



# External validation and longitudinal application of the DO-GAP index to individualise survival prediction in idiopathic pulmonary fibrosis

Abhimanyu Chandel <sup>1</sup>, Christopher S. King<sup>2</sup>, Rosalinda V. Ignacio<sup>3</sup>, Jean Pastre<sup>4</sup>, Oksana A. Shlobin<sup>2</sup>, Vikramjit Khangoora<sup>2</sup>, Shambhu Aryal<sup>2</sup>, Alan Nyquist <sup>2</sup>, Anju Singhal<sup>2</sup>, Kevin R. Flaherty<sup>5</sup> and Steven D. Nathan <sup>2</sup>

<sup>1</sup>Department of Pulmonary and Critical Care, Walter Reed National Military Medical Center, Bethesda, MD, USA. <sup>2</sup>Advanced Lung Disease and Transplant Program, Inova Fairfax Hospital, Falls Church, VA, USA. <sup>3</sup>Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA. <sup>4</sup>Service de Pneumologie et Soins Intensifs, Hôpital Européen Georges Pompidou, APHP, Paris, France. <sup>5</sup>Division of Pulmonary and Critical Care Medicine, University of Michigan, Ann Arbor, MI, USA.

Corresponding author: Steven D. Nathan ([Steven.Nathan@inova.org](mailto:Steven.Nathan@inova.org))



Shareable abstract (@ERSpublications)

The DO-GAP index is a simple validated tool that improves baseline prognostication compared to the GAP index. Serial change in the DO-GAP index can further improve and individualise survival prediction in patients with IPF. <https://bit.ly/3Jnp0wO>

Cite this article as: Chandel A, King CS, Ignacio RV, 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: 00124-2023 [DOI: 10.1183/23120541.00124-2023].

Copyright ©The authors 2023

This version is distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0. For commercial reproduction rights and permissions contact [permissions@ersnet.org](mailto:permissions@ersnet.org)

Received: 25 Feb 2023  
Accepted: 7 March 2023

## Abstract

**Background** The Distance-Oxygen-Gender-Age-Physiology (DO-GAP) index has been shown to improve prognostication in idiopathic pulmonary fibrosis (IPF) compared to the Gender-Age-Physiology (GAP) score. We sought to externally validate the DO-GAP index compared to the GAP index for baseline risk assessment in patients with IPF. Additionally, we evaluated the utility of serial change in the DO-GAP index in predicting survival.

**Methods** We performed an analysis of patients with IPF from the Pulmonary Fibrosis Foundation-Patient Registry (PFF-PR). Discrimination and calibration of the two models were assessed to predict transplant-free survival and IPF-related mortality. Joint longitudinal time-to-event modelling was utilised to individualise survival prediction based on DO-GAP index trajectory.

**Results** There were 516 patients with IPF from the PFF-PR with available demographics, pulmonary function tests, 6-min walk test data and outcomes included in this analysis. The DO-GAP index (C-statistic: 0.73) demonstrated improved discrimination in discerning transplant-free survival compared to the GAP index (C-statistic: 0.67). DO-GAP index calibration was adequate, and the model retained predictive accuracy to identify IPF-related mortality (C-statistic: 0.74). The DO-GAP index was similarly accurate in the subset of patients taking antifibrotic medications. Serial change in the DO-GAP index improved model discrimination (cross-validated area under the curve: 0.83) enabling the personalised prediction of disease trajectory in individual patients.

**Conclusion** The DO-GAP index is a simple, validated, multidimensional score that accurately predicts transplant-free survival in patients with IPF and can be adapted longitudinally in individual patients. The DO-GAP may also find use in studies of IPF to risk stratify patients and possibly as a clinical trial end-point.

## Introduction

Idiopathic pulmonary fibrosis (IPF) is a chronic progressive fibrotic lung disease with an uncertain trajectory and a highly variable rate of clinical decline [1]. Given this heterogeneity, risk prediction models have been developed to predict the course of the disease based on clinical parameters [2–6]. The Gender-Age-Physiology (GAP) index is a well-known, validated and simple risk prediction model to estimate disease prognosis based on baseline patient factors. This model has been modified to incorporate



additional factors over a 6-month time period to predict clinical trajectory more precisely [3]. However, emerging evidence has demonstrated that the baseline GAP index lacks predictive precision and overestimates risk in some cohorts [3, 7, 8].

An extended baseline clinical risk prediction model for IPF, the Distance-Oxygen-Gender-Age-Physiology (DO-GAP) index has been developed and validated in IPF (supplementary table S1). This model adds measures of exercise capacity (distance ambulated during a 6-min walk test (6MWT) and exertional hypoxaemia) to the GAP index and has demonstrated improved predictive performance from baseline measures in real-world patient cohorts [8]. Despite being previously validated in a small external geographically distinct cohort of patients with IPF (n=108), further validation is essential to provide additional support and confidence of generalisability, thus enabling wider-spread adoption in clinical practice [8].

The primary aim of this study was to compare the performance of the GAP and DO-GAP indices in patients enrolled in the Pulmonary Fibrosis Foundation-Patient Registry (PFF-PR), a well-defined, real-world cohort of patients with IPF [9]. In addition to predicting transplant-free survival, we sought to evaluate model performance to specifically predict IPF-related mortality. Finally, we examined the DO-GAP index trajectory over time and evaluated if this longitudinal change could be used to enable the dynamic prediction of transplant-free survival in individual patients with IPF.

### Material and methods

Data from the PFF-PR, a large multicentre registry of patients with interstitial lung disease, which has collected data on >2000 patients across the USA since 2016, was collated and analysed through January 2022 [9]. Registered patients with IPF were included if baseline spirometry, diffusion capacity and 6MWT distance was performed within 6 months of their enrolment in the PFF-PR. Important specifics related to the design and methods utilised to enrol patients into the PFF-PR, including diagnostic considerations related to IPF, have previously been described in detail [9].

The DO-GAP and GAP indices were calculated at baseline (the date of baseline spirometry) and repeated testing was collected where available for up to 1 year after initial baseline measures. Derivation and validation of the DO-GAP index has also previously been described and additional details related to the component predictor variables are included within the supplementary materials [8].

### Statistical analysis

For the external validation of the DO-GAP index, survival analysis to predict transplant-free survival was performed using the Kaplan–Meier method and the log-rank test was used to compare groups. The Cox proportional hazards model was used to calculate hazard ratios and relevant 95% confidence intervals (CI). All evaluated models complied with the assumptions of proportional hazards.

Model discrimination was assessed by the change in Harrell's C-statistic. "Strict" calibration of the DO-GAP index was assessed by comparing predicted survival from the original DO-GAP derivation cohort to the observed survival in the PFF-PR database [10]. Calibration was evaluated over 3 years of follow-up by examination at multiple time points comparing observed to predicted model values and by risk groups of the staging system.

To assess the utility of the DO-GAP index to predict ILD-related mortality, Fine and Gray competing-risk regression estimating the subdistribution hazard ratio (sHR) was utilised treating non-IPF-related death as a competing risk. Indicators for whether cause of death was IPF-related was adjudicated and recorded by the PFF-PR. Bootstrap resampling with 500 repetitions was used to calculate 95% bias-corrected CIs of the C-statistic.

Finally, we evaluated whether the serial change and trajectory of the DO-GAP score over time was associated with a change in predicted transplant-free survival. This was accomplished by utilising joint longitudinal and time-to-event modelling through a simultaneous estimation of a random-intercept-and-slope longitudinal model, Cox proportional hazards model for the time-to-event component and Markov chain Monte Carlo parameter estimation. The longitudinal and survival sub-models included a covariate for treatment with an antifibrotic medication (at the time of initial spirometry after enrolment in the PFF-PR) as emerging evidence supports these having a survival benefit [11–13]. The primary output of interest from this modelling method is the association parameter  $\alpha$  (the association between longitudinal DO-GAP score trajectory and the hazard of death or transplantation expressed as a hazard ratio).

Time-dependent area under the curve (AUC) at 3 years of follow-up of the joint model was estimated and internally validated utilising 10-fold cross-validation [14].

Though the DO-GAP index was originally derived to predict overall survival, censoring patients as event free at the time of lung transplantation (to evaluate overall survival) may result in informative censoring and introduce bias in the model by not accounting for changes in the risk of death related to lung transplantation. To more appropriately account for this risk and to ease implementation and interpretation of subsequent survival and joint modelling procedures, mortality and lung transplantation were considered as a composite outcome in all analyses (transplant-free survival). Relatedly, several sensitivity analyses were performed. Details regarding these measurements and other statistical considerations including sample size calculation and the handling of missing data are described in detail within the supplementary material.

## Results

### *PFF-PR cohort and baseline models*

There were 516 patients with IPF in the PFF-PR identified with lung function testing and 6MWT data completed within 6 months of enrolment. The median age of the cohort was 71 years (25th percentile (Q1), 75th percentile (Q3): 66–75), 122 (23.6%) were women and 484 (96.2%) were white. The median forced vital capacity (FVC) % predicted and uncorrected diffusion capacity of the lung for carbon monoxide ( $D_{LCO}$ ) % predicted were 65.6 (Q1–Q3: 56.1–78.1) and 38.3 (Q1–Q3: 29.1–48.0), respectively. Median 6MWT distance was 365 m (Q1–Q3: 274–441) and a total of 270 patients (52.3%) exhibited exertional hypoxaemia. Median follow-up was 2.4 years (Q1–Q3: 1.2–3.4) during which 149 deaths and 79 lung transplantations occurred. Demographics, baseline characteristics and relevant pulmonary metrics of the cohort are summarised in table 1.

The discrimination of the original GAP index was measured by a C-statistic of 0.67 (95% CI: 0.64–0.71). Calibration of the staging system of the original GAP was assessed by comparing previously published mortality by GAP index stage to that observed in this cohort and generally overestimated the observed risk (supplementary table S2) [2]. The DO-GAP index demonstrated good outcome discrimination compared to the GAP index (C-statistic: 0.73; 95% CI: 0.70–0.76, difference in C-statistic compared with the GAP index 0.06; 95% CI: 0.03–0.09,  $p < 0.001$ ). Overall transplant-free survival in the cohort based on DO-GAP

**TABLE 1** Baseline characteristics of all 516 patients in the cohort

Demographics	
Age years	71 (66–75)
Gender, female	122 (23.6)
Race, white	484 (96.2)
Body mass index $\text{kg}\cdot\text{m}^{-2}$	28.4 (25.5–32.2)
COPD	52 (10.1)
Pulmonary hypertension	34 (6.6)
Lung cancer	7 (1.4)
Extrapulmonary malignancy	101 (19.6)
Coronary artery disease	122 (23.6)
Congestive heart failure	17 (3.3)
Diabetes mellitus	106 (20.5)
History of tobacco use	324 (62.8)
Pack-years of smoking history (prior smokers)	22 (14–31)
Measures of IPF severity and management considerations	
FVC (%) predicted	65.6 (56.1–78.1)
$D_{LCO}$ (%) predicted	38.3 (29.1–48.0)
6MWT distance m	365 (274–441)
Exertional hypoxaemia	
Desaturation during 6MWT	55 (20.4)
Prescribed oxygen	215 (79.6)
Antifibrotic therapy	
Pirfenidone	188 (36.4)
Nintedanib	167 (32.4)

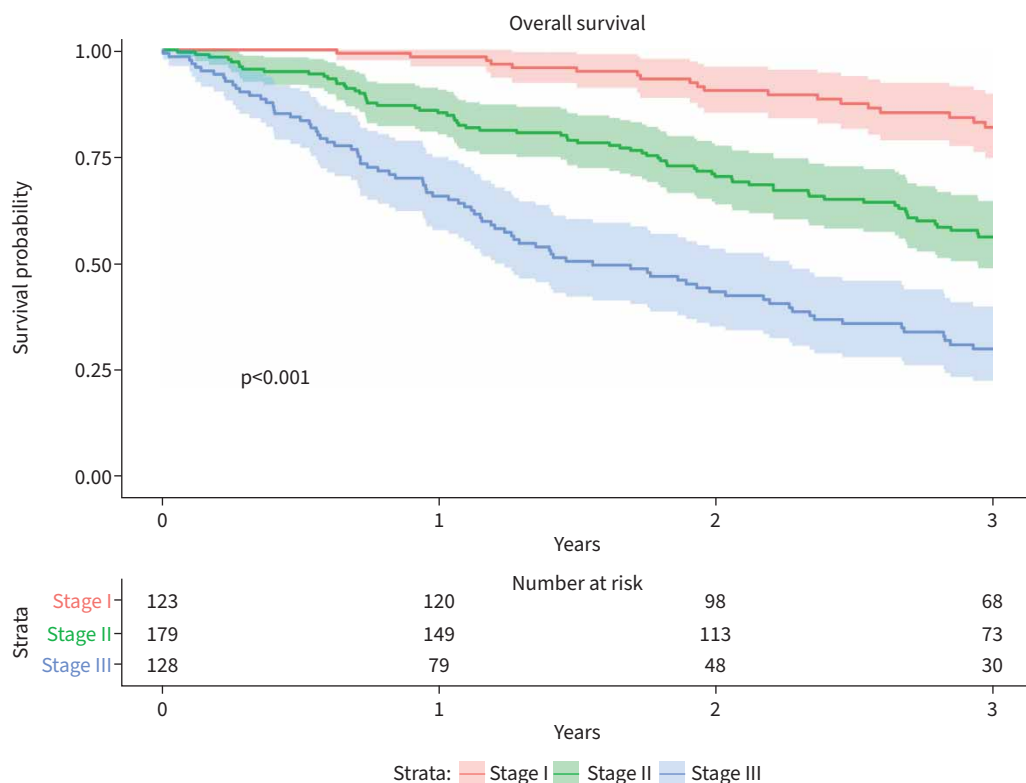
Data are presented as median (25th percentile–75th percentile) or n (%). FVC: forced vital capacity;  $D_{LCO}$ : single breath diffusion capacity of the lung for carbon monoxide; 6MWT: 6-min walk test.

stage is displayed in figure 1. In complete cases, the DO-GAP staging system resulted in stage reclassification from the original GAP stage (for example, GAP stage II → DO-GAP stage I) in n=136 (31.6%) of patients. A complete breakdown of stage reclassification is provided in supplementary table S3.

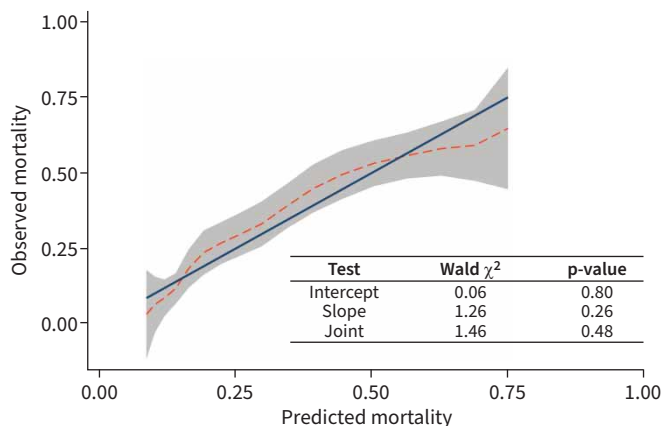
Calibration of the DO-GAP index was assessed by comparing the predicted survival from the original DO-GAP derivation set to the observed transplant-free survival in the PFF-PR database. This relationship is graphically displayed overall and by DO-GAP stage in figures 2 and 3, respectively. Tabulated data by DO-GAP stage is presented in supplementary table S4. Results of the intercept, slope, and joint intercept and slope test were all non-significant, indicative of good overall model calibration.

Sensitivity analyses evaluating DO-GAP index performance when treating lung transplantation as a competing risk, predicting transplant-free survival in patients taking antifibrotic medications, and when missing  $D_{LCO}$  measurements were considered as “unable to perform” were performed and included in supplementary table S5. Of note, DO-GAP model discrimination in these instances remained good and overall consistent with estimation derived in the primary analysis. Assessment of the calibration of the DO-GAP model to predict survival considering lung transplantation as a competing risk demonstrated some evidence of miscalibration (overestimation of observed risk) in patients with the highest predicted event probabilities (supplementary figure S1). Miscalibration of the GAP staging system when applied in this manner was substantial and evident over all stages and time periods (supplementary table S6).

The GAP and DO-GAP indices were then applied to predict death related to IPF (or lung transplantation) treating death from other cause as a competing risk. The DO-GAP index and, separately, the related staging system were both significantly associated with IPF-related mortality (sHR: 1.11; 95% CI: 1.09–1.12,  $p < 0.001$  and sHR: 1.85; 95% CI: 1.65–2.07,  $p < 0.001$ , respectively). The DO-GAP index continued to provide improved model discrimination compared to the GAP index in this competing-risk analysis (C-statistic: 0.74; 95% CI: 0.69–0.79 versus 0.70; 95% CI: 63.5–75.7). The cumulative incidence of IPF-related transplant-free survival based on DO-GAP stage is displayed in figure 4.



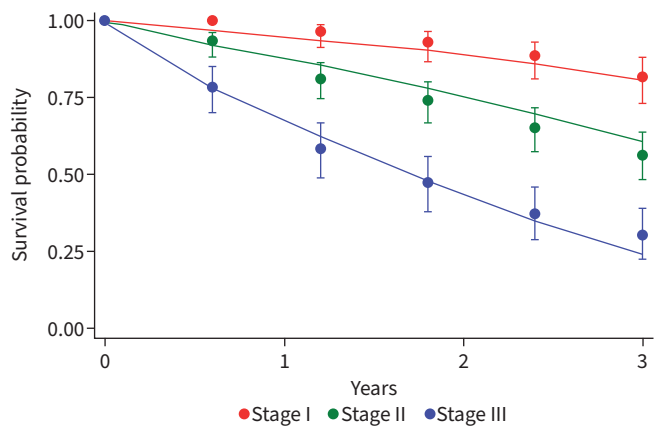
**FIGURE 1** Kaplan–Meier survival curves from time of initial pulmonary function testing based on idiopathic pulmonary fibrosis stage defined by the Distance-Oxygen-Gender-Age-Physiology (DO-GAP) index.



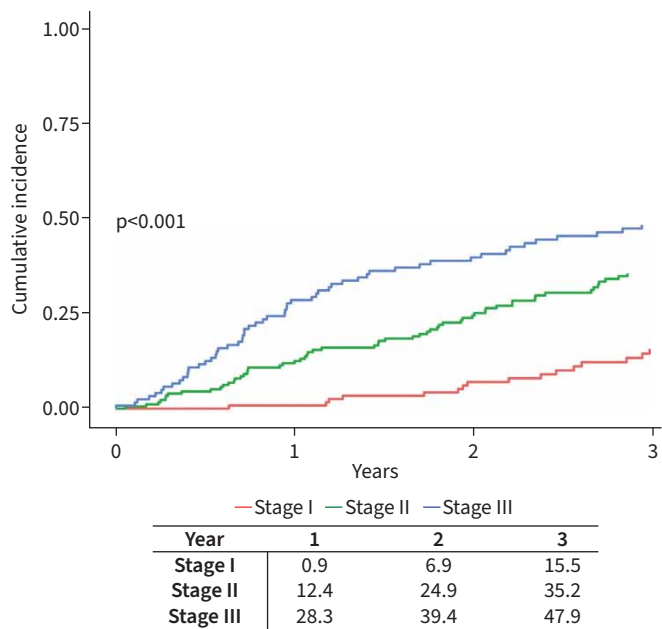
**FIGURE 2** Calibration of the Distance-Oxygen-Gender-Age-Physiology (DO-GAP) index by stage in the Pulmonary Fibrosis Foundation-Patient Registry (PFF-PR). Smoothed pseudovalues (dashed lines) with pointwise 95% confidence intervals (shaded area) are plotted against predicted event probabilities at 2 years of follow-up. The solid line is the line of identity, denoting perfect calibration. Tables beneath the figures are the results of statistical tests of overall model calibration over multiple time points through 3 years of follow-up. The null hypothesis of these tests assumes the intercept is equal to 0 and the slope of the model is equal to 1.

**DO-GAP index trajectory**

After baseline assessment, 207 patients returned at least once for lung function and 6MWT testing within 1 year of follow-up resulting in 275 additional instances where repeat DO-GAP index was calculable. Of those that returned for repeat testing, the median duration from baseline assessment to repeated testing was 153 days (Q1–Q3: 99–189). In these individuals, an increased longitudinal DO-GAP index was associated with a higher rate of mortality at any time of follow-up (table 2). This association persisted when the longitudinal and survival sub-model were adjusted for baseline receipt of antifibrotic medication. When DO-GAP index was modelled as a continuous variable on an ordinal scale, every 1-unit increase in score results in a 31.5% higher rate of transplant or death. Discrimination of the joint model incorporating serial measures of the DO-GAP index over time markedly improved compared to assessment based solely on baseline functional testing (cross-validated AUC: 0.83). A unique benefit of joint longitudinal and time-to-event modelling is the ability to obtain personalised predictions of mortality risk based on each patient’s DO-GAP index trajectory. Figure 5 provides a representation of four patients in the PFF-PR that returned for follow-up lung function testing after initial testing and depicts how each individual’s risk



**FIGURE 3** Calibration of the Distance-Oxygen-Gender-Age-Physiology (DO-GAP) index in the Pulmonary Fibrosis Foundation-Patient Registry (PFF-PR). Lines represent predicted survival probabilities based on the original model derivation cohort and vertical capped lines denote the Kaplan-Meier estimates with 95% confidence interval observed in the validation set.



**FIGURE 4** Cumulative incidence curves for idiopathic pulmonary fibrosis-related mortality from the time of initial pulmonary function testing based on idiopathic pulmonary fibrosis stage defined by the Distance-Oxygen-Gender-Age-Physiology (DO-GAP) index. Table beneath figure includes the estimated observed overall mortality (%) by DO-GAP stage.

profile differs based on repeated scoring. Figure 6 provides a more detailed examination of the clinical course of one typical patient in the PFF-PR and how that patient’s predicted survival changed based on repeated functional evaluation.

**Discussion**

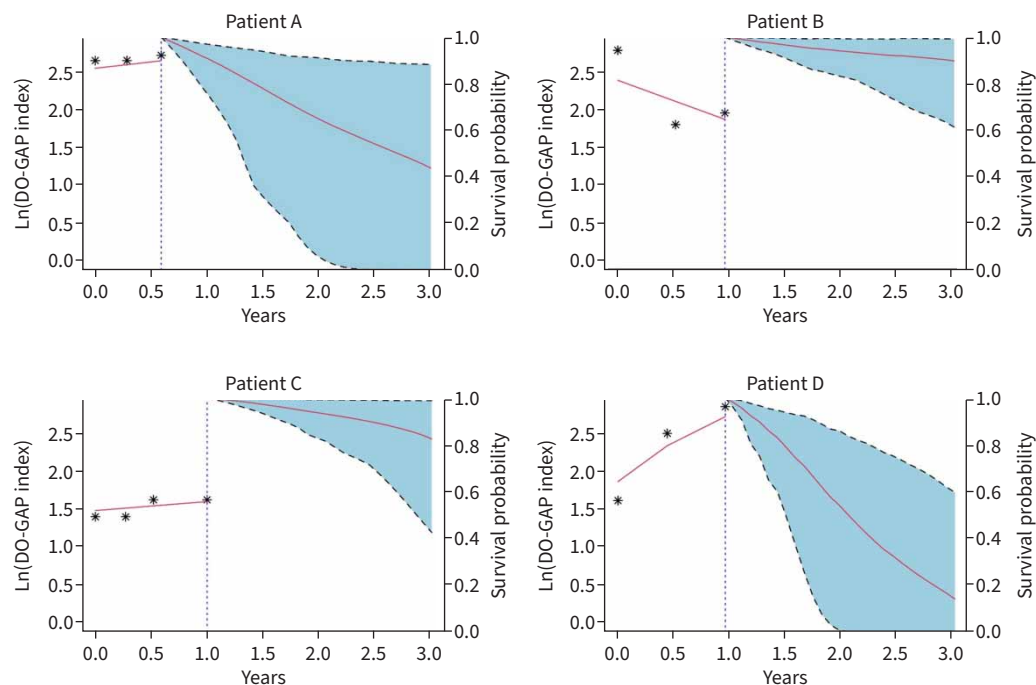
In this evaluation of IPF prognosis among patients enrolled in the PFF-PR, we assessed the performance of the DO-GAP index in a well-characterised registry cohort of real-world patients with IPF. We found that the DO-GAP index significantly outperforms the GAP index in the baseline prediction of transplant-free survival, and this finding was persistent over several subgroup analyses including when prediction was confined to only patients taking antifibrotic medications. Further, the DO-GAP index accurately identified the risk of IPF-related transplant-free survival and finally, we demonstrated that DO-GAP index trajectory over time is reflective of overall prognosis and can be incorporated to estimate prognosis accurately on an individual level.

Several models exist to predict survival in IPF [2–6]. Of these models, some attempt to estimate prognosis incorporating baseline data, while others include clinical factors over a 6-month period to improve risk prediction [3, 5]. The GAP index is the most commonly used model to predict prognosis in IPF; however, significant changes in clinical care have occurred resulting in improvement in overall survival of patients with IPF since it was developed in 2011 [15, 16]. In addition, no baseline or longitudinal clinical

**TABLE 2** Association of Distance-Oxygen-Gender-Age-Physiology (DO-GAP) index trajectory and transplant-free survival

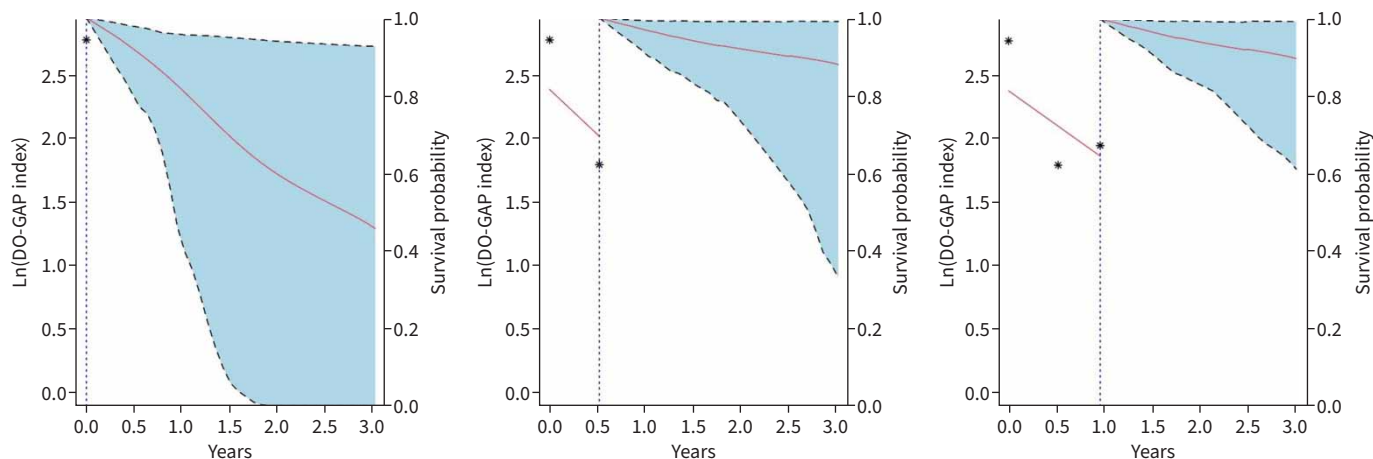
Trajectory variable	Association parameter $\alpha$ (hazard ratio)	95% CI	p-value
<b>Model 1<sup>#</sup></b> DO-GAP Index (ordinal scale)	1.31	1.19–1.46	<0.001
<b>Model 2<sup>†</sup></b> DO-GAP Index (per doubling)	3.03	1.71–6.59	<0.001

<sup>#</sup>: joint model with DO-GAP index modelled on an ordinal scale to provide easy to understand association parameter of the hazard ratio associated with a 1-unit increase in current true value of DO-GAP index; <sup>†</sup>: joint model based on log-transformed DO-GAP index, hazard ratio represents risk per doubling of DO-GAP index.



**FIGURE 5** Personalised prediction of 3-year overall transplant-free survival of four study patients. Predictions are based on model 2 in table 2. Asterisks represent Distance-Oxygen-Gender-Age-Physiology (DO-GAP) score (expressed as log-transformed value) at baseline and then subsequent follow-up encounters up to 1 year after enrolment in the Pulmonary Fibrosis Foundation-Patient Registry. Patient A (deceased after 1.6 years of follow-up) has a high score (DO-GAP stage III) consistently over time and, thus, has lower survival probability compared to patient B (alive at 3 years), who starts at a similarly high score and then improves over time. Likewise, Patient C (alive at 3 years) has a moderate DO-GAP index which remains stable over time and has a much lower risk of death compared to patient D (deceased after 1.8 years of follow-up), who starts at the same score but experiences a progressive worsening over the course of 1 year.

prediction model has demonstrated sufficient clinical utility to be included in clinical practice guidelines for the care of IPF patients [15, 17]. Furthermore, although the original GAP index has been updated to incorporate change in clinical status at 6 months, this adaptation is limited by its development among clinical trial participants and the utility of trends in this score beyond the 6-month period is unclear [3]. A further limitation is that the baseline and longitudinal versions of the GAP index include different parameters, which makes the incorporation of these models in day-to-day practice challenging [2, 3]. The relative improvement in the C-statistic of the DO-GAP index compared to the original GAP index of 0.06 for baseline transplant-free survival prediction observed in this cohort is consistent with a large overall improvement in model discrimination and is similar to the improvement noted in the initial derivation and validation of the DO-GAP index [8, 18]. Further, the DO-GAP index demonstrated better overall calibration, which may reflect its construction in a more contemporary cohort. Notably, the GAP index was developed to predict overall survival considering lung transplantation as a competing risk [2]. To ensure appropriate comparison, as a sensitivity analysis, we applied both models in this manner, and discrimination remained improved utilising the DO-GAP index compared to the GAP index. Model calibration when applied for this outcome (overall survival with lung transplantation as a competing risk) was worse for both models. Specifically, the GAP index overpredicted risk across all stages and time points (“miscalibration in the large”), while the DO-GAP index overpredicted mortality only for patients at the highest observed risk levels. This may reflect an expected change in risk of death at the time of lung transplantation in patients that underwent this procedure. A significant strength of this study was the ability to compare predicted to observed event frequency on an individual basis (“strict calibration”) given access to the original DO-GAP index derivation cohort [19]. We chose not to refit the DO-GAP index as we judged the magnitude of observed miscalibration to be acceptable, and further, the model was well calibrated to predict transplant-free survival, which represents a clinically meaningful end-point in the care of patients with IPF.



**FIGURE 6** Change in personalised risk prediction of transplant-free survival for a representative patient (Patient B). Plots from left to right demonstrate predictions and their associated 95% confidence intervals based on available functional testing over the course of 1 year after registry enrolment. Asterisks represent Distance-Oxygen-Gender-Age-Physiology (DO-GAP) score (expressed as log-transformed value) at baseline and then subsequent follow-up encounters. With repeated testing, predicted survival and the confidence in this prediction improve dramatically.

IPF is a progressive disease with a worse prognosis than many forms of cancer [20]. Prior evidence suggests clinical judgement in estimating disease trajectory in such diseases is often quite poor [21, 22]. In IPF, accurate prognostication is critical to prepare patients for their expected disease course and further, to ensure identification and referral of individuals with IPF who may be candidates for lung transplantation. Relatedly, existing IPF clinical prediction tools based on proportional hazards models apply at the group level, and such predictions can be inaccurate on an individual patient basis [21]. Further, a number of studies that have evaluated the change in functional measures of IPF over a longitudinal time frame have employed statistical techniques that do not account for endogenous covariates [3, 23–25]. The joint longitudinal and time-to-event modelling framework simultaneously estimates both longitudinal and time-to-event components, making it better suited for evaluating longitudinal parameters contingent upon the longitudinal outcome [14]. This modelling technique has been applied previously in diseases related to respiratory health; however, application of this modelling technique to personalise disease prediction is rare overall, but becoming increasingly common in other fields [26–29]. This technique holds exciting opportunities to improve the care of patients with IPF as well as other respiratory diseases. Conceptually, tools to individualise event prediction based on trajectory of clinical risk scores or biomarkers could be incorporated into electronic health record decision tools as relevant technology improves [30].

The International Society for Heart and Lung Transplantation (ISHLT) suggests that patients with a chronic end-stage lung condition with a >50% 2-year mortality risk without transplant may be acceptable candidates for consideration of lung transplantation [17]. In figure 6, we demonstrate how the risk profile of a patient (“Patient B”) in the PFF-PR cohort changes over time. Individual prediction confidence is quite poor based on a single “snapshot” assessment at the time of enrolment in the PFF-PR, but the confidence in disease trajectory improves dramatically as additional data are collected over the course of a year. Based on a single assessment of the DO-GAP index at entry in the PFF-PR, this patient would be classified as DO-GAP stage III disease with an associated estimated survival at 2 years of roughly 50%. Based solely on this initial value, the patient may be deemed to meet the ISHLT threshold for consideration of lung transplantation. With the benefit of additional DO-GAP index assessment over the course of 1 year, this individual patient’s predicted survival for an additional 2 years is estimated to be >80% and our confidence in this latter prediction improves dramatically.

The GAP index has been used in prior clinical trials to risk stratify patients but has never been employed in a serial fashion to provide better longitudinal prognostication. In fact, its performance in this regard is likely to be suboptimal as there are only two components of the GAP index that might be subject to meaningful change ( $D_{LCO}\%$  and FVC%). This current analysis demonstrates the excellent performance characteristics of serial change in the DO-GAP index. As such, the DO-GAP index might not only be useful in baseline risk stratification, but also possibly as a clinic trial end-point as there are four modifiable components (FVC%,  $D_{LCO}\%$ , 6MWT distance and oxygen needs).



Our study has a few limitations. First, the race of patients included from the PFF-PR was almost exclusively white. Though the age of non-white patients in the original external validation cohort of the DO-GAP was higher, the overall number of non-white patients thus far assessed by the model is small, and as a result, further study in different geographic and racially diverse populations is important. Additionally, the DO-GAP index includes factors such as categories of age and gender which are unlikely to change over time. Joint modelling of scores based on continuous factors or scores with different parameters may produce even more accurate estimates of prognosis; however, this was beyond the scope of our current research. Likewise, the DO-GAP index improved upon the GAP index by the inclusion of exercise capacity parameters. In some settings, it may be challenging or impractical to collect necessary lung function testing and 6MWT on a routine longitudinal basis. Additionally, though the administration of antifibrotics at the time of initial spirometry was included in the longitudinal model, the effect of changes in treatment over time or of individual antifibrotic medications on mortality were not examined in this current work. Further, it may be possible that changes in other factors, such as the inclusion of radiographic parameters or biomarkers, may further improve the model. Finally, this study was a retrospective evaluation of an existing patient registry. Notably, the PFF-PR included patients with comorbidities in addition to IPF. For instance, nearly 20% of the patients in this cohort were recorded as having extrapulmonary malignancies. Though the inclusion of patients in model validation with comorbidities is valuable to ensure a broadly applicable model, it remains possible that the model may not accurately reflect the population prognosis in regions where the prevalence of such comorbidities is different. Relatedly, comorbidities with prognostic value in IPF, such as pulmonary hypertension, were based on reported data from registry centres. This reporting may reflect use of various diagnostic strategies and case definitions which were likely not standardised across all reporting centres. Though updated frequently to attempt to capture patient outcomes, the PFF-PR requires timely reporting by enrolled sites to ensure accuracy. Bias may have been introduced if outcome data related to patient deaths was not provided in all instances. Patient follow-up for subsequent testing after the initial encounter is complicated by many factors including but not limited to geographic proximity to treatment centres, patient compliance, inconsistent follow-up testing and disease severity. Though joint longitudinal and time-to-event modelling is well suited for application in this situation and allows for valid inferences in the setting of missing data, the utility of our joint model should be externally validated [31].

In summary, we compared two existing baseline multidimensional models for the prediction of transplant-free survival in IPF. In a large real-world patient registry, we found that the DO-GAP index, which incorporates exercise capacity into the GAP index, provides significantly better baseline risk prediction compared to the GAP index. Further, we extended this model to the baseline prediction of IPF-related survival and found a similar improvement in predictive performance. Finally, we demonstrated that the DO-GAP trajectory over time can be incorporated to adjust and more accurately estimate prognosis on an individual patient basis. These results confirm the DO-GAP index as a validated tool that can be incorporated into clinical practice for the baseline risk assessment of IPF. The DO-GAP index may also find utility in the context of clinical trials to risk stratify patients, while the joint model may prove to have a role as a clinical trial end-point. Further studies are still necessary and strongly encouraged to confirm the utility of longitudinal DO-GAP index assessment for individualised IPF survival prediction thus enabling dynamic, real-time patient counselling.

Provenance: Submitted article, peer reviewed.

Acknowledgements: We thank all patients who participated in the Pulmonary Fibrosis Foundation (PFF) Patient Registry. We also thank the investigators, clinical research coordinators and other staff at participating PFF Care Centers (see supplementary material) for providing clinical data; the Pulmonary Fibrosis Foundation, which established and has maintained the PFF-PR since 2016; and lastly, the many generous donors. The views expressed in this article are those of the author and do not necessarily reflect the official policy of the Department of Defense or the US Government.

Conflict of interest: C.S. King is a consultant for United Therapeutics, Actelion, Altavant and Boehringer Ingelheim, and serves on the advisory boards for Actelion, United Therapeutics, Merck and Boehringer Ingelheim. O.A. Shlobin is a consultant for United Therapeutics, Janssen, and Altavant, and serves on the speaker bureau of United Therapeutics, Bayer and Janssen. S.D. Nathan is a consultant for United Therapeutics, Roche, Bellerophon and Merck. He is on the speakers' bureaus for United Therapeutics and Boehringer Ingelheim. A. Chandel, R.V. Ignacio, J. Pastre, V. Khangoora, S. Aryal, A. Nyquist, A. Singhal and K.R. Flaherty have no conflicts of interest to report.

## References

- 1 Khor YH, Ng Y, Barnes H, *et al.* Prognosis of idiopathic pulmonary fibrosis without anti-fibrotic therapy: a systematic review. *Eur Respir J* 2020; 29: 190158.
- 2 Ley B, Ryerson CJ, Vittinghoff E, *et al.* A multidimensional index and staging system for idiopathic pulmonary fibrosis. *Ann Intern Med* 2012; 156: 684–691.
- 3 Ley B, Bradford WZ, Weycker D, *et al.* Unified baseline and longitudinal mortality prediction in idiopathic pulmonary fibrosis. *Eur Respir J* 2015; 45: 1374–1381.
- 4 Wells AU, Desai SR, Rubens MB, *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: 962–969.
- 5 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: 459–466.
- 6 Lettieri CJ, Nathan SD, Browning RF, *et al.* The distance-saturation product predicts mortality in idiopathic pulmonary fibrosis. *Respir Med* 2006; 100: 1734–1741.
- 7 Robbie H, Daccord C, Chua F, *et al.* Evaluating disease severity in idiopathic pulmonary fibrosis. *Eur Respir J* 2017; 26: 170051.
- 8 Chandel A, Pastre J, Valery S, *et al.* Derivation and validation of a simple multidimensional index incorporating exercise capacity parameters for survival prediction in idiopathic pulmonary fibrosis. *Thorax* 2023; 78: 368–375.
- 9 Wang BR, Edwards R, Freiheit EA, *et al.* The pulmonary fibrosis foundation patient registry. Rationale, design, and methods. *Ann Am Thorac Soc* 2020; 17: 1620–1628.
- 10 Royston P. Tools for checking calibration of a Cox model in external validation: approach based on individual event probabilities. *Stata J* 2014; 14: 738–755.
- 11 Kang J, Han M, Song JW. Antifibrotic treatment improves clinical outcomes in patients with idiopathic pulmonary fibrosis: a propensity score matching analysis. *Sci Rep* 2020; 10: 15620.
- 12 Jo HE, Glaspole I, Grainge C, *et al.* Baseline characteristics of idiopathic pulmonary fibrosis: analysis from the Australian Idiopathic Pulmonary Fibrosis Registry. *Eur Respir J* 2017; 49: 1601592.
- 13 Dempsey TM, Sangaralingham LR, Yao X, *et al.* Clinical effectiveness of antifibrotic medications for idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2019; 200: 168–174.
- 14 Rizopoulos D. The R package JMBayes for fitting joint models for longitudinal and time-to-event data using MCMC. *J Stat Softw* 2016; 72: 1–46.
- 15 Raghu G, Rochweg 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: e3–19.
- 16 Jeganathan N, Smith RA, Sathananthan M. Mortality trends of idiopathic pulmonary fibrosis in the United States from 2004 through 2017. *Chest* 2021; 159: 228–238.
- 17 Leard LE, Holm AM, Valapour M, *et al.* Consensus document for the selection of lung transplant candidates: an update from the International Society for Heart and Lung Transplantation. *J Heart Lung Transplant* 2021; 40: 1349–1379.
- 18 Pencina MJ, D’Agostino RB, Sr, D’Agostino RB, Jr., *et al.* Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008; 27: 157–172; discussion 207–112.
- 19 Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol* 2013; 13: 33.
- 20 Vancheri C, Failla M, Crimi N, *et al.* Idiopathic pulmonary fibrosis: a disease with similarities and links to cancer biology. *Eur Respir J* 2010; 35: 496–504.
- 21 Henderson R, Keiding N. Individual survival time prediction using statistical models. *J Med Ethics* 2005; 31: 703–706.
- 22 Hirji A, Zhao H, Ospina MB, *et al.* Clinical judgment versus lung allocation score in predicting lung transplant waitlist mortality. *Clin Transplant* 2020; 34: e13870.
- 23 Jacob J, Bartholmai BJ, van Moorsel CHM, *et al.* Longitudinal prediction of outcome in idiopathic pulmonary fibrosis using automated CT analysis. *Eur Respir J* 2019; 54: 1802341.
- 24 Wuyts WA, Dahlqvist C, Slabbynck H, *et al.* Longitudinal clinical outcomes in a real-world population of patients with idiopathic pulmonary fibrosis: the PROOF registry. *Respir Res* 2019; 20: 231.
- 25 Caminati A, Bianchi A, Cassandro R, *et al.* Walking distance on 6-MWT is a prognostic factor in idiopathic pulmonary fibrosis. *Respir Med* 2009; 103: 117–123.
- 26 Mason SE, Moreta-Martinez R, Labaki WW, *et al.* Longitudinal association between muscle loss and mortality in ever smokers. *Chest* 2022; 161: 960–970.
- 27 Goudsmit BFJ, Braat AE, Tushuizen ME, *et al.* Development and validation of a dynamic survival prediction model for patients with acute-on-chronic liver failure. *JHEP Rep* 2021; 3: 100369.
- 28 Posch F, Riedl J, Reitter E-M, *et al.* Dynamic assessment of venous thromboembolism risk in patients with cancer by longitudinal D-Dimer analysis: a prospective study. *J Thromb Haemost* 2020; 18: 1348–1356.

- 29 Sudell M, Kolamunnage-Dona R, Tudur-Smith C. Joint models for longitudinal and time-to-event data: a review of reporting quality with a view to meta-analysis. *BMC Med Res Methodol* 2016; 16: 168.
- 30 Long JD, Mills JA. Joint modeling of multivariate longitudinal data and survival data in several observational studies of Huntington's disease. *BMC Med Res Methodol* 2018; 18: 138.
- 31 Ibrahim JG, Chu H, Chen LM. Basic concepts and methods for joint models of longitudinal and survival data. *J Clin Oncol* 2010; 28: 2796–2801.