Risk Management and Healthcare Policy

Dovepress

open access to scientific and medical research

Open Access Full Text Article

REVIEW

Bias and Reporting Quality of Clinical Prognostic Models for Idiopathic Pulmonary Fibrosis: A Cross-Sectional Study

Jiaqi Di[1], Xuanlin Li[1], Jingjing Yang[1], Luguang Li[2], Xueqing Yu[2]

[1]Co-Construction Collaborative Innovation Center for Chinese Medicine and Respiratory Diseases by Henan & Education Ministry of P.R, Henna University of Chinese Medicine, Zhengzhou, 450046, People's Republic of China; [2]Department of Respiratory Diseases, The First Affiliated Hospital of Henan University of Chinese Medicine, Zhengzhou, Henan, 450000, People's Republic of China

Correspondence: Xueqing Yu, Email yxqshi@163.com

Objective: This study aims to evaluate the risk of bias (ROB) and reporting quality of idiopathic pulmonary fibrosis (IPF) prediction models by assessing characteristics of these models.

Methods: The development and/or validation of IPF prognostic models were identified via an electronic search of PubMed, Embase, and Web of Science (from inception to 12 August, 2021). Two researchers independently assessed the risk of bias (ROB) and reporting quality of IPF prediction models based on the Prediction model Risk Of Bias Assessment Tool (PROBAST) and Transparent Reporting of a multivariable prognostic model for Individual Prognosis or Diagnosis (TRIPOD) checklist.

Results: Twenty prognostic model studies for IPF were included, including 7 (35%) model development and external validation studies, 8 (40%) development studies, and 5 (25%) external validation studies. According to PROBAST, all studies were appraised with high ROB, because of deficient reporting in the domains of participants (45.0%) and analysis (67.3%), and at least 55% studies were susceptible to 4 of 20 sources of bias. For the reporting quality, none of them completely adhered to the TRIPOD checklist, with the lowest mean reporting score for the methods and results domains (46.6% and 44.7%). For specific items, eight sub-items had a reporting rate ≥80% and adhered to the TRIPOD checklist, and nine sub-items had a very poor reporting rate, less than 30%.

Conclusion: Studies adhering to PROBAST and TRIPOD checklists are recommended in the future. The reproducibility and transparency can be improved when studies completely adhere to PROBAST and TRIPOD checklists.

Keywords: idiopathic pulmonary fibrosis, PROBAST, reporting quality, risk of bias, TRIPOD

Introduction

Idiopathic pulmonary fibrosis (IPF) is the most common type of interstitial lung disease and is characterized by dyspnea and progressive deterioration of lung function,[1,2] with median survival of 2 to 3 years from time of diagnosis.[3,4] The acute exacerbations and the associated complications often lead to hospitalization and death,[5–8] resulting in significant economic and health-care burdens.[5,9] At present, there are some challenges in the diagnosis and treatment of IPF, for instance, the complicated diagnosis process, limited and expensive interventions, and corresponding side effects.[1,10] Therefore, to mitigate the risks and burdens of IPF and to improve the perceptions of best practices of care, more efficient prognosis predictions are needed.[11]

Multivariable prediction models in which multiple characteristics or pieces of information are applied can estimate an individual's risk of a current condition in future.[12] When a patient's score passes a certain threshold, an alarm may be sent to the appropriate clinicians for further evaluation and intervention. In recent years, interest in using prediction models has increased and the models are more often recommended in clinical practice guidelines[13,14] for individual prognosis and diagnosis. At present, there are more prognostic models for IPF.[15,16] However, because there is a surplus of IPF prognostic models with widely variable quality, it is important to identify IPF prognostic models of high quality.

Received: 8 January 2022
Accepted: 9 May 2022
Published: 8 June 2022

Risk of bias (ROB) is usually defined as the presence of a systematic error that may affect the study's validity. The Prediction model Risk of Bias Assessment Tool (PROBAST)[17] guideline has been developed to assess the ROB of model development and model validation, including updating of a prediction model.[18] A full reporting is essential to evaluate the validation and applicability of a multivariable prediction model;[19,20] therefore, a protocol was developed for a guided, Transparent reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) statement.[21] Previous systematic reviews from other research teams showed that there are high ROB and suboptimal reporting quality in prediction models in oral health and preterm birth.[22,23] To date, there has been limited data for evaluating the quality of existing IPF prognostic models; therefore, in this cross-sectional study, we assessed IPF prognostic models with PROBAST and TPRIPOD checklists, aiming to identify the ROB and reporting quality of these studies and highlight the strengths and limitations of the methodologies of IPF prognostic models.

Materials and Methods

Study Design

We conducted a critical cross-sectional appraisal on ROB and reporting quality of IPF prediction models using the PROBAST[24] and TRIPOD checklists,[21] respectively. Although this review was not a typical system review (SR), we did strictly adhere to the guidelines of conducting and reporting a SR,[25] the details of reported PRISMA-Checklists are listed in Supplementary Table 1. The review protocol was registered on INPLASY.[26]

Inclusion and Exclusion Criteria

Studies were included if they met the following criteria: 1) A prognostic multivariable prediction model of IPF describing development, validation, or both, and 2) They should predict events at the probability of future outcomes (prognosis) related to IPF.

Studies were excluded if they 1) examined independent prognostic factors and did not aim to develop a model; 2) combined IPF with other diseases; 3) were not original research (such as review, methodological articles, conference abstracts, protocols); and 4) performed comparison models.

Literature Search

A comprehensive literature search was conducted in PubMed, Embase, and Web of Science between inception to 12 August, 2021. The details of the search strategy can be found in Supplementary Table 2. All relevant articles in the reference list of all included articles were also retrieved.

Study Selection

We created a database in EndNote X9 software. After eliminating duplicates, we read the titles and abstracts for a preliminary screening. Then, we downloaded the full text and filtered it again until all relevant prediction models of IPF were confirmed. Two researchers selected the literature, and if there were discrepancies between them, it was addressed by the discussion with the third researcher.

Data Extraction

We focused our research on study design, outcome measurement, modeling methodology, and validation strategy. Therefore, one researcher extracted the above key information, including author, publication year, population characteristics, follow-up time, etc. Another researcher checked the extracted data, and if there existed different opinions between two researchers, they would refer to the original text and revise it.

Application of Evaluation Tools

We classified each study into model development with or without external validation in the same publication and external validation study of a previously developed model only. Two of us used the PROBAST tool to assess the ROB for each included study. Following four PROBAST domains (Participants, Predictors, Outcome, Analysis), we assessed 20

signaling questions for development models and 17 signaling questions for validation models within each domain with yes/probably yes, no/probably no, or no information. We rated domain-level and applicability assessments using "low risk of bias," "high risk of bias," and "unclear risk of bias" according to PROBAST suggestion.[17]

Two researchers assessed the reporting quality of the included models according to the TRIPOD statement.[21] The checklist covered 37 sub-items in 6 domains, including title and abstract (items 1 and 2), introduction (item 3), methods (items 4 to 12), results (items 13 to 17), discussion (items 18 and 19), and other information (items 21 and 22). We rated items as "reported" if the relevant information was fully presented, "unreported" if all relevant information were lacking, and "not applicable" for inappropriate data, as reported by previous studies.[22,27]

Statistical Methods

The reporting rates of PROBAST and TRIPOD were calculated in a descriptive manner using proportions (%). The PROBAST items for development studies only (items 4.5, 4.8, 4.9) and the TRIPOD "if done" item (item 5c), the validation items (items 10c, 10e, 13c, 17, 19a) were used as both the numerator and denominator when the overall adherence rate was calculated. In each evaluation, we completed two rounds of pilot evaluation, SPSS 25.0 software was used to calculate the Kappa value of the internal consistency coefficient, and the formal assessment was made when Kappa $\geq$0.8. Our pre-evaluation results (Kappa = 0.818, $P \leq 0.001$) demonstrated a good agreement between our reviewers.

Results

Literature Selection

A total of 1670 records were collected. Of those, 46 eligible full-text articles were reviewed after removal of duplicates and irrelevant articles. We excluded 26 studies for reasons shown in Figure 1. Finally, 20 IPF clinical prognostic model studies[15,16,28–45] met the inclusion criteria in this process.

Overview of Characteristics for IPF Prognostic Model Studies

Of the 20 prognostic model studies, 7 (35%) publications[15,28–33] reported model development and external validation, 8 (40%) publications[34–41] reported model development only, and 5 (25%) publications[16,42–45] reported external validation with one updated.[42] According to these studies, 20 prognostic models were identified but more than 50% models did not undergo any external validation. A summary of characteristics for the included studies is shown in Table 1.

ROB of IPF Prediction Model Studies

All models[15,16,28–45] were at high ROB in development or validation. Eight models[16,35–37,39,40,42,45] are rated as high ROB in applicability. Figures 2 and 3 show the proportion assessment for each PROBAST item and proportion of studies with potential bias using PROBAST, respectively. Overall, 55% of the studies were identified with at least four sources of bias (sub-items 1.1, 2.2, 4.2, and 4.8).

As for participants, predictors, outcomes, and analysis domains, there were 12, 12, 6, and 18 studies that had a high ROB, respectively (The "biased" domain, applicability identified in each study is provided in Supplementary Figure 1). Of the included studies, 55.0% resulted in a high risk of bias because of the inclusion of retrospective studies (sub-item 1.1). For predictors of definition (item 2.1) and assessment (item 2.2), one study[39] proposed to exclude the predictor of diffusing capacity of lung for carbon monoxide (DLCO), and one study[41] included a question as a predictor. In addition, 45.0% studies may not report any actions to blind assessment of the predictors. However, unreasonable predictor selection methods may limit the use of the model. Most sub-items in outcomes domains have low ROB, especially inconsistency in defining outcome (sub-item 3.4) and time interval between predictor and outcome assessment (sub-item 3.6). This is probably because most studies used outcomes that are easy to assess (eg, death, survival time). In addition, in the analysis domain, continuous variables were converted into dichotomous variables (sub-item 4.2) or there was lack of overfitting consideration (item 4.8) in most studies, which are the main reasons leading to the high ROB in analysis domain.

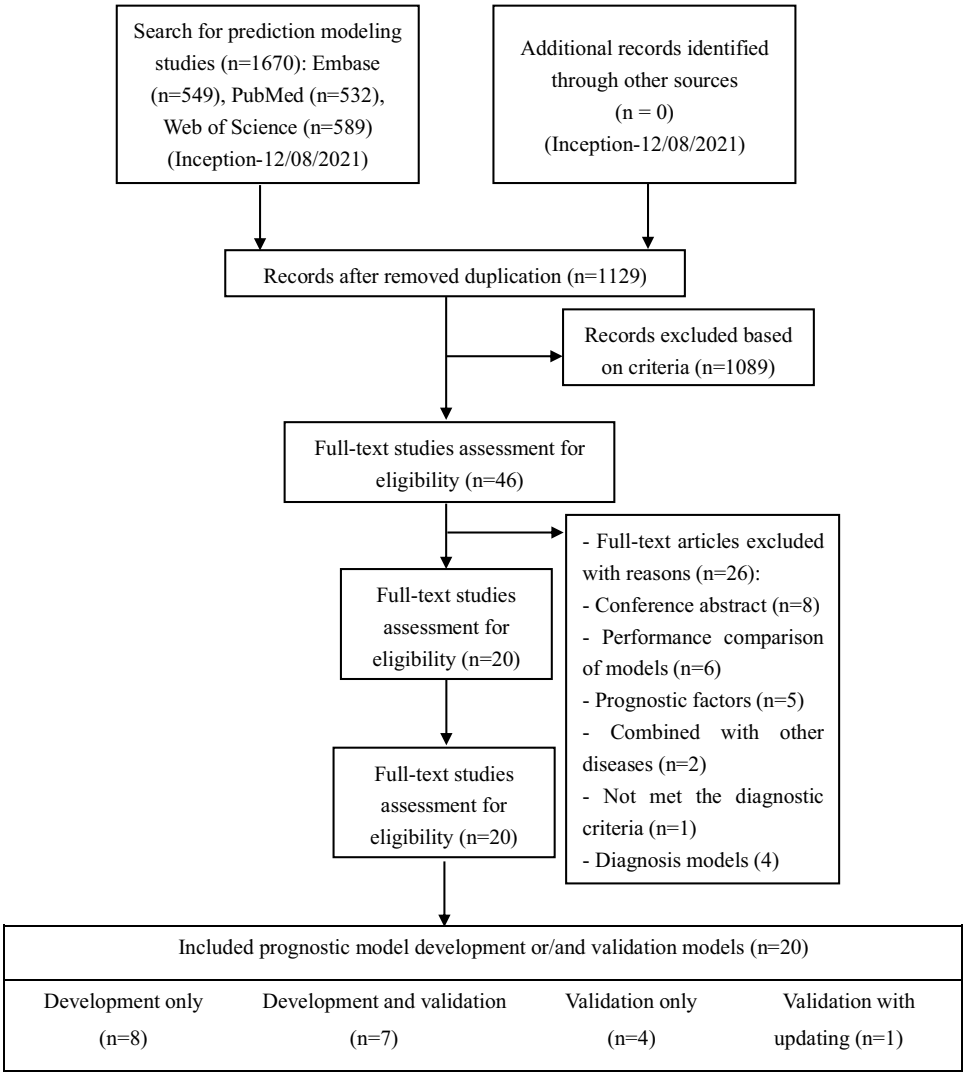


Figure 1 Schema of literature selection process.

Bias Related to Applicability

There was high ROB about applicability, indicating that these studies were not well aligned with the search question. This is most commonly due to concerns about applicability in domain 1 (population) for included randomized clinical trials or specific groups. In addition, two[37,40] out of 20 studies' outcomes included exacerbation, but the definition of acute exacerbation may have varied between studies and may have been expanded to include events for which a trigger can be identified, such as infection.[46]

Reporting Quality of IPF Prediction Model Studies

Of the six domains, the introduction part of TRIPOD with scores of 90.0% is the domain where IPF prediction model studies had the highest mean score reporting rate. The prediction model of IPF had the lowest mean score for the domains of methods and results (46.6% and 44.7%). The mean reporting rate of each domain according to the TRIPOD checklist is less than 65%. The details of reporting rates of each of the TRIPOD domains are shown in Figure 4.

For specific items, eight sub-items had a reporting rate ≥80% adherence to the TRIPOD checklist, and nine sub-items showed a low reporting rate with less than 30%. None of the included IPF prediction model studies reported any actions to blind assessment of the outcome/predictors (items 6b, 7b). In addition, with respect to validation models, few of them

Table 1 The Characteristics of IPF Prediction Model Studies

Author (Years)	Data Sources	Participant Sample Size (Events)	Time of Follow-Up	Predictors in Final Model	Modeling Method	Internal Validation Method	Outcomes	Name of the Model	Model Performance Measure		Overall Risk of Bias Using PROBAST
									Discrimination	Calibration	
Model development and validation studies											
Ley 2012[15]	The patients enrolled in three hospitals from two countries (US and Italy)	D: 228(89); V: 330(186)	3 years	Gender, age, FVC%, DLCO%	Based on Fine–Gray models for survival	10-fold cross-validation	Survival time	GAP models (Calculator and Index)	C-index: 0.708 95% CI [0.637–0.75] (calculator); 0.693 95% CI [0.622–0.731] (index)	NP	High
Huang 2015[28]	Patients from UCMC and UPMC	D: 45(NP); V1: 21(NP); V2: 75(NP)	D: 18.8 months; V1: 43.8 months; V2: 23.5 months	118 prognostic predictor genes within IL1R2, ERAF, CEACAM8, ARG1, FOXO3	Cox proportional hazard regression	10-fold cross validation	The progression of IPF (survival)	Prognostic index (PI) score	AUC: 0.96	Kaplan-Meier plot	High
Torrisi 2019[29]	Patients from four internationals academic ILD centres (Italy, Germany, Netherland, US)	D: 476(NP); V: 461(NP)	28 months	Age, p%FVC, p%DLCO, diabetes mellitus, systemic hypertension, major depressive disorder, valvular heart disease, atrial arrhythmias	Cox proportional hazard regression	NP	Survival time	TORVAN models (full and sparse)	C-index: 0.71 95% [67.8–74.2] (full model); 0.725 95% [69.5–75.6] (sparse model)	Calibration plot	High
Nishikiori 2020[30]	Patients from 3 regions of Japan and Korea	D: 326(NP); V1: 117(48); V2: 1262 (415)	29 months	Gender, age, VC%, DLCO%	Cox proportional hazard regression	NP	Survival time	Modified GAP model (Index)	Harrell's C-index: 0.67 95% CI (0.601–0.739) (1 year); 0.698 (0.638–0.758) (2 year); 0.738 (0.782–0.795) (3 year)	NP	High
Li 2021[31]	IPF patients from the Gene Expression Omnibus database	D: 176(121); V1: 57(NP); V2: 120(NP); V3: 119(NP)	5 years	Hypoxia-Immune-Related genes†††	Cox proportional hazard regression	Three-fold cross validation	The progression of IPF (survival)	Hypoxia-immune-related prediction model	AUC: 0.789(1 year); 0.768(2 year); 0.754(3 year); 0.798(4 year); 0.913(5 year)	Kaplan–Meier plot	High
Lu 2021[32]	The microarray expression matrix dataset of 75 IPF from GSE28042	D: 75(NP); V: 45(NP)	3 years	Inflammation-Related gene: S100A12, CCR7, TNFSF4	Cox proportional hazard regression	NP	The progression of IPF (survival)	Inflammation-Related Prognostic Model	AUC: 0.611(1 year); 0.695(2 year); 0.681(3 year)	NP	High

(*Continued*)

https://doi.org/10.2147/RMHP.S357606 DovePress

https://doi.org/10.2147/RMHP.S357606 DovePress

Table 1 (Continued).

Author (Years)	Data Sources	Participant Sample Size (Events)	Time of Follow-Up	Predictors in Final Model	Modeling Method	Internal Validation Method	Outcomes	Name of the Model	Model Performance Measure		Overall Risk of Bias Using PROBAST
									Discrimination	Calibration	
Xia 2021[33]	The gene expression data of BAL cells and clinical information for IPF patients came from the GEO database (Freiburg, Siena, and Leuven)	D: 176(NP); V:64(NP)	3 years	Bronchoalveolar lavage cell-associated gene: TLR2, CCL2, HTRA1, SFN	Cox proportional hazard regression	NP	Survival time	NP	AUC: 0.773(1 year); 0.772(2 year); 0.752(3 year)	Calibration plots	High
Model development studies											
King 2001[34]	Patients enrolled into a Specialized Center of Research Study at the National Jewish Medical and Research Center	238 (155)	110 months	Age, smoking status, clubbing, profusion (radiographic abnormality), pulmonary hypertension, TLC, PaO2 at maximal exercise	Cox proportional hazard regression	NP	One-year mortality	CRP score (Complete and Abbreviated)	NP	NP	High
du Bois, 2011[35]	Patients from two clinical trials of IFN-g1b and GIPF 007	1099 (152)	1 year	Age, history of respiratory hospitalization, %pFVC, 24-week in % pFVC	Methodology set forth by Wilson and coworkers	NP	One-year mortality	Mortality risk scoring system (Comprehensive model and Clinical model)	C-statistic:0.77 95% CI [0.72–0.81] (Comprehensive model); 0.75 95% CI [0.71–0.79] (Clinical model)	Chi-square statistic	High
Soares 2015[36]	Patients from three reference centers for interstitial lung diseases (ILD) in São Paulo	120 (80)	37.5 months	Dyspnea, %pDLCO, %pFVC, FEV1/FVC	Cox proportional hazard regression	NP	Survival time	DDS index	C-statistic: 0.78 95% CI [0.705–0.865]	NP	High
Ashley, 2016[37]	Patients come from an observational cohort study of a multi-center	60 (35)	80 weeks	Biomarkers: ICOS, LGMN, FCN2, TRY3, VEGF sR2, Cathepsin S	Multivariable logistic and Cox regression models	Boot-strap	Progression-free survival†	Six-SOMAmer Index	AUC: 0.91	NP	High
Lee, 2018[38]	Retrospective cohort study in Asan Medical Center	144 (106)	57.9 months (range, 13–131 months)	Age, desaturation, fibrosis score††, interval changes in fibrosis score	Cox proportional hazard regression	NP	Survival time	NP	C-index: 0.768 95% CI [0.707–0.829]	NP	High

Fukuda 2020[39]	Patients from 3 reference centers for ILD in São Paulo	173 (154)	43 months	Dyspnea, FVC%, ExSpO2 (oxygen desaturation during exercise)	Cox proportional hazard regression	Bootstrap	Survival time	DOS score	C- statistic: 0.7	NP	High
Tang 2020[40]	Patients participated in INPULSIS-1 and 2	1061 (63)	372 days	Age, decline in FVC to week 52, baseline % pFVC, supplement oxygen use	Laplace estimation method	Bootstrap	Exacerbation risk	Time-to Event (TTE) model	NP	NP	High
Moor 2021[41]	Patients come from a prospective cohort study in Netherlands	140 (28)	1 year	Surprise question, MRC score, %pDLCO	Multivariable logistic regression model	NP	One-year mortality	NP	C-statistic: 0.82 95% [0.73–0.91]	NP	High
Model validation studies with or without updating											
Ley 2015[42]	One study of interferon γ1b and two studies of pirfenidone in IPF	1109 (128)	1.1 years (0.01–2.36 years)	RH, UCSD SOBQ, 6 MWD, FVC 24-week change	Cox proportional hazard regression	Bootstrap	Survival time	Longitudinal GAP model	C-statistic: 0.785 95% CI [0.78–0.79]	H-L test	High
Kim, 2015[43]	Patients come from Seoul National University Hospital in Korean	268 (157)	4.64 years	Gender, age, FVC%, DLCO%	NA	NA	Survival time	GAP model	C statistic: GAP calculator 0.74 95% CI [0.35–1] (1 year), 0.71 95% [0.44–0.92] (2 year), 0.68 95% [0.46–0.87] (3 year); GAP index 0.72 95% CI [0.34–1] (1 year), 0.69 95% [0.42–0.91] (2 year), 0.66 95% CI [0.44–0.85] (3 year)	H-L test	High

(*Continued*)

https://doi.org/10.2147/RMHP.S357606 DovePress

Table 1 (Continued).

Author (Years)	Data Sources	Participant Sample Size (Events)	Time of Follow-Up	Predictors in Final Model	Modeling Method	Internal Validation Method	Outcomes	Name of the Model	Model Performance Measure		Overall Risk of Bias Using PROBAST
									Discrimination	Calibration	
Lee 2016[44]	Patients come from 54 university and teaching hospitals in Korean	1228 (NP)	19±16 months	Gender, age, FVC%, DLCO%	NA	NA	Survival time	GAP model	C-statistic: GAP calculator 0.61 95% CI [0.559–0.653] (1 year), 0.61 95% [0.566–0.649] (2 year), 0.59 95% [0.549 −0.627] (3 year); GAP index 0.59 95% CI [0.537–0.638] (1 year), 0.59 95% CI [0.544–0.631] (2 year), 0.57 95% [0.53–0.611] (3 year)	NP	High
Harari 2019[16]	Patients treated with pirfenidone in 12 interstitial lung disease centers across Italy	68 (22)	2.4 years	Gender, age, FVC%, DLCO%	NA	NA	Survival time	GAP model	C-index: 0.74 95% CI [0.57–0.93] (GAP index); 0.77 95% CI [0.59–0.93] (GAP calculator)	H-L test	High
Abe 2020[45]	Patients treated with nintedanib in the Chiba University Hospital	89 (18)	16.4 months	Gender, age, FVC%, DLCO%	NA	NA	Survival time	GAP model	NP	NP	High

Notes: †Progression-free survival as determined by the time until any of the following: death, acute exacerbation of IPF, lung transplant, or relative decrease in forced vital capacity (FVC, liters) of 10% or DLCO (mL/min/mmHg) of 15%; ††Fibrosis score: The fibrosis score was defined as the sum of the extent of honeycombing and reticular opacity; †††NALCN, IL1R2, S100A12, PROK2, CCL8, RAB15, MARCKSL1, TPCN1, HS3ST.

Abbreviations: NA, not applicable; UCMC, University of Chicago Medical Center; UPMC, University of Pittsburgh Medical Center; NP, not provided; D, derivation; V, validation; BAL, bronchoalveolar lavage; RH, respiratory hospitalization in the preceding 24 weeks; UCSD SOBQ, University of California San Diego Shortness of Breath Questionnaire; ILD, interstitial lung disease; FVC, forced vital capacity; DLCO, diffusing capacity of carbon monoxide; VC, vital capacity; FEV1/FVC, forced expiratory volume in one second/forced vital capacity; AUC, area under curve; GAP, gender-age-physiology; H-L, Hosmer Lemeshow.

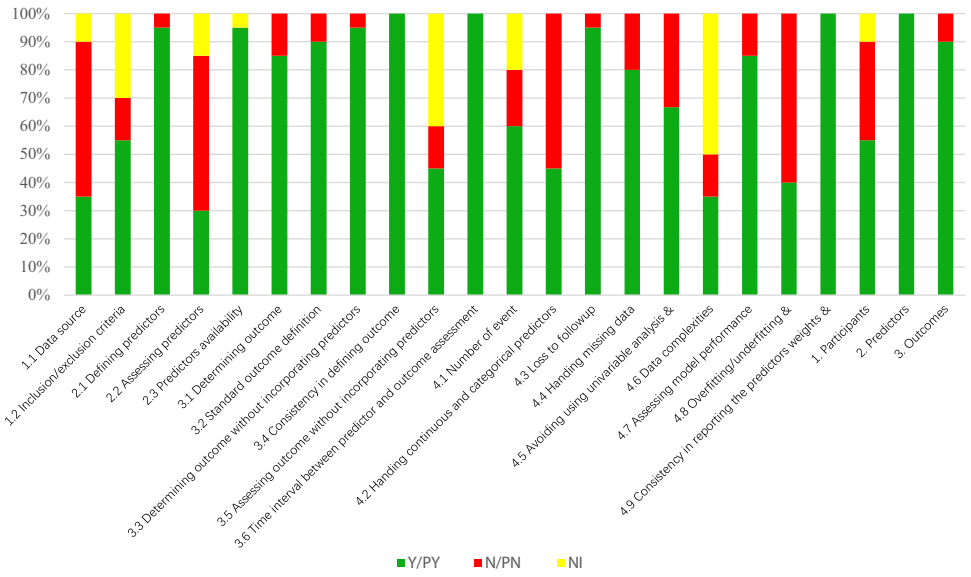
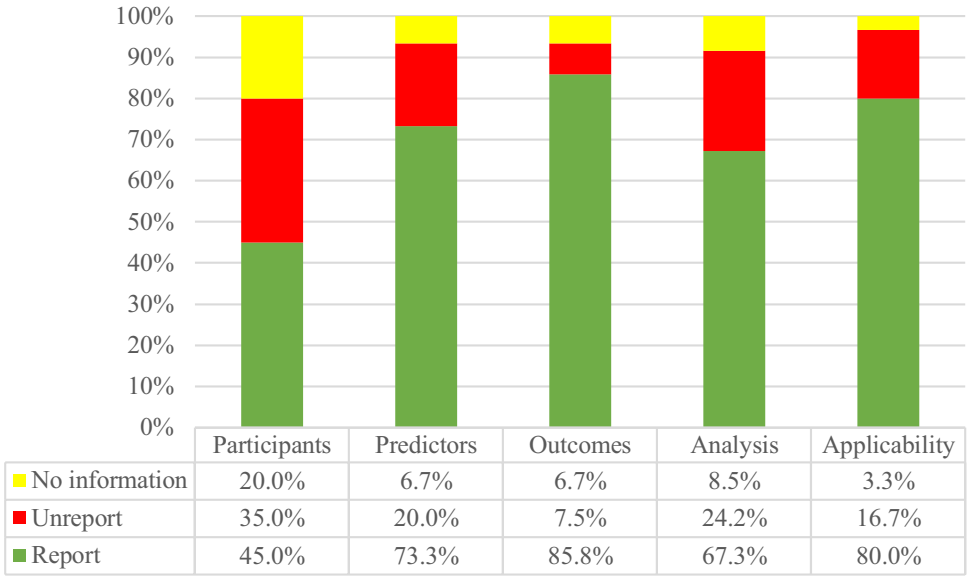


Figure 2 The proportion assessment for each PROBAST item.



	Participants	Predictors	Outcomes	Analysis	Applicability
No information	20.0%	6.7%	6.7%	8.5%	3.3%
Unreport	35.0%	20.0%	7.5%	24.2%	16.7%
Report	45.0%	73.3%	85.8%	67.3%	80.0%

Figure 3 Proportion of studies with potential bias using PROBAST.

identified any difference or showed the comparison from the data between development and validation (items 12, 13c). The details of reporting for each item adherence to the TRIPOD checklist are shown in Figure 5.

Discussion

Main Findings

In this cross-sectional study of prognostic risk models related to the IPF, we identified and critically appraised 20 studies that were described 20 models. All models reported good to excellent predictive performance, but all of them had high ROB according to PROBAST, and none of them completely adhered to the TRIPOD checklist for reporting, which demonstrates deficiencies in applicability and reporting transparency. First, as for ROB, we identified the mean reporting rate of each domain, which was less than 70%. Similarly, using the TRIPOD standards, we found that the mean reporting

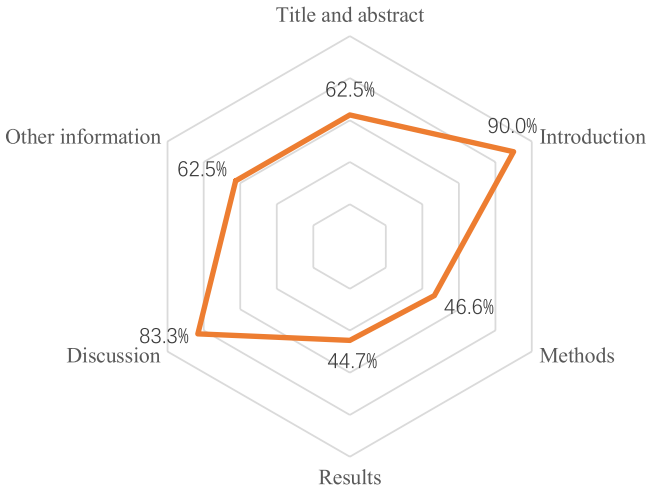


Figure 4 The detail of reporting rates of each TRIPOD domains.

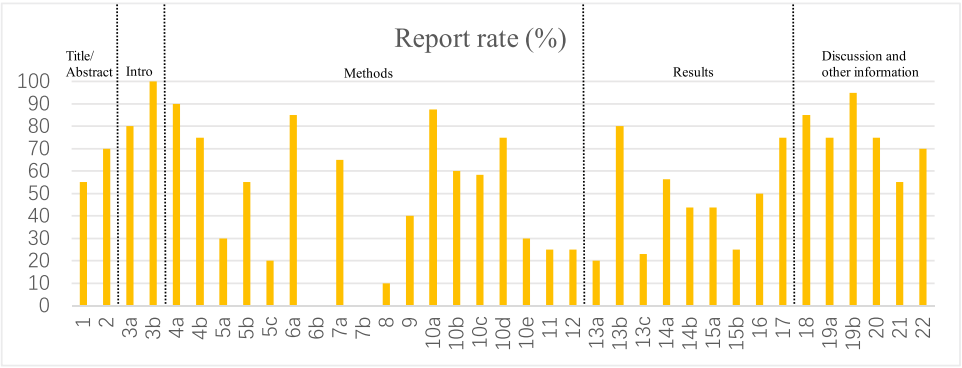


Figure 5 The detail of reporting for each item adherence to TRIPOD checklist.

rate of each domain was less than 65%, especially the methods and results domains. The two tools suggest that there was a general lack of transparent reporting and identification of bias across the studies of IPF progression models.

The ROB of Included Studies According PROBAST

The main aim of prediction models is to support medical decision-making, and a high ROB implies that these models will probably perform worse in practice than in the studies reported by researchers.[11] We identified the reporting rate in participants domain was 45.0% and 67.3% in analysis, which means researchers should pay attention to the ROB of IPF prediction models.

In the participants domain, most studies did not select the appropriate data source, and there were a large number of retrospective studies in our study. The Gender-Age-Physiology (GAP) system, which might help inform decision-making, has been externally validated five times with good discrimination, but the ROB for the external validation studies of the retrospective research data was high. While prospective cohort studies are recommended for model development, there are practical issues for using prospective cohort studies given that IPF is relatively rare. In addition, there are many well-designed prospective national IPF registries that could serve this purpose.

In the analysis domain, continuous predictors should not be dichotomized or categorized, as this will result in loss of information, which in turn may lead to imprecise risk estimates.[47] In addition, we suggest that continuous predictors should appear as original data in future research on models. Calibration, which can assess the fit of models, is one of the essential features in the assessment of the usability of a predictive scale.[48] However, few studies considered calibration in

our review. Although some studies evaluated this indicator, most of them used the statistical test of Hosmer-Lemeshow (H-L).[49] However, the H-L test cannot retain the most information on possible miscalibration, so a calibration plot or other evaluation method is recommended.[50,51] Lack of internal validation may lead to overfitting because quantifying the predictive performance of a model on the same data from which the model was developed tends to give optimistic estimates of performance,[22] so that internal validation is recommended when there is no external validation.

The Reporting Quality of Included Studies According to TRIPOD

Full reporting of studies facilitates reproducibility of models, appraisal of model validity, and judgment of model generalizability to other clinical settings.[20] We found there is room for improvement in the reporting quality of IPF prediction models.

In the methods domain, a large number of these prediction models skipped estimating sample size, mainly because of a lack of consensus in estimating sample size requirements for derivation and validation.[52] However, a reasonable sample size is necessary, so until a canonical method of calculating sample size is available, we can use 10 or 15 times the flat number of events to initially calculate the sample size. It was brought to our attention that the GAP system has been externally validated several times but has rarely been updated. Models will be more generalizable when the case mix of the new population is within the case mix range of the development population.[53] Therefore, in future studies, it is recommended to continuously update the GAP models according to the characteristics of the study population.

In the results domain, the reporting rates of participants' engagement process and method of using the prediction model are lower than 30%. To make the results of prediction model research more complete and more transparent, we should consider not only the performance of the prediction model but also how to use the developed model for individuals. Therefore, we recommend that items such as how to use the model be reported clearly and directly.

Strengths and Limitations

To our knowledge, this is the first review to assess the ROB and reporting quality of multivariable prognostic models for IPF based on the PROBAST and TRIPOD checklists. We systematically searched the IPF prognostic models to assess the ROB and explore report quality of these studies and more importantly, to summarize the characteristics of the existing IPF prediction model studies to provide necessary references for future research. Before the formal assessment, we conducted a pilot experiment, requiring that the evaluator's consistency coefficient was higher than 0.8 before the evaluation was carried out to further reduce the subjectivity of the researchers and improve the credibility of the results of our review.

Our study also has several limitations. First, we did not assess the relationship among ROB, reporting rates, and forecast accuracy. In addition, our study only considered the important factor of insufficient reporting, but other factors need to be further explored to promote the scientific implementation of the IPF prediction model in clinical practice.

Conclusion

According to the study, the rates of reporting adhering to PROBAST and TRIPOD checklists are low, especially in the domains of participants and analysis for PROBAST and the domains of methods and results for TRIPOD. Moreover, there is a general lack of identification of bias and transparent reporting, which can decrease the reproducibility rate of IPF models. However, the low rate of reporting, according to the two evaluation tools, does not mean low prediction accuracy, but the reproducibility and transparency can be improved when studies completely adhere to PROBAST and TRIPOD checklists. Therefore, studies adhering to PROBAST and TRIPOD checklists are recommended in the future.

Funding

This study was supported by National Natural Science Foundation of China (grant number 81973779, 82174307); Special Scientific Research Project of National Clinical Research Base of Traditional Chinese Medicine of China (grant number 2018JDZX006); Central Plains Thousand People Program (grant number 194200510018).

Disclosure

Jiaqi Di and Xuanlin Li are co-first authors. The authors declare that there are no conflicts of interest.

References

1. Raghu G, Rochwerg B, Zhang Y, 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(5):e3–e19. doi:10.1164/rccm.201506-1063ST
2. Richeldi L, Collard HR, Jones MG. Idiopathic pulmonary fibrosis. *Lancet*. 2017;389(10082):1941–1952. doi:10.1016/S0140-6736(17)30866-8
3. Ley B, Collard HR, King TE. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. 2011;183(4):431–440. doi:10.1164/rccm.201006-0894CI
4. Martinez FJ, Collard HR, Pardo A, et al. Idiopathic pulmonary fibrosis. *Nat Rev Dis Primers*. 2017;3:1–19. doi:10.1038/nrdp.2017.74
5. Cottin V, Schmidt A, Catella L, et al. Burden of idiopathic pulmonary fibrosis progression: a 5-year longitudinal follow-up study. *PLoS One*. 2017;12(1):e0166462. doi:10.1371/journal.pone.0166462
6. Kim S, Myong J, Yoon H, et al. Health care burden and medical resource utilisation of idiopathic pulmonary fibrosis in Korea. *Int J Tuberc Lung Dis*. 2017;21(2):230–235. doi:10.5588/ijtld.16.0402
7. Yu YF, Wu N, Chuang -C-C, et al. Patterns and economic burden of hospitalizations and exacerbations among patients diagnosed with idiopathic pulmonary fibrosis. *J Manag Care Spec Pharm*. 2016;22(4):414–423. doi:10.18553/jmcp.2016.22.4.414
8. Mooney JJ, Raimundo K, Chang E, et al. Hospital cost and length of stay in idiopathic pulmonary fibrosis. *J Med Econ*. 2017;20(5):518–524. doi:10.1080/13696998.2017.1282864
9. Xuan JW, Lu YJ, Ren MD. Direct economic burden of patients in idiopathic pulmonary fibrosis in China. *Chin J Pharm Econ*. 2019;14(6):9–12.
10. Raghu G, Remy-Jardin M, Myers JL, 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. doi:10.1164/rccm.201807-1255ST
11. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ*. 2020;369:m1328. doi:10.1136/bmj.m1328
12. Steyerberg EW, Moons KG, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med*. 2013;10(2):e1001381. doi:10.1371/journal.pmed.1001381
13. Camm AJ, Kirchhof P, Lip G, et al.; Developed with the Special Contribution of the European Heart Rhythm Association (EHRA), Endorsed by the European Association for Cardio-Thoracic Surgery (EACTS), Authors/Task Force Members. Guidelines for the management of atrial fibrillation: the task force for the management of atrial fibrillation of the European Society of Cardiology (ESC). *Eur Heart J*. 2010;31(19):2369–2429. doi:10.1093/eurheartj/ehq278
14. Ravdin PM, Siminoff LA, Davis GJ, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J Clin Oncol*. 2001;19(4):980–991. doi:10.1200/JCO.2001.19.4.980
15. Ley B, Ryerson CJ, Vittinghoff E, et al. A multidimensional index and staging system for idiopathic pulmonary fibrosis. *Ann Intern Med*. 2012;156(10):684–691. doi:10.7326/0003-4819-156-10-201205150-00004
16. Harari S, Caminati A, Confalonieri M, et al. The prognostic role of gender-age-physiology system in idiopathic pulmonary fibrosis patients treated with pirfenidone. *Clin Respir J*. 2019;13(3):166–173. doi:10.1111/crj.12999
17. Wolff RF, Moons KG, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med*. 2019;170(1):51–58. doi:10.7326/M18-1376
18. Moons KG, de Groot JA, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med*. 2014;11(10):e1001744. doi:10.1371/journal.pmed.1001744
19. Choi YJ, Chung MS, Koo HJ, et al. Does the reporting quality of diagnostic test accuracy studies, as defined by STARD 2015, affect citation? *Korean J Radiol*. 2016;17(5):706–714. doi:10.3348/kjr.2016.17.5.706
20. Heus P, Damen JA, Pajouheshnia R, et al. Poor reporting of multivariable prediction model studies: towards a targeted implementation strategy of the TRIPOD statement. *BMC med*. 2018;16(1):1–12. doi:10.1186/s12916-018-1099-2
21. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Br J Cancer*. 2015;112(2):251–259. doi:10.1038/bjc.2014.639
22. Du M, Haag D, Song Y, et al. Examining bias and reporting in oral health prediction modeling studies. *J Dent Res*. 2020;99(4):374–387. doi:10.1177/0022034520903725
23. Kim JI, Lee JY. Systematic review of prediction models for preterm birth using CHARMS. *Biol Res Nurs*. 2021;23(4):708–722. doi:10.1177/10998004211025641
24. Moons KG, Wolff RF, Riley RD, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med*. 2019;170(1):W1–W33. doi:10.7326/M18-1377
25. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg*. 2010;8(5):336–341. doi:10.1016/j.ijsu.2010.02.007
26. Di JQ, Li XL, Yang JJ, et al. Critical appraisal of the reporting quality of risk prediction models for idiopathic pulmonary fibrosis. which can be achieved by. Available from: https://inplasy.com/inplasy-2020-11-0105/. Accessed May 20, 2022.
27. Jiang MY, Dragnev NC, Wong SL. Evaluating the quality of reporting of melanoma prediction models. *Surgery*. 2020;168(1):173–177. doi:10.1016/j.surg.2020.04.016
28. Huang Y, Ma S-F, Vij R, et al. A functional genomic model for predicting prognosis in idiopathic pulmonary fibrosis. *BMC Pulm Med*. 2015;15:1–12. doi:10.1186/s12890-015-0142-8
29. Torrisi SE, Ley B, Kreuter M, et al. The added value of comorbidities in predicting survival in idiopathic pulmonary fibrosis: a multicentre observational study. *Eur Respir J*. 2019;53(3):1801587. doi:10.1183/13993003.01587-2018
30. Nishikiori H, Chiba H, Lee SH, et al. A modified GAP model for East-Asian populations with idiopathic pulmonary fibrosis. *Respir Investig*. 2020;58(5):395–402. doi:10.1016/j.resinv.2020.04.001

31. Li X, Zhang Q, Cai Y, et al. Investigation of a hypoxia-immune-related microenvironment gene signature and prediction model for idiopathic pulmonary fibrosis. *Front Immunol*. 2021;12:2244. doi:10.3389/fimmu.2021.629854
32. Lu Y, Chen J, Tang K, et al. Development and validation of the prognostic index based on inflammation-related gene analysis in idiopathic pulmonary fibrosis. *Front Mol Biosci*. 2021;8:667459. doi:10.3389/fmolb.2021.667459
33. Xia Y, Lei C, Yang D, et al. Construction and validation of a bronchoalveolar lavage cell-associated gene signature for prognosis prediction in idiopathic pulmonary fibrosis. *Int Immunopharmacol*. 2021;92:107369. doi:10.1016/j.intimp.2021.107369
34. King TE, Tooze JA, Schwarz MI, et al. Predicting survival in idiopathic pulmonary fibrosis: scoring system and survival model. *Am J Respir Crit Care Med*. 2001;164(7):1171–1181. doi:10.1164/ajrccm.164.7.2003140
35. du Bois RM, Weycker D, Albera C, et al. Ascertainment of individual risk of mortality for patients with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. 2011;184(4):459–466. doi:10.1164/rccm.201011-1790OC
36. Soares MR, Pereira C, Ferreira R, et al. A score for estimating survival in idiopathic pulmonary fibrosis with rest SpO2>88. *Sarcoidosis Vasc Diffuse Lung Dis*. 2015;32(2):121–128.
37. Ashley SL, Xia M, Murray S, et al. Six-SOMAmer index relating to immune, protease and angiogenic functions predicts progression in IPF. *PLoS One*. 2016;11(8):e0159878. doi:10.1371/journal.pone.0159878
38. Lee SM, Seo JB, Oh SY, et al. Prediction of survival by texture-based automated quantitative assessment of regional disease patterns on CT in idiopathic pulmonary fibrosis. *Eur Radiol*. 2018;28(3):1293–1300. doi:10.1007/s00330-017-5028-0
39. 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*. 2020;99(25):e20739. doi:10.1097/MD.0000000000020739
40. Tang F, Weber B, Stowasser S, et al. Parametric time-to-event model for acute exacerbations in idiopathic pulmonary fibrosis. *CPT Pharmacometrics Syst Pharmacol*. 2020;9(2):87–95. doi:10.1002/psp4.12485
41. Moor CC, van Jaarsveld NCT, Owusuaa C, et al. The value of the surprise question to predict one-year mortality in idiopathic pulmonary fibrosis: a prospective cohort study. *Respiration*. 2021;100(8):780–785. doi:10.1159/000516291
42. Ley B, Bradford WZ, Weycker D, et al. Unified baseline and longitudinal mortality prediction in idiopathic pulmonary fibrosis. *Eur Respir J*. 2015;45(5):1374–1381. doi:10.1183/09031936.00146314
43. Kim ES, Choi SM, Lee J, et al. Validation of the GAP score in Korean patients with idiopathic pulmonary fibrosis. *Chest*. 2015;147(2):430–437. doi:10.1378/chest.14-0453
44. Lee SH, Kim SY, Kim DS, et al. Predicting survival of patients with idiopathic pulmonary fibrosis using GAP score: a nationwide cohort study. *Respir Res*. 2016;17(1):1–9. doi:10.1186/s12931-016-0454-0
45. Abe M, Tsushima K, Yoshioka K, 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):369–376. doi:10.5603/ARM.a2020.0137
46. Collard HR, Ryerson CJ, Corte TJ, et al. Acute exacerbation of idiopathic pulmonary fibrosis. An international working group report. *Am J Respir Crit Care Med*. 2016;194(3):265–275. doi:10.1164/rccm.201604-0801CI
47. de Jong Y, Ramspek CL, Zoccali C, et al. Appraising prediction research: a guide and meta-review on bias and applicability assessment using the Prediction model Risk Of Bias ASsessment Tool (PROBAST). *Nephrology*. 2021;26(12):939–947. doi:10.1111/nep.13913
48. Petre M-A, Saha B, Kasuya S, et al. Risk prediction models for emergence delirium in paediatric general anaesthesia: a systematic review. *BMJ open*. 2021;11(1):e043968. doi:10.1136/bmjopen-2020-043968
49. Altman DG, Vergouwe Y, Royston P, et al. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338:b605–b605. doi:10.1136/bmj.b605
50. Moons KG, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio) marker. *Heart*. 2012;98(9):683–690. doi:10.1136/heartjnl-2011-301246
51. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014;35(29):1925–1931. doi:10.1093/eurheartj/ehu207
52. Altman DG. Prognostic models: a methodological framework and review of models for breast cancer. *Cancer Invest*. 2009;27(3):235–243. doi:10.1080/07357900802572110
53. Vergouwe Y, Moons KG, Steyerberg EW. External validity of risk models: use of benchmark values to disentangle a case-mix effect from incorrect coefficients. *Am J Epidemiol*. 2010;172(8):971–980. doi:10.1093/aje/kwq223