1-, 2-, and 3-year OS	Time dependent ROC curve analysis AUC 1,3 Y: VC1-0.82, 0,78; VC2 - 0,71;0,86	Time dependent ROC curve analysis AUC: 1,3,5 Y: 0.81; 0.79; 0.91	Absent
Fatty acid metabolism-related genes model	Lyu Y, 2022 (GEO: GSE70866)(80)	3	Cox regression/Nomogram	112; 69.5 (62, 76), 52.8		76	64/24	time-to-event	1,2,3 y OS	Time dependent ROC curve analysis AUC 1,2,3 Y 0.622, 0.679, and 0.723	Harrell's C-index Nomogram 0.7642 (0.7125 –0.7962)	Calibration Plot
Mogulkoc N, 2001 ((81)	Mogulkoc N, 2001 (UK)(81)	1a	Logistic Regression	115; 56 (8),70		46		Binary outcome	2 y OS		Time dependent ROC curve analysis AUC 2 Y: 0.907 (CI 0.831 to 0.983)	Absent
4MGS or slow 4MGS model	Nolan G, 2019 (UK; 2015-2016)(82)	1a	Cox regression	130: 72(7), 83		23		time-to-event	1 year		Time dependent ROC curve analysis 1 Y 0.81 (0.72-0.91)	Absent

Legend: AUC-area under the curve; DC-derivation cohort; GEO - Gene Expression Omnibus; GSE -Geo Series Experiment; OS-overall survival; ROC -Receiver operating Characteristic; NA – Not applicable; NI – No information; NS- not specified; VC-Validation cohort; 4MGS – four-meter gate speed

Original Models

Model	Author,Year (Sites, enrolment dates)	TRIPOD type model	Modelling method/algorithm/alternative presentation	Development cohort: Sample (N; Age (SD), % male)	Analysis sample	Events (n)	Validation cohort: Sample (n)/Events (n)	Outcome (binary or time-to-event)	Prediction horizon	Discrimination (C-statistic, 95% CI) VC	Discrimination (C-statistic, 95% CI) DC	Calibration
Polat G 2024	Polat G, 2024 (Turkey; 2016-2021)(83)	1a	Cox regression	119; 67.9(7.07), 80.7		46		time-to-event	NI		Time-dependent ROC curve analysis AUC 0.82	Absent
Prasse A, 2019	Prasse A, 2019 (GEO)(84)	3	Cox regression	62; 67.46(9.1), 85		45	VC1 50/31; VC2 64/24	time-to-event	18 months, OS	C-statistic NS VC1 (0.66, 0.55–0.76) VC2 (0.63, 0.54–0.72)	C-statistic NS 0.67 (0.62–0.71)	Absent
PCMI model	Richards T, 2012 (USA)(85)	2b	Cox regression	140; 67.2(8.3), 72.1		73	101/43	time-to-event	1-year mortality	NI	Uno C statistic (from 0.73–0.84)	Absent
DDS index and staging system	Soares MR 2015 (Brazil - multicentric; 1996-2011) (86)	1a	Cox regression	120; 68.6(7)		80		time-to-event	2, 4 Y OS		Uno C statistic 0.785 (0.705-0.865).	Absent
Song 2013	Song, 2013 (Korea; 2004-2008)(87)	1b/ Bootstrap	Cox regression	118; 62.8 (8.1), 85		64		time-to-event	1 y OS		C-statistic NS 0.731 (0.683-0.818)	Absent
Tian Y 2025	Tian Y 2025 (China; 2017-2020)(88)	2a	Cox regression/Nomogram	140, NI		NI	71/NI	time-to-event	1,2 ,3 Y OS	Time dependent ROC curve analysis AUC 1,2 Y:0.754, 0.759	Time dependent ROC curve analysis AUC 1,2 Y: 0.866; 0.916	Calibration plot

Legend: AUC-area under the curve; DDS – Dyspnoea, Diffusion, Spirometry; DC-derivation cohort; GEO - Gene Expression Omnibus; OS-overall survival; ROC - Receiver operating Characteristic; NA – Not applicable; NI – No information; NS-not specified; PCMI -personal clinical and molecular mortality prediction index; Y-Year; VC-Validation cohort

Original Models

Model	Author,Year (Sites, enrolment dates)	TRIPOD type model	Modelling method/algorithm/alternative presentation	Development cohort: Sample (N; Age (SD), % male)	Analysis sample	Events (n)	Validation cohort: Sample (n)/Events (n)	Outcome (binary or time-to-event)	Prediction horizon	Discrimination (C-statistic, 95% CI) VC	Discrimination (C-statistic, 95% CI) DC	Calibration
TORVAN	Torrisi SE, 2019 (Italy, Netherlands, Germany; 2004-2016)(89)	3	Cox regression	476; 68.8 (8.41), 76.7		272	190	time-to-event	1-5 Y OS	Harrell's C-index 70.3 (66.1–74.5	Harrell's C-index 71.3 (68.1–74.6)	Calibration plot
Tseng JM 2022	Tseng JM 2022 (Tawain; 2017-2020)(90)	1a	Cox regression/score system	40; 75.6 (8.3), 77.5		14		time-to-event	1 Y Mortality		None.. Global Performance R^2	Calibration like comparisions instead
Yu WS 2018	Yu WS 2018 (South Korea; 2004-2014)(91)	1a	Cox regression	59; 56.4(7.6), 69.5		22		time-to-event	30-day and 6-month mortality		C-statistic NS 0.758 (0.653-0.863)	Absent
Wang E 2022	Wang E 2022 (GEO-GSE70866)(92)	3	Cox regression/nomogram	112; 68.0(10.0), 83		76	64/24	time-to-event	1,2,3 Y OS	Time dependent ROC curve analysis AUC 1,2,3 Y: 0.795, 0.782, and 0.794	Harrell's C-index 0.76 (0.73 – 0.79)	Absent
Wang X 2024	Wang X 2024 (GEO: GSE28042)(93)	3	Cox regression/ nomogram	75; 69.0 (8.2), 69		NI	112/NI	time-to-event	1-, 2- 3-year OS	Time dependent ROC curve analysis AUC 1,2,3 Y (0.66, 0.73; 0.66)	Time dependent ROC curve analysis AUC 1,2,3 Y: 0.76; 0.87; 0.88	Calibration Plot
Wang Z, 2022	Wang Z 2022 (GEO-GSE28042)(94)	3	Cox regression	75; NI		NI	60/NI	time-to-event	1,2,3,4 Y OS	Time dependent Roc curve analysis AUC 1,2 Y 0,73; 0.70	Time dependent Roc curve analysis AUC 1,2,3,4 Y: 0.63; 0.77, 0.85, 0.95	Absent

Legend: AUC-area under the curve; DC-derivation cohort; GEO - Gene Expression Omnibus; OS-overall survival; ROC -Receiver operating Characteristic; NA – Not applicable; NI – No information; NS-not specified; TORVAN -*TOR* Torrisi, *VAN* Vancheri Y-Year; VC-Validation cohort

Original Models

Model	Author,Year (Sites, enrolment dates)	TRIPOD type model	Modelling method/algorithm/alternative presentation	Development cohort: Sample (N; Age (SD), % male)	Analysis sample	Events (n)	Validation cohort: Sample (n)/Events (n)	Outcome (binary or time-to-event)	Prediction horizon	Discrimination (C-statistic, 95% CI) VC	Discrimination (C-statistic, 95% CI) DC	Calibration
CTPF Model	Wu X 2022 (China; 2011-2020)(95)	2b	Cox regression/Nomogram	165; 64.1 (7.9), 95		65	41/Nl	time-to-event	1,2,3 Y OS	Time dependent Roc curve analysis AUC 1,2,3 Y: 92.0 [83.4,100.0]; 75.0 [57.1,92.9]; 76.0 [56.8,95.2	Time dependent Roc curve analysis AUC 1,2,3 Y: 74.3 [63.2,85.4]; 78 [70.2,85.9]; 72.8 [58.3,87.3]	Calibration Plot
BAL cell-associated gene signature model	Xia Y 2021 (GEO - GSE70866)(96)	3	Cox regression/nomogram	122; 67.97(10.07), 8.3		NI	64/NI	time-to-event	1,2,3 Y OS	Time dependent Roc curve analysis AUC 1,2,3 Y 0.760, 0.717, 0.748	C-statistic NS 0.72 (0.66–0.77) Time dependent Roc curve analysis AUC 1,2,3 Y 0.773, 0.772, 0.752	Calibration plot
Xiang Y 2024	Xiang Y 2024 (GEO: GSE70866)(97)	3	Cox regression	176; NI		NI	60/NI	time-to-event	NI	Time dependent Roc curve analysis AUC:0.66	Time dependent Roc curve analysis AUC: 0.741[0.668,0.810]	Calibration plot
Inflammation related genes model	Yin Y-Q 2022 (GEO: GSE70866, GSE28042)(98)	3	Cox regression/nomogram	132; NI		NL	VC1-64/NI; VC2-75/NI	time-to-event	1,3,5 YOS	Time dependent Roc curve analysis AUC 1, 3, 4 year 0.786, 0.713, 0.793	C-statistic NS C-index (>0.7) Time dependent Roc curve analysis AUC 1, 3,5 Y 0.784, 0.835, 0.921	Absent
CPB index and CPB stage	Zhang X 2022 (China; 2000-2015) (99)	3	Cox regression/ stage system /nomogram	374; NI, 84		321	324/75	time-to-event	1,2,3 Y OS	c-statistic NS : C-index 0.635 (0.558–0.712) CPB stage: C-index0.619 (0.544–0.694)	c-statistic NS CPB index: C-index 0.632 (0.574 – 0.690) CPB stage: C-index: 0.589 (0.537–0.642)	Calibration plot
Zhao J2023	Zhao J 2023 (GEO: GSE70866)(100)	3	Cox regression/nomogram	112; NI		/NI	64/NI	time-to-event	1,2,3 Y OS	Time dependent Roc curve analysis AUC 1,2,3 Y: 0.814, 0.814; 0.808	Time dependent Roc curve analysis AUC 1,2,3 Y: 0.872; 0.905; 0.868	Calibration Plot

Legend: AUC-area under the curve; BAL-bronchoalveolar lavage cell; CPB - Clinical-Physiological-Biomarker; CTPF – Computed Tomography fibrotic stages and the pulmonary function and physiological features grades; DC-derivation cohort; GEO - Gene Expression Omnibus; OS-overall survival; ROC -Receiver operating Characteristic; NA – Not applicable; NI – No information; NS-not specified; Y-Year; VC-Validation cohort

Original Models

Model	Author,Year (Sites, enrolment dates)	TRIPOD type model	Modelling method/algorithm/alternative presentation	Development cohort: Sample (N; Age (SD), % male)	Analysis sample	Events (n)	Validation cohort: Sample (n)/Events (n)	Outcome (binary or time-to-event)	Prediction horizon	Discrimination (C-statistic, 95% CI VC	Discrimination (C-statistic, 95% CI) DC	Calibration
3 Epithelial-Mesenchymal Transition and Immune-Related model	Zheng J 2022 (GEO: GSE7 0866)(101)	3	Cox regression/nomogram	176; NI		NI	118/NI	time-to-event	1.2.3 Y OS	Time dependent Roc curve analysis AUC 1,2,3 Y:0.691, 0.734, 0.684	Time dependent Roc curve analysis AUC 1,2,3 Y: 0.789; 0.743; 0.722	Calibration Plot
TRIM family gene signature from BAL-based Model	Zhou M 2022 (GEO-GSE70866)(102)	3	Cox regression/nomogram	62; NI		NI	VC 1-50/NI VC264/NI-	time-to-event	1.2.3 Y OS	AUC 1,2,3 Y: VC1 0.742, 0.805, 0.818; VC2: 0.912, 0.867, 0.881	AUC 1,2,3 Y: 0.832, 0.786; 0.750	Calibration Plot
Five Senescence-Related Genes model	Zhong C 2024 (GEO: GSE70867)(103)	2b	Cox regression/nomogram	123; NI		NI	53/NI	time-to-event	1.2.3 Y OS	AUC 1,2,3 Y 0.721, 0.802 0.739	AUC 1,2,3 Y: 0.797, 0.778, 0.751	Calibration Plot
Endoplasmic reticulum stress-related gene prediction model	Zhu H 2024 (GEO; GSE70866, GSE28042, GSE110147)(104)	1 a	Cox regression	209; NI	-	113	-	time-to-event	1.2.3 Y OS		AUC 1,3,5 Y: 0.636; 0.676; 0.867	Calibration Plot
IC4	Zinellu A 2022 (France, Italy; 2006-2015)(105)	1a	Multiplicative combination	90; 70.1(6.3), 88	-	62	-	time-to-event	1.2.3,4 Y mortality		AUC 1,2,3,4Y (0.864; 0.776-0.927); (0.793 0.695 - 0.871); (0.856 ,0.766 - 0.921); (0.859, 0.770-0.924)	Absent

Legend: AUC-area under the curve; BAL-bronchoalveolar lavage cell; DC-derivation cohort; GEO - Gene Expression Omnibus; IC4 - combined non-dimensional index incorporating FVC%, DLCO%, BMI and 6MWT; OS-overall survival; ROC -Receiver operating Characteristic; NA – Not applicable; NI – No information; Y-Year; VC -Validation cohort

External validation studies and updated models

Table 3 - External validation studies and updated models

Model	Author,Year (Sites, enrolment dates)	TRIPOD type model	Modeling method/algorithm/alternative presentation	Development cohort: Sample (N; Age (SD), % male)	Analysis sample	Events (n)	Validation cohort: Sample (n)/Events (n)	Outcome (binary or time-to-event)	Prediction horizon	Discrimination (C-statistic, 95% CI) VC	Discrimination (C-statistic, 95% CI) DC	Calibration
du Bois RM, 2014 (106)	du Bois RM, 2014	4	Cox regression	748	1156 visits	79		time-to-event	1-year all-cause mortality		C-statistic NS (0.8; 0.76-0.85)	Absent
CPI	Wells A, (Uk,2003)(53)	3	Cox regression	106;61.8 (10.9),79	-	NI	VC (split sample)-106/NI; Ext VC-36/NI	Time to event	5 years mortality	NI	NI	Absent
	Fisher JH, 2017 (107)	4	Fine-Gray competing risk regression model	302; 61 (8), 72	-	63	-	Time to event	1-year mortality	Time-dependent ROC analysis (AUC 1 y - 0.75)		Absent
	Lee SH, 2018 (108)	4	Cox regression	832; 66.4 (9.3),72		272		time-to-event	1,2,3-year mortality	Time-dependent ROC curve analysis ;(AUC1,2,3 y - 0.647 (0.6-07), 0.647(0.6-0.69), 0638 (0.6-0.68))		Absent
	Sharp 2017 (UK; 2007-2014)(109)	4	Cox regression	104; Median 73.5 [71.9–75.0], 82		56		time-to-event	1,2 years	Time-dependent ROC curve analysis (AUC 1, 2 Y - 0.657; 0.761)		Absent
LAS	Fisher JH, 2017 (Canada-multicentric; 2003-2014)(107)	4	Fine-Gray competing risk regression model	302; 61 (8), 72		63		time-to-event	1 year mortality	Time-dependent ROC curve analysis (AUC 1 y - 0.78)		Absent
RISE	Fisher JH, 2017 (Canada-multicentric; 2003-2014)(107)	4	Fine-Gray competing risk regression model	302; 61 (8), 72		63		time-to-event	1-year mortality	Time-dependent ROC curve analysis (AUC 1 y - 0.78)		Absent

Legend: AUC-area under the curve; ; CPI -Composite physiologic index ; DC-derivation cohort; GEO - Gene Expression Omnibus; LAS – Lung Allocation Score; OS-overall survival; RISE -RIsk Stratification score; ROC -Receiver operating Characteristic; NA – Not applicable; NI – No information; NS- not specified; Y-Year; VC - Validation cohort

External validation studies and updated models

Model	Author,Year (Sites, enrolment dates)	TRIPOD type model	Modeling method/algorithm/alternative presentation	Development cohort: Sample (N; Age (SD), % male)	Analysis sample	Events (n)	Validation cohort: Sample (n)/Events (n)	Outcome (binary or time-to-event)	Prediction horizon	Discrimination (C-statistic, 95% CI VC	Discrimination (C-statistic, 95% CI) DC	Calibration
GAP	Ley B, 2012 (San Francisco, 2001-2010)(69)	3	Cox regression/score and stage system	228; 69.7 (8.7), 72.8		89	330/186	time-to-event	1,2,3 years mortality	Harrell's C-index (69.1,64.3–73.5)	Harrell's C-index (70.8, 63.7-75.0)	Table O/E
	Abe M, 2020 (Japan; 2015-2018)(110)	4	Cox regression	89; 71.3 (6.2), 76	-	23		time-to-event		NI		Table O/E
	AlDalilah Y, 2022 (UK; 2007-2017)(111)	4	Cox regression	59; 70 (8.8), 83.1		30		time-to-event		NI		Absent
	Bocchino ML, 2023 (Italy:2012-2019)(112)	4	Cox regression	235; 69.8 (7.1), 76.2		116		time-to-event	OS	Harrell's C-index (0.635; 0.575-0.686)		Absent
	Chandel A, 2023 (USA; 2007-2020) (113)	4	Cox regression	281; 72, 71.5		NI		time-to-event	OS	Harrell's C-index (0.676; 0.635-0.717		CITL
	Fisher JH, 2017 (Canada-multicentric; 2003-2014)(107)	4	Cox regression	302; 61 (8), 72		63		time-to-event	1 year	Time-dependent ROC curve analysis AUC 1 y -0.67		Absent
	Harari S, 2019 (Italy-Multicentric; 2011-2013)(114)	4	Cox regression	68; NI, 76		22		time-to-event	1,2,3-year mortality	Harrell's C-index (0.74, 0.57-0.93)	.	Calibration plot/ HL test - p=0.088, 0.218, and 0.778 at 1, 2, and 3 years.

Legend: AUC-area under the curve; CITL-calibration in the large; GAP- gender age and physiology; DC-derivation cohort; OS-overall survival; HL test -- Hosmer-Lemeshow test; ROC -Receiver operating Characteristic; NA – Not applicable; NI – No information; Y-Year; VC -Validation coh

External validation studies and updated models

Model	Author, Year (Sites, enrolment dates)	TRIPOD type model	Modeling method/algorithm/ alternative presentation	Development cohort: Sample (N; Age-Media (SD) or Median [IQR], % male))	Analysis sample	Events (n)	Validation cohort: Sample (n)/Events (n)	Outcome (binary or time-to-event)	Prediction horizon	Discrimination (C-statistic, 95% CI) VC	Discrimination (C-statistic, 95% CI) DC	Calibration
GAP	Kim ES 2015 (Korea; 2005 - 2013)(115)	4	Cox regression	268; 65.9 (9.6), 67.5		157		time-to-event	1,2,3 Y mortality	Time-dependent ROC analysis (AUC 1,2,3 Y: 0.74, 0.71, 0,68)		Calibration plot/HL test
	Kreuter M 2024 (Multicentric: ASCEND; CAPACITY 004/ 006) (116)	4	Cox regression	1247; 67.2(7.5), 74.4		64		time-to-event	1Y mortality	Harrell's C-index (0.730, 0.705–0.725)		Absent
	Ley B, 2015 (Multicentric; 2000-2001; 2008-2012)(117)	4	Cox regression	1109; 66.2 (7.7), 71.3		128		time-to-event	1, 2 Y mortality	Harrell's C-index (0.676; 0.625–0.728)		HL test
	Lee SH, 2018 (Korea-Multicentric; 2003-2007)(108)	4	Cox regression	832; 66.4 (9.3), 72		272		time-to-event	1,2,3 Y	time -dependent ROC curve analysis 1,2,3 Y 0.619 (0.57-0.67); 0.625(0.58-0.67); 0.610(0.57-0.65)		Absent
	Lee J H, 2023 (Korea-multicentric; 2016-2021)(118)	4	Cox regression	966; 71.23 (7.69), 81.1		257		time-to-event	1,2,3 Y	C-statistic NS (0.674; 0,650–0,698)		Calibration plot/HL test
	Lee T, 2024 (Korea-multicentric: 2011–2021; 2007–2010)(119)	4	Cox regression	1007; 70 [64–75], 76	187	NI		time-to-event	OS	Harrell's C-index (:0,73, 0,67–0,78)		Absent
	Mora-Cuesta VM ,2025 (Spain: 2017 - 2023)(120)	4	Cox regression	413; 61.4 [56.1-64.4], 75.5		135		time-to-event		NI		Absent

Legend: AUC-area under the curve; CITL-calibration in the large; GAP- gender age and physiology; DC-derivation cohort; OS-overall survival; HL test -- Hosmer-Lemeshow test; ROC -Receiver operating Characteristic; NA – Not applicable; NI – No information; Y-Year; VC -Validation cohort

External validation and updated studies

Model	Author, Year (Sites, enrolment dates)	TRIPOD type model	Modeling method/algorithm/ alternative presentation	Development cohort: Sample (N; Age-Media (SD) or Median [IQR], % male)	Analysis sample	Events (n)	Validation cohort: Sample (n)/Events (n)	Outcome (binary or time-to-event)	Prediction horizon	Discrimination (C-statistic, 95% CI) VC	Discrimination (C-statistic, 95% CI) DC	Calibration
GAP	Nam J G 2023 (Korea; 2005-2009)(121)	4	Cox regression	130; 68 [62–73]; 65		23		time-to-event		Uno C-statistic (0.72; 0.68–0.76)		Absent
	Nishikiori H 2020 (Japan; 2003-2007)(122)	4	Cox regression	326; 69.5 [64.0-76.0]; 75		NI	VC1:326/NI	time-to-event	OS	Harrell's C-index 0.696		Absent
	Song, 2019 (China; 2002-2015)(123)	4	Cox regression	380; 63(9.25), 86.3		274		time-to-event		NI		Absent
	Ohkubo H ,2018 (Japan;2008-2013)(124)	4	Cox regression	175; 66.1(7.9), 81.1		94		time-to-event	OS	Harrell's C-index (0.727, 0.719-0.738)		Absent
	Park J ,2021 (South Korea; 2006-2016)(125)	4	Cox regression	104; 68.2 (9.4), 68		56		time-to-event	OS	Harrell's C-index (0.79, 0.71-0.86)		Absent
	Sharp, 2017 (UK; 2007-2014)(109)	4	Cox regression	104;73.5 [71.9–75.0], 82		56		time-to-event	1,2 Y	Time-dependent ROC curve analysis AUC 1,2 Y: 0.636, 0.722		Absent
	Suzuki Y, 2021 (Japan-multicentric; 2009-2020)(126)	4	Cox regression	229; 72.0 [67.5–72.0], 81.2		122		time-to-event	OS	C-statistic NS (0.675; 0.630–0.720)		Absent

Legend: AUC-area under the curve; CITL-calibration in the large; GAP- gender age and physiology; DC-derivation cohort; OS-overall survival; HL test -- Hosmer-Lemeshow test; ROC -Receiver operating Characteristic; NA – Not applicable; NI – No information; NS-not specified; Y-Year; VC -Validation cohort

External validation and updated studies

Model	Author, Year (Sites, enrolment dates)	TRIPOD type model	Modeling method/algorithm/ alternative presentation	Development cohort: Sample (N; Age-Media (SD)or Median [IQR], % male)	Analysis sample	Events (n)	Validation cohort: Sample (n)/Events (n)	Outcome (binary or time-to-event)	Prediction horizon	Discrimination (C-statistic, 95% CI) VC	Discrimination (C-statistic, 95% CI) DC	Calibration
GAP	Takuma S 2024 (Japan-multicentric; 2009-2022)(127)	4	Cox regression	416; 71 [66–76], 73.8		189		time-to-event	6 months, 1,2,3,4,5 OS	Harrell's C-index 6 months, 1,2,3,4,5Y (0.658, 0.715, 0.713,0.680, 0.701, 0.734)		Absent
	Torrisi SE, 2019 (Italy, Netherlands, Germany; 2004-2016)(89)	4	Cox regression	476; 68.8 (8.41) /76,73		272		time-to-event	OS	Harrell's C-index (68.8; 65.3-74.2)		Absent
	Whi W 2025 (Korea; 2007-2020)(128)	4	NI	224; 66.87(10.1), 74.6		NI		time-to-event	OS	C-statistic NS 0.59		Absent
	Win T 2018(UK; 2008-2017)(129)	4	Cox regression	113; 79(9), NI		54		time-to-event binary outcome	OS	C-statistic NS 0,667 (0,567-0.767)		Absent
	Zhang X,2022 (China:2015-2019)(130)	4	Cox regression	212; 65 [60.00-71.00], 83.49		58		time-to-event	OS	Harrell's C-index (0.736,0.667–0.864)	Calibration plot	Calibration plot
GAP_qCT-T_ qCT-P	AlDalilah Y, 2022 (UK; 2007-2017)(111)	4	Cox regression/stage score	59; 70 (8.8)		30		time-to-event	9 Y mortality		Absent	Absent
GAP_qCT-T	AlDalilah Y, 2022 (UK; 2007-2017)(111)	4	Cox regression/stage score	59; 70 (8.8), 83.1		30		time-to-event	9 Y mortality		Absent	Absent

Legend: AUC-area under the curve; CITL-calibration in the large; GAP- gender age and physiology; DC-derivation cohort; OS-overall survival; qCT-T - Quantitative Computed Tomography texture analysis; qCT-P - Quantitative CT pattern analysis; ROC-Receiver operating characteristic; NA – Not applicable; NI – No information; NS-not specified; Y-Year; VC -Validation

External validation studies and updated models

Model	Author,Year (Sites, enrolment dates)	TRIPOD type model	Modeling method/algorithm/alternative presentation	Development cohort: Sample (N; Age-Media (SD)or Median [IQR], % male))	Analysis sample	Events (n)	Validation cohort: Sample (n)/Events (n)	Outcome (binary or time-to-event)	Prediction horizon	Discrimination (C-statistic, 95% CI) VC	Discrimination (C-statistic, 95% CI) DC	Calibration
DO-GAP	Chandel A, 2023 (USA; 2007-2020)(131)	4/ Split sample/external validation	Cox regression/ staging system	281; 72, 71.5		NI	VC 281(NI; External cohort 108/NI	time-to-event	3-year mortality	Internal VC Harrell´s C-index (0.752; 0.701-0.802), Extern VC - (0.768; 0.664 to 0.873)	Harrell's C-index (0.756; 0.706-0.806)	Calibration plot; slope Wald χ^2 test p = 0.316,
TMNC-GAP	Desai UD, 2018 (India; 2012-2013)(132)	4	NI/Score System	70; 62, NI		49		Binary outcome	1,2,3-year mortality		NI	Absent
GAP_TBR_qCT	Fraioli F,2019 (UK; 2008-2017)(133)	4	Cox regression	91; 70 (8.9), 86		47		time-to-event	9 Y		NI	Absent
GAP _CIRBP	Hozumi H, 2021 (Japan-multicentric; 2004-2016)(134)	4	Cox regression	95; 70.3 (8), 87		51		time-to-event	2 Y mortality		Time dependente ROC curve analysis 0.747	Absent
ILD-GAP/normogram	Kam MLW 2019 (Singapore; 2012-2017)(70)	4	Cox regression	181; 66 (11.3), 60		48		time-to-event	1,2,3 Y		Harrell's C-index Normgram (0.777, NI)	Calibration Plot
mGAP	Kondoh Y, 2017 (Japan - Multicentric; 2005-2007)(67)	4	NA	199; NI/NI		152		time-to-event	OS		Harrell's C-index (0.645, 0.600–0.686)	Absent
ILD GAP-O2	Khor Y H, 2021 (Australia; Canada; Switzerland)(135)	4	Cox regression	669; NI[NI]/NI		NI		time-to-event	1 Y		Harrell's C-index 0.86. NI	Goodness-of-fit likelihood ratio test

Legend: AUC-area under the curve; DO-GAP – Distance Oxygen Gender Age physiology; DC-derivation cohort; GAP- gender age and physiology; GAP_CIRBP – GAP, Cold-inducible RNA-binding protein; ILD-GAP: Interstitial Lung Disease GAP; ILD GAP-O2: Interstitial Lung Disease GAP, oxyhemoglobin saturation; GAP_TBR_qCT: GAP, Target to background ratio, quantitative computed tomography; mGAP-modified GAP; NA – Not applicable; NI – No information, ROC -Receiver operating Characteristic; TMNC-GAP - The New Modified Classification of Gender Age Physiology; Y-Year; VC -Validation cohort

Model	Author,Year (Sites, enrolment dates)	TRIPOD type model	Modeling method/algorithm/alternative presentation	Development cohort: Sample (N; Age - Media (SD)or Median [IQR], % male)	Analysis sample	Events (n)	Validation cohort: Sample (n)/Events (n)	Outcome (binary or time-to-event)	Prediction horizon	Discrimination (C-statistic, 95% CI) VC	Discrimination (C-statistic, 95% CI) DC	Calibration
Modified blood cell GAP	Kreuter M 2024 (Multicentric: ASCEND; CAPACITY 004/ 006) (116)	4	Cox regression	1247; 67.2(7.5), 74.4		64		time-to-event	1 Y		Harrell's C-index (0.765, 0.730–0.755)	Absent
GAP index_BMI	Lacedonia D 2023(Italy-multicentric)(136)	4	Machine learning techniques	211; 69.56(7.5), 88		78		Binary outcome	3Y		AUC 0.78	Absent
GAP 6model	Lee J H, 2023 (Korea-multicentric; 2016-2021)(137)	4	Cox regression	966; 71.23 (7.69), 78		257		time-to-event	3 Y		C-statistic NS (0.691; 0.667-0.715)	Calibration plot/HL test
GAP-CR	Lee T, 2024 (Korea-multicentric:2011–2021; 2007–2010)(119)	4	Neural Network	1007; 70 [64–75], 76		NI		time-to-event	3-year survival		Harrell's C-index (0.74, 0.68 –0.79)	Calibration plot
CT-GAP Categorical model	Ley B,2014 (USA-multicentric; 2000-2010)(138)	4/3	Cox regression	176; 69.5(9), 72.7		83	172; 68,1(8,3),68.6	time-to-event	1,2,3 Y mortality	Wolbers competing-risk C = 0.668 (0.594–0.712)	Wolbers competing-risk C-index 0.699 (0.624–0.748)	Calibration plot
longitudinal GAP model	Ley B, 2015 (Multicentric; 2000-2001; 2008-2012) (117)	4	Cox regression	1109; 66.2((7.7), 71.3		128		time-to-event	OS		Harrell's C-index (0.785, 0.780–0.790)	Hosmer-Lemeshow test
GA-FVC-CT index	Nam J G 2023 (Korea; 2005-2009)(121)	4	Cox regression	161; 68 [62–73], 10		140		time-to-event	OS		Uno C-statistic (0.72, 0.68–0.76)	Absent

Legend: AUC-area under the curve; BMI- Body mass index; CT-GAP: Computed tomography fibrosis score, GAP; GA_FVC_CT: Gender Age, Forced Vital Capacity, CT fibrosis quantification; GAP- gender age and physiology; GAP_CR: GAP, Chest radiographs; GAP6 model: GAP, nadir saturation of percutaneous oxygen during six-minute walk test; DC-derivation cohort; OS-overall survival; HL test -- Hosmer-Lemeshow test; ROC -Receiver operating Characteristic; NA – Not applicable; NI – No information; Y-Year; VC -Validation cohor

External validation studies and updated models

Model	Author,Year (Sites, enrolment dates)	TRIPOD type model	Modeling method/algorithm/alternative presentation	Development cohort: Sample (N; Age-Media (SD)or Median [IQR], % male)	Analysis sample	Events (n)	Validation cohort: Sample (n)/Events (n)	Outcome (binary or time-to-event)	Prediction horizon	Discrimination (C-statistic, 95% CI) VC	Discrimination (C-statistic, 95% CI) DC	Calibration
modified GAP model	Nishikiori H 2020 (Japan; 2003-2007)(122)	4	Cox regression	326; 69.5 [64.0-76.0]; 75		205	VC1-117/48; VC 2-1262/415	time-to-event	1,2,3 Y	1,3 Y: VC1-(0.817,0715-0.902); (0.726,0.610-0.843) VC2 –(0.605, 0.556-0.653), (0.588, 0.549-0.627)	Harrell's C-index 1,2,3 Y (0.670,0.601-0.739); (0.698,0.638-0.758) ;(0.738,0.682-0.795)	Absent
NL%-GAP	Ohkubo H ,2018 (Japan;2008-2013)(124)	4	Cox regression	175; 66.1(7.9); 81.1		94		time-to-event	OS		Harrell's C-index (0.775, 0.767-0.785	Absent
mGAP - (DLCO predicted by CT)	Park J ,2021 (South Korea; 2006-2016)(125)	4	Cox regression	104; 68.2 (9.4), 68		56		time-to-event	OS		Harrell's C-index (0.76, 0.69–0.84)	Absent
GAP-ILD	Ryerson CJ 2014 (USA; 2001-2012)(139)	4	Cox regression	1012; NI[NI]/NI		218		time-to-event	OS		Harrell's C-index 0.0.746(IC95% NI)	Calibration plot
GAP _BMI	Suzuki Y, 2021 (Japan-multicentric; 2009-2020)(126)	4	Cox regression	229; 72.0 [67.5–72.0], 81.2		122		time-to-event	OS		C-statistic NS Index (0.698, 0.652–0.743)	Absent
A combined GAPindex–NLR	Takuma S 2024 (Japan-multicentric; 2009-2022)(127)	4	Cox regression	416; 71 [66–76], 73.8		189		time-to-event	OS		Harrell's C-index 0.743	Absent
HE4-based nomogram (GAP+HE4gene)	Tian M 2024 (GEO-GSE70866)(140)	4	Cox regression	118; 68.0 (10.1), 79		NI		time-to-event	OS		C- statistic NS (0.64, NI)	Calibration plot

Legend: AUC-area under the curve; BMI- Body mass index; GAP- gender age and physiology; GAP-ILD: GAP- interstitial lung diseases; GAP-NLR: GAP, Neutrophil–lymphocyte ratio; DC- derivation cohort; HE4 - Human epididymis protein 4; NL %-GAP: Percentage of normally attenuated lung volume on CT to whole-lung volume – GAP; mGAP- modified GAP; NA – Not applicable; NI – No information; NS-not specified; OS-overall survival; ROC -Receiver operating Characteristic; Y-Year; VC - Validation cohort

External validation studies and updated models

Model	Author,Year (Sites, enrolment dates)	TRIPOD type model	Modelling method/algorithm/alternative presentation	Development cohort: Sample (N; Age-Media (SD)or Median [IQR], % male))	Analysis sample	Events (n)	Validation cohort: Sample (n)/Events (n)	Outcome (binary or time-to-event)	Prediction horizon	Discrimination (C-statistic, 95% CI VC	Discrimination (C-statistic, 95% CI) DC	Calibration
PET-modified GAP score-based staging system	Whi W 2025 (Korea; 2007-2020)(128)	4	NI	224; 66.87(10.1) /74.6		67		time-to-event	os		C-statistic NS (0.66,NI)	Absent
PET modified GAPscore	Win T 2018(UK; 2008-2017)(129)	4	Cox regression	113; 79(9) /NI		54		time-to-event	OS		Time-dependent ROC curve analysis AUC (0.694, 0.598–0.791)	Absent
Fibresolve score	Moran-Mendoza O, 2024 (Multicentric; 2015-2018)(141)	4	Cox regression	228; 60 [23–84], 56		89		time-to-event	OS		C-statistic NS (0.58, 0.52–0.64)	Absent
52-Gene Risk Score	Söllner JF, 2024 (USA:2014-2017)(142)	4	NA	245; 70 [65, 75] , 75		42		time-to-event		NI		Absent
CCI	Sonaglioni A, 2023 (Italy; 2016-2018)(143)	4	NA	70; 70.2(7.4), 74		19		time-to-event		NI		Absent
CHA2DS2-VASc score	Sonaglioni A, 2023 (Italy; 2016-2018)(143)	4	NA	109; 70.7 (7.3), 79.6		23		time-to-event	OS	Time-dependent ROC curve analysis AUC 0.91 (0.94-1)		Absent
Lung Immune Prognostic Index	Suzuki T, 2023 (japan-multicentric; 2001-2020)(144)	4	Cox regression	NI; 70[65–75], 88.7		NI	VC1-460/NI; VC2-414/NI	time-to-event	OS	C-index NS 0.587		Absent

Legend: AUC-area under the curve; CCI -Charlson Comorbidity Index; CHASDS2 - Congestive heart failure, Age, Diabetes mellitus, Stroke/Vascular disease; age, sex category; GAP- gender age, and physiology; NA – Not applicable; NI – No information; NS-not specified; OS-overall survival; PET - positron emission tomography; ROC -Receiver operating Characteristic; Y-Year; VC -Validation cohort

External validation studies and updated models

Model	Author, Year (Sites, enrolment dates)	TRIPOD type model	Modelling method/algorithm/ alternative presentation	Development cohort: Sample (N; Age-Media (SD)or Median [IQR], % male)	Analysis sample	Events (n)	Validation cohort: Sample (n)/Events (n)	Outcome (binary or time-to-event)	Prediction horizon	Discrimination (C-statistic, 95% CI VC	Discrimination (C-statistic, 95% CI) DC	Calibration
Bao Y, 2024 (47)	Bao Y, 2024	3	Cox regression/ Nomogram	75; 69 (8.1), 69	-	NI	45/NI	time-to - event	1,2,3 Y OS	time-dependent ROC curve analysis (AUC 1y ,2y ,3y - 0,72;0,76; 0,66)	time-dependent ROC curve analysis (AUC 1y ,2y ,3y - 0,68;0,71; 0,68)	Calibration Plot
Fan G, 2022 (51)	Fan G, 2022	3	Cox regression/ Nomogram	75; NI, NI	-	NI	45/NI	time-to-event	1,2,3 Y OS	time-dependent ROC curve analysis (AUC 1 y, 2y, 3y - 0.742, 0.729, 0.634)	time-dependent ROC curve analysis (AUC 1 y, 2y ,3y - 0.670, 0.787, 0.86)	Calibration plot
Feng A, 2023 (52)	Feng A, 2023 (Genes data set) (52)	3	Cox regression/Risk score system	45; NI, 84.4	-	13	VC1-51/31; VC2 76/51	time-to-event	1,2,3 Y OS	time-dependent ROC curve analysis (VC1 AUC 1,2 y-0.709; 0.715; VC2 AUC 1 y, 2y, 3y - 0.638; 0.629;0.709)	time-dependent ROC analysis (AUC 1 y, 2y,3y - 0.882; 0.814; 0.758)	Absent
Gudmundsson E, 2021 (55)	Gudmundsson E, 2021 (Turkey; 2008-2015) (55)	3	Cox regression/Sum Score	142; 66[42-86], 80		70	145/81	time-to-event	OS	Harrell's C-index (0.72, NI)	Harrell's C-index (0.79, NI)	Absent
Functional Genomic Model (57)	Huang Y 2015 (USA; 2007) (57)	3	Cox regression	45; 67.1 (8.2), 90		17	VC1-21; VC2-75	time-to-event	1,3,5 Y OS	NI	Time-dependent ROC curve analysis (AUC 1,3, 5 y - 0.889, 0.816, 0.725)	Absent
He Y,2022 (58)	He Y,2022 (Gene Expression Omnibus (GEO)) (58)	3	Cox regression (Internal validation- Cross-validation)	62; 65.7 (10.1), 83.9		45	VC1-64/24; VC2-50/31	time-to-event	1,2,3,5 Y OS	VC1 (AUC 1,2,3,5,7, y-0.813, 0.806, 0.835, 0.635, 0,632); VC2 (AUC1,2,3 y-0,815, 0,849; 0,75)	Time-dependent ROC curve analysis (AUC 1,2,3,5, y-0.869, 0.845, 0.83, 0.936)	Absent
Hou Z, 2022(59)	Hou Z, 2022 (GEO))(59)	3/bootstrap	Cox regression/ Nomogram	112; 69.5 [62.0, 76.0], 83		72	64/41	time-to-event	1,2,3 Y OS	C-statistic NS (0.812, 0.703-0.921)	C-statistic NS (0.815, 0.769 0.861)	Calibration Plot

Legend: AUC-area under the curve; NA – Not applicable; NI – No information;NS-not specified; OS-overall survival; PET - positron emission tomography; ROC - Receiver operating Characteristic; Y-Year; VC -Validation cohort

External validation studies and updated models

Model	Author,Year (Sites, enrolment dates)	TRIPOD type model	Modelling method/algorithm/alternative presentation	Development cohort: Sample (N; Age-Media (SD)or Median [IQR], % male	Analysis sample	Events (n)	Validation cohort: Sample (n)/Events (n)	Outcome (binary or time-to-event)	Prediction horizon	Discrimination (C-statistic, 95% CI) VC	Discrimination (C-statistic, 95% CI) DC	Calibration
Karampitsakos T 2024 (64)	Karampitsakos T 2024 (GEO: GSE28221) (64)	3	Cox regression	75; 69.0(8.2), 69		NI	VC1-45/NI, VC2 176/NI; VC3 120/NI	time-to-event	3, 4 years	NI	time-dependent ROC curve analysis (NI)	Absent
Inflammation and Hypoxia-Related Genes model	Liu J 2022 (GEO - GSE70866) (72)	3	Cox regression/Nomogram	112; NI, 83		76	64/24	time-to-event	1,3,5 y OS	time-dependent ROC curve analysis (AUC 1,2,3,5 Y 0.78, 0.79, 0.80, 0.91)	time-dependent ROC curve analysis (AUC 1,2,3,5 Y- 0.83, 0.82, 0.80, 0.81,0.94)	Calibration Plot
Plasma cells-related gene signature ST6GAL1 model	Lin F 2025 (GEO: GSE70866) (74)	3	Machine learning techniques	112;67.97 (10.07), 83		76	64/24	time-to-event	OS	Harrell's C-index (0.76, NI)	Harrell's C-index 0.72, NI)	Calibration Plot
Li X, 2021(75)	Li X, 2021 (GEO-GSE70866) (75)	3	Cox regression/ Nomogram	176; NI, 66.5		NI	VC1 57/NI; VC2121/NI; VC3-179/NI	time-to-event	1, 2, 3, 4, 5-year OS	NI	Time dependent ROC curve analysis AUC 1,2,3 Y: 0.789, 0.768, 0754, 0798, 0.913	Absent
Li Y2023 (76)	Li Y 2023 (GEO dataset - GSE28042) (76)	3	Cox regression/Nomogram	75; 69.00 (8.16)		NI	VC1 -75/NI; VC2-45/NI; VC3 -176/NI	Time-to-event	1, 2, 3 year OS	Time dependent ROC curve analysis VC 2 AUC 1,2,3 Y :0.600, 0.671, 0.683	Time dependent ROC curve analysis AUC 1,2,3 Y :0.666, 0.651, 0.717	Absent
Prognostic Index Based on Inflammation-Related Gene Analysis	Lu Y 2021(GEO; GSE28042) (77)	3	Cox regression/Nomogram	75; 75; NI		NI	45/NI	time-to-event	1, 2, 3-year OS	Time dependent ROC curve analysis AUC 1,2,3 Y:0.678, 0.6956, 0.671	Time dependent ROC curve analysis AUC 1,2,3 Y:0.611, 0.695, 0.681	Absent
Extracellular matrix-related genes model	Luo H 2023 (GEO; GSE70866) (78)	3	Cox regression/Nomogram	112; 69.5 [62, 76], 52.8		76	64/24	time-to-event	OS	C-statistic NS (0.727, 0.619–0.834)	C-statistic NS (0.707, 0.649 – 0.76)	Calibration Plot

Legend: AUC-area under the curve; NA – Not applicable; NI – No information; NS – not specified; OS-overall survival; PET - positron emission tomography; ROC - Receiver operating Characteristic; Y-Year; VC -Validation cohort

External validation studies and updated models

Model	Author,Year (Sites, enrolment dates)	TRIPOD type model	Modelling method/algorithm/alternative presentation	Development cohort: Sample (N; Age-Media (SD)or Median [IQR], % male	Analysis sample	Events (n)	Validation cohort: Sample (n)/Events (n)	Outcome (binary or time-to-event)	Prediction horizon	Discrimination (C-statistic, 95% CI VC	Discrimination (C-statistic, 95% CI) DC	Calibration
Cold exposure-related genes (CERGs) model	Luo P 2025 (GEO)(79)	3	Cox regression/Nomogram	62; NI		NI	VC1- 50/NI; VC2-64/NI	time-to-event	1, 2, 3-year OS	Time dependent ROC curve analysis AUC 1,3 Y: VC1-0.82, 0,78; VC2 - 0,71;0,86	Time dependent ROC curve analysis AUC: 1,3,5 Y: 0.81; 0.79; 0.91	Absent
Fatty acid metabolism-related genes model	Lyu Y, 2022 (GEO: GSE70866)(80)	3	Cox regression/Nomogram	112; 69.5 (62. 8), 52.8		76	64/24	time-to-event	1,2,3 Y OS	C-statistic NS 0,6906 (0,5578–0,8234)	C-statistic NS Nomogram (0.7642, 0.7125-0.7962)	Calibration Plot
Prasse A, 2019	Prasse A, 2019 (GEO) (84)	3	Cox regression	62; 67.46(9.1), 85		45	VC1 50/31; VC264/24	time-to-event	18 months, OS	C-statistic NS VC1 – (0.66. 0.55–0.76) VC2-(0.63. 0.54–0.72)	C-statistic NS (0.67, 0.62–0.71)	Absent
TORVAN	Torrisi SE, 2019 (Italy, Netherlands, Germany; 2004-2016)(89)	3	Cox regression	476; 68.8 (8.41) /76.7		272	190	time-to-event	1-5 Y OS	Harrell's C-index 70.3 (66.1–74.5	Harrell's C-index (71.3, 68.1–74.6)	Calibration plot
	Bocchino ML, 2023 (Italy:2012-2019)(112)	4	NA	235; 69.8 (7.1), 76		116		time-to-event	OS	Harrell's C-index (0.608, 0.562-0.619)		Absent
Wang E 2022	Wang E 2022 (GEO-GSE70866)(92)	3	Cox regression/nomogram	112; 68.0(10.0), 83		76	64/24	time-to-event	1,2,3 Y OS	Time dependent ROC curve analysis AUC 1,2,3 Y: 0.795, 0.782, and 0.794	Harrell's C-index (0.76, 0.73 - 0.79)	Absent
Wang X 2024	Wang X 2024 (GEO: GSE28042)(93)	3	Cox regression/ nomogram	75; 69.0 (8.2), 69		NI	112/NI	time-to-event	1,2,3 Y OS	Time dependent ROC curve analysis AUC 1,2,3 Y (0.66, 0.73; 0.66)	Time dependent ROC curve analysis AUC 1,2,3 Y: 0.76; 0.87; 0.88	Calibration Plot

Legend: AUC-area under the curve; NA – Not applicable; NI – No information; OS-overall survival; PET - positron emission tomography; ROC -Receiver operating Characteristic; Y-Year; VC -Validation cohort

External validation studies and updated models

Model	Author,Year (Sites, enrolment dates)	TRIPOD type model	Modelling method/algorithm/alternative presentation	Development cohort: Sample (N; Age (SD), % male)	Analysis sample	Events (n)	Validation cohort: Sample (n)/Events (n)	Outcome (binary or time-to-event)	Prediction horizon	Discrimination (C-statistic, 95% CI) VC	Discrimination (C-statistic, 95% CI)	Calibration
m 6A genes model	Wang Z 2022 (GEO-GSE28042)(94)	3	Cox regression	75; NI		NI	60/NI	time-to-event	1,2,3,4 Y OS	Time-dependent ROC curve analysis AUC 1,2 Y 0,73; 0,70	Time-dependent ROC curve analysis AUC 1,2,3,4 Y: 0.63; 0.77, 0.85, 0.95	Absent
BAL cell-associated gene signature model	Xia Y 2021 (GEO - GSE70866)(96)	3	Cox regression/nomogram	122; 67.97(10.07), 8.3		NI	64/NI	time-to-event	1,2,3 Y OS	Time-dependent ROC curve analysis AUC 1,2,3 Y 0.760, 0.717, 0.748	Harrell's C-index (0.72, 0.66–0.77) Time-dependent ROC curve analysis AUC 1,2,3 Y 0.773, 0.772, 0.752	Calibration plot
Xiang Y 2024	Xiang Y 2024 (GEO: GSE70866)(97)	3	Cox regression	176; NI		NI	60/NI	time-to-event	NI	Time dependent Roc curve analysis AUC:0.66	Time-dependent ROC curve analysis AUC: 0.741[0.668,0.810]	Calibration plot
Inflammation related genes model	Yin Y-Q 2022 (GEO: GSE70866, GSE28042)(98)	3	Cox regression/nomogram	132; NI		NL	VC1-64/NI; VC2-75/NI	time-to-event	1,3,5 YOS	Time-dependent ROC curve analysis AUC 1, 3, 4 year 0.786, 0.713, 0.793	Harrell's C-index (>0.7) time-dependent ROC curve analysis AUC 1, 3,5 Y 0.784, 0.835, 0.921	Absent
CPB index and CPB stage	Zhang X 2022 (China; 2000-2015) (99)	3	Cox regression/ stage system /nomogram	374; NI, 84		321	324/75	time-to-event	1,2,3 Y OS	C-statistic NS CPB index: C-index 0.635 (0.558–0.712) CPB stage: C-index0.619 (0.544–0.694)	C-statistic NS CPB index: C-index (0.632, 0.574 –0.690) CPB stage: C-index:(0.589, 0.537–0.642)	Calibration plot
Zhao J2023	Zhao J 2023 (GEO: GSE70866)(100)	3	Cox regression/nomogram	112; NI		/NI	64/NI	time-to-event	1,2,3 Y OS	Time dependent Roc curve analysis AUC 1,2,3 Y: 0.814, 0.814; 0.808	Time dependent Roc curve analysis AUC 1,2,3 Y: 0.872; 0.905; 0.868	Calibration Plot

Legend: AUC-area under the curve; CPB - Clinical-Physiological-Biomarker; m 6A - N6-methyladenosine; NA – Not applicable; NI – No information; OS-overall survival; PET - positron emission tomography; ROC -Receiver operating Characteristic; Y-Year; VC -Validation cohort

Model	Author,Year (Sites, enrolment dates)	TRIPOD type model	Modelling method/algorithm/alternative presentation	Development cohort: Sample (N; Age (SD), % male)	Analysis sample	Events (n)	Validation cohort: Sample (n)/Events (n)	Outcome (binary or time-to-event)	Prediction horizon	Discrimination (C-statistic, 95% CI) VC	Discrimination (C-statistic, 95% CI) DC	Calibration
3 Epithelial-Mesenchymal Transition and Immune-Related Genes based model	Zheng J 2022 (GEO: GSE70866)(101)	3	Cox regression/nomogram	176; NI		NI	118/NI	time-to-event	1.2.3 Y OS	Time dependent Roc curve analysis AUC 1,2,3 Y:0.691, 0.734, 0.684	Time dependent Roc curve analysis AUC 1,2,3 Y: 0.789; 0.743; 0.722	Calibration Plot
TRIM family gene signature from bronchoalveolar lavage cells-based Model	Zhou M 2022 (GEO-GSE70866)(102)	3	Cox regression/nomogram	62; Nl		NI	VC 1-50/NI VC264/NI-	time-to-event	1.2.3 Y OS	AUC 1,2,3 Y: VC1 0.742, 0.805, 0.818; VC2: 0.912, 0.867, 0.881	AUC 1,2,3 Y: 0.832, 0.786; 0.750	Calibration Plot

Legend: AUC-area under the curve; NA – Not applicable; NI – No information; OS-overall survival; ; ROC -Receiver operating Characteristic; TRIM - Tripartite motif family genes-Year; VC -Validation cohort

12.2 Original models identified in the included studies

Fifty-nine original prognostic models for Idiopathic Pulmonary Fibrosis (IPF) were identified.

The prognostic models were predominantly developed using genomic datasets (n=32; 54.2 %). Geographically, the United States (n=6; 10.2%) and the United Kingdom (n=4; 6.8 %) were the most represented countries. Seven models (11.9%) originated from multicenter investigations, four of which involved international collaborations across multiple countries.

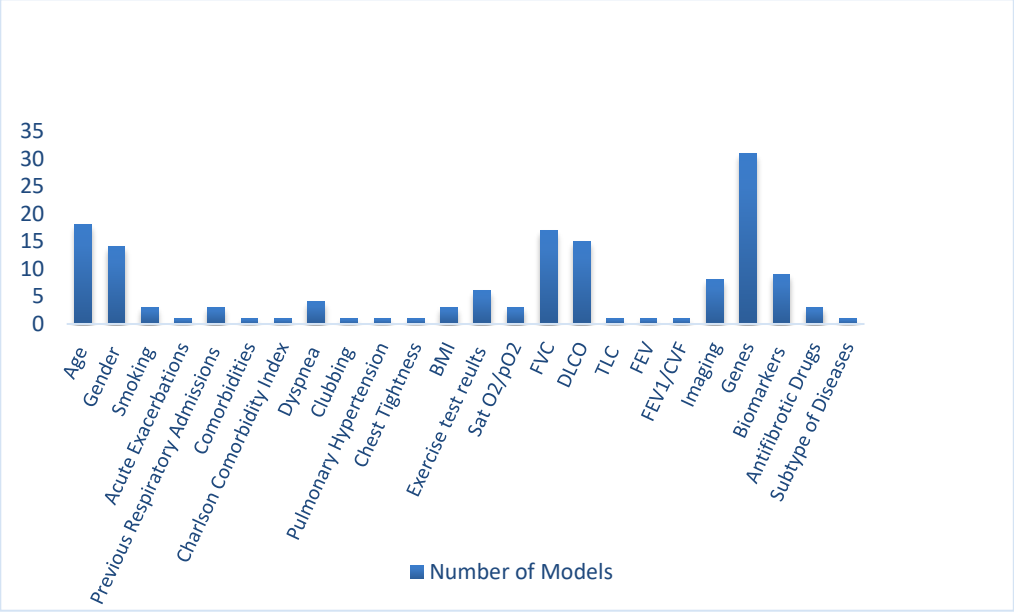
Regarding the derivation cohorts, the median sample size was 115 [IQR: 75–175], and the median number of events was 72.5 [IQR: 42–100]. In type 2 and type 3 studies, the validation cohorts yielded a median sample size of 73 [IQR: 60.75–136] and a median event count of 42 [IQR: 24–79.50].

Table 4 - Methodological characteristics of the original models

Original models	n (%) (59/100)
Validation	
Apparent performance	14 (23.7)
Internal validation: Bootstrap	5 (8.5)
Internal validation: Cross-validation	1 (1.7)
Random split	3 (5.1)
Non-random split	3 (5.1)
Validation in an external dataset	30 (50.8)
Combination of methods	1 (1.7)
Not reported	2 (3.4)
Modelling method	
Cox regression	54 (91.5)
Logistic regression	2 (3.4)
Machine learning	1 (1.7)
Multiplicative combination	1 (1.7)
Not reported	1 (1.7)
Handling of missing data	
Complete case analysis	13 (22.0)
Not reported	45 (76.3)
Model discrimination	
Time-dependent ROC curve analysis	28 (47.5)
Harrell's C-index statistic	10 (16.9)
Uno C-statistic	3 (5.1)
C-statistic not specified	14 (23.7)
Not reported	4 (6.8)
Model calibration	
Calibration plot	22 (37.3)
Hosmer–Lemeshow (HL) test	2 (3.4)
Other	1 (1.7)
Not reported	34 (57.6)
Model presentation	
Risk score system	8 (13.6)
Stage system	1 (1.7)
Nomogram	23 (39.0)
Combination methods	4 (6.8)
Equation	31 (52.5)

In terms of mathematical approach, various modeling methods were applied across the 59 prognostic models. The vast majority applied Cox proportional hazards regression (n=54; 91.5%), followed by logistic regression (n=2; 3.4%) (table 4).

Figure 2 - Included predictors within the original models



Legend- BMI -Body mass index, FVC - forced vital capacity; DLCO -diffusing capacity of carbon monoxide; FEV1- forced expiratory volume in the 1st second; TLC- Total Lung Capacity

The most frequently included predictors across these models were genomic variables (n=31, 52.5%), age (n=18, 30.5%), forced vital capacity (FVC; n=17, 28.8%), diffusing capacity of carbon monoxide (DLCO; n=15, 25.4%), sex (n=14, 23.7%), imaging findings (n=8, 13.6%), and walked distance on the 4- or 6-minute walk tests (n=6, 10.2%) (Fig.2)

Regarding model validation during the development phase, 14 (23.7%) prognostic models reported only apparent performance. Internal or external validation was conducted for the remaining models: 30 (50.1%) underwent external validation, while

internal validation was performed using bootstrapping in 6 models (10.2%), random split-sample methods in 4 (6.8%), and non-random split-sample methods in 3 (5.1%).

The remaining 2 models (5.1%) did not report any validation method. (Table 4)

Discrimination was assessed in 55 (93.2%) of the 59 prognostic models. Time-dependent discrimination measures, such as time-dependent ROC curve analysis, were the most reported metric, appearing in 47.5% (n=28) of the models, followed by Harrell's C-index in 16.9% (n=10) of the models. Five models (8.5%) did not report any discrimination metric. (Table 4)

Methodological evaluation of model calibration was reported for a subset of 25 models (42.4%). To assess agreement between predicted and observed outcomes, calibration plots were used most frequently (n=22; 37.3%), followed by the Hosmer–Lemeshow test (n=2; 3.4%). Most of the developed models (n=34; 57.6%) lacked reported calibration assessments. (Table 4)

A considerable proportion of the developed models (n=22; 37.3%) did not include a formal presentation format. Among those that did, the most prevalent method was a nomogram (n=23; 39.0%), followed by a sum score (n=8; 13.6%) and a risk stage score (n=2; 3.4%). Overall model performance was rarely evaluated (n=4); when it was, the Brier score (n=3) and R-squared (n=1) were used. Finally, only seven studies addressed clinical utility through formal decision analysis. (Table 4)

12.3 Models with external validation and identification of models that were updated.

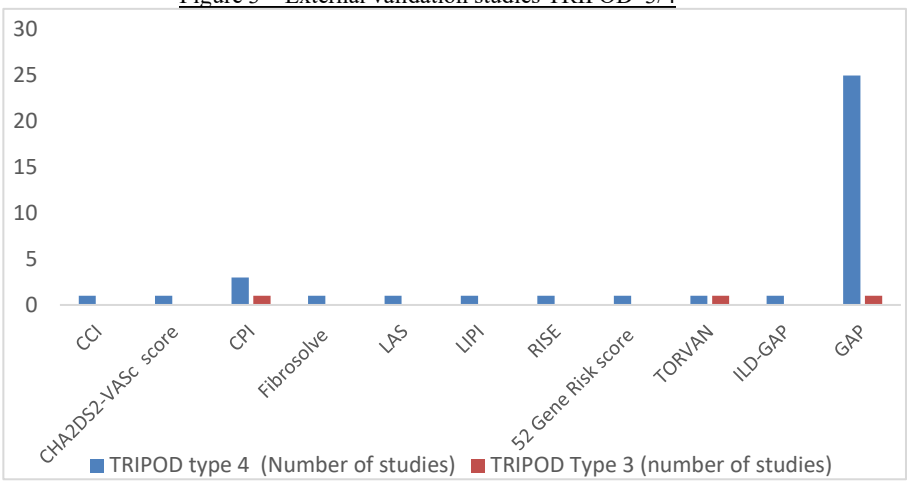
Of the 66 models (59 original models and 7 originally developed to assess health outcomes/diseases other than IPF mortality), 38 were externally validated at least once (Fig 3, 4); however, only 11 were externally validated by an independent research team (Fig 3).

Of these 11 models, the 7 originally developed to assess health outcomes or diseases other than mortality in IPF patients were the Lung Allocation Score (LAS), the Risk Stratification Score (RISE), the FibroSolve score, the Charlson Comorbidity Index (CCI), the Lung Immune Prognostic Index (LIPI), the 52-Gene Risk score, and the CHA2DS2-VASc score. The remaining 4 models, originally developed to evaluate mortality in IPF patients and validated by an independent research team, were the GAP index, ILD-GAP, the Composite Physiologic Index (CPI), and the TORVAN model (Fig 3).

Regarding the validation cohorts (TRIPOD type 3 and 4 studies), the median sample size was 106 [IQR: 64-242], and the median number of events was 57 [IQR: 25,50-120,50].

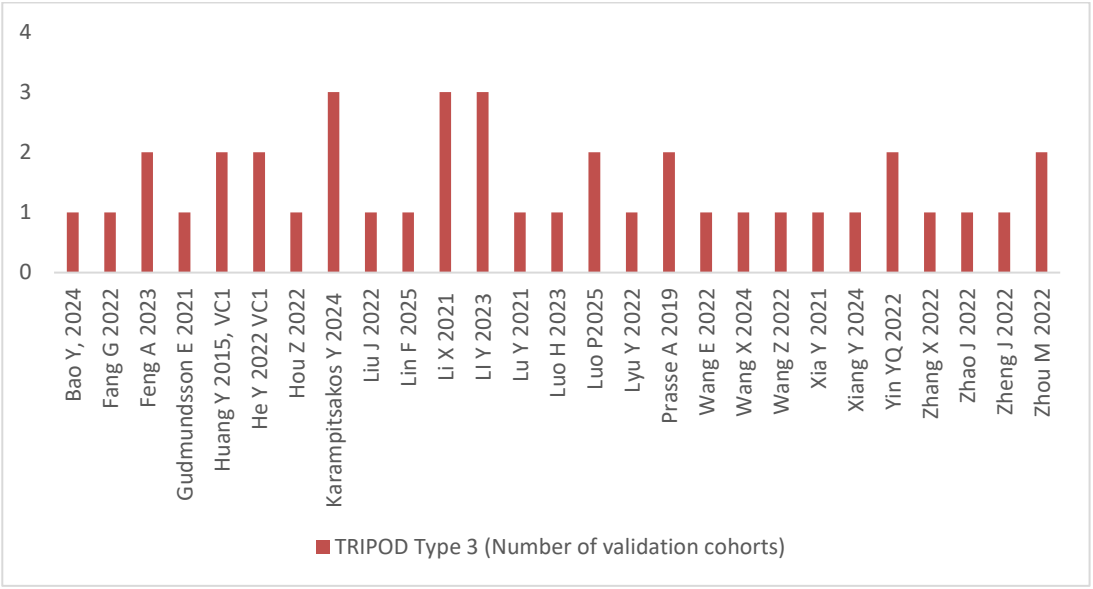
Prognostic models externally validated in more than two cohorts included the GAP index (25 cohorts), CPI (4 cohorts), and the models by Karampitsakos et al. (2024; 3 cohorts), Li X et al. (2021; 3 cohorts), and Li Y et al. (2023; 3 cohorts) (Fig 4).

Figure 3 – External validation studies TRIPOD 3/4



Legend – CCI - Charlson Comorbidity Index; LAS - Lung Allocation Score (LAS); LIPI- Lung Immune Prognostic Index; RISE - Risk Stratification Score; GAP – Gender Age Physiology; ILD- Interstitial lung diseases

Figure 4 - External validation studies TRIPOD type 3



The prognostic models were predominantly validated using genomic datasets (n = 29, 36.3%). The most represented geographic locations were South Korea (n = 7, 8.6%), United States (n = 5, 6.3%), Italy (n = 5, 6.3%) and United Kingdom (n = 5, 6.3%),

followed by Canada (n = 4, 5.0 %). Thirteen models (16.3%) originated from multicenter investigations, six of which involved international collaborations across multiple countries.

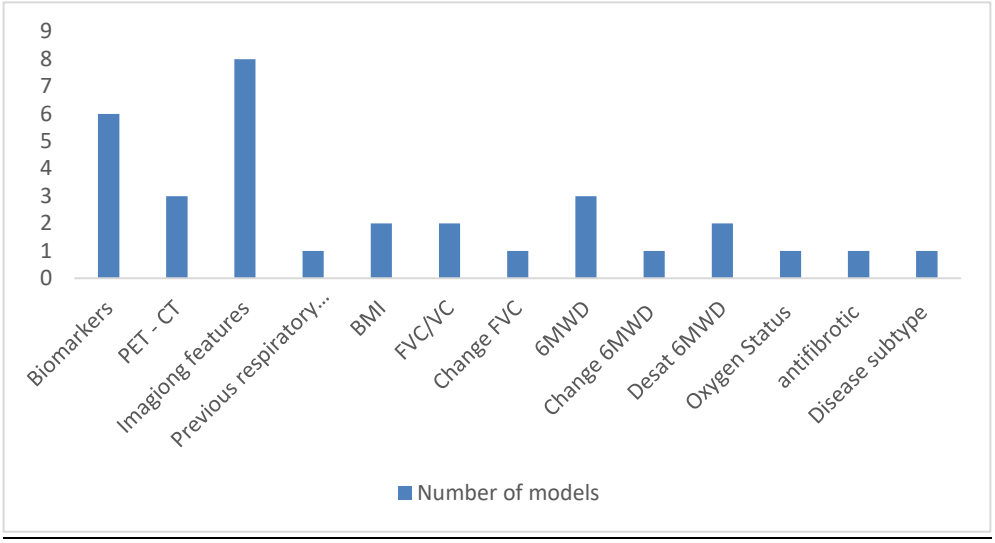
Discrimination was assessed in 65 of the 80 external validation models (81.0 %). Among the reported metrics, time-dependent measures, particularly time-dependent ROC curve analysis, were the most frequently used, applied in 41% of models (n = 33). This was followed by Harrell's C-index, which was reported in 25% of models (n =20). Notably, 16 models (20%) did not report any discrimination measure.

Calibration was inconsistently reported across studies. Calibration plots were the most used method (8.6%, n = 7), followed by the Hosmer–Lemeshow (HL) test (2.5%, n = 2) and the observed-to-expected (O/E) ratio (2.5%, n = 2). Calibration-in-the-large (CITL) was reported in only one study (1.3%). Two studies (2.5%) reported both calibration plots and HL tests. Most external validation models (n = 66; 82.5%) lacked reported calibration assessments.

For the updated models, the median cohort sample size was 226.5 (IQR: 110.8–711.3). The median number of events was 78.5 (IQR: 53.3–131.0), and the median proportion of male participants was 75.5% (IQR: 71.5–82.6).

The most frequently included new predictors across these models were imaging features (n=8), biomarkers (n=6), and measures from the 6-minute walk test—specifically desaturation or walking distance (n=6) (Fig.5).

Figure 5 - New Predictors included in the Updated Prognostic Models



Legend: BMI -Body mass index; FVC - forced vital capacity; DLCO -diffusing capacity of carbon monoxide; PET-CT – Positron Emission Tomography- Computed Tomography; 6 MWT- six-minute walk test;

A total of 26 updated prognostic models were identified, the majority of which were based on the GAP model (n=23). Only one model was derived from each of the ILD-GAP, Du Bois RM (2014), and the Japanese revised system. (Table 5)

Table 5 - Methodological characteristics of the updated models

Updated models	**N (%) - 26 (100)**
GAP	23 (88)
Du Bois RM 2014	1 (3.9)
Japanese revised system	1 (3.9)
ILD- GAP	1 (3.9)
Validation	
None	13 (50)
Internal Validation: Bootstrap/ Cross validation	5/2 ((19.2/7.7)
Random split	1 (3.9)
Non-Random split	1 (3.9)
Validation in an external data set	2 (7.7)
Combination of methods	2 (7.7)
Refit	1 (3.9)
Replacement 1 variable	6 (23)
Add variables	
1	10 (38.5)
2	5 (19.2)
3	3 (11.5)
Modelling method	
Cox regression	20 (76.9)
Machine Learning	2 (7.7)
Not reported	3 (11.5)
Handling of missing data	
Complete Case Analysis	21 (80.8)
Multiple imputation	1 (3.9)
Not reported	4 (15.4)
Model discrimination	
Time-dependent ROC curve analysis	3 (11.5)
Harrell´s C-index statistic	12 (46.2)
C-statistic not specified	5 (19.2)
Uno C statistic	1 (3.9)
Wolbers competing-risk C-index	1 (3.9)
Not reported	4 (15.4)
Model calibration	
Calibration Plot	7 (26.9)
HL-test	2 (7.7)
Goodness of fit	1 (3.9)
Confusion Matrix	1 (3.9)
Not reported	15 (57.7)
Model presentation	
Sum Score	1 (3.9)
Stage System	8 (30.8)
Algorithm form	1 (3.9)
Equation	0

Legend - GAP- Gender age physiology; HL- Hosmer-Lemeshow test; ILD – interstitial lung diseases; ROC- Receiver operating characteristic

Regarding validation, 13 models did not report any validation. Internal validation was performed using bootstrap (n=5) or cross-validation (n=2), and random and non-random data splitting methods were each used in one study. External validation was conducted in two models, and two studies reported a combination of validation methods (Table 5).

For updating strategies, adding variables was the most common approach (n=10), followed by replacing a single variable (n=6). Complete refitting of the model was rare (n=1). Among models that added predictors, five included one additional variable, three included two variables, and two included three variables (Table 5).

Cox proportional hazards regression was the predominant modeling method (n=20), whereas machine learning approaches were used in only two studies. Concerning missing data handling, complete case analysis was most frequently applied (n=21), while multiple imputation was used in just one study; four studies did not report their approach (Table 5).

Model discrimination was primarily assessed using Harrell's C-index (n=12), with time-dependent ROC curve analysis employed in only three cases. Model calibration was poorly reported, with 15 models lacking any description of calibration assessment. When reported, calibration plots were the most frequently used method (n=7), followed by the Hosmer–Lemeshow test (n=2), goodness-of-fit tests (n=1), and confusion matrices (n=1) (Table 5).

Finally, most models were presented as stage-based systems (n=8), while only a small number were expressed as sum scores (n=1) or algorithm-based formats (n=1), and none presented the final model in the form of an explicit equation (Table 5).

12.4 Risk of bias assessment

The PROBAST tool was used to assess the risk of bias in studies that developed or externally validated prognostic models.

Development Models:

A total of 59 development models were assessed using PROBAST.

Graph 7 summarizes the risk-of-bias assessment for all development models.

Figure 6- Risk of Bias for Development Models

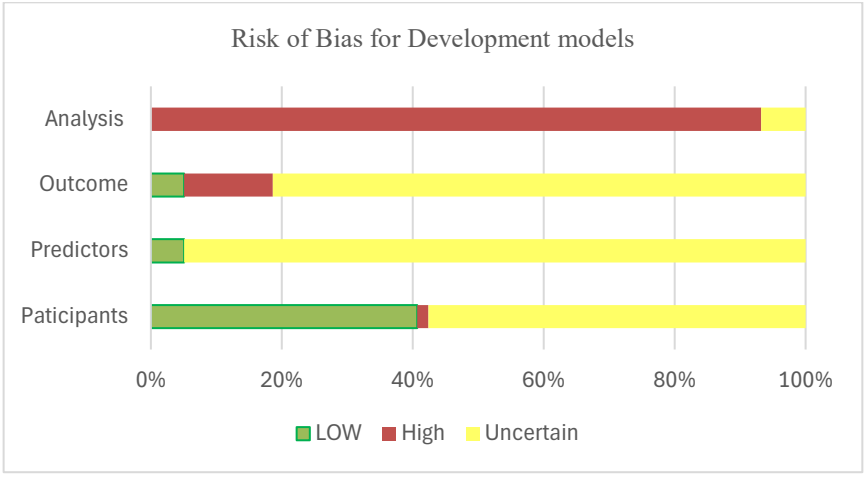


Figure 7 - Applicability for Development models

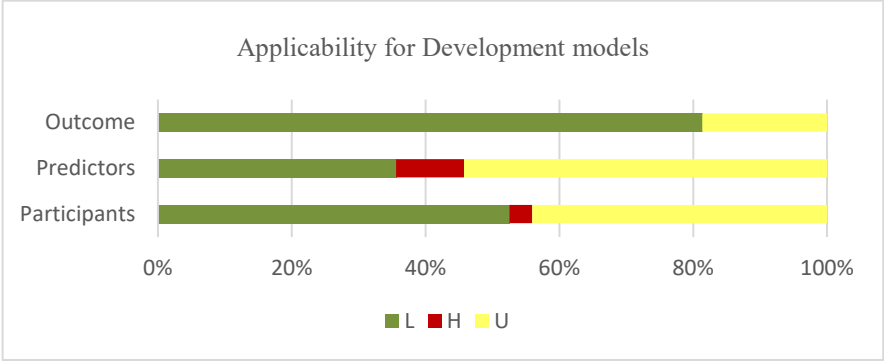
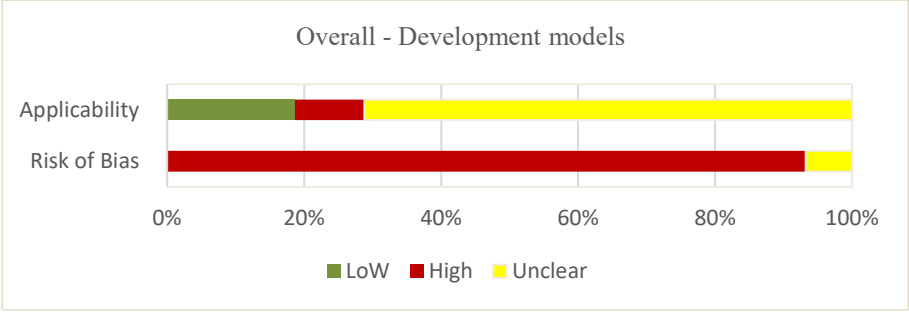


Figure 8 - Overall Development models



Regarding risk of bias, in the participants domain, most models were classified as unclear risk (57.6%, 34/59), followed by low risk (40.7%, 24/59), with only one model (1.7%) rated as high risk. In the predictors domain, almost all models were rated as unclear risk (94.9%, 56/59), with the remaining 5.1% (3/59) classified as low risk. No models were considered high risk. In the outcome domain, most models were also rated as unclear risk (81.4%, 48/59), while 13.6% (8/59) were classified as high risk and 5.1% (3/59) as low risk. The analysis domain showed a markedly different pattern. Most models were classified as high risk (93.2%, 55/59), while the remainder were rated as unclear risk (6.8%, 4/59). No models were at low risk in this domain. (Fig 6)

Applicability concerns were generally rated as low or unclear across all domains. Among participants, 50.8% (30/59) of models were judged to have low concern, while 49.2% (29/59) showed unclear or high concern, mainly due to the predominance of single-center studies and limited representativeness. For predictors, only 25.4% (15/59) of models were classified as low concern, whereas 64.4% (38/59) had unclear concern and 10.2% (6/59) high concern, largely reflecting heterogeneity in predictor definitions and the frequent use of genomic variables (52.5%, 31/59) that limit clinical applicability. In contrast, the outcome domain demonstrated the best applicability, with 79.7% (47/59) of models rated as low concern and 20.3% (12/59) as unclear or high concern, as mortality was consistently used across studies despite some variation in definitions and follow-up periods (Fig 7).

Overall, these findings indicate moderate applicability, with predictor selection representing the main limitation to generalizability and clinical implementation. The overall risk of bias across studies was high, largely reflecting ratings in the analysis domain (Fig 8).

Updated Models

Twenty-six updated models were assessed using PROBAST

Figure 9 summarises the risk-of-bias assessment for the updated models.

.

Figure 9- Risk of Bias Updated Models

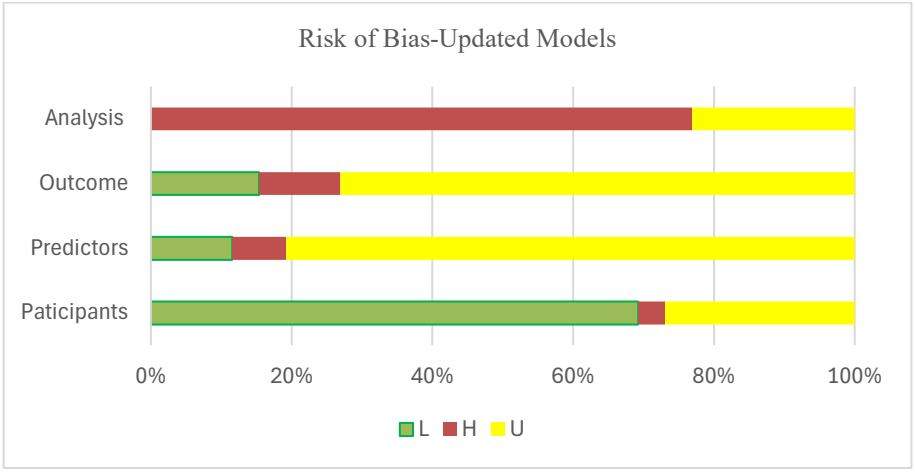


Figure 10 - Applicability-updated models

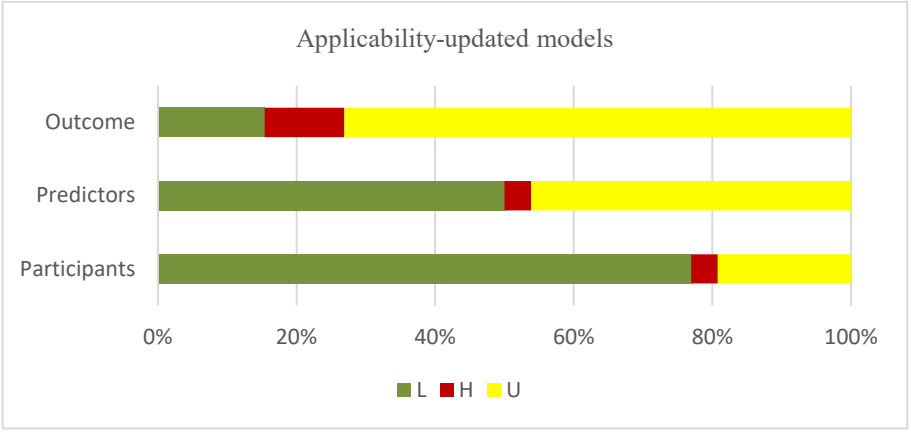
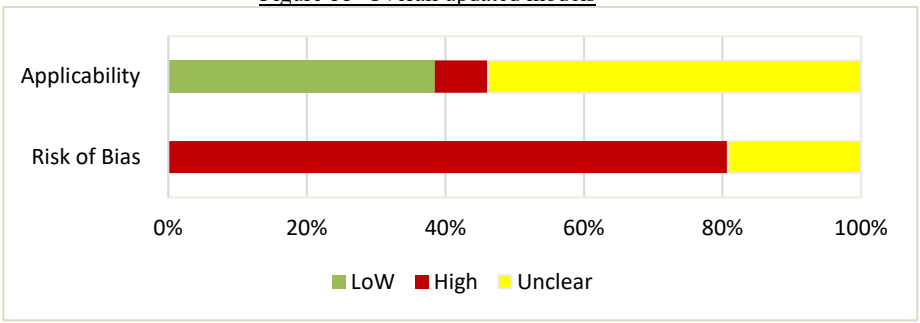


Figure 11 -Overall updated models



Regarding risk of bias, in the participants domain, 69.2% (18/26) were rated as low risk, 26.9% (7/26) as unclear risk, and 3.8% (1/26) as high risk. In the predictors domain, 76.9% (20/26) were rated as unclear risk, 11.5% (3/26) as low risk, and 11.5% (3/26) as high risk. In the outcome domain, 73.1% (19/26) were rated as unclear risk, 15.4% (4/26) as low risk, and 11.5% (3/26) as high risk. In the analysis domain, 80.8% (21/26) were rated as high risk, 19.2% (5/26) as unclear risk, and none were rated as low risk (Fig 9).

Applicability concerns were generally low or unclear. In the participants domain, 76.9% (20/26) were rated as low concern, 19.2% (5/26) as unclear, and 3.8% (1/26) as high concern. In the predictors domain, 53.8% (14/26) were rated as low concern, 42.3% (11/26) as unclear, and 3.8% (1/26) as high concern. In the outcome domain, 69.2% (18/26) were rated as unclear concern, 19.2% (5/26) as low concern, and 11.5% (3/26) as high concern (Fig. 10).

Overall, these findings indicate low or unclear applicability concerns, whereas the risk of bias in the updated models was high, largely reflecting ratings in the analysis domain (Fig. 11).

External validation studies

Eighty external validations were assessed using PROBAST. Figure 12 summarizes the risk-of-bias assessment for external validation studies.

Figure 12 - Risk of Bias for External Validation studies

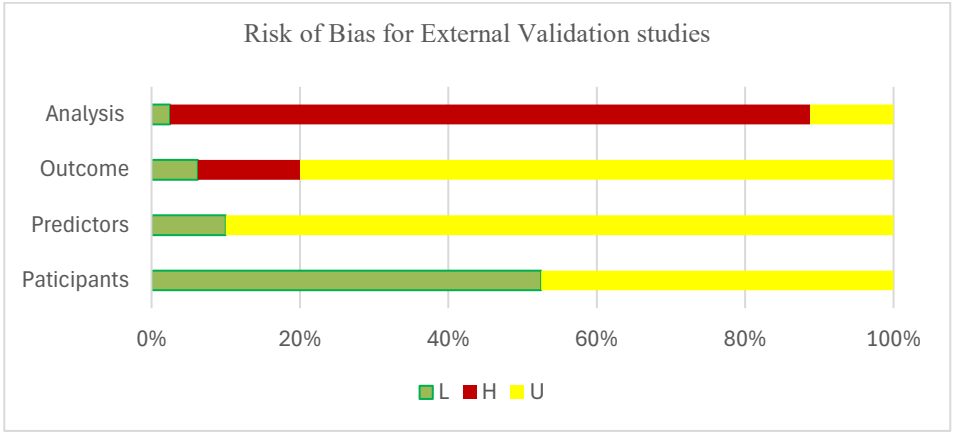


Figure 13 - Applicability for External Validation studies

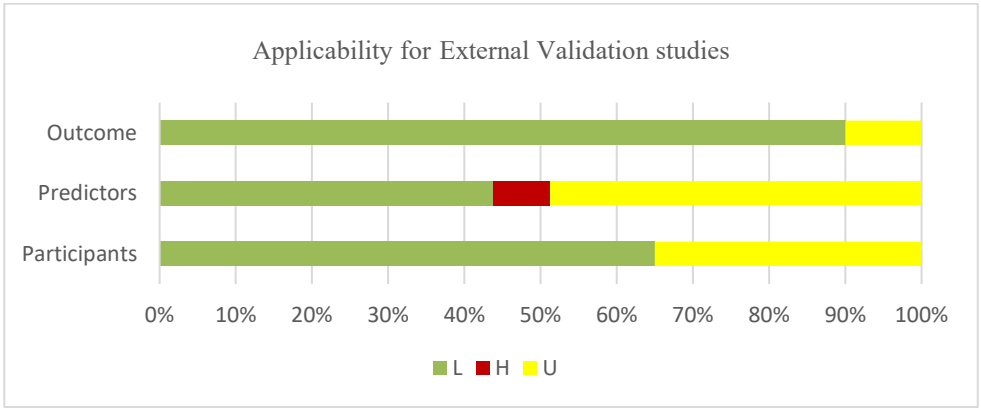
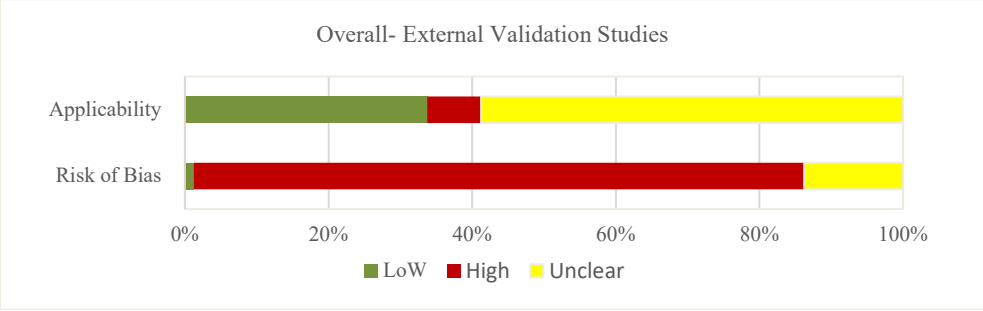


Figure 14 - Overall- External Validation Studies



Regarding risk of bias, in the participants domain, 52.5% (42/80) were rated as low risk and 47.5% (38/80) as unclear risk. In the predictors domain, 90% (72/80) were rated as unclear risk and 10% (8/80) as low risk; no models were rated as high risk. In the outcome domain, 80% (64/80) were rated as unclear risk, 13.8% (11/80) as high risk, and 6.2% (5/80) as low risk. In the analysis domain, 86.3% (69/80) were rated as high risk, 11.3% (9/80) as unclear risk, and 2.5% (2/80) as low risk (Fig. 12).

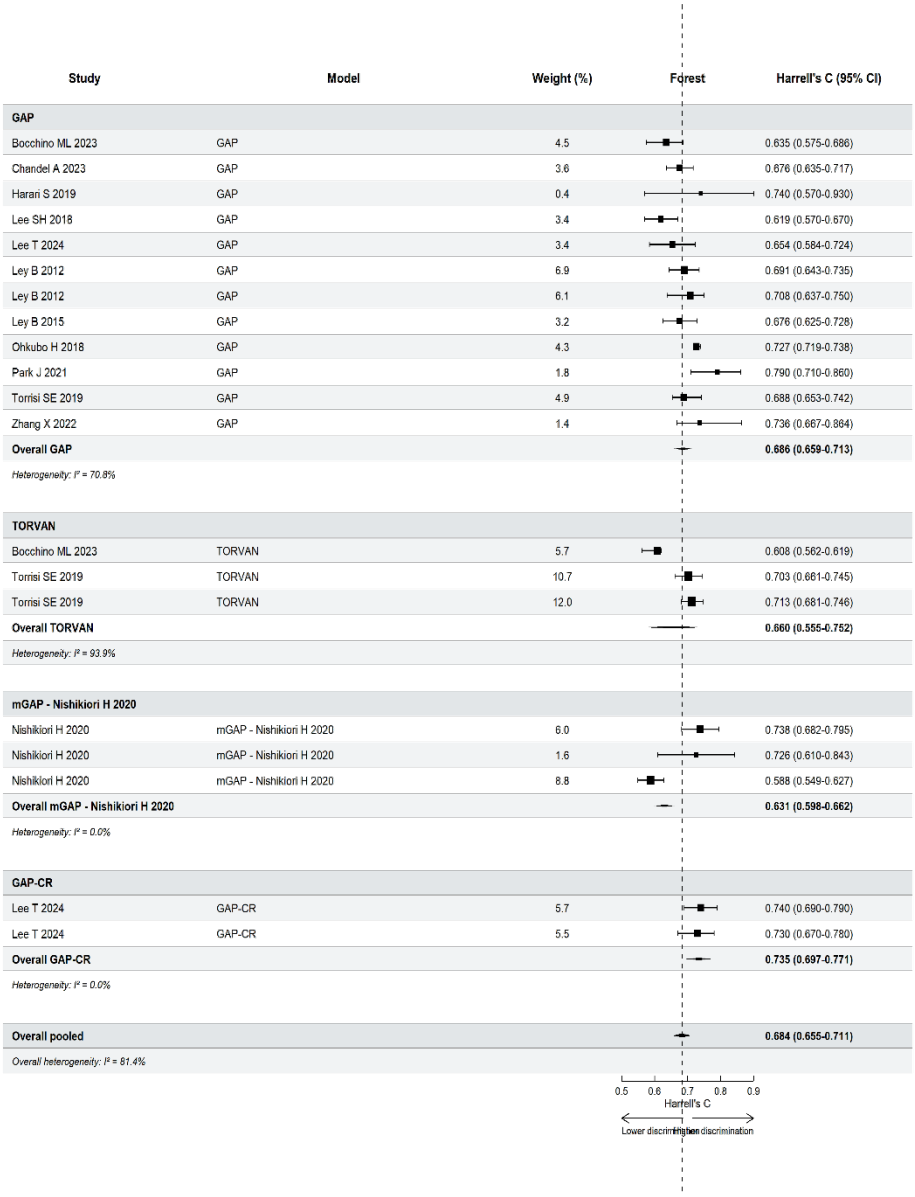
Applicability concerns were generally low or unclear. In the participants domain, 65% (52/80) were rated as low concern and 35% (28/80) as unclear concern. In the predictors domain, 43.6% (35/80) were rated as low concern, 48.8% (39/80) as unclear, and 7.5% (6/80) as high concern. In the outcome domain, 90% (72/80) were rated as low concern and 10% (8/80) as unclear concern (Fig. 13).

Overall, these findings suggest minimal or uncertain applicability concerns, while the risk of bias in external validation studies was generally high, primarily due to ratings in the analysis domain (Fig. 14).

12.5 Meta-analysis of prognostic mortality models

Measures of discrimination and calibration were extracted to assess model performance. Although multiple performance measures were extracted, only Harrell's C-index, as a measure of discrimination, was reported with sufficient consistency across studies to enable quantitative synthesis. Meta-analysis was performed when at least two estimates of Harrell's C-index with corresponding 95% confidence intervals were available for the same statistical model.

Figure 15 - Forest plot of pooled Harrell's C-index estimates for prognostic mortality models in IPF



Legend. Forest plot showing the pooled Harrell's C-index estimates for the prognostic mortality models evaluated in at least two datasets. Squares represent individual effect estimates weighted by inverse variance, horizontal lines indicate 95% confidence intervals, and diamonds represent pooled estimates from random-effects meta-analyses; GAP-Gender age Physiology; GAP-CR: GAP - Chest radiographs; mGAP-modified GAP

Figure 15 shows the pooled discriminative performance of prognostic models with at least two eligible Harrell's C estimates. The GAP model was the most extensively evaluated, with 12 performance estimates from 11 studies, yielding a pooled Harrell's C of 0.686 (95% CI 0.659–0.713).

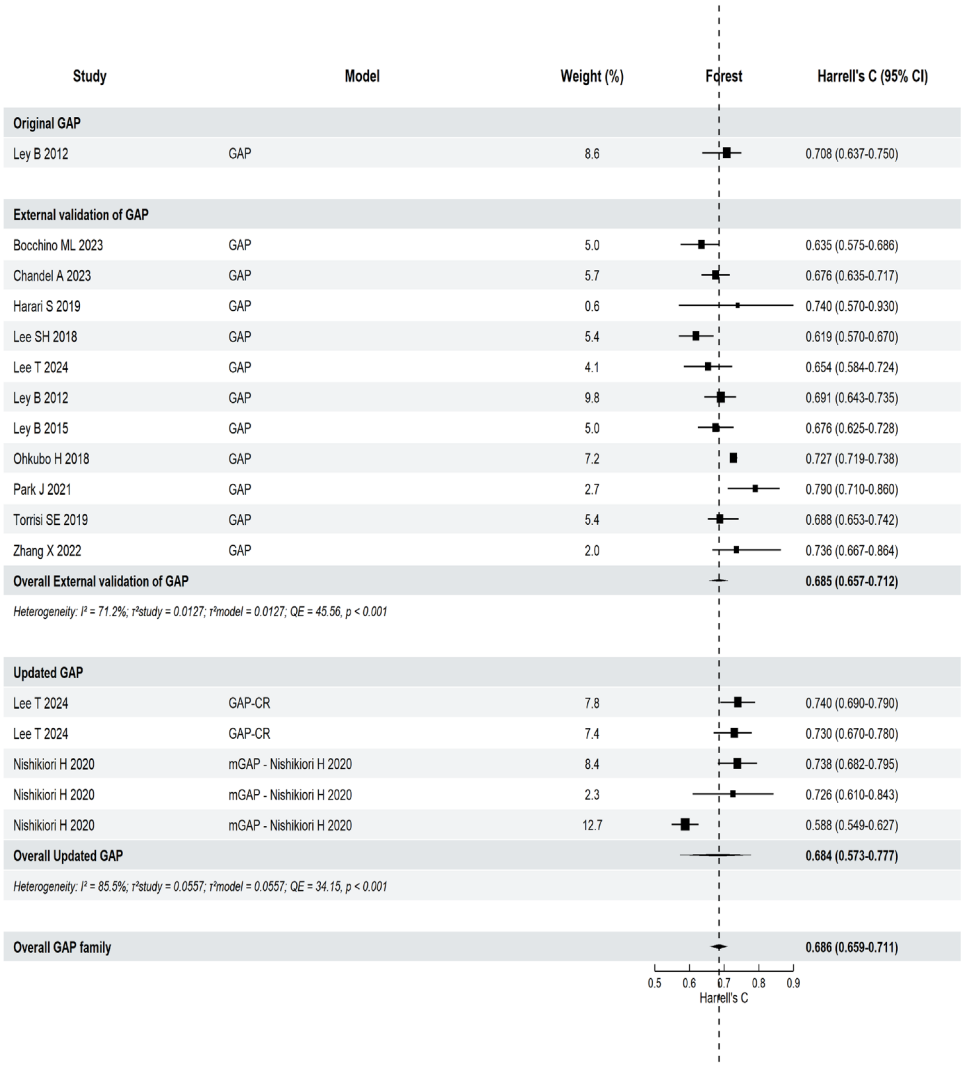
Among the remaining eligible models, GAP-CR showed the highest pooled discrimination (Harrell's C = 0.735, 95% CI 0.697–0.771), followed by TORVAN (0.660, 95% CI 0.555–0.752) and mGAP–Nishikiori H 2020 (0.631, 95% CI 0.598–0.662). The overall pooled Harrell's C across all eligible models was 0.684 (95% CI 0.655–0.711).

Heterogeneity was substantial for GAP (I^2 = 70.8%) and high for TORVAN (I^2 = 93.9%), whereas I^2 was 0% for both mGAP–Nishikiori H 2020 and GAP-CR. However, the interpretation of these heterogeneity estimates should be cautious, given the small number of performance estimates and the fact that they came from the same individual studies. Overall heterogeneity across models was high (I^2 = 81.4%).

Although GAP-CR showed the highest pooled discrimination, this estimate was based on only two performance estimates from a single study. Similarly, the evidence base for TORVAN and mGAP–Nishikiori H 2020 was considerably smaller than that for GAP. Overall, GAP remained the model supported by the most extensive evidence base across independent studies

To assess the discriminative performance of the GAP model across different stages of model development and validation, subgroup analyses were performed according to the original GAP model, external validation studies, and updated GAP-derived models. The pooled Harrell's C-index estimates and corresponding 95% confidence intervals are presented in Fig 16.

Figure 16 - Forest plot of pooled Harrell's C-index estimates for the original GAP model, external validation studies, updated GAP-derived models



Legend: Forest plot of pooled Harrell's C-index estimates for the original GAP model, external validation studies, updated GAP-derived models, and the overall GAP family. Squares represent study-specific estimates weighted by inverse variance, horizontal lines indicate 95% confidence intervals, and diamonds indicate pooled random-effects estimates. GAP-Gender, Age, Physiology model

Seventeen performance estimates from GAP and GAP-derived prognostic models were included (Figure 16). The pooled C-index across the GAP family was 0.686 (95% CI 0.659–0.711). External validations of the original GAP models showed a pooled Harrell's C-index of 0.685 (95% CI 0.657-0.712) with substantial heterogeneity (I^2 = 71.2%). Updated GAP models showed similar pooled discrimination (Harrell's C-index = 0.684; 95% CI 0.573–0.777), with marked heterogeneity (I2 = 85.5%), likely reflecting variability in both study populations and models.

13 Discussion

This systematic review identified 59 original and 26 updated mortality prediction models for IPF across 99 eligible studies, along with 80 external validations of existing models. Predictor selection varied widely: genomic variables were included in 52.5% of models, while age (30.5%), FVC (28.8%), DLCO (25.4%), and gender (23.7%) were the most used clinical and physiological predictors.

Discrimination was assessed in 93.2% of original models and remained the most frequently evaluated performance measure in external validation studies (81%). In contrast, calibration was assessed in only 42.4% of original models and was not evaluated in 80.8% of external validation studies, limiting assessment of the accuracy of absolute risk predictions.

Although 58% of eligible models had undergone external validation, only four models originally developed for mortality prediction in IPF were externally validated by independent research groups. Furthermore, 23.7% of original models reported only apparent performance, which may lead to optimistic estimates of predictive performance, particularly when internal validation is inadequate. The GAP index had the most extensive validation evidence, followed by the CPI and the models by Karampitsakos et al., Li X et al., and Li Y et al.

Previous literature reviews of prognostic models examined narrower subsets of the available evidence. Rozanski and Mura (2014) identified eight studies evaluating multidimensional prognostic indices and provided an early synthesis of clinical prognostic tools for IPF (8). More recently, Wu et al. identified 17 mortality prediction models; however, their eligibility criteria excluded studies focused exclusively on external validation and models based solely on genetic predictors (44). Khalid et al. addressed a more specific area, including eight studies focused on artificial intelligence,

machine learning, and radiomics approaches using CT imaging for IPF diagnosis and prognosis (145).

In comparison, the present review considered models incorporating a wide range of predictors, including clinical, physiological, imaging, biomarker, genomic, and AI-derived predictors. By including studies of model development, updating, and external validation within the same review, the present study provides a more comprehensive overview of available mortality prediction models and the extent to which their performance has been evaluated beyond the original development populations.

This broader assessment revealed considerable clinical and methodological variability across the IPF prognostic modeling literature, particularly in study populations, prediction horizons, and methods. Only seven development studies (12%) were conducted in multicenter cohorts, while several genomic models relied on repeatedly analyzed public datasets, including GSE70866 and GSE28042. Prediction horizons varied considerably across studies, so performance estimates did not necessarily refer to the same period of mortality risk and were therefore not directly comparable. Methodological variability was also evident in predictor selection, missing-data handling, and validation strategies. These differences should be considered when comparing reported performance across prognostic models.

Model performance assessment remained largely focused on discrimination, including in external validation studies, while other dimensions of predictive performance received considerably less attention. Overall predictive performance was rarely evaluated, with the Brier score reported for only three models and R^2 for one. Formal assessment of clinical utility was limited to seven studies. This is important because discrimination alone does not establish whether predicted mortality probabilities are accurate or whether using a model would improve clinical decision-making.

Among models with sufficient comparable external validation data for meta-analysis, the original GAP model had the largest evidence base. Across 11 external

validation estimates, Harrell's C-index ranged from 0.619 to 0.790, with a pooled C-index of 0.685 (95% CI 0.657-0.712), compared with the 0.708 reported for the original GAP model. These findings indicate that the GAP model retained discriminative ability across multiple external validation cohorts.

Comparisons of discrimination across different models should nevertheless be interpreted cautiously, as observed differences may reflect variation in study populations, prediction horizons, and evaluation methods rather than differences in predictive performance alone. More broadly, prognostic model evaluation should integrate discrimination with calibration, overall predictive performance, and clinical utility rather than relying on discrimination alone.

Methodological quality was a major limitation across the evidence base, with a high risk of bias particularly common in the analysis domain. High risk of bias in the analysis domain was identified in 55 of the 59 original models, 21 of the 26 updated models, and 69 of the 80 external validation studies. Several factors may have contributed to these findings, including reliance on apparent performance in some development studies, inadequate handling of missing data, predictor-selection methods, and sample size relative to model complexity. Missing-data reporting was especially poor, with 45 original studies not reporting how missing observations were handled and 13 relying on complete-case analysis. These methodological features may increase the risk of overfitting and optimism in reported performance estimates and reduce reproducibility when models are applied to new data.

The generalizability and clinical applicability of many models remain uncertain because the extent and independence of external validation varied considerably across the evidence base. Practical implementation also depends on predictor requirements, as models based on routinely available demographic and physiological variables may be more readily applicable than those requiring genomic assays, specialized biomarkers, or advanced imaging analysis. Moreover, the limited assessment of calibration and clinical utility means that evidence remains insufficient to determine whether many

models provide accurate individual risk estimates or improve clinical decision-making in practice. Future research should therefore prioritize independent external and temporal validation of existing models rather than continued development of new models without adequate evaluation. Prediction horizons should be clearly defined, and where sufficient comparable evidence is available, horizon-specific meta-analyses of the same model should be considered. Temporal transportability should also be examined, particularly whether models developed before widespread antifibrotic use retain their performance in contemporary treated populations, with recalibration or updating considered where performance deteriorates.

This review has several strengths, including the inclusion of 99 studies, consideration of diverse predictor domains, and assessment of models across development, updating, and external validation. This approach enabled examination of prognostic models across different stages of evaluation. Quantitative synthesis was performed when sufficiently comparable performance estimates were available, allowing model-specific performance to be evaluated without treating fundamentally different prognostic models as estimates of a common effect

Nevertheless, several limitations remain. Clinical and methodological variability across studies, including differences in study populations, predictor composition, prediction horizons, and analytical approaches, limited the comparability of reported performance estimates. Incomplete reporting in primary studies also restricted the assessment of several methodological and performance-related characteristics. Furthermore, the limited number of models with sufficiently comparable external validation data constrained the extent of quantitative synthesis. EMBASE was not searched because institutional access was unavailable, and conference proceedings and abstract books were not comprehensively searched, which may have led to the omission of relevant evidence.

Despite these limitations, this review provides a comprehensive overview of mortality prediction modelling in IPF and indicates a need for greater emphasis on rigorous validation, comprehensive performance assessment, and clinical applicability rather than the continued development of new prognostic models.

14 Conclusions:

This systematic review demonstrates that mortality prediction modelling in IPF is an extensive and evolving field, encompassing models based on diverse clinical, physiological, imaging, biomarker, genomic, and AI-derived predictors. Although many models have undergone external validation, the extent and independence of validation varied considerably, with the GAP index having the most extensive validation evidence. Model performance assessment remained predominantly focused on discrimination, whereas calibration, overall predictive performance, and clinical utility were less frequently evaluated. Furthermore, methodological limitations and the high risk of bias identified across the evidence base restrict confidence in the performance and generalisability of many models.

Future research should prioritise rigorous independent and temporal validation of existing models, comprehensive assessment of predictive performance, and evaluation of clinical utility rather than the continued development of new models without adequate validation. Particular attention should be given to clearly defined and clinically relevant prediction horizons, as differences in the time frame of mortality prediction affect the comparability and interpretation of model performance. Where sufficient evidence is available, performance should be evaluated and synthesised at specific prediction horizons.

15 References

1. Wijsenbeek M, Cottin V. Spectrum of Fibrotic Lung Diseases. N Engl J Med. 2020;383(10):958-68.
2. Nathan SD, Shlobin OA, Weir N, Ahmad S, Kaldjob JM, Battle E, et al. Long-term course and prognosis of idiopathic pulmonary fibrosis in the new millennium. Chest. 2011;140(1):221-9.
3. King TE, Jr., Bradford WZ, Castro-Bernardini S, Fagan EA, Glaspole I, Glassberg MK, et al. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. N Engl J Med. 2014;370(22):2083-92.
4. King TE, Jr., Bradford WZ, Castro-Bernardini S, Fagan EA, Glaspole I, Glassberg MK, et al. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. N Engl J Med. 2014;370(22):2083-92.
5. Richeldi L, Azuma A, Cottin V, Kreuter M, Maher TM, Martinez FJ, et al. Nerandomilast in Patients with Idiopathic Pulmonary Fibrosis. N Engl J Med. 2025;392(22):2193-202.
6. Podolanczuk AJ, Thomson CC, Remy-Jardin M, Richeldi L, Martinez FJ, Kolb M, Raghu G. Idiopathic pulmonary fibrosis: state of the art for 2023. Eur Respir J. 2023;61(4).
7. Tomassetti S, Ryu JH, Poletti V. Staging systems and disease severity assessment in interstitial lung diseases. Curr Opin Pulm Med. 2015;21(5):463-9.
8. Rozanski C, Mura M. Multi-dimensional indices to stage idiopathic pulmonary fibrosis: a systematic review. Sarcoidosis Vasc Diffuse Lung Dis. 2014;31(1):8-18.
9. Wijsenbeek M, Suzuki A, Maher TM. Interstitial lung diseases. Lancet. 2022;400(10354):769-86.
10. Cottin V, Hirani NA, Hotchkin DL, Nambiar AM, Ogura T, Otaola M, et al. Presentation, diagnosis and clinical course of the spectrum of progressive-fibrosing interstitial lung diseases. Eur Respir Rev. 2018;27(150).
11. Travis WD, Costabel U, Hansell DM, King TE, Jr., Lynch DA, Nicholson AG, et al. An official American Thoracic Society/European Respiratory Society statement: Update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. Am J Respir Crit Care Med. 2013;188(6):733-48.
12. Ryerson CJ, Adegunsoye A, Piciucchi S, Hariri LP, Khor YH, Wijsenbeek MS, et al. Update of the international multidisciplinary classification of the interstitial pneumonias: an ERS/ATS statement. Eur Respir J. 2025;66(6).
13. American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. This joint statement of the American Thoracic Society (ATS), and the European Respiratory Society (ERS) was adopted by the ATS board of directors, June 2001 and by the ERS Executive Committee, June 2001. Am J Respir Crit Care Med. 2002;165(2):277-304.
14. 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-e68.
15. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med. 2011;183(6):788-824.

16. Raghu G, Remy-Jardin M, Richeldi L, Thomson CC, Inoue Y, Johkoh T, 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):e18-e47.
17. Cottin V, Tomassetti S, Valenzuela C, Walsh SLF, Antoniou KM, Bonella F, et al. Integrating Clinical Probability into the Diagnostic Approach to Idiopathic Pulmonary Fibrosis: An International Working Group Perspective. Am J Respir Crit Care Med. 2022;206(3):247-59.
18. Ma X, Zhu L, Kurche JS, Xiao H, Dai H, Wang C. Global and regional burden of interstitial lung disease and pulmonary sarcoidosis from 1990 to 2019: results from the Global Burden of Disease study 2019. Thorax. 2022;77(6):596-605.
19. Shah Gupta R, Koteci A, Morgan A, George PM, Quint JK. Incidence and prevalence of interstitial lung diseases worldwide: a systematic literature review. BMJ Open Respir Res. 2023;10(1).
20. Hilberg O, Hoffmann-Vold AM, Smith V, Bouros D, Kilpelainen M, Guiot J, et al. Epidemiology of interstitial lung diseases and their progressive-fibrosing behaviour in six European countries. ERJ Open Res. 2022;8(1).
21. Diamantopoulos A, Wright E, Vlahopoulou K, Cornic L, Schoof N, Maher TM. The Burden of Illness of Idiopathic Pulmonary Fibrosis: A Comprehensive Evidence Review. Pharmacoeconomics. 2018;36(7):779-807.
22. Rodriguez-Nieto MJ, Cano-Jimenez E, Romero Ortiz AD, Villar A, Morros M, Ramon A, Armengol S. Economic Burden of Idiopathic Pulmonary Fibrosis in Spain: A Prospective Real-World Data Study (OASIS Study). Pharmacoeconomics. 2023;41(8):999-1010.
23. Raghu G. Idiopathic pulmonary fibrosis: lessons from clinical trials over the past 25 years. Eur Respir J. 2017;50(4).
24. Nathan SD, Albera C, Bradford WZ, Costabel U, du Bois RM, Fagan EA, et al. Effect of continued treatment with pirfenidone following clinically meaningful declines in forced vital capacity: analysis of data from three phase 3 trials in patients with idiopathic pulmonary fibrosis. Thorax. 2016;71(5):429-35.
25. Noble PW, Albera C, Bradford WZ, Costabel U, Glassberg MK, Kardatzke D, et al. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. Lancet. 2011;377(9779):1760-9.
26. Kolb M, Richeldi L, Behr J, Maher TM, Tang W, Stowasser S, et al. Nintedanib in patients with idiopathic pulmonary fibrosis and preserved lung volume. Thorax. 2017;72(4):340-6.
27. Cosgrove GP, Bianchi P, Danese S, Lederer DJ. Barriers to timely diagnosis of interstitial lung disease in the real world: the INTENSITY survey. BMC Pulm Med. 2018;18(1):9.
28. Hewson T, McKeever TM, Gibson JE, Navaratnam V, Hubbard RB, Hutchinson JP. Timing of onset of symptoms in people with idiopathic pulmonary fibrosis. Thorax. 2017.
29. Collard HR, Ryerson CJ, Corte TJ, Jenkins G, Kondoh Y, Lederer DJ, et al. Acute Exacerbation of Idiopathic Pulmonary Fibrosis. An International Working Group Report. Am J Respir Crit Care Med. 2016;194(3):265-75.
30. Song JW, Hong SB, Lim CM, Koh Y, Kim DS. Acute exacerbation of idiopathic pulmonary fibrosis: incidence, risk factors and outcome. Eur Respir J. 2011;37(2):356-63.
31. Qiu M, Chen Y, Ye Q. Risk factors for acute exacerbation of idiopathic pulmonary fibrosis: A systematic review and meta-analysis. Clin Respir J. 2018;12(3):1084-92.
32. 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.
33. Raghu G, Rochwerg B, Zhang Y, Garcia CA, Azuma A, Behr J, et al. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline: Treatment of Idiopathic Pulmonary Fibrosis. An Update of the 2011 Clinical Practice Guideline. Am J Respir Crit Care Med. 2015;192(2):e3-19.

34. Petnak T, Lertjitbanjong P, Thongprayoon C, Moua T. Impact of Antifibrotic Therapy on Mortality and Acute Exacerbation in Idiopathic Pulmonary Fibrosis: A Systematic Review and Meta-Analysis. Chest. 2021;160(5):1751-63.
35. Ley B, Swigris J, Day BM, Stauffer JL, Raimundo K, Chou W, Collard HR. Pirfenidone Reduces Respiratory-related Hospitalizations in Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med. 2017;196(6):756-61.
36. Cilli A, Uzer F. Disease progression in idiopathic pulmonary fibrosis under anti-fibrotic treatment. Sarcoidosis Vasc Diffuse Lung Dis. 2023;40(3):e2023034.
37. Ley B, Collard HR, King TE, Jr. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2011;183(4):431-40.
38. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill 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.
39. Snell KIE, Levis B, Damen JAA, Dhiman P, Debray TPA, Hooft L, et al. Transparent reporting of multivariable prediction models for individual prognosis or diagnosis: checklist for systematic reviews and meta-analyses (TRIPOD-SRMA). BMJ. 2023;381:e073538.
40. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med. 2014;11(10):e1001744.
41. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. 2015;162(1):55-63.
42. Debray TP, Damen JA, Snell KI, Ensor J, Hooft L, Reitsma JB, et al. A guide to systematic review and meta-analysis of prediction model performance. BMJ. 2017;356:i6460.
43. C R. <Multi-dimensional indeces to stage idiopathic pulmonary.pdf>. SARCOIDOSIS VASCULITIS AND DIFFUSE LUNG DISEASES 2014; 31; 8-18. 2014;Vol. 31 No. 1 (2014).
44. Wu Y, Li P, Wang M, Liu Y, Leng J, Li X, et al. A Systematic Review of Mortality Risk Prediction Models for Idiopathic Pulmonary Fibrosis. Br J Hosp Med (Lond). 2025;86(4):1-22.
45. Fernandez-Felix BM, Lopez-Alcalde J, Roque M, Muriel A, Zamora J. CHARMS and PROBAST at your fingertips: a template for data extraction and risk of bias assessment in systematic reviews of predictive models. BMC Med Res Methodol. 2023;23(1):44.
46. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med. 2019;170(1):51-8.
47. Bao Y, Yang S, Zhao H, Wang Y, Li K, Liu X, et al. A prognostic model of idiopathic pulmonary fibrosis constructed based on macrophage and mitochondria-related genes. BMC Pulm Med. 2024;24(1):176.
48. Cai H, Chen S, Li X, Liu H, Zhang Y, Zhuang Q. The Combined Model of CX3CR1-Related Immune Infiltration Genes to Evaluate the Prognosis of Idiopathic Pulmonary Fibrosis. Front Immunol. 2022;13:837188.
49. Chen HZ, Xia ZK, Qing B, Wang W, Gu LG, Chen Y, et al. Analysis of necroptosis-related prognostic genes and immune infiltration in idiopathic pulmonary fibrosis. Front Immunol. 2023;14:1119139.
50. du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, et al. Ascertainment of individual risk of mortality for patients with idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2011;184(4):459-66.

51. Fan G, Liu J, Wu Z, Li C, Zhang Y. Development and validation of the prognostic model based on autophagy-associated genes in idiopathic pulmonary fibrosis. Front Immunol. 2022;13:1049361.
52. Feng A, Caro YM, Gardner C, Grischo G, Liang Y, Wickremasinghe PD, et al. PTK2-associated gene signature could predict the prognosis of IPF. Respir Res. 2023;24(1):304.
53. Wells AU, Desai SR, Rubens MB, Goh NS, Cramer D, Nicholson AG, et al. Idiopathic pulmonary fibrosis: a composite physiologic index derived from disease extent observed by computed tomography. Am J Respir Crit Care Med. 2003;167(7):962-9.
54. Fukuda CY, Soares MR, de Castro Pereira CA. A score without diffusion capacity of the lung for carbon monoxide for estimating survival in idiopathic pulmonary fibrosis. Medicine (Baltimore). 2020;99(25):e20739.
55. Gudmundsson E, Zhao A, Mogulkoc N, Stewart I, Jones MG, Van Moorsel CHM, et al. Pleuroparenchymal fibroelastosis in idiopathic pulmonary fibrosis: Survival analysis using visual and computer-based computed tomography assessment. EClinicalMedicine. 2021;38:101009.
56. Hamai K, Iwamoto H, Ishikawa N, Horimasu Y, Masuda T, Miyamoto S, et al. Comparative Study of Circulating MMP-7, CCL18, KL-6, SP-A, and SP-D as Disease Markers of Idiopathic Pulmonary Fibrosis. Dis Markers. 2016;2016:4759040.
57. Huang Y, Ma SF, Vij R, Oldham JM, Herazo-Maya J, Broderick SM, et al. A functional genomic model for predicting prognosis in idiopathic pulmonary fibrosis. BMC Pulm Med. 2015;15:147.
58. He Y, Shang Y, Li Y, Wang M, Yu D, Yang Y, et al. An 8-ferroptosis-related genes signature from Bronchoalveolar Lavage Fluid for prognosis in patients with idiopathic pulmonary fibrosis. BMC Pulm Med. 2022;22(1):15.
59. Hou ZL, Peng D, Yang JJ, Zhang S, Wang JX. A novel seven-gene risk profile in BALF to identify high-risk patients with idiopathic pulmonary fibrosis. J Thorac Dis. 2022;14(5):1450-+.
60. Huang G, Xu X, Ju C, Zhong N, He J, Tang XX. Identification and validation of autophagy-related gene expression for predicting prognosis in patients with idiopathic pulmonary fibrosis. Front Immunol. 2022;13:997138.
61. Jacob J, Bartholmai BJ, Rajagopalan S, Kokosi M, Nair A, Karwoski R, et al. Mortality prediction in idiopathic pulmonary fibrosis: evaluation of computer-based CT analysis with conventional severity measures. Eur Respir J. 2017;49(1).
62. Jacob J, Aksman L, Mogulkoc N, Procter AJ, Gholipour B, Cross G, et al. Serial CT analysis in idiopathic pulmonary fibrosis: comparison of visual features that determine patient outcome. Thorax. 2020;75(8):648-54.
63. Kalluri M, Cui Y, Wang T, Bakal JA. Validation of a Novel Clinical Dyspnea Scale – A Retrospective Pilot Study. American Journal of Hospice and Palliative Medicine. 2024;41(3):253-61.
64. Karampitsakos T, Tourki B, Jia M, Perrot CY, Visinescu B, Zhao A, et al. The transcriptome of CD14 (+) CD163 (-) HLA-DR (low) monocytes predicts mortality in Idiopathic Pulmonary Fibrosis. medRxiv. 2024.
65. King TE, Jr., Tooze JA, Schwarz MI, Brown KR, Cherniack RM. Predicting survival in idiopathic pulmonary fibrosis: scoring system and survival model. Am J Respir Crit Care Med. 2001;164(7):1171-81.
66. Kishaba T, Nagano H, Nei Y, Yamashiro S. Body mass index-percent forced vital capacity-respiratory hospitalization: new staging for idiopathic pulmonary fibrosis patients. J Thorac Dis. 2016;8(12):3596-604.
67. Kondoh Y, Taniguchi H, Kataoka K, Furukawa T, Ando M, Murotani K, et al. Disease severity staging system for idiopathic pulmonary fibrosis in Japan. Respirology. 2017;22(8):1609-14.

68. Lee SM, Seo JB, Oh SY, Kim TH, Song JW, Lee SM, Kim N. Prediction of survival by texture-based automated quantitative assessment of regional disease patterns on CT in idiopathic pulmonary fibrosis. Eur Radiol. 2018;28(3):1293-300.
69. 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.
70. Kam MLW, Li HH, Tan YH, Low SY. Validation of the ILD-GAP Model and a Local Nomogram in a Singaporean Cohort. Respiration. 2019;98(5):383-90.
71. Liao YW, Chen YM, Liu MC, Wu YC, Hsu CY, Fu PK, et al. Multidisciplinary-derived clinical score for accurate prediction of long-term mortality in fibrotic lung disease patients. Eur J Med Res. 2024;29(1):69.
72. Liu J, Gu LM, Li WL. The Prognostic Value of Integrated Analysis of Inflammation and Hypoxia-Related Genes in Idiopathic Pulmonary Fibrosis. Front Immunol. 2022;13:730186.
73. Li M, Wang K, Zhang YP, Fan M, Li AQ, Zhou JJ, et al. Ferroptosis-Related Genes in Bronchoalveolar Lavage Fluid Serves as Prognostic Biomarkers for Idiopathic Pulmonary Fibrosis. Frontiers in Medicine. 2021;8:693959.
74. Lin F, Lin K, Li D, Kong W, Zhuang J, He W, et al. Machine learning-based integration identifies plasma cells-related gene signature ST6GAL1 in idiopathic pulmonary fibrosis. BMC Pulm Med. 2025;25(1):295.
75. Li X, Cai H, Cai Y, Zhang Q, Ding Y, Zhuang Q. Investigation of a Hypoxia-Immune-Related Microenvironment Gene Signature and Prediction Model for Idiopathic Pulmonary Fibrosis. Front Immunol. 2021;12:629854.
76. Li Y, Deng Y, He J. A novel prognostic index based on the analysis of glycolysis-related genes in idiopathic pulmonary fibrosis. Medicine (Baltimore). 2023;102(11):e33330.
77. Lu YJ, Chen JK, Tang K, Wang SS, Tian Z, Wang MJ, et al. Development and Validation of the Prognostic Index Based on Inflammation-Related Gene Analysis in Idiopathic Pulmonary Fibrosis. Frontiers in Molecular Biosciences. 2021;8:667459.
78. Luo H, Yan J, Zhou X. Constructing an extracellular matrix-related prognostic model for idiopathic pulmonary fibrosis based on machine learning. BMC Pulm Med. 2023;23(1):397.
79. Luo P, Gu Q, Wang J, Meng X, Zhao M. Developing an IPF Prognostic Model and Screening for Key Genes Based on Cold Exposure-Related Genes Using Bioinformatics Approaches. Biomedicines. 2025;13(3).
80. Lyu Y, Guo C, Zhang H. Fatty acid metabolism-related genes in bronchoalveolar lavage fluid unveil prognostic and immune infiltration in idiopathic pulmonary fibrosis. Front Endocrinol (Lausanne). 2022;13:1001563.
81. Mogulkoc N, Brutsche MH, Bishop PW, Greaves SM, Horrocks AW, Egan JJ. Pulmonary function in idiopathic pulmonary fibrosis and referral for lung transplantation. Am J Respir Crit Care Med. 2001;164(1):103-8.
82. Nolan CM, Maddocks M, Maher TM, Banya W, Patel S, Barker RE, et al. Gait speed and prognosis in patients with idiopathic pulmonary fibrosis: a prospective cohort study. Eur Respir J. 2019;53(2).
83. Polat G, Özdemir Ö, Ermin S, Serçe Unat D, Demirci Üçsular F. Predictive factors of mortality in patients with idiopathic pulmonary fibrosis treated with antifibrotics: a novel prognostic scoring system. Sarcoidosis Vasc Diffuse Lung Dis. 2024;41(2):e2024021.
84. Prasse A, Binder H, Schupp JC, Kayser G, Bargagli E, Jaeger B, et al. BAL Cell Gene Expression Is Indicative of Outcome and Airway Basal Cell Involvement in Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med. 2019;199(5):622-30.

85. Richards TJ, Kaminski N, Baribaud F, Flavin S, Brodmerkel C, Horowitz D, et al. Peripheral blood proteins predict mortality in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2012;185(1):67-76.
86. Soares MR, Pereira C, Ferreira R, Nei Aparecida Martins Coletta E, Silva Lima M, Muller Storrer K. A score for estimating survival in idiopathic pulmonary fibrosis with rest SpO2>88. Sarcoidosis Vasc Diffuse Lung Dis. 2015;32(2):121-8.
87. Song JW, Do KH, Jang SJ, Colby TV, Han S, Kim DS. Blood biomarkers MMP-7 and SP-A: predictors of outcome in idiopathic pulmonary fibrosis. Chest. 2013;143(5):1422-9.
88. Tian YQ, Zhou X, Tian M, Ren LJ, Zou RY, Jiang HY, et al. Nomogram model using serum Club cell secretory protein 16 to predict prognosis and acute exacerbation in patients with idiopathic pulmonary fibrosis. European Journal of Medical Research. 2025;30(1).
89. Torrisi SE, Ley B, Kreuter M, Wijsenbeek M, Vittinghoff E, Collard HR, Vancheri C. The added value of comorbidities in predicting survival in idiopathic pulmonary fibrosis: a multicentre observational study. Eur Respir J. 2019;53(3).
90. Tseng CM, Chen MY, Kao CY, Tao CW. Investigation of clinical predictors of survival in idiopathic pulmonary fibrosis patients: A cohort study in Taiwan. J Chin Med Assoc. 2022;85(5):578-83.
91. Yu WS, Lee JG, Paik HC, Kim SJ, Lee S, Kim SY, et al. Carcinoembryonic antigen predicts waitlist mortality in lung transplant candidates with idiopathic pulmonary fibrosis. Eur J Cardiothorac Surg. 2018;54(5):847-52.
92. Wang E, Wang Y, Zhou S, Xia X, Han R, Fei G, et al. Identification of three hub genes related to the prognosis of idiopathic pulmonary fibrosis using bioinformatics analysis. Int J Med Sci. 2022;19(9):1417-29.
93. Wang X, Yang L, Wang Y, Dou X, Li Y, Wang K, Zhang H. Establishment and validation of a prognostic model for idiopathic pulmonary fibrosis based on mitochondrial-related genes. J Thorac Dis. 2024;16(11):7427-45.
94. Wang Z, Shen L, Wang J, Huang J, Tao H, Zhou X. Prognostic analysis of m6A-related genes as potential biomarkers in idiopathic pulmonary fibrosis. Front Genet. 2022;13:1059325.
95. Wu X, Yin C, Chen X, Zhang Y, Su Y, Shi J, et al. Idiopathic Pulmonary Fibrosis Mortality Risk Prediction Based on Artificial Intelligence: The CTPF Model. Front Pharmacol. 2022;13:878764.
96. Xia Y, Lei C, Yang D, Luo H. Construction and validation of a bronchoalveolar lavage cell-associated gene signature for prognosis prediction in idiopathic pulmonary fibrosis. Int Immunopharmacol. 2021;92:107369.
97. Xiang Y, Huang G, Luo C, Jiang J, Zhang T, Zeng Q, et al. Investigates the Role of PANoptosis in Idiopathic Pulmonary Fibrosis and Potential Therapeutic Targets. Journal of Inflammation Research. 2024;17:11605-29.
98. Yin YQ, Peng F, Situ HJ, Xie JL, Tan L, Wei J, et al. Construction of prediction model of inflammation related genes in idiopathic pulmonary fibrosis and its correlation with immune microenvironment. Front Immunol. 2022;13.
99. Zhang X, Ren Y, Xie B, Ye Q, Ban C, Zhang S, et al. Blood monocyte counts as a prognostic biomarker and predictor in Chinese patients with idiopathic pulmonary fibrosis. Front Med (Lausanne). 2022;9:955125.
100. Zhao J, Wang C, Fan R, Liu X, Zhang W. A prognostic model based on clusters of molecules related to epithelial-mesenchymal transition for idiopathic pulmonary fibrosis. Front Genet. 2022;13:1109903.

101. Zheng J, Dong H, Zhang T, Ning J, Xu Y, Cai C. Development and Validation of a Novel Gene Signature for Predicting the Prognosis of Idiopathic Pulmonary Fibrosis Based on Three Epithelial-Mesenchymal Transition and Immune-Related Genes. Front Genet. 2022;13:865052.
102. Zhou M, Ouyang J, Zhang G, Zhu X. Prognostic value of tripartite motif (TRIM) family gene signature from bronchoalveolar lavage cells in idiopathic pulmonary fibrosis. BMC Pulm Med. 2022;22(1):467.
103. Zhong C, Lei YQ, Zhang JY, Zheng Q, Liu ZY, Xu YL, et al. Prognostic Function and Immunologic Landscape of a Predictive Model Based on Five Senescence-Related Genes in IPF Bronchoalveolar Lavage Fluid. Biomedicines. 2024;12(6).
104. Zhu H, Zhou A, Zhang M, Pan L, Wu X, Fu C, et al. Comprehensive analysis of an endoplasmic reticulum stress-related gene prediction model and immune infiltration in idiopathic pulmonary fibrosis. Front Immunol. 2023;14:1305025.
105. Zinellu A, Collu C, Zinellu E, Ahmad K, Nasser M, Traclet J, et al. IC4: a new combined predictive index of mortality in idiopathic pulmonary fibrosis. Panminerva Med. 2022;64(2):228-34.
106. du Bois RM, Albera C, Bradford WZ, Costabel U, Leff JA, Noble PW, et al. 6-Minute walk distance is an independent predictor of mortality in patients with idiopathic pulmonary fibrosis. Eur Respir J. 2014;43(5):1421-9.
107. Fisher JH, Al-Hejaili F, Kandel S, Hirji A, Shapera S, Mura M. Multi-dimensional scores to predict mortality in patients with idiopathic pulmonary fibrosis undergoing lung transplantation assessment. Respir Med. 2017;125:65-71.
108. Lee SH, Park JS, Kim SY, Kim DS, Kim YW, Chung MP, et al. Comparison of CPI and GAP models in patients with idiopathic pulmonary fibrosis: a nationwide cohort study. Sci Rep. 2018;8(1):4784.
109. Sharp C, Adamali HI, Millar AB. A comparison of published multidimensional indices to predict outcome in idiopathic pulmonary fibrosis. ERJ Open Res. 2017;3(1).
110. Abe M, Tsushima K, Yoshioka K, Sakayori M, Suzuki K, Hirasawa Y, et al. The Gender-Age-Physiology system as a prognostic model in patients with idiopathic pulmonary fibrosis treated with nintedanib: a longitudinal cohort study. Adv Respir Med. 2020;88(5):369-76.
111. AlDalilah Y, Ganeshan B, Endozo R, Bomanji J, Porter JC, Machado M, et al. Filtration-histogram based texture analysis and CALIPER based pattern analysis as quantitative CT techniques in idiopathic pulmonary fibrosis: head-to-head comparison. Br J Radiol. 2022;95(1134):20210957.
112. Bocchino M, Bruzzese D, Scioscia G, Capitelli L, Tondo P, Rea G, et al. Disease stage-related survival in idiopathic pulmonary fibrosis patients treated with nintedanib and pirfenidone: An exploratory study. Respir Med Res. 2023;84:101013.
113. Chandel A, Pastre J, Valery S, King CS, Nathan SD. Derivation and validation of a simple multidimensional index incorporating exercise capacity parameters for survival prediction in idiopathic pulmonary fibrosis. Thorax. 2023;78(4):368-75.
114. Harari S, Caminati A, Confalonieri M, Poletti V, Vancheri C, Pesci A, et al. The prognostic role of Gender-Age-Physiology system in idiopathic pulmonary fibrosis patients treated with pirfenidone. Clin Respir J. 2019;13(3):166-73.
115. Kim ES, Choi SM, Lee J, Park YS, Lee CH, Yim JJ, et al. Validation of the GAP score in Korean patients with idiopathic pulmonary fibrosis. Chest. 2015;147(2):430-7.
116. Kreuter M, Lee JS, Tzouvelekis A, Oldham JM, Molyneaux PL, Weycker D, et al. Modified blood cell GAP model as a prognostic biomarker in idiopathic pulmonary fibrosis. ERJ Open Res. 2024;10(4).
117. Ley B, Bradford WZ, Weycker D, Vittinghoff E, du Bois RM, Collard HR. Unified baseline and longitudinal mortality prediction in idiopathic pulmonary fibrosis. Eur Respir J. 2015;45(5):1374-81.

118. Lee JH, Jang JH, Jang HJ, Kim SY, Chung MP, Yoo H, et al. New prognostic scoring system for mortality in idiopathic pulmonary fibrosis by modifying the gender, age, and physiology model with desaturation during the six-minute walk test. Front Med (Lausanne). 2023;10:1052129.
119. Lee T, Ahn SY, Kim J, Park JS, Kwon BS, Choi SM, et al. Deep learning-based prognostication in idiopathic pulmonary fibrosis using chest radiographs. Eur Radiol. 2024;34(7):4206-17.
120. Mora-Cuesta VM, Zuazaga-Fuentes J, Iturbe-Fernández D, Tello-Mena S, Izquierdo-Cuervo S, Alonso-Lecue P, Cifrián-Martínez JM. Prognostic Utility of the GAP Score in Interstitial Lung Disease Patients Evaluated for Lung Transplantation: A Single-Center Study. Clin Transplant. 2025;39(3):e70136.
121. Nam JG, Choi Y, Lee SM, Yoon SH, Goo JM, Kim H. Prognostic value of deep learning-based fibrosis quantification on chest CT in idiopathic pulmonary fibrosis. Eur Radiol. 2023;33(5):3144-55.
122. Nishikiori H, Chiba H, Lee SH, Kondoh S, Kamo KI, Nakamura K, et al. A modified GAP model for East-Asian populations with idiopathic pulmonary fibrosis. Respir Investig. 2020;58(5):395-402.
123. Song HF, Sun DJ, Ban CJ, Liu Y, Zhu M, Ye Q, et al. Independent Clinical Factors Relevant to Prognosis of Patients with Idiopathic Pulmonary Fibrosis. Med Sci Monit. 2019;25:4193-201.
124. Ohkubo H, Taniguchi H, Kondoh Y, Yagi M, Furukawa T, Johkoh T, et al. A Volumetric Computed Tomography Analysis of the Normal Lung in Idiopathic Pulmonary Fibrosis: The Relationship with the Survival. Intern Med. 2018;57(7):929-37.
125. Park J, Jung J, Yoon SH, Hong H, Kim H, Kim H, et al. CT quantification of the heterogeneity of fibrosis boundaries in idiopathic pulmonary fibrosis. Eur Radiol. 2021;31(7):5148-59.
126. 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.
127. Takuma S, Suzuki Y, Kono M, Hasegawa H, Hashimoto D, Yokomura K, et al. Neutrophil–lymphocyte ratio being associated with mortality risk in patients receiving antifibrotic therapy. Respiratory Medicine. 2024;223 C7 - 107542.
128. Whi W, Lee H, Cho YS, Lee H, Yoo H, Chung MP, Choi JY. Prognostic value of [(18)F]FDG PET/CT in patients with idiopathic pulmonary fibrosis. Sci Rep. 2025;15(1):143.
129. Win T, Screaton NJ, Porter JC, Ganeshan B, Maher TM, Fraioli F, et al. Pulmonary (18)F-FDG uptake helps refine current risk stratification in idiopathic pulmonary fibrosis (IPF). Eur J Nucl Med Mol Imaging. 2018;45(5):806-15.
130. 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 Respir J. 2023;17(9):831-40.
131. Chandel A, King CS, Ignacio RV, Pastre J, Shlobin OA, Khangoora V, et al. External validation and longitudinal application of the DO-GAP index to individualise survival prediction in idiopathic pulmonary fibrosis. ERJ Open Res. 2023;9(3).
132. Desai UD, Karkhanis VS, Joshi JM. TNMC-GAP (The New Modified Classification of Gender Age Physiology)-A novel risk assessment system in idiopathic pulmonary fibrosis. Medical Journal of Dr DY Patil Vidyapeeth. 2018:115-20.
133. Fraioli F, Lyasheva M, Porter JC, Bomanji J, Shortman RI, Endozo R, et al. Synergistic application of pulmonary (18)F-FDG PET/HRCT and computer-based CT analysis with conventional severity measures to refine current risk stratification in idiopathic pulmonary fibrosis (IPF). Eur J Nucl Med Mol Imaging. 2019;46(10):2023-31.
134. Hozumi H, Kataoka K, Kondoh Y, Isayama T, Okada J, Sugiura K, et al. Clinical Significance of Cold-Inducible RNA-Binding Protein in Idiopathic Pulmonary Fibrosis. Chest. 2021;160(6):2149-57.

135. Khor YH, Gutman L, Abu Hussein N, Johannson KA, Glaspole IN, Guler SA, et al. Incidence and Prognostic Significance of Hypoxemia in Fibrotic Interstitial Lung Disease: An International Cohort Study. Chest. 2021;160(3):994-1005.
136. 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 (Basel). 2023;10(2).
137. Lee JH, Jang JH, Jang HJ, Kim SY, Chung MP, Yoo H, et al. New prognostic scoring system for mortality in idiopathic pulmonary fibrosis by modifying the gender, age, and physiology model with desaturation during the six-minute walk test. Frontiers in Medicine. 2023;10:1052129.
138. Ley B, Elicker BM, Hartman TE, Ryerson CJ, Vittinghoff E, Ryu JH, et al. Idiopathic pulmonary fibrosis: CT and risk of death. Radiology. 2014;273(2):570-9.
139. Ryerson CJ, Vittinghoff E, Ley B, Lee JS, Mooney JJ, Jones KD, et al. Predicting survival across chronic interstitial lung disease: the ILD-GAP model. Chest. 2014;145(4):723-8.
140. Tian M, Zhu X, Ren L, Zhou X, Gu L, Meng K, et al. HE4-based nomogram for predicting overall survival in patients with idiopathic pulmonary fibrosis: construction and validation. Eur J Med Res. 2024;29(1):238.
141. Moran-Mendoza O, Singla A, Kalra A, Muelly M, Reicher JJ. Computed tomography machine learning classifier correlates with mortality in interstitial lung disease. Respir Investig. 2024;62(4):670-6.
142. Söllner JF, Bentink S, Hesslinger C, Leonard TB, Neely ML, Patel NM, et al. Utility of the 52-Gene Risk Score to Identify Patients with Idiopathic Pulmonary Fibrosis at Greater Risk of Mortality in the Era of Antifibrotic Therapy. Lung. 2024;202(5):595-9.
143. Sonaglioni A, Caminati A, Re M, Elia D, Trevisan R, Granato A, et al. Prognostic role of CHA(2)DS(2)-VASc score for mortality risk assessment in non-advanced idiopathic pulmonary fibrosis: a preliminary observation. Intern Emerg Med. 2023;18(3):755-67.
144. Suzuki T, Karayama M, Aoshima Y, Mori K, Yoshizawa N, Ichikawa S, et al. Association of the lung immune prognostic index with the survival of patients with idiopathic interstitial pneumonias. Respirology. 2024;29(2):136-45.
145. Khalid A. Radiomics-Based Artificial Intelligence and
Machine Learning Approach for the Diagnosis
and Prognosis of Idiopathic Pulmonary Fibrosis:
A Systematic Review. Cureus. 2025.