SARCOIDOSIS VASCULITIS AND DIFFUSE LUNG DISEASES 2014; 31; 8-18

Multi-dimensional indeces to stage idiopathic pulmonary fibrosis: a systematic review

Camilla Rozanski, Marco Mura
Division of Respirology, Western University, London, Canada

Abstract. Idiopathic pulmonary fibrosis (IPF) has the highest mortality rate among all interstitial lung diseases, with a mean survival time of 2 to 3 years from the time of diagnosis. Prognosis is difficult to determine, due to the recognised heterogeneous progression of the disease, as well as lack of a well recognized staging system. This in turn raises challenges when considering therapeutic options for IPF patients, such as lung transplantation, versus a more conservative approach. Multiple independent factors have been identified as prognostic indicators in IPF, and a number of studies have proposed multi-dimensional indices that combine several parameters in order to achieve a more accurate delineation of disease progression. In this systematic review, the Medline database was used to identify studies in the area of multi-dimensional staging of IPF. We analysed and critically appraised previously proposed prognostic scoring systems. Our aim is to encourage research developments in this area to help identify an optimal multi-dimensional staging system for IPF. *(Sarcoidosis Vasc Diffuse Lung Dis 2014; 31: 8-18)*

Key words: Idiopathic pulmonary fibrosis, multi-dimensional, staging, prognosis, survival, scoring system.

Introduction

Diagnosis of idiopathic pulmonary fibrosis (IPF) is generally associated with a very poor prognosis. Mean survival after diagnosis typically ranges between 2-3 years (1,2). However, there is a significant population of patients who have a slower and less aggressive disease course, with longer survival times (3-5). Patients with IPF can remain stable for years regardless of medical treatment. However, the course of disease can change, with patients who originally display a slow and stable disease course progressing to a rapid decline in lung function. Delays in time to diagnosis can also occur, particularly amongst slow progressors, allowing for highly variable survival rates. Variability in the disease course makes staging and prognosis difficult, which in turn raises challenges when considering therapeutic options, such as lung transplantation (LTx), versus a more conservative approach.

Several studies have looked at prognosis and staging of IPF, however, a widely accepted scoring system is yet to be determined (6-8). Multi-dimensional scoring of disease is generally superior to single predictors alone, as seen in well accepted staging systems for other pulmonary diseases, such as the BODE index (BMI, Obstruction, Distance walked in 6-min, Exercise, Medical Research Council dyspnea score) for chronic obstructive pulmonary disease (9) and the CURB-65 score (Confusion, Urea, Respiratory rate, low Blood pressure, age over 65) for pneumonia (10). The predictive value towards mortality of the BODE score is in fact much higher than the value of forced

Received: 11 December 2013
Accepted after revision: 24 February 2014
Address for correspondences:
Marco Mura, MD, PhD
London Health Science Centre, Victoria Hospital
800 Commissioners Road East Room E6-203
N6A 5W9 London, Ontario, Canada
Tel. 001-519-667-6744 - Fax 001-519-685-8406
e-mail marco.mura@lhsc.on.ca

expiratory volume in 1 second (FEV_1) alone. Similarly, in pneumonia the prognostic strength of the CURB-65 predictors in combination are much stronger than when considered individually. More recently, a FACED score (FEV1, Age, chronic Colonisation with Pseudomonas Aeruginosa, radiologic Extension, Dyspnea) for cystic fibrosis was developed (11). The area under the receiver operating systems characteristics curve for FACED was 0.87, correlating with excellent prognositic value for 5-year all cause mortality in patients with CF. Similarly, a "Vienna prediction model" was developed in order to estimate probability of venous thromboembolism recurrence. Again, the hazard ratio for the developed model was higher than the hazard ratios for independently significant variables (male sex, positive D-dimer, history of proximal deep vein thrombosis, and history of previous pulmonary embolism) (12).

Due to the complex pathophysiology of IPF, single variable pulmonary function tests (PFTs) are unlikely to be accurate or representative of disease prognosis. Different studies showed only a moderate correlation between PFTs, dyspnea and exercise tolerance in patients with IPF (13,14). This suggests that exertional dyspnea and decreased exercise tolerance in IPF patients are multifactorial, and could not be adequately represented by PFTs alone. Similarly, the relationship between physiologic and clinical parameters with disease extent on high resolution computed tomography (HRCT) is not very strong (15-17). The presence of concomitant emphysema (CE) can also have confounding effects on the PFTs and on the severity of symptoms (18).

A variety of prognostic indicators have been used as part of proposed multi-dimensional scoring systems, such as age, gender, PFTs (including changes in results over time), quantitative scoring of the extent of disease on HRCT, exercise tests and dyspnea scores (6,8,19-21).

There is a critical need for a well recognised staging system in IPF patients. Staging of IPF would facilitate the prognostic evaluation as well as patient entry and future study design in IPF treatment research. Optimisation for lung transplant referral could also be improved, as the highest mortality rate amongst patients awaiting transplantation is observed in IPF (22). A precise staging system would facilitate more accurate and prompt referral, optimising the number of patients who would benefit from this intervention (23).

The objective of this review is to critically appraise the main studies in the area of multi-dimensional staging and prognosis of IPF, and summarise the variety of proposed prognostic scoring systems. Important considerations include generalisation of scores among patients with mild through to severe disease, the use of variables that are accessible and non-invasive, assessment of both baseline predictors and longitudinal predictors (change in data over time), exclusion or adjustment for patients with co-existing lung pathology, time-point of disease diagnosis, study design, and statistical methods.

With a better understanding of the approach to multi-dimensional scores in this area, and the pros and cons of previously designed studies, a single, more widely accepted staging system could be determined.

Methods

Data sources and searches

A literature search was finalised on November 30th, 2013. Studies were identified using PubMed (http://www.ncbi.nlm.nih.gov/pubmed) (April 1998 to November 2013). The diagnostic criteria for usual interstitial pneumonia (UIP)/IPF and other types of idiopathic interstitial pneumonia were established in April 1998 (24). Keywords, used alone or in combination, included: "idiopathic pulmonary fibrosis", "usual interstitial pneumonia", "multidimensional", "score", "stage", "index", "model", "scale", "prognosis", "outcome", "survival" and "mortality". Restrictions imposed included languages other than English and review articles. Titles and abstracts were screened to identify studies reporting multi-dimensional scoring/staging systems predicting survival in patients diagnosed with IPF. Eight original articles were identified.

Results

A total of 8 studies were analysed and critiqued. Findings regarding the methodology of multi-dimensional indeces and their predictive values were summarised in Tables 1 and 2, respectively.

Table 1. Summary of multi-dimensional indeces developed in patients with IPF.

Author	Population	N	Methodology	Variables in the Model	Endpoint	Reference
Gay et. al.	Prospective	38	- ANOVA - Fisher exact test for dichotomous variables - Student's t-test - ROC analysis	- HRCT score	Survival (average 34 months follow-up)	Am J Respir Crit Care Med 1998; 157: 1063-72
King et. al.	Retrospective	91 (available for survival analysis; model was derived from 228 subjects)	- Kaplan-Meier curves/log rank test - Cox univ./multiv. proportional hazards model	- Age - Smoking status - Clubbing - HRCT score - HRCT score for PH - TLC % pred - PaO_2 at max exercise	Survival (average 20 months follow-up)	Am J Respir Crit Care Med 2001; 164: 1171-81
Mogulkoc et. al.	Retrospective	95	- Cox univ/multiv. proportional hazards model - Univ/multiv. logistic regression analysis - ROC analysis	- HRCT score - DLCO % pred	2-year survival	Am J Respir Crit Care Med 2001; 164: 103-08
Wells et. al.	Retrospective	212	- Wilcoxon's rank-sum test - Chi-squared testing - Pearson's product-moment correlation	- DLCO % pred - FVC % pred - FEV1 % pred	5-year survival	Am J Respir Crit Care Med 2003; 167: 962-69
du Bois et. al.	Prospective (Clinical trial)	830	Univariate/ multivariate Cox proportional hazards model	- Age - Respiratory hospitalization - FVC % pred - 24wk in FVC % pred	1-year survival	Am J Respir Crit Care Med 2011; 184: 459-66
Richards et. al.	Prospective	241 (103, derivation cohort; 80 validation cohort)	-Kaplan-Meier curves/Log-rank test -Cox-proportional hazards model - ROC analysis	-gender -FVC %pred -DLCO %pred -MMP-7 (plasma protein)	Survival (average 1.8 yrs follow-up)	Am J Respir Crit Care Med 2012; 185: 67-76
Ley et. al.	Retrospective	558 (228, derivation cohort, 330, validation cohort)	- Fine-Gray survival model - Harrell c-index of discrimination - Fine-Gray competing risks model	- Gender - Age - FVC % pred - DLCO % pred	1,2, 3-year survival (average follow-up)	Ann Intern Med. 2012; 156: 684-91
Mura et. al.	Prospective (from time of diagnosis) / Retrospective	138 (70, prospective cohort; 68, retrospective cohort)	- Kaplan-Meier curves/log rank test - Cox univ./multiv. proportional hazard model - ROC analysis	- MRCDS - 6MWD % pred - CPI	3-year survival	Eur Respir J 2012; 40: 101-09

Table chronologically summarising study design, methodology, variables and endpoints for all papers described in review, with corresponding references. ANOVA = Analysis of Variance. HRCT=High resolution computed tomography. DLCO = Diffusing capacity for carbon monoxide. TLC = Total Lung Capacity. FVC = Forced Vital Capacity. FEV1 = Forced Expiratory Volume in 1 second. 6MWD = 6 Minute Walking Distance. MRCDS = Medical Research Council Dyspnea Score. CPI = Composite Physiologic Index.

Table 2. Summary of the predictive value of each multi-dimensional index developed in patients with IPF..

Author	Variables in the Model	Predictive Value	Reference
Gay et. al.	- CT-fibrosis score: 0-no interstitial disease 1-interlobe septal thickening 2-honeycombing (HC) 25% lobe 3-HC 25-49% lobe 4-HC 50-75% lobe 5-HC > 75% lobe	HRCT fibrotic score≥2: 80% sensitive and 85% specific in predicting death Pathology fibrotic score ≥ 16.0: sensitivity 84%, specificity 67% in predicting survival	Am J Respir Crit Care Med 1998; 157: 1063-72
King et. al.	- Age - Smoking Status - Clubbing - Radiologic evidence of profusion - Radiologic evidence of PHTN - TLC % pred - PaO_2 at max exercise	*Clubbing: HR=2.53, p=0.0001 Radiologic evidence of profusion: HR=2.48, p<0.0001 Radiologic evidence of PHTN: HR= 1.96, p=0.0017 TLC % pred: HR=0.70, p<0.0001 PaO_2 at max exercise: HR=0.74, p<0.0001	Am J Respir Crit Care Med 2001; 164: 1171-81
Mogulkoc et. al.	- HRCT-fibrosis score - DLCO % pred	HRCT fibrosis score: HR=0.957, p=0.005 DLCO% pred: HR=2.067, p=0.026 Combined model: AUC 0.91; sensitivity 84%, specificity 82% in predicting survival	Am J Respir Crit Care Med 2001; 164: 103-08
Wells et. al.	- CPI (based on FVC%, FEV1, DLCO%)	CPI: p<0.0005 (no HR provided)	Am J Respir Crit Care Med 2003; 167: 962-69
du Bois et. al.	- Age - Respiratory hospitalization - FVC% pred - 24wk Δ in FVC% pred	Age: >60-HR=2.19, p<0.009; 60-69 HR=1.64, p<0.10 , <60 HR=1.00 Respiratory hospitalization: HR=2.82, p<0.001 FVC% pred: <50 HR=3.90, p<0.006; 51-65 HR=2.35, p<0.016; 66-79 HR=1.46, p<0.291; >80 HR=1.00 ΔFVC% pred: ≤-10 HR=3.65, p<0.001; -5 to -9.9 HR=1.95, p<0.004 >-5 HR=1.00 Risk scoring system: AUC 0.75	Am J Respir Crit Care Med 2011; 184: 459-66
Richards et. al.	-gender -FVC%pred -DLCO%pred -MMP-7 (plasma protein)	MMP-7: HR=2.1, p=0.0021 (derivation); HR 1.7, p=0.17 (validation) PCMI≥330: AUC 0.74-0.84 (for survival at 0.4 to 1.75 years)	Am J Respir Crit Care Med 2012; 185: 67–76
Ley et. al.	- Gender - Age (≤60, 61-65, >65) - FVC % pred (>75, 50-75, <50) - DLCO % pred (>55, 36-55, ≤55) - Cannot perform DLCO	†Gender: Male HR=1.66, p=0.072 ‡Age: Age1 HR=1.15, p=0.001; Age2 HR=0.91, p=0.020 FVC% pred: HR=0.98, p=0.001 Inverse DLCO% pred: HR=1.03, p=0.002 Cannot perform DLCO: HR=17.5, p<0.001 GAP: c-index 69.3 (68.7 in the validation cohort)	Ann Intern Med 2012; 156: 684-91
Mura et. al.	- MRCDS>3 - 6MWD ≤72% predicted - CPI >41	MRCDS>3: HR=6.77, p<0.0005 6MWD <72% predicted: HR=3.25, p<0.0162 CPI > 41: HR=5.36, p<0.0071 ROSE>2: HR 11.4, p<0.0001; AUC 0.76; sensitivity 39%, specificity 100% in predicting survival	Eur Respir J. 2012; 40: 101-19

Table summarising the predictive value of multi-dimensional indices developed in each corresponding study for patients with IPF. Predictive values of independent variables are also included.

AUC = area under the curve (receiver operating characteristics analysis). HRCT=High resolution computed tomography. PH = Pulmonary Hypertension. DLCO = Diffusing capacity for carbon monoxide. TLC = Total Lung Capacity. FVC = Forced Vital Capacity. FEV1 = Forced Expiratory Volume in 1 second. 6MWD = 6 Minute Walking Distance. MRCDS = Medical Research Council Dyspnea Score. CPI = Composite Physiologic Index. PCMI = The Personal Clinical and Molecular Mortality Index. GAP = Gender Age Physiology score. ROSE = Risk Stratification Score.

**Hazard Ratios were adjusted for age and smoking status*

†The Fine-Gray competing risk model was used to evaluate categorical predictors for survival

‡Age1 and Age2 are calculated age variables following 3-know restricted cubic spline after centering age on its mean (67.68y)

Clinical, radiological and physiological (CRP) scoring system

A prospective study by Gay et al. (15) tested the hypothesis that initial diagnostic studies can identify patients with biopsy-proven IPF who are more likely to respond to corticosteroid therapy, predicting a better survival. Specifically, the study looked at the ability of clinical, functional and radiographic characteristics to identify patients with greater disease severity and greater responsiveness to therapy. Thirty-eight patients from a US tertiary centre were included and all had biopsy-confirmed UIP. Physiologic assessment included spirometry, lung volumes, diffusing capacity for carbon monoxide (DLCO), alveolar-arterial oxygen tension differences ($A\text{-}aO_2$) and cardio-pulmonary exercise testing (CPET). Clinical severity was assessed by a CRP scoring system developed by Watters et al., before the definition of clear diagnostic criteria for usual interstitial pneumonia/IPF (25). The original 1986 study included 26 patients, all diagnosed with biopsy-proven IPF and followed prospectively for 6 months. The CRP score was based on seven parameters: dyspnea (assessed by a subjective questionnaire ranked from 0 (no distress) to 20 (max distress)), quantitative chest radiograph, spirometry (forced vital capacity (FVC), FEV_1), lung volume (estimated by thoracic gas volume), DLCO, resting $A\text{-}aO_2$, and O_2 saturation corrected for maximal achieved VO_{2max}. Scores ranged from 0 to 100 (100 being the most severe disease). HRCTs were evaluated by four experienced thoracic radiologists independently and a HRCTs fibrosis score was obtained. All patients were then treated with prednisone for 3 months after lung biopsy and baseline studies were completed. The primary endpoint was mortality, with change in CRP score from baseline as a secondary endpoint. Receiver operating characteristics (ROC) analysis was used to qualify the value of all predictors in identifying individuals most likely to die during follow-up (26.0±2.9 months, 20.9±3.0 months, and 39.4±3.9 months in responders, non-responders, and stable individuals respectively). Only the HRCT fibrotic score ($p<0.009$) and the fibrotic pathology score ($p<0.03$) independently predicted survival. Addition of physiologic measures, CRP score, or pathologic findings to the HRCT fibrotic score did not improve its predictive value. Unfortunately, no hazard ratio, Kaplan-Meier or sensitivity/specificity analysis was provided for the CRP *per se* in this study.

Updated Clinical, Radiological and Physiological scoring system

King et. al. looked at the clinical, radiological and physiological predictors of survival in order to develop an updated CRP scoring system as a predictor of survival in newly diagnosed cases of IPF (6). This retrospective study in the US included 183 patients diagnosed between 1982 and 1996. IPF was diagnosed based on clinical and histological criteria, as all diagnoses were biopsy-confirmed. Patients were further stratified according to smoking status: current smokers, former smokers, and never smokers. Clinical assessment included demographics, past medical history, degree of shortness of breath, duration of illness and radiographic assessment of the extent of fibrosis and pulmonary hypertension (PH). The endpoint was survival and the average follow-up was 20 months. A CRP scoring system was derived from cox proportional hazards regression adjusted for age and smoking status. Independent, significant predictors of survival included age, smoking history, clubbing, extent of profusion of interstitial opacities, presence or absence of PH, % predicted total lung capacity (TLC) and PaO_2 at the end of maximal exercise. Maximal CRP score was 100, where a score of 100 corresponds to severe disease. Since lung mechanics and exercise testing may be difficult to obtain in common practice, an abbreviated model which excluded pulmonary mechanics and exercise data was also developed. The present model represents an evolution of a previously described scoring system (25). The updated, complete CRP model was superior to the abbreviated model in predicting survival, and both were superior to the original CRP score. However, an overall hazard ratio for the CRP score was not provided in the paper.

This study included a large number of subjects and a comprehensive group of variables. Limitations include its retrospective nature, the relative complexity of the score and the use of cardio-pulmonary exercise testing in the complete model. The exclusive use of biopsy-proven cases of usual interstitial pneumonia (UIP), although it strengthens the diagnostic criteria, may also have introduced a selection bias given that subjects amenable for surgical biopsy are much less likely to have severe disease.

Combined DLCO and HRCT disease score

The study of Mogulkoc et. al. (20) was prompted by the high death rates among patients with IPF listed for lung transplant (LTx). This study evaluated the relationship between baseline PFTs, HRCT fibrosis scores and survival in patients with IPF, in order to develop a non-invasive estimate of survival and optimize the timing of LTx referral in IPF patient. The patient population included 115 patients from a British tertiary medical centre. Forty-four (38%) patients underwent a surgical lung biopsy. All patients were less than 65 years of age. Spirometry, lung volumes, and DLCO parameters were measured. Based on the HRCT scanning, a mean score incorporating fibrotic and ground glass changes was obtained for each patient. Median follow-up was 26.2 months, and median survival was 55 months. The endpoint was 2-year survival. Univariate Cox regression was used to identify potential predictors of survival, which were then used as covariates in the multivariate Cox regression analysis to identify independent predictors of survival. Twelve different factors were found to be significantly associated with survival. Using multivariate stepwise regression analysis, only DLCO percent predicted (HR = 0.957, p<0.005) and HRCT-fibrosis score (HR = 2.076, p<0.026) were independent factors of survival. ROC analysis was used to identify the optimal cut-off values for DLCO (39%) and HRCT fibrosis score (2.25) towards 2-year mortality. With these cutoffs, multivariate analysis showed that DLCO (HR=0.923, p<0.021) and HRCT-fibrosis score (HR=6.274, p<0.021) were significant independent predictors. The model combining these two parameters yielded a specificity and sensitivity of 84% and 82%, respectively, to predict 2-year survival.

Relative strengths of this study are the combined use of radiographic and functional variables and the use of ROC analysis to compare the sensitivity and specificity of different models to predict survival. The limitations of this study are the exclusion of patients older than 65 years, the retrospective nature of the analysis and the relatively short follow-up. The authors could also have considered defining a specific scoring system with the independent variables.

The Composite Physiologic Index

A Composite Physiologic Index (CPI) derived from disease extent on HRCT was developed in a British study from Wells et al (8). The CPI is a score that provided an accurate estimate of the overall disease extent on HRCT, as it was derived against a quantitative radiographic score of pulmonary fibrosis. A combination of PFTs was used. The CPI was derived in a development group of 106 patients (36 with biopsy-proven UIP) and was tested in a validation group of the same consistency. In the development group, stepwise regression was used to generate a weighted combination of lung function variables fitting best with the extent of pulmonary fibrosis on computed tomography (CT). Parameters examined in the model included FEV_1, FVC, TLC, residual volume (RV), DLCO, carbon monoxide transfer coefficient (KCO), PO_2, and $A\text{-}aO_2$. The CPI was calculated using the formula derived from stepwise regression analyses against the radiographic extent on CT of pulmonary fibrosis. The extent of IPF on CT was independently related to percent predicted DLCO (p<0.0005), percent predicted FVC (p<0.0005), and percent predicted FEV_1 (p<0.02). Therefore the final derived CPI formula was:

Extent of disease on CT = 91.0
− (0.65 x percent predicted DLCO)
− (0.53 x percent predicted FVC)
+ (0.34 x percent predicted FEV_1)
− to account for (CE).

The CPI was then tested in a validation group. Correlation between the CPI and the extent of IPF (r=0.71, r^2=0.51, p<0.0005) was much stronger than any other functional-morphologic correlation. Regarding survival, the endpoint was 5-year survival, retrospectively assessed. Survival rate was 23%. Univariate analysis showed increased mortality associated with increasingly extensive disease on CT (p<0.0005), increased CPI scores (p<0.0005), and greater functional impairment. The CPI had the greatest prognostic power (p<0.005) in the combined cohort and both groups. Hazard ratios for the CPI were not provided.

The CPI provides both a reliable estimate of the extent of disease without the necessity of using HRCT at each follow-up, and a multi-dimensional physiologic tool to predict survival. Compared to the CRP score, the use of DLCO levels rather than exercise data is a strong advantage, as many patients have very poor tolerance to exercise. However, patients with advanced disease may not be able to perform a DLCO measurement. Finally, the CPI takes

into account the confounding effect of CE in the PFTs (18), but it is still a valid predictor of survival in patients without CE.

Mortality Risk Scoring System with Ascertainable Predictors

du Bois et. al. sought to develop a risk scoring system for 1-year mortality (26). The population consisted of patients from two international clinical trials of IFN-y1b (27,28), from both the placebo and treatment groups. Patients with concomitant emphysema were excluded. The study population included 830 patients with mild to moderate disease (FVC 50 to 90 percent predicted, DLCO greater than or equal to 25 percent predicted, and PaO_2 of more than 55 mm Hg at rest on room air). Many patients progressed to severe disease during the course of observation. Baseline predictors at the initial visit, and longitudinal predictors (i.e. changes over a 24-week period) measured at the week-24 visit and week-72 visit, were assessed. Patient sex, race, smoking status, history of cardiovascular disease, presence of honeycombing on HRCT scan, use of supplemental oxygen, and past surgical lung biopsy were evaluated at the baseline visit. PFTs, dyspnea - as assessed by the University of California at San Diego Shortness of Breath Questionnaire (UCSD SOBQ), health related quality of life (HRQL), and the occurrence of respiratory hospitalisation, were evaluated as longitudinal predictors. The endpoint was 1-year survival. One year mortality was found to be 9.7%. In the comprehensive, multivariate model developed for this trial, independent predictors of all-cause mortality were identified as age (HR=2.19, p<0.009), history of respiratory hospitalisations (HR=2.82, p<0.001), percent predicted FVC (HR=3.90, p<0.006), 24-week change in percent predicted FVC (HR=3.65, p<0.001), percent predicted DLCO (HR=1.74, p<0.046), 24-week change in percent predicted DLCO (HR=2.41, p<0.015), and 24-week change in HRQL (HR=3.63, p<0.001). A more applicable clinical model, which included only age, 24-week history of respiratory hospitalisation, percent predicted FVC, and 24-week change in percent predicted FVC was also derived. The comprehensive model and the clinical model had comparable discriminatory power. Finally, a mortality risk scoring system for patients with IPF was developed based on the clinical model predictive factors, providing an estimation of the 1-year risk of mortality for an individual patient. The model for the mortality risk scoring system is shown in Table 3.

Strengths of this study included the firm inclusion criteria in regard to the diagnosis of IPF, the large number of patients and the prospective nature of follow-up. Limitations are related to the fact that the population was derived from a clinical trial, and therefore only patients with mild to moderate disease were included and subjects were recruited at different time points from diagnosis, and follow-up was limited to 1 year. The use of hospitalisation for respiratory causes as a predictor of mortality has also been criticised, as the reasons why patients are hospitalised are not universally defined and can be difficult to discern (29).

The Personal Clinical and Molecular Mortality Index

Richards et al. investigated the predictive value of 95 different peripheral blood proteins in patients with IPF (30). A personal clinical molecular mortality index (PCMI), a multi-dimensional index, was derived incorporating peripheral blood matrix metalloproteinase-7 (MMP-7), a protein identified as

Table 3. The Mortality Risk Scoring System with Ascertainable Predictors (8).

Risk Factors	Score
Age	
>70	8
60-69	4
<60	0
History of respiratory hospitalization	
Yes	14
No	0
% Predicted FVC	
<50	18
51-65	13
66-79	8
>80	0
24-week change in % predicted FVC	
≤-10	21
-5 to -9.9	10
>-4.9	0

Table summarising the multi-dimensional risk scoring system developed by du Bois et al. to estimate 1-year mortality in patients with IPF. Individual scores for each risk factor are to be summed for a given patient with IPF. The total score then corresponds to 1-year probability of death.
FVC = Forced Vital Capacity

predictive of survival, with PFTs. This prospective study included 241 patients, divided into a derivation cohort (n=140) and a validation cohort (n=101). Eighty-five patients from the derivation cohort and 41 patients from the validation cohort had a histological diagnosis. The primary endpoints were mortality, transplant-free survival, and progression-free survival, defined as the first relative decline of 10% or more in FVC % predicted within a 1-year interval. Average follow-up was 1.8 years in the derivation cohort and 1.4 years in the validation cohort. For derivation of the personal clinical and molecular mortality index (PCMI), the Akaike information criterion (AIC), which provides a means for model selection, was applied for variable selection in the Cox proportional hazards model. With univariate analysis, 5 proteins were found to be associated with mortality or disease progression after Bonferroni correction (MMP-7, ICAM-1, interleukin-8, VCAM-1, S100A12). Overall, concentrations higher then calculated threshold plasma concentrations were associated with significantly lower median survival times (1.4-2.1 yr), while lower concentrations were associated with significantly higher median survival times (3.0-4.6 yr). The index was derived in a subset of 103 patients with non-missing biomarkers and all PFTs. The PCMI equation was derived by the stepAIC approach, as follows:

PCMI = 114 x I(Male) + 2(100% - FVC% Predicted) + 3(100% - DLCO%Predicted) + 111 x I(MMP-7≥4.3ng/ml) where I equals 1 if and only if the condition inside the parentheses is true.

The predictive value of FVC and DLCO towards the endpoints was not reported. With a PCMI cut-off of 330, low-risk patients (PCMI<330) had a median survival of 5.13 years after blood draw, while high risk subjects (PCMI≥330) had a median survival of 1.56 years after blood draw. With ROC analysis, the areas under the curve for mortality prediction ranged from 0.736 to 0.835 at different points, between 0.4 and 1.75 years.

Including serum biomarkers in a staging system for IPF is an important innovation, but some methodological considerations limit the strengths of conclusions. Although this study was prospective in nature, the length of follow-up from baseline is limited to assess outcome and no longitudinal assessment of biomarkers was done. Furthermore, although MMP7 was selected to be included in the model by AIC analysis, it is concerning that this biomarker did not predict survival in the validation cohort. Finally, it would have been interesting to learn about the prognostic power of PFTs in this study, since these variables were included in the final index.

The Gender, Age, Physiology (GAP) score

Ley et. al. (19) used 3 year retrospective data to develop a multi-dimensional staging system for prognosis of IPF. The index was termed GAP (gender, age, physiology [including 2 variables]), and aimed at developing a scoring system to predict mortality in IPF. Data was collected from retrospective cohorts in two US and one Italian centres. Median follow-up for the derivation and validation cohorts was 1.7 (range, 0.03 to 9.1) and 2.4 (range, 0.01 to 9.0) years respectively. Patients were divided into three cohorts: the derivation cohort (n=228), the validation cohort (n=330), and a second validation cohort (n=325) meant to validate the multi-dimensional index in follow-up. 44.3% and 54.7% of patients had biopsy-proven UIP, in the derivation and validation cohorts respectively. A competing-risk regression model was used to retrospectively screen potential predictors of mortality in the derivation cohort. Candidates for predictor variables included age, sex, BMI, smoking status, supplemental oxygen use, FVC, FEV_1, TLC and DLCO. The Fine-Gray models for survival were used to predict mortality risk. The primary endpoint was time to death or lung transplantation. Mortality was 49% in the derivation cohort, and 62% in the validation cohort. Four independent predictors were identified: age, sex (male HR = 1.66, p<0.072), FVC% pred (HR = 0.98, p<0.001), and DLCO% pred (HR = 1.03, p<0.002), which were then used to develop the GAP individual risk calculator towards mortality and staging system. A continuous model was developed, and each continuous predictor was divided into clinical categories, then used as categorical predictors in the Fine-Gray competing risks model. A staging system was developed: stage I - low risk; stage II - intermediate risk; stage III - high risk (Table 4). Three-year mortality was estimated to be 16.3%, 42.1% and 76.8% in stages I, II, and III respectively. The predictive accuracy of the continuous and point-score models was assessed during follow-up at 6, 12, 18, and 24 months after baseline. C-index towards survival for the GAP calculator was 70.8 and 69.1 in the deriva-

Table 4. The Gender, Age, Physiology (GAP) index (19).

	Predictor	Points
G	Gender	
	Female	0
	Male	1
A	Age (years)	
	≤60	0
	61-65	1
	≥65	2
P	Physiology	
	FVC % predicted	
	75	0
	50-75	1
	<50	2
	DLCO % predicted	
	>55	0
	36-55	1
	≤35	2
	Cannot perform	3

Table summarising the multidimensional GAP score (Gender, Age, Physiology) developed by Ley et al. to predict 1-, 2-, and 3-year mortality in patients diagnosed with IPF. Individual scores for each risk factor are summed. The total score then classifies patients into stage I, II, or III (not shown), which then corresponds to 1-, 2-, and 3-year mortality.
FVC = Forced Vital Capacity
DLCO = Diffusing Capacity for Carbon Monoxide.

tion and validation cohorts, respectively. The c-index for the GAP index was 69.3 and 68.7 in the derivation and validation cohorts, respectively. Model performance was evaluated 6 to 24 months after baseline in 325 patients (second validation cohort). The c-index for this group was 71.9 and 72.3 for the GAP calculator and index respectively.

The inclusion of a validation cohort and the large number of variables included are definite strengths of this study, although no exercise test was considered. The main limitation is the retrospective analysis for both the derivation and validation cohort, which can limit data quality and the strength of the conclusions, as some patients who died quickly from aggressive disease might have been excluded. Certainly it would be very interesting to investigate whether a longitudinal change of the GAP score can also predict disease progression and outcome.

The Risk Stratification Score

In this Italian study a cohort of 70 patients newly diagnosed with IPF were prospectively followed for a period of at least 3 years (21). A retrospective cohort of 68 patients from a different centre was used for comparative analysis. Of the total 138 patients, 55 (40%) had the diagnosis of UIP confirmed by surgical pathology. The remaining cases had the radiographic pattern reviewed by a panel of 3 expert radiologists. 27% had CE. Clinical indicators were collected at the time of diagnosis and then re-evaluated 6 months later. The incidence of acute IPF exacerbations was also addressed. Survival was defined as time to death or transplant. Mean survival after diagnosis in the prospective cohort was 30±21 months and the incidence of acute exacerbations (AEs) in this cohort was 18.6%. The independent predictors of 3-yr survival were identified using ROC and multivariate analysis: (1) Medical research council dyspnea score (MRCDS) >3 (HR=6.77, p<0.0005), (2) 6 minute walking distance (6MWD) ≤72% pred (HR=3.27, p<0.0162) and (3) CPI>41 (HR=5.36, p<0.0071). A Risk stratificatiOn ScorE (ROSE) was derived (Table 5).

3-yr mortality was 19%, 42% and 100% in the low, intermediate, and high risk groups, respectively. The ROSE predicted 3-yr survival with 39% sensitivity and 100% specificity. A ROSE of 3 carried a hazard ratio of 11.4 towards 3-year mortality. Importantly, advancement in the ROSE 6 months after diagnosis predicted 3-yr mortality with 94% sensitivity and 41% specificity in subjects with disease

Table 5. The Risk stratificatiOn ScorE (ROSE) (21).

Low Risk (I) - all of the below:	Intermediate Risk (II) - one of the below:	High Risk (III) - all of the below:
MRCDS≤3	MRCDS>3	MRCD>3
6MWD>72% pred	6MWD≤72% pred	6MWD≤72% pred
CPI≤41	CPI>41	CPI>41

Table summarising the multidimensional Risk stratificatiOn ScorE (ROSE) developed by Mura et al. to predict 3-year survival for patients newly diagnosed with IPF. 3 year mortality was 19% in the low risk group, 42% in the intermediate risk group, and 100% in the high risk group.
MRCDS = Medical Research Council dyspnoea score. CPI = Composite Physiologic Index
6MWD = 6-minute walking distance

initially considered mild to moderate (ROSE 1 or 2). These results suggest that clinically significant disease progression could be detected with these indicators. Other variables that significantly impacted 3-yr survival, but were not retained in the final model, included body mass index (BMI), FVC, DLCO, alveolar-arterial oxygen gradient, desaturation during the 6MWT, HRCT radiographic disease extent, and bronco-alveolar lavage total cell count.

Strengths of this study included rigorous diagnostic criteria, the exclusive inclusion of newly diagnosed patients, the longitudinal assessment of the score and the adequate length of prospective follow-up. However, the score needs validation in a larger population of patients with IPF. The use of a dyspnea score has also been criticised due to the subjective nature of the symptom (31). However, the MRCDS is the most objective available scale for dyspnea (32), it is largely independent from functional and exercise variables (14) and has be shown to strongly and independently predict survival in this and other studies (33,34).

Conclusions

The pathophysiology of IPF is complex and characterized by gas exchange and ventilatory impairments, insufficient increase in heart rate during exercise, increased elastic inspiratory load thus disturbing the balance with the elastic forces of the chest wall, and excessive dead space ventilation, which leads to an increase of the metabolic energy requirement of respiratory muscles and insufficient energy delivery to non-respiratory muscles that substain locomotion (35).

Studies have shown a less then optimal correlation between pulmonary function tests, dyspnea, exercise capacity and disease extent on the HRCT. Not surprisingly, functional tests alone resulted to be insufficient in staging the disease. This represents a serious limitation in conducting new clinical trials in IPF, which are urgently needed. Multi-dimensional scores showed an improved predictive power towards outcome in IPF. Due to their nature, multi- dimensional scores are able to capture different domains of the disease's pathophysiology and provide a broader extent of prognostic information.

However, the optimal multi-dimensional index to stage IPF is yet to be defined, as all scores published present some limitations either in design, methodology, type of population, length of follow-up, or number of patients. Indices derived from prospective clinical trials usually enroll patients with mild to moderate disease, and may underestimate the real prevalence of acute exacerbations, which are more common in advanced IPF (21,36). On the other hand, retrospective studies tend to miss patients with rapidly progressing disease, underestimating the real mortality rate, and are deficient in the longitudinal assessment of each variable's changes. Therefore, these important data are often missing from retrospective studies.

The optimal multi-dimensional score would probably originate from a multi-centre effort, as a prospective study that would include a large number of newly diagnosed patients and a comprehensive group of variables, would be of adequate length to capture AEs and mortality, and would implement a systematic, longitudinal assessment of parameters considered. The ideal multi-dimensional index would also be simple to calculate and include parameters that are easy to measure, safe and inexpensive.

A reliable multi-dimensional index with enhanced predictive power towards survival could likely represent a practical, sensitive and specific endpoint for clinical trials in IPF, thus enormously contributing to the advancement of therapeutic options for this devastating disease.

References

1. 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;165:277-304.
2. Rudd RM, Prescott RJ, Chalmers JC, Johnston ID. British Thoracic Society Study on cryptogenic fibrosing alveolitis: Response to treatment and survival. Thorax 2007;62:62-6.
3. Martinez FJ, Safrin S, Weycker D, et al. The clinical course of patients with idiopathic pulmonary fibrosis. Ann Intern Med 2005;142:963-7.
4. Raghu G, Collard HR, Egan JJ, 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;183:788-824.
5. Selman M, Carrillo G, Estrada A, et al. Accelerated variant of idiopathic pulmonary fibrosis: clinical behavior and gene expression pattern. PLoS One 2007;2:e482.

6. 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;164:1171-81.
7. Latsi PI, du Bois RM, Nicholson AG, et al. Fibrotic idiopathic interstitial pneumonia: the prognostic value of longitudinal functional trends. Am J Respir Crit Care Med 2003;168:531-7.
8. 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;167:962-9.
9. Celli BR, Cote CG, Marin JM, et al. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. N Engl J Med 2004;350:1005-12.
10. Lim WS, van der Eerden MM, Laing R, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax 2003;58:377-82.
11. Martinez-Garcia MA, de Gracia J, Vendrell Relat M, et al. Multidimensional approach to non-cystic fibrosis bronchiectasis. The FACED score. Eur Respir J 2013, published online, November 13.
12. Eichinger S, Heinze G, Jandeck LM, Kyrle PA. Risk assessment of recurrence in patients with unprovoked deep vein thrombosis or pulmonary embolism: the Vienna prediction model. Circulation;121:1630-6.
13. Papiris SA, Daniil ZD, Malagari K, et al. The Medical Research Council dyspnea scale in the estimation of disease severity in idiopathic pulmonary fibrosis. Respir Med 2005;99:755-61.
14. Mura M, Ferretti A, Ferro O, et al. Functional predictors of exertional dyspnea, 6-min walking distance and HRCT fibrosis score in idiopathic pulmonary fibrosis. Respiration 2006;73:495-502.
15. Gay SE, Kazerooni EA, Toews GB, et al. Idiopathic pulmonary fibrosis: predicting response to therapy and survival. Am J Respir Crit Care Med 1998;157:1063-72.
16. Wells AU, King AD, Rubens MB, Cramer D, du Bois RM, Hansell DM. Lone cryptogenic fibrosing alveolitis: a functional-morphologic correlation based on extent of disease on thin-section computed tomography. Am J Respir Crit Care Med 1997;155:1367-75.
17. Xaubet A, Agusti C, Luburich P, et al. Pulmonary function tests and CT scan in the management of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 1998;158:431-6.
18. Mura M, Zompatori M, Pacilli AM, Fasano L, Schiavina M, Fabbri M. The presence of emphysema further impairs physiologic function in patients with idiopathic pulmonary fibrosis. Respir Care 2006;51:257-65.
19. Ley B, Ryerson CJ, Vittinghoff E, et al. A multidimensional index and staging system for idiopathic pulmonary fibrosis. Ann Intern Med;156:684-91.
20. 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;164:103-8.
21. Mura M, Porretta MA, Bargagli E, et al. Predicting survival in newly diagnosed idiopathic pulmonary fibrosis: a 3-year prospective study. Eur Respir J 2012;40:101-9.
22. Hosenpud JD, Bennett LE, Keck BM, Edwards EB, Novick RJ. Effect of diagnosis on survival benefit of lung transplantation for end-stage lung disease. Lancet 1998;351:24-7.
23. Egan JJ, Martinez FJ, Wells AU, Williams T. Lung function estimates in idiopathic pulmonary fibrosis: the potential for a simple classification. Thorax 2005;60:270-3.
24. Katzenstein AL, Myers JL. Idiopathic pulmonary fibrosis: clinical relevance of pathologic classification. Am J Respir Crit Care Med 1998;157:1301-15.
25. Watters LC, King TE, Schwarz MI, Waldron JA, Stanford RE, Cherniack RM. A clinical, radiographic, and physiologic scoring system for the longitudinal assessment of patients with idiopathic pulmonary fibrosis. Am Rev Respir Dis 1986;133:97-103.
26. 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;184:459-66.
27. Raghu G, Brown K, Bradford WZ, et al. A placebo-controlled trial of interferon gamma-1b in subjects with idiopathic pulmonary fibrosis. N Engl J Med 2004;350:125-33.
28. King TE, Jr., Albera C, Bradford WZ, et al. Effect of interferon gamma-1b on survival in patients with idiopathic pulmonary fibrosis (INSPIRE): a multicentre, randomised, placebo-controlled trial. Lancet 2009;374:222-8.
29. Wells AU, Behr J, Costabel U, Cottin V, Poletti V, Richeldi L. Hot of the breath: mortality as a primary end-point in IPF treatment trials: the best is the enemy of the good. Thorax 2012;67:938-40.
30. Richards TJ, Kaminski N, Baribaud F, et al. Peripheral blood proteins predict mortality in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 2012;185:67-76.
31. Cottin V, Camus P. Practical issues and challenges in idiopathic pulmonary fibrosis. Respiration;86:1-4.
32. Sahebjami H, Sathianpitayakul E. Influence of body weight on the severity of dyspnea in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2000;161:886-90.
33. Nishiyama O, Taniguchi H, Kondoh Y, et al. A simple assessment of dyspnoea as a prognostic indicator in idiopathic pulmonary fibrosis. Eur Respir J;36:1067-72.
34. Manali ED, Stathopoulos GT, Kollintza A, et al. The Medical Research Council chronic dyspnea score predicts the survival of patients with idiopathic pulmonary fibrosis. Respir Med 2008;102:586-92.
35. Leblanc P, Bowie DM, Summers E, Jones NL, Killian KJ. Breathlessness and exercise in patients with cardiorespiratory disease. Am Rev Respir Dis 1986;133:21-5.
36. 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;37:356-63.