External validation and longitudinal application of the DO-GAP index to individualize survival prediction in IPF

Abhimanyu Chandel, MD, Christopher S. King, MD, Rosalinda V. Ignacio, MS, Jean Pastre, MD, Oksana A. Shlobin, MD, Vikramjit Khangoora, MD, Shambhu Aryal, MD, Alan Nyquist, MD, Anju Singhal, MD, Kevin R. Flaherty, MD, Steven D. Nathan, MD

Supplementary material

Table of Contents: **Page number**

Contributors

Assisted in patient care and data collection for patients enrolled in the Pulmonary Fibrosis Foundation Patient Registry

Columbia University Medical Center - Anna Podolanczuk

Dignity Health St. Joseph's Hospital and Medical Center - Walia Rajat

Duke University Medical Center - Lake Morrison

John Hopkins University - Danoff Sonye

Massachusetts General Hospital - Sydney Montesi

Mayo Clinic - Moua Teng

Medical University of South Carolina - Courtney Rowley

National Jewish Health - Tristan Huie

Piedmont Healthcare - Amy Case

Regents of the University of Minnesota Twin Cities - Hyun Kim

Stanford University - Joshua Mooney

Stony Brook University Hospital - Alpa Desai

Temple University Health System - Gerard Criner

The Ohio State University - Nitin Bhatt

The Pennsylvania State University - Rebecca Bascom

The University of Arizona - Sachin Chaudhary

The University of Kansas Medical Center - Mark Hamblin

The University of Texas HSC Houston - Rodeo Abrencillo

Tulane University - Joseph Lasky

UCLA, David Geffen School of Medicine at UCLA - Stephen Weigt

University of Alabama at Birmingham - Joao de Andrade

University of California at San Francisco - Paul Wolters

University of Chicago - Mary Strek

University of Cincinnati - Francis McCormack

University of Louisville - Saad Mohamed

University of Maryland - Nevins Todd

University of Miami School of Medicine - Marilyn Glassberg

University of Michigan - Elizabeth Belloli

University of Pennsylvania - Maryl Kreider

University of Pittsburgh - Daniel Kass

University of Rochester - R. Matthew Kottmann

University of Texas Southwestern Medical Center - Craig Glazer

University of Utah - Mary Beth Scholand

University of Virginia Interstitial Lung Disease Clinic - Paul Tessy

University of Washington - Ganesh Raghu

University of Texas Health Science Center at San Antonio- Anoop Nambiar

Vanderbilt University - Lisa Lancaster

Washington University School of Medicine - Adrian Shifren

Weill-Cornell - Robert Kaner

Yale University - Mridu Gulati

Methods

Exertional hypoxemia

Exertional hypoxemia within the DO-GAP index was defined by either an active prescription for resting or exertional supplemental oxygen or if desaturation (oxygen saturation <88%) was observed during six-minute walk testing (6MWT).

Sample size calculation:

Approximate necessary sample size for external validation (N=475) was estimated using the simulation-based method described by Riley et al. for time-to-event analysis and was based on a time point of interest of 3 years, C-statistic of 0.752, and overall mortality of approximately 35% at 3 years [1, 2].

Handling of missing data:

Missing data for several model parameters required consideration. Specifically, FVC% (2.3%), DLCO% (2.3%), and 6MWT distance (14.3%) were missing in some instances where patients were recorded to have completed this testing. To reduce bias introduced by listwise deletion of these cases, multiple imputations using chained equations was used for missing data in the primary analysis. Twenty sets of imputed data were created to replace missing values.

Joint modeling procedure:

Joint longitudinal and time-to-survival modeling 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 was performed. By design,

DO-GAP index has a right skewed distribution and natural logarithm transformation of this variable in the longitudinal model resulted in a better fit (Akaike information criterion: 712 vs. 2551 [lower values indicate improved fit]). Various model association structures were also tested and compared by the Deviance Information Criterion (DIC). The joint model formed by current value parameterization is conceptually easy to understand and comparably minimized DIC compared to shared random effects parameterization and thus, was selected for use.

Description of sensitivity analyses:

Sensitivity analyses were performed to supplement overall comparison with the existing GAP index to the DO-GAP index. First, the model performance was compared based on prediction of overall survival treating lung transplantation as a competing risk. Comparisons were performed via Fine and Gray competing risk regression with time-dependent estimates of Harrell's C-statistic made via the method outlined by Wolbers et al. to estimate the C-statistic in the presence of competing risks [3]. Model calibration was assessed as described in the original analysis. Additionally, as antifibrotic medication has the potential to change baseline survival, model performance was compared in the subset of patients taking these medications at the time of enrollment in the PFF-PR [4].

Finally, as discussed above, missing data in the PFF-PR was present in some instances for patients documented to have performed a DLCO maneuver (N=12). In the primary analysis, as inability to perform the DLCO maneuver was not recorded in the PFF-PR database, this category was eliminated from the examined models. However, as it was uncertain if this data was missing due to inability to perform the maneuver or for other reasons, repetition of the primary analysis categorizing these missing maneuvers as having been "unable to perform" was completed. The

results of tests of calibration related to the later two of these sensitivity analyses were unchanged from the primary analysis and this data has been omitted to avoid redundancy.

Software utilized:

Statistical analysis was performed with Stata/SE 17.0 (Stata Corp.), R version 4.1.2 (R Foundation for Statistical Computing), and the *JMbayes* package.

Table S1. DO-GAP index and staging system[2]

Predictor	Category	Points
Distance	**6MWT distance**	
	≥250 meters	0
	<250 meters	5
Oxygen	No hypoxemia	0
	Exertional hypoxemia[a]	5
Gender	Female	0
	Male	1
Age (years)	≤60	0
	61-65	1
	>65	2
Physiology	**FVC, % predicted**	
	>75	0
	50-75	1
	<50	2
	DLCO, % predicted	
	>55	0
	36-55	1
	≤35	2
	Cannot perform	3
Total possible points		**18**

Points	0-4	5-10	11-18
Stage	**I**	**II**	**III**

[a]Defined by either an active prescription for resting or exertional supplemental oxygen or if desaturation (oxygen saturation <88%) was observed during six-minute walk testing

Table S2. Predicted versus observed mortality in the entire cohort based on original GAP index.

Year	Predicted event frequency[a] (%)	Observed event frequency[b] (%)	Ratio[c]
Stage I[d]			
1	5.6	4.1	0.73
2	10.9	11.7	1.07
3	16.3	21.9	1.34
Stage II[d]			
1	16.2	12.5	0.77
2	29.9	26.1	0.87
3	42.1	41.0	0.97
Stage III[d]			
1	39.2	34.1	0.87
2	62.1	56.2	0.90
3	76.8	69.8	0.91

[a]Based on reported values by Ley et al. Note, this model was originally developed treating transplantation as a competing risk [5].
[b]Based on Kaplan–Meier estimates of transplant-free survival in the PFF-PR cohort
[c]Ratio of observed to predicted mortality
[d]Stages as defined by Ley et al [5].

Table S3. Change in classified stage in the validation set based on application of the DO-GAP index staging system compared to the original GAP index staging system in complete cases

GAP Stage	Patients (N)	DO-GAP Stage (N)			Percent reclassified (%)
		I	II	III	
I	88	60 (68.2)	25 (28.4)	3 (3.4)	31.8
II	232	63 (27.2)	139 (59.9)	30 (12.9)	40.1
III	110	0 (0)	15 (13.6)	95 (86.4)	13.6
Overall	430	123 (28.6)	179 (41.6)	128 (29.8)	31.6

Data expressed as n (%)
GAP, gender-age-physiology; DO-GAP, distance-oxygen-gender-age-physiology

Table S4. Predicted versus observed mortality (%) in the entire cohort based on the DO-GAP index.

Year	Predicted event frequency[a] (%)	Observed event frequency[b] (%)	Ratio[c]
Stage I[d]			
1	4.1	1.6	0.39
2	9.4	9.5	1.01
3	16.0	18.1	1.13
Stage II[d]			
1	10.3	14.7	1.43
2	21.4	29.5	1.38
3	35.7	43.5	1.22
Stage III[d]			
1	38.0	34.0	0.89
2	58.7	56.3	0.96
3	72.6	69.6	0.96

[a]Based on reported values by Chandel et al [2].
[b]Based on Kaplan–Meier estimates of transplant-free survival in the PFF-PR cohort
[c]Ratio of observed to predicted mortality
[d]Stages as defined by Chandel et al [2].

Table S5. Model discrimination of the GAP index and DO-GAP index based on the results of sensitivity analyses

	C-statistic (95% CI)	
Analysis	**GAP Index**	**DO-GAP Index**
Lung transplantation as competing risk	0.69 (0.65-0.74)	0.74 (0.70-0.77)
Patients taking antifibrotics	0.67 (0.64-0.71)	0.72 (0.69-0.75)
Missing DLCO treated as "unable to perform"	0.67 (0.63-0.71)	0.73 (0.69-0.74)

Table S6. Predicted versus observed mortality in the entire cohort based on original GAP index treating transplantation as a competing risk.

Year	Predicted mortality[a] (%)	Observed mortality[b] (%)	Ratio[c]
Stage I[d]			
1	5.6	3.1	0.55
2	10.9	5.2	0.48
3	16.3	8.1	0.50
Stage II[§]			
1	16.2	8.1	0.50
2	29.9	16.9	0.57
3	42.1	26.9	0.64
Stage III[§]			
1	39.2	17.8	0.45
2	62.1	35.7	0.57
3	76.8	47.5	0.62

[a]Based on reported values by Ley et al [5].
[b]Based on Kaplan–Meier estimates of transplant-free survival in the PFF-PR cohort
[c]Ratio of observed to predicted mortality
[d]Stages as defined by Ley et al [5].

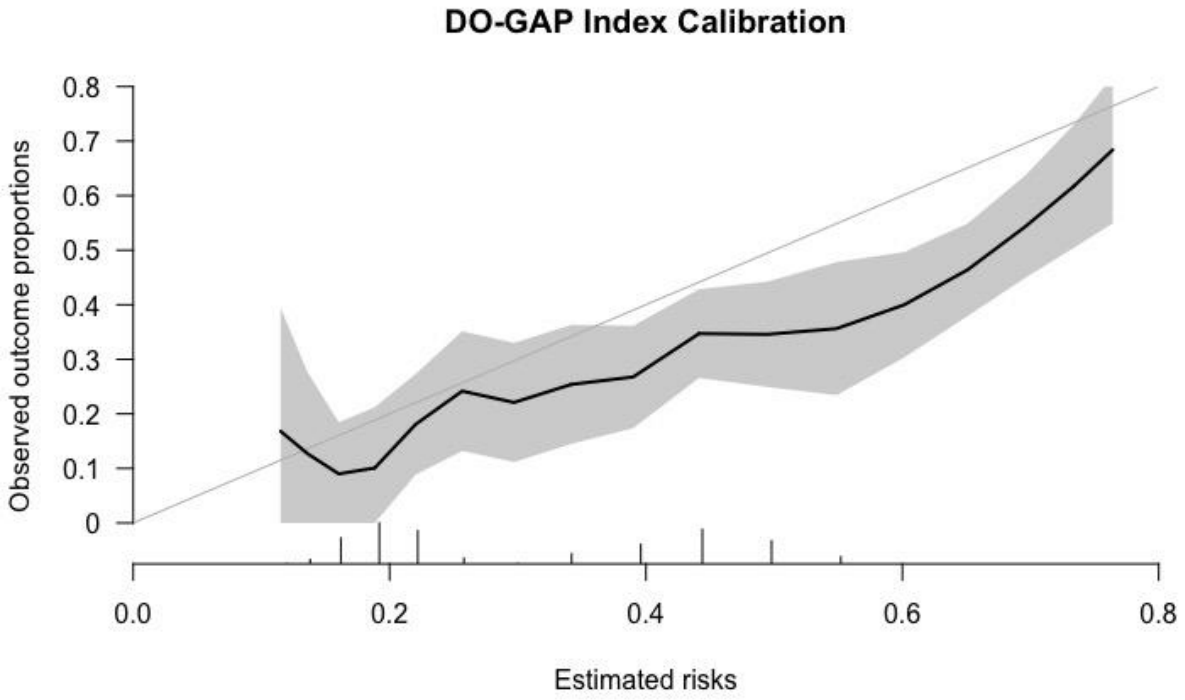

Figure S1. Calibration of the Distance-Oxygen-Gender-Age-Physiology (DO-GAP) index in the Pulmonary Fibrosis Foundation-Patient Registry (PFF-PR) for predicting overall survival treating lung transplantation as a competing risk. Smoothed pseudovalues (dark solid line) with pointwise 95% confidence intervals (shaded area) are plotted against predicted probabilities at 3 years of follow-up. The light solid line is the line of identity, denoting perfect calibration. The spike histogram below the plot shows the distribution of predicted risks. Evidence of miscalibration (overestimation of observed risk) in patients with the highest observed overall mortality is observed.

Supplementary references

1. Riley RD, Collins GS, Ensor J, Archer L, Booth S, Mozumder SI, Rutherford MJ, van Smeden M, Lambert PC, Snell KIE. Minimum sample size calculations for external validation of a clinical prediction model with a time-to-event outcome. *Stat Med* 2022: 41(7): 1280-1295.
2. Chandel A, Pastre J, Valery S, King CS, Nathan SD. Derivation and validation of a simple multidimensional index incorporating exercise capacity parameters for survival prediction in idiopathic pulmonary fibrosis. *Thorax* 2022.
3. Wolbers M, Blanche P, Koller MT, Witteman JCM, Gerds TA. Concordance for prognostic models with competing risks. *Biostatistics* 2014: 15(3): 526-539.
4. Kang J, Han M, Song JW. Antifibrotic treatment improves clinical outcomes in patients with idiopathic pulmonary fibrosis: a propensity score matching analysis. *Scientific Reports* 2020: 10(1): 15620.
5. Ley B, Ryerson CJ, Vittinghoff E, Ryu JH, Tomassetti S, Lee JS, Poletti V, Buccioli M, Elicker BM, Jones KD, King TE, Jr., Collard HR. A multidimensional index and staging system for idiopathic pulmonary fibrosis. *Annals of internal medicine* 2012: 156(10): 684-691.