



The effect of CALIPER-derived parameters for idiopathic pulmonary fibrosis in predicting prognosis, progression, and mortality: a systematic review

Xin-yao Jiao¹ · Han Song¹ · Wei-wu Liu¹ · Jun-ling Yang² · Zhi-wei Wang¹ · Dan Yang¹ · Sa Huang¹

Received: 3 September 2022 / Revised: 7 May 2023 / Accepted: 3 June 2023 / Published online: 1 August 2023
© The Author(s), under exclusive licence to European Society of Radiology 2023

Abstract

Background High-resolution computed tomography (HRCT), as the main tool for monitoring idiopathic pulmonary fibrosis (IPF), is characterized by subjective variability among radiologists and insensitivity to subtle changes. Recently, a few studies have aimed to decrease subjective bias by assessing the severity of IPF using computer software, i.e., Computer-Aided Lung Informatics for Pathology Evaluation and Rating (CALIPER). However, these studies had diverse research directions. In this review, we systematically assess the effect of CALIPER in the management of IPF.

Methods A systematic review was conducted through a search of published studies in PubMed, Web of Science, Cochrane, Embase, Scopus, and CNKI databases from database inception through February 28, 2022. The methodological quality would be evaluated by using Methodological Index for Non-Randomized Studies (MINORS). Narrative synthesis summarized findings by participant characteristics, study design, and associations with outcomes.

Results Ten studies were included. They evaluated the relationship between CALIPER-derived parameters and pulmonary function test (PFT) and mortality. CALIPER-derived parameters showed a significant correlation with PFT and mortality. Two studies reported that CALIPER could be used to stratify outcomes.

Conclusion CALIPER-derived parameters can be used to evaluate prognosis and mortality. CALIPER-derived parameters combined with composite physiologic index (CPI) or Gender-Age-Physiology (GAP) could help clinicians implement targeted management by refining prognostic stratification. However, research has been constrained by small number of retrospective investigations and sample sizes. Therefore, it is essential to design prospective controlled studies and establish the staging system by CALIPER-derived parameters and combining them with CPI, FVC, or GAP.

Clinical relevance statement It is beneficial for clinic to provide objective, sensitive, and accurate indicators of disease progression. It also helps the clinic to develop individualized treatment plans based on the stage of disease progression and provides evaluation of efficacy in drug trials.

Key Points

- *Computer-Aided Lung Informatics for Pathology Evaluation and Rating (CALIPER) is a quantitative CT analysis software that can be used to evaluate the progression of disease on CT.*
- *The CALIPER-derived vessel-related structure shows great performance in the management of idiopathic pulmonary fibrosis.*
- *CALIPER-derived parameters combined with composite physiologic index or Gender-Age-Physiology can be used to refine prognostic stratification.*

Keywords CALIPER · Idiopathic pulmonary fibrosis · Prognosis · Mortality

✉ Sa Huang
huangsa@jlu.edu.cn

¹ Department of Radiology, The Second Hospital of Jilin University, Changchun 130041, People's Republic of China

² Department of Respiratory, The Second Hospital of Jilin University, Changchun 130041, People's Republic of China

Abbreviations

CALIPER	Computer-Aided Lung Informatics for Pathology Evaluation and Rating
CPI	Composite physiologic index
CT	Computed tomography
DLCO	Diffusion lung capacity for carbon monoxide
FVC	Forced vital capacity

GAP	Gender-Age-Physiology
GGO	Ground glass opacity
H	Honeycombing
HRCT	High-resolution computed tomography
ILD	Interstitial lung disease
IPF	Idiopathic pulmonary fibrosis
MINORS	Methodological Index for Non-Randomized Studies
MZ	Middle zone
PFT	Pulmonary function test
QCT	Quantitative CT analysis
R	Reticulation
UZ	Upper zones
VRS	Vessel-related structure

Introduction

Idiopathic pulmonary fibrosis (IPF) is a chronic and progressive interstitial lung disease. IPF has an indistinct etiology, a poor prognosis, and an unpredictable course [1–3]. The incidence and mortality rates of IPF are consistently increasing [4–7]. Existing antifibrotic treatments are not curative. Early and timely intervention is essential to preserve pulmonary function furthest [8, 9]. The heterogeneous disease course and the importance of pharmacological interventions in the early stages of IPF make it difficult to assess IPF disease progression and to predict the prognosis. Therefore, it is necessary to explore accurate and sensitive medical imaging indicators to assess condition stage and guide physicians in early pharmacological intervention.

Early and effective antifibrotic treatment reduces the risk of respiratory-related hospitalization and death in IPF patients. Some studies found that patients with faster progression appear to have greater benefit than those with slower progression within 6–12 months of drug initiation [10]. And studies have shown that in patients treated with pirfenidone, transplant-free survival is significantly lower in patients with advanced IPF than in those with mild to moderate IPF [11, 12]. Therefore, the ability to accurately monitor disease progression is of considerable clinical importance in the development of treatment plans for patients with IPF. Pulmonary function tests (PFTs), such as forced vital capacity (FVC) and diffusion lung capacity for carbon monoxide (DLCO), are considered the most standardized methods for estimating deterioration conditions in IPF [13]. A greater than 10% reduction in FVC has been shown to predict a higher mortality over the next 6 or 12 months [14, 15]. Compared to the baseline, a 15% decline in DLCO was a sign of disease progression [13, 16]. Nevertheless, a 10–15% measurement deviation for each PFT test inevitably decreased efficacy. The presence of emphysema weakens the impact of fibrosis on pulmonary function indices and conceals real

disease progression [17–21]. Therefore, PFT indicators may not be useful or timely indicators of disease progression for clinicians [22, 23]. Recently, capturing the change in characteristic HRCT signs (e.g., ground glass opacity, reticulation, honeycombing) by radiology has been the preferred method of estimating disease progression [20]. However, there is subjective variability when these signs are assessed by the radiologist alone. Given the lack of discrepancy in professional practice, radiologists show moderate agreement regarding honeycombing when it is combined with emphysema, traction bronchiectasis, and thin-walled large cysts [24–26]. Such a level of agreement could lead to errors in assessing disease progression based on HRCT signs.

Quantitative CT analysis is one area of the rapidly developing field of radiomics. Radiomics can extract, analyze, and interpret data from medical images to aid disease diagnosis and prognosis [27]. With the extensive application of machine learning in the medical field, quantitative analysis techniques have been used in quantitative CT analysis (QCT) [27–29]. Machine learning is a branch of artificial intelligence. When the researcher provided a set of data, the machine learning algorithm system could be used to extract image features that were considered important for prediction or diagnosis. The features found in the training dataset could then be used to make predictions and classify future inputs [30, 31]. The results of the entire study were summarized after classification and could then be statistically analyzed in terms of absolute volumes per region or percentages per feature. The Computer-Aided Lung Informatics for Pathology Evaluation and Rating (CALIPER) software, which can be used as a QCT method, was trained by an expert radiologist. Currently, CALIPER is always used in diagnostic or prognostic studies of COPD and IPF. CALIPER can be used to segment and extract the anatomical region in HRCT. Then, through the machine learning algorithms, CALIPER could label the interstitial lung abnormalities of ILD subjects automatically in 3D CT and quantify the characteristic imaging signs of IPF [32, 33]. Then, it can be used to calculate the percentage of characteristic imaging signs on total lung volume, such as ground glass opacity (GGO), reticulation (R), honeycombing (H), low attenuation areas, normal tissue, and vessel-related structure (VRS). The extent of these signs and longitudinal change reflected disease condition [34].

Currently, several studies have examined the performance of CALIPER in disease surveillance. Different studies have had diverse research directions when examining the application of CALIPER for IPF. Some studies also had smaller sample sizes. In this systematic review, we reviewed literature reporting the use of CALIPER for analyzing IPF with respect to its disease progression, prognosis, and mortality. The benefits, limitations, and several future directions are also discussed in this review.

Methods

Search strategy

We searched the PubMed, Web of Science, Cochrane, Embase, Scopus, and CNKI databases up to February 28, 2022, to identify relevant studies. We applied no language restrictions. We used the following search terms: “CALIPER,” “Pulmonary Fibrosis,” “Idiopathic Pulmonary Fibrosis,” and “IPF” in combination with Boolean operators “OR” “AND.” The complete search strategy used in PubMed was as follows: (CALIPER) AND (((IPF) OR (“Pulmonary Fibrosis”[Mesh])) OR (“Idiopathic Pulmonary Fibrosis”[Mesh])). In addition, reference lists of all eligible studies about the topic were manually searched for relevant studies.

Study selection

The included studies met the following eligibility criteria: (1) the target disease was IPF; (2) the detection method was computed tomography (CT); (3) the CT patterns were evaluated by CALIPER; and (4) the study assessed the association of CALIPER with PFT and mortality.

The exclusion criteria were as follows: (1) duplicate publications; (2) REVIEWS and conference abstracts; (3) FULL text could not be obtained; and (4) studies in which valid data were not available.

Data extraction

The articles and relevant information were independently inspected and extracted by 2 investigators. The extracted information was as follows: first author’s name, publication year, study location, study design, sample capacity, patient characteristics (mean age), diagnosis criteria, endpoint definition, years of follow-up, number of events, statistical model used, and multivariate adjustments. The mean and standard deviation values of baseline PFT and CALIPER indices, and survival analysis of HR and 95% CI were extracted. To minimize confounding, we used adjusted HRs from the multivariate regression analysis as possible. When disagreements arose, they were resolved by seeking a senior investigator to discuss and re-examine the appropriate research together.

Quality assessment

Two reviewers used the Methodological Index for Non-Randomized Studies (MINORS) tool to independently evaluate the non-RCT studies’ quality. The MINORS tool comprised

12 items, the first eight of which were designed for noncomparative investigations, and the maximum score was 16 or 24. Each item was scored from 0 to 2; scores of 0 indicated that the item was not reported in the article, 1 indicated that it was reported but insufficiently, and 2 indicated that it was reported sufficiently.

Statistical analysis

Owing to the enormous heterogeneity between studies, we intentionally aimed not to perform a formal quantitative synthesis.

Results

Search results

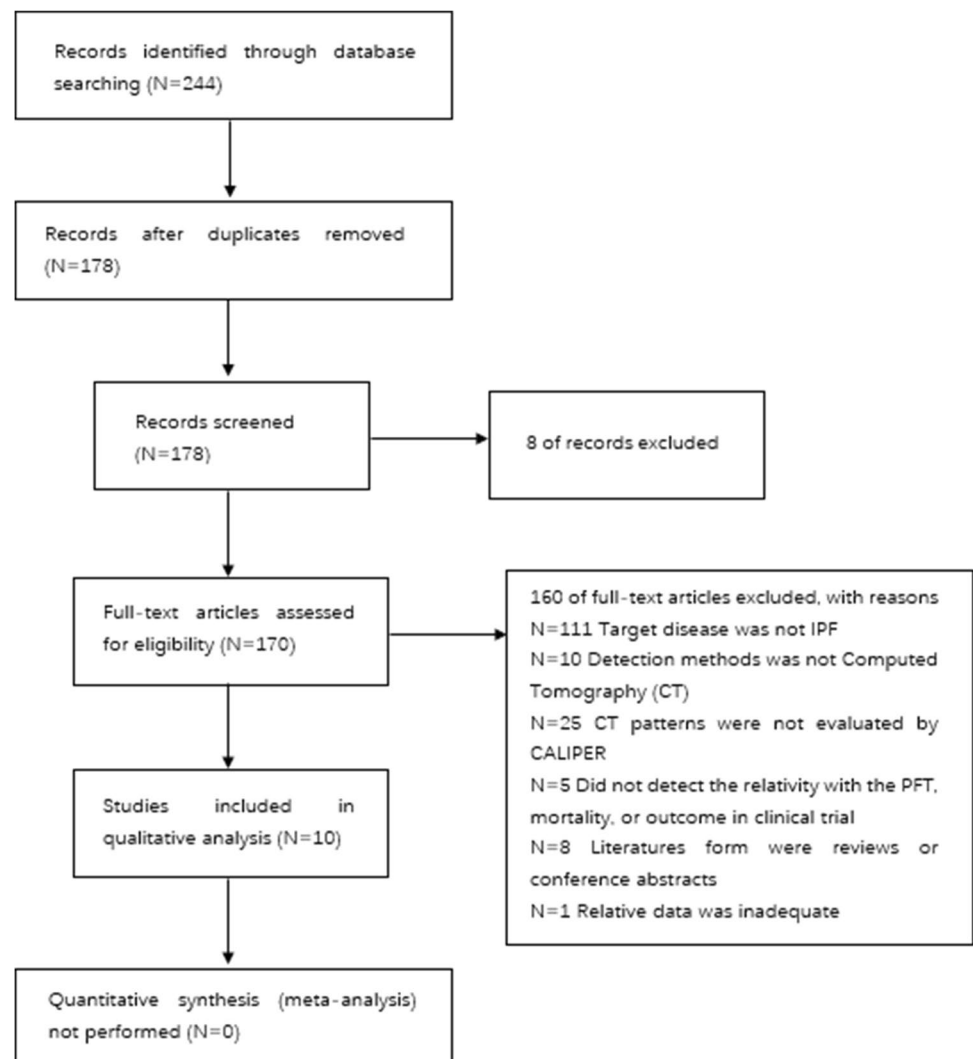
The electronic literature search yielded 244 potential studies. After duplicates were removed by Endnote 20, 178 studies remained. Eight studies were excluded because the full text could not be obtained. After reviewing 170 full texts, 111 studies were excluded because the target disease was not IPF, 10 studies were excluded because the detection method was not computed tomography (CT), 25 studies were excluded because CT patterns were not evaluated by CALIPER, 5 studies were excluded because they did not detect the relationship with PFT, mortality, or outcome in clinical trials, and 8 studies were excluded because they were reviews or conference abstracts. One Chinese study was not included because there was inadequate interpretation of the data. Ultimately, 10 studies met the inclusion criteria. Figure 1 presents a flow diagram of the search.

Study characteristics

Ten studies were included and evaluated. Most of the included studies used a single center design (7 out of 10 studies). They were conducted in Europe or the USA (3 and 7 respectively). Of the ten studies, there were one prospective study, two case–control studies, and seven retrospective studies. The included studies were published from 2014 to 2021. Our analysis contained 1312 participants. In particular, 2 studies used the same data. The number of subjects taking part in these studies ranged from 38 to 531 (Table 1). Supplement Table 1 is a summary of the inclusion and exclusion criteria for all studies. We summarize the duration of follow-up and number of events for each study in Table 2.

Five of the ten studies measured CALIPER-derived ILD, R, H, GGO, and VRS; three studies measured CALIPER-derived ILD and VRS; one measured CALIPER-derived R, H, GGO, and VRS; and one study measured CALIPER-derived VRS only. Four studies had all-cause mortality as

Fig. 1 Flow diagram of selected papers including identified number of records (included and excluded) and the reasons for exclusions



an endpoint, and three studies used both endpoints (all-cause death and lung transplantation). One study reported mortality but was not applied for prognostic analysis.

The statistical methods used for assessment vary considerably due to differences in study design. Among the included studies, 7 studies evaluated the correlation between the CALIPER-derived parameters and PFT and four cohort studies conducted Pearson's correlation analysis. Three cohort studies conducted linear correlation analyses. Seven studies assessed the association between parameters and mortality, 3 cohort studies performed Cox regression analyses and calculated HRs, and 2 cohorts calculated $-\log_{10} p$ values for covariates in multivariate Cox mortality models. One cohort study used Kaplan–Meier analysis to assess survival, and one cohort study was conducted by a chi-square test in relation to mortality. Four studies conducted segmental analysis for CALIPER-VRS, and 2 studies evaluated the performance of CALIPER with respect to stratification. One study used receiver operating characteristic (ROC) curve

analysis to confirm the best threshold. One cohort study identified independent predictors of mortality by Cox survival analysis and then combined it with CPI to obtain a new formula that represents a per-patient mortality estimate. The characteristics of the articles are summarized in Table 1, Table 3, and Table 4.

Descriptive summary of results

CALIPER-derived parameters and PFT at the baseline level

Five studies involving 872 subjects reported the association of CALIPER-derived parameters with PFT at baseline (the time point of diagnosis or pre-treatment) [29, 35–38]. In particular, 2 studies used the same data, so in this part they were deemed to be one study. The total of GGO, R, and H features constituted the extent of CALIPER-derived interstitial lung disease (ILD). A majority of results showed that CALIPER-ILD and CALIPER-VRS appeared to be strongly

Table 1 Basic characteristics of patient of included studies

Trial	Year	Country	Design	Groups (<i>n</i>)	Age (years)	Male/female
Koo, C. W. et al [35]	2021	USA	Prospective	<i>n</i> = 38	76	28/10
Sverzellati, N. et al [36]	2020	Italy	Retrospective	<i>n</i> = 58	66.4	40/18
Romei, C. et al [37]	2020	Italy	Retrospective	<i>n</i> = 105	72.1	84/21
Crews, M. S. et al [42]	2020	USA	Retrospective	<i>n</i> = 58	65.8	45/13
Jacob, J. et al [40]	2018	UK	Retrospective	Discovery cohort <i>n</i> = 71 Validation cohort <i>n</i> = 47	NA	NA
Jacob, J. et al [29]	2018	UK	Retrospective	Discovery cohort <i>n</i> = 247 Validation cohort <i>n</i> = 284	*Discovery cohort 67 Validation cohort 69	Discovery cohort 192/55 Validation cohort 225/59
Jacob, J. et al [39]	2018	UK	Retrospective	<i>n</i> = 66	*65.5	53/13
Jacob, J. et al [43]	2017	UK	Retrospective	<i>n</i> = 283	*67	219/64
Jacob, J. et al [38]	2016	UK	Retrospective	<i>n</i> = 283	*67	219/64
Maldonado, F. et al [41]	2014	USA	Retrospective	<i>n</i> = 55	72.4	28/27

Data are presented as *n* or mean, unless otherwise stated. *Data represents median values
NA, not assign

related to FVC [29, 35–38]. Only one study analyzing 38 patients identified that there was inadequate evidence to prove that CALIPER-derived GGO, R, ILD, and VRS were correlated with DLCO [35], and CALIPER-H was not significantly related to FVC.

CALIPER-derived parameters and PFT in longitudinal follow-up

Five studies involving 324 subjects reported the association of CALIPER-derived parameters with PFT in longitudinal follow-up. A majority of results showed a statistically significant association of FVC with CALIPER-VRS and CALIPER-ILD at the 1-year follow-up time point [35–37, 39, 40]. Both the studies of Sverzellati, N. et al and Jacob, J. et al found that

this association of FVC with CALIPER-derived GGO, R, H, ILD, and VRS also existed at the 2-year follow-up time point [36, 39]. Only one study reported that CALIPER-GGO and CALIPER-R were significantly related to FVC at the 3-month and 6-month follow-up time point [35]. Both the studies of Jacob, J. et al and Maldonado, F. et al provided the annual variation of each CALIPER-derived parameter. Among the individual CALIPER-derived parameter, CALIPER-GGO and R were the patterns that increased the most per year [39, 41].

CALIPER-derived parameters in predicting mortality at the baseline level

Five studies involving 991 subjects evaluated the prognostic value of CALIPER-derived parameters at baseline

Table 2 Adverse events by study

Trial	Year	Follow-up, y	Total, <i>n</i>	Endpoint	Events, <i>n</i>
Koo, C. W. et al [35]	2021	1	20	Death	5
Sverzellati, N. et al [36]	2020	3.6	58	Death or lung transplantation	17
Romei, C. et al [37]	2020	3	105	Death	21
Crews, M. S. et al [42]	2020	NA	58	Death	NA
Jacob, J. et al [40]	2018	2.5	118	Death or lung transplantation	98
Jacob, J. et al [29]	2018	NA	558	Death	375
Jacob, J. et al [39]	2018	NA	66	NA	46
Jacob, J. et al [43]	2017	2.5 ± 1.79	283	Death	210
Jacob, J. et al [38]	2016	NA	NA	NA	NA
Maldonado, F. et al [41]	2014	2.4	55	Death or lung transplantation	39

Data are presented as *n* or mean, unless otherwise stated
NA, not assign

Table 3 Main characteristics of selected studies evaluating CALIPER-derived parameters as the biomarkers for predicting prognosis in patients with IPF

Trial	Year	Diagnosis	PFT mean (SD)		CALIPER mean (SD)		Changes	CALIPER
			Time	PFT	Time	CALIPER		
Koo, C. W. et al [35]	2021	Multidisciplinary diagnosis	Pre-treatment	FVC, 2.625 (0.773) FVC%, 71.9 (18.9) FEV1, 2.131 (0.562) FEV1%, 77 (18.2) DLCO, 10.136 (3.435) DLCO%, 42.5 (12.7)	Pre-treatment	Pct ILD, 25.507 (16.169) Pct VRS, 5.398 (1.872)	N	N
			Baseline	FVC% 80.1 (18.0) DLco % 57.8 (16.5)	Baseline	Pct total lung fibrosis extent, 17.86 (13.46) Pct ground glass opacity, 8.39 (10.30) Pct reticular, 8.02 (6.21) Pct honeycombing, 1.44 (2.77) Pct VRS, 4.18 (1.69)	N	N
Sverzellati, N. et al [36]	2020	2011 Guidelines	Baseline	FVC% 80.1 (18.0) DLco % 57.8 (16.5)	Baseline	Pct ILD, 19.81 (14.69) Pct VRS, 4.97 (1.67)	N	N
Romei, C. et al [37]	2020	Multidisciplinary diagnosis	Baseline	FVC%, 78.27 (17.60) DLCO%, 54.37 (16.19)	Baseline	Mean VRS volume, 195.7	N	N
Crews, M. S. et al [42]	2020	Multidisciplinary diagnosis	N	N	Baseline	N	N	N
Jacob, J. et al [40]	2018	Multidisciplinary diagnosis	N	N	N	N	N	N
Jacob, J. et al [29]	2018	Multidisciplinary diagnosis	Baseline	FVC % Discovery 69.1 (20.1) Validation 74.2 (19.5) DLco % Discovery 36.3 (12.9) Validation 39.1 (14.4)	Baseline	Pct ILD Discover 25.4 (17.0) Validation 23.5 (16.3) Pct VRS Discover 5.0 (1.6) Validation 5.3 (1.7)	N	N

Table 3 (continued)

Trial	Year	Diagnosis	PFT mean (SD)		CALIPER mean (SD)		Changes	CALIPER	CALIPER
			Time	Time	Time	Time			
Jacob, J. et al [39]	2018	Multidisciplinary diagnosis	Relative change per year	* FEV1, -8.7 (11.9) * FVC, -10.4 (16.3) * DLco, -18.7 (18.3)	Relative change per year	*ILD, 7.8 (15.1) *Ground glass opacity, 5.3 (12.6) *Reticular pattern, 2.3 (4.0) *Honeycombing, 0.2 (0.6) *VRS, 0.9 (1.3)	Annual change	Pct ground glass opacity, 5.3 (12.6) Pc reticular pattern, 2.3 (4.0) Pct honeycombing, 0.2 (0.6) Pct VRS, 0.9 (1.3)	<i>N</i>
Jacob, J. et al [43]	2017	Multidisciplinary diagnosis	Baseline	FEV1%, 70.8 (19.1) FVC %, 68.8 (20.5) DLCO%, 36.1 (12.9)	Baseline	Pct total extent of ILD, 26.5 (18.1) Pct ground glass opacity, 17.0 (14.7) Pct reticular pattern, 8.5 (6.0) Pct honeycombing, 1.0 (1.7) Pct VRS, 5.1 (1.7)			<i>N</i>
Jacob, J. et al [38]	2016	Multidisciplinary diagnosis	Baseline	FEV1%, 70.8 (19.1) FVC %, 68.8 (20.5) DLCO%, 36.1 (12.9)	Baseline	Pct total extent of ILD, 26.5 (18.1) Pct ground glass opacity, 17.0 (14.7) Pct reticular pattern, 8.5 (6.0) Pct honeycombing, 1.0 (1.7) Pct VRS, 5.1 (1.7)			<i>N</i>
Maldonado, F. et al [41]	2014	2011 Guidelines	Time points 1 to 2	FVC %, -3.4 (8.2) FVC, -0.1 (0.3) FEV1%, -3.1 (8.5) FEV1, -0.1 (0.2) DLCO %, -7.0 (11.6) DLCO, -1.5 (2.6)	Time points 1 to 2	Pct total GGO, 3.2 (9.8) Pct total reticular, 3.0 (6.4) Pct total honeycombing, -0.1 (6.3) Pct total ILD, 6.1 (13.0)	Annual change	Pct ground glass opacity, 3.2 (9.8) Pc reticular pattern, 3.0 (6.4) Pct honeycombing, -0.1 (6.3)	<i>N</i>

Definition of abbreviations: CALIPER, Computer-Aided Lung Informatics for Pathology Evaluation and Rating; DLCO, diffusion lung capacity for carbon monoxide; FEV1, forced expiratory volume in 1 s; FVC, forced vital capacity; ILD, interstitial lung disease; *N*, not reported; PFT, pulmonary function test; SD, standard deviation; VRS, vessel-related structure

*: Data represents non-baseline level; %: percent predicted; Pct: percent CALIPER-derived parameters defined as (CALIPER ILD and VRS) extent volume in milliliters divided by total lung volume, then multiplied by 100

Time points 1 to 2: at least two serial HRCTs had been obtained within a 3–15-month interval, representing time points 1 and 2

Table 4 Prognostic value of CALIPER-derived parameters

Trial	Year	Prognostic indicator (CALIPER-derived parameters)	HR (95%CI)	Multivariate adjustments
Koo, C. W. et al [35]	2021	NA	NA	NA
Sverzellati, N. et al [36]	2020	Pct Total lung fibrosis extent Pct Ground glass opacity Pct Reticular pattern Pct Honeycombing VRS	Pct Total lung fibrosis extent, 1.034 (1.005–1.064) Pct Ground glass opacity, 1.022 (0.983–1.062) Pct Reticular pattern, 1.112 (1.045–1.182) Pct Honeycombing, 0.771 (0.487–1.221) Pct VRS, 1.643 (1.270–2.125)	GAP score and pack-years of smoking
Romei, C. et al [37]	2020	Survival analysis curve	Survival analysis curve	NA
Crews, M. S. et al [42]	2020	NA	NA	Age and sex
Jacob, J. et al [40]	2018	Pct Total lung fibrosis extent Pct Ground glass opacity Pct Reticular pattern Pct Honeycombing VRS	NA	Age and sex
Jacob, J. et al [29]	2018	Survival analysis curve	Survival analysis curve	Sex, age, and smoking status
Jacob, J. et al [39]	2018	NA	NA	NA
Jacob, J. et al [43]	2017	Pct Total ILD extent Pct Ground glass opacity Pct Reticular pattern Pct Honeycombing Pct VRS	Pct Total ILD extent, 1.03 (1.03–1.04) Pct Ground glass opacity, 1.03 (1.02–1.04) Pct Reticular pattern, 1.10 (1.07–1.12) Pct Honeycombing, 1.11 (1.03–1.20) Pct VRS, 1.52 (1.40–1.65)	NA
Jacob, J. et al [38]	2016	NA	NA	NA
Maldonado, F. et al [41]	2014	Pct Total Reticular Pct Total Honeycombing Pct Total ILD	Pct Total Reticular, 1.38 (1.03–1.84) Pct Total Honeycombing, 0.94 (0.74–1.20) Pct Total ILD, 1.30 (0.99–1.71)	Sex, pack-years, baseline FVC % pred and baseline DLCO % pred

Definition of abbreviations: *CALIPER*, Computer-Aided Lung Informatics for Pathology Evaluation and Rating; *CI*, confidence interval; *GAP*, Gender-Age-Physiology; *HR*, hazard ratio; *DLCO*, diffusion lung capacity for carbon monoxide; *FVC*, forced vital capacity; *ILD*, interstitial lung disease; *NA*, not assign; *SD*, standard deviation; *VRS*, vessel-related structure

Pct: percent CALIPER-derived parameters defined as (CALIPER ILD and VRS) extent volume in milliliters divided by total lung volume, then multiplied by 100

[29, 36, 37, 42, 43]. All five studies indicated that CALIPER-VRS was the strongest independent predictor of survival, and CALIPER-ILD also appeared to have great performance in predicting mortality. Of these studies, one was adjusted for GAP score and pack-years of smoking, one for age and sex, and one for sex, age, and smoking status. The third study analyzed 105 patients and found that 3-year survival was poor in patients with CALIPER-VRS% $\geq 5\%$ and CALIPER-ILD% $\geq 20\%$. However, in the eighth study analyzing 283 patients when the CALIPER-ILD range was combined with the CALIPER-VRS, only the CALIPER-VRS remained an independent predictor of mortality. Two studies analyzing 341 patients showed that CALIPER-R and CALIPER-H could also predict mortality [36, 43]. One study was adjusted for GAP score and pack-years of smoking.

CALIPER-derived parameters in predicting mortality in longitudinal follow-up

Three studies involving 231 subjects reported the correlation between CALIPER-derived parameters and mortality at an average follow-up time point of 1 or 2 years. The study of Sverzellati, N. et al and Jacob, J. et al indicated that the CALIPER-VRS had the best correlation with the long-term outcome [36, 40]. One study was adjusted for GAP score and pack-years of smoking. The long-term outcome was any all-cause death or lung transplantation more than 2 years from the baseline level. In addition, Maldonado, F. et al analyzed 55 patients and showed that CALIPER-R and CALIPER-ILD had a statistically significant association with the 1- or 2-year survival rate [41]. This analysis was adjusted for sex, pack-years, baseline FVC % pred, and baseline DLCO % pred. The study of Sverzellati, N. et al found that the amount

of change in CALIPER-H, VRS showed more strong correlation with prognosis during the 1-year follow-up period than the amount of change during the 2-year follow-up period [36].

Subgroup analysis of the CALIPER-VRS

Compared with PFT which can only evaluate the global lung, CALIPER can assess regional features. Four studies involving 998 subjects performed the segmental analysis [29, 39, 40, 43]. CALIPER-VRS could be divided into upper zone (UZ) VRS, middle zone (MZ) VRS, and lower zone (LZ) VRS according to the vessel location, and it could be divided into $< 5 \text{ mm}^2$, $5\text{--}10 \text{ mm}^2$, $10\text{--}15 \text{ mm}^2$, $15\text{--}20 \text{ mm}^2$, and $> 20 \text{ mm}^2$ according to the cross-sectional area. Jacob, J. analyzed 649 patients and found that the UZ VRS had a superior effect to others in predicting FVC reduction, or prognostic value [29, 40]. One study was adjusted for age and sex. Another study was adjusted for sex, age, and smoking status. The eighth study containing 283 patients analyzed that when the CALIPER-VRS $> 5 \text{ mm}^2$ range was combined with CALIPER-H, CPI, and GAP, only CALIPER-VRS $> 5 \text{ mm}^2$ was still an independent predictor of mortality [43]. In another study by Jacob, J. et al, only $10\text{--}15 \text{ mm}^2$ appeared independently correlated with the change in FVC [39].

Multivariate model in stratification outcome

Two studies including 341 patients reported the utility of CALIPER in refining the evaluation of outcome. In the study of Sverzellati, N. et al according to the ROC analysis, CALIPER-ILD $\geq 20\%$ and CALIPER-VRS $\geq 20\%$ were the optimal thresholds for stratification of results. Three risk categories were defined by combining FVC% reduction $\geq 10\%$ and relative rise of CALIPER-ILD $\geq 20\%$, at the 12–14-month follow-up. The risk categories were as follows: low risk meant that neither condition was included, intermediate risk meant that any one of these conditions was included, and high risk meant that both conditions were included. At 23–26 months of follow-up, the model integrating FVC% decline trend and relative rise of CALIPER-VRS $\geq 20\%$ generated similar effect [36].

Another study by Jacob, J. used CALIPER-VRS and CALIPER-H, which could independently predict mortality, to modify the formula of the composite physiologic index (CPI). The formula was presented as follows: CALIPER–CPI score = (CALIPER PVV $\times 23.0904$) + (CALIPER honeycombing $\times 18.3795$) + (CPI $\times 4.5065$) [43]. Cox regression analysis showed that the CALIPER–CPI not only had similar prognostic strength to the GAP Index staging system, but also showed better goodness of fit. Additionally, Cox regression analysis showed that the multivariate model

integrating CALIPER-H, CALIPER-VRS, and Gender-Age-Physiology (GAP) demonstrated improved performance compared to using GAP only [43].

Risk of bias in trials

The quality assessment included 10 studies and is shown in Supplement Table 2. Two articles did not mention the follow-up period, and 1 article did not give adequate interpretation for the follow-up period. All articles did not perform the blinding of outcome assessments or carry out the prospective calculation of the study size. The included studies had immense lost proportion. One study included a validation cohort. All studies collected data that were established before the beginning of the study, and all patients who met the inclusion criteria were included. The included studies carried out definite interpretation for endpoint indices.

Discussion

Our study provided a systematic review of the literature on the relationship between CALIPER-derived parameters and lung function and prognosis. In all nine studies, we found that CALIPER-derived parameters served as significant predictors of mortality and improved prediction of future adverse events (death or lung transplantation) by FVC, CPI, or GAP.

Our systematic review showed a strong correlation between CALIPER-derived parameters and FVC, as well as a significant value in predicting prognosis. The condition of prognosis and survival could be estimated by PFT at the baseline level and vertical change over time [44, 45]. Additionally, the vertical change in PFT maintained a superior effect to a single moment as well as CALIPER-derived parameters [46, 47]. Most of the studies in our collection, after finding a strong correlation between longitudinal changes in CALIPER-VRS, ILD, and lung function indicators, went on to develop multiparametric models or explore the role of this parameter in disease prognostic stratification through multivariate Cox regression analyses. Then, we found that the baseline level or amount of longitudinal change in CALIPER-derived parameters showed great performance in predicting prognosis. We find that the variation of CALIPER-derived parameters was different over a 1-year time interval. The two studies we included showed that CALIPER-GGO and R were the patterns with the highest annual growth [39, 41]. Given that most of the studies we included did not provide the annual growth of CALIPER-derived parameters and discuss their value in evaluating prognosis, the use of these two models could be further explored in future studies to find the value of

detecting subtle advances at shorter intervals for estimating survival times and potentially as markers of treatment response in future clinical trials. In cases where HRCT was routinely reviewed in IPF patients, CALIPER analysis did not require additional image acquisition time. CALIPER analysis was semiautomatic and independent of the severity of the patient's disease, with good accuracy and reproducibility [27]. CALIPER analysis therefore reduced to a certain extent the level of expertise required of the operator.

In the included studies, only two studies distinguish between patients with and without antifibrotic treatment, and they showed that patients with IPF who received treatment had slower CT progression compared to untreated patients. Given that such studies are still in the minority, in the future we should explore whether CALIPER-derived parameters are still able to assess prognosis after a period of antifibrotic treatment. And then, we should explore whether a stable or reduced CALIPER-derived parameter after antifibrotic treatment is associated with a good prognosis. Future studies should examine the increase in CALIPER-derived parameters before and after antifibrotic therapy, or the growth rate, and then explore the role in assisting clinical evaluation of drug efficacy.

The results of our analysis suggest the need for a staging system for the quantitative assessment of CALIPER-derived parameters, similar to the creation of standardized quantitative categories in the GAP index. The GAP index, a staging model based on clinical (e.g., sex, age) and physiologic (e.g., FVC, DLCO) parameters, classified patients into three phases, and different staging corresponded to different rates of patient mortality [48]. Currently, most subjects in clinical trials are in a mild to moderate stage (according to the GAP system), and these subjects might experience more noticeable reactions to antifibrotic treatment. The establishment of a staging system facilitated the identification of early stages of the disease and timely intervention. Early intervention with antifibrotic therapy can preserve pulmonary function to the greatest extent, decrease hospitalization risk, and prolong lifespan [12, 49]. Two of our included studies found CALIPER-ILD $\geq 20\%$ and CALIPER-VRS $\geq 20\%$ to be the optimal thresholds for outcome stratification, and the optimal cut-off value to distinguish between the presence and absence of FVC reduction for CALIPER-ILD was 20%, and for CALIPER-VRS was 5%. Due to the small size of the relevant cohort, the corresponding indicators should be confirmed in a larger cohort. A case–control study needs to be set up to further validate the accuracy of the thresholds and to refine the corresponding rating scale.

Our analysis suggests that the results should be stratified more precisely by a staging system established jointly with other parameters. Given that PFT was insensitive to less extensive disease [50], GAP was less powerful at predicting

future pulmonary function decline [51]. Combined with CPI-, GAP-, and FVC-related indicators, CALIPER quantification could provide patients and clinicians with a more accurate estimate of risk than other methods. However, these factors should be harmonized in future studies due to differences in endpoint events and the duration of follow-up.

Our analysis showed that the CALIPER-VRS was more valuable than other indicators in predicting adverse events. Five of our included studies used multivariate Cox regression analysis to conclude that CALIPER-VRS was the strongest independent predictor of mortality. CALIPER-VRS enhanced power in longitudinal variations for “silent” disease progression. One of our included studies discovered that the relative rise in CALIPER-VRS $\geq 20\%$ was the optimum threshold for outcome refined categorization and the change of VRS showed superior efficiency than FVC. Therefore, the subregional analysis of CALIPER-VRS was also more valuable. Given IPF was more severe in the inferior and dorsal regions of the histopathology, the abnormal signs always emerged in the basal and peripheral lung parenchyma on imaging [52]. Global analysis that integrates severe fibrotic areas with milder areas would conceal real disease progression. CALIPER-derived regional analysis could perform targeted analysis of different parts of the pulmonary system. By incorporating more precise and fine-grained location data into the multiparameter model, the accuracy and sensitivity of the staging system created would be further improved and would be more targeted. For example, two of our included studies used the method of adding the UZ VRS and CALIPER-VRS $> 5 \text{ mm}^2$, which are strongly correlated with adverse events, to the multiparameter model. Existing studies differ in their approach to CALIPER-VRS segmentation and in the creation of multiparameter models, and the harmonization of standards was necessary.

Our review has a few limitations. First, the study was limited by a small sample size. The indicators exhibited high heterogeneity and variability. Some studies provided insufficient data. Therefore, we could not conduct a meta-analysis. Second, most of the studies were retrospective, and only one study was prospective. Because the subjects were selected from a single center or two centers, we could not verify the diversity of subjects. Third, the imaging protocols were different among studies, such as reconstruction kernel and section thickness; these slight differences may have led to bias. Finally, the inclusion criteria were different for the included studies, such as the disease severity or follow-up time.

In conclusion, CALIPER-derived parameters appear to have a considerable effect on assessing prognosis, progression, and mortality, especially CALIPER-VRS. The CALIPER–CPI score formula and the multivariate model of CALIPER–H, CALIPER–VRS, and GAP could refine prognostic stratification, which would support clinicians in implementing targeted management strategies. CALIPER-VRS-derived

regional analysis could provide a more precise assessment of the disease condition and improve evaluation of the therapeutic effect. It is essential to establish the staging system by CALIPER-derived parameters and combining them with CPI, FVC, or GAP.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00330-023-10010-w>.

Funding This study has received funding by the National Key R&D program of China (2021YFC2500703), and Jilin Province Science and Technology Development Project (20210204100YY).

Declarations

Guarantor The scientific guarantor of this publication is Professor Sa Huang.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Informed consent Written informed consent was not required for this study because this study is a systematic review of literature.

Ethical approval Institutional Review Board approval was not required because this study is a systematic review of literature.

Methodology

- Systematic review

References

- Flaherty KR, Kolb M, Vancheri C, Tang W, Conoscenti CS, Richeldi L (2018) Stability or improvement in forced vital capacity with nintedanib in patients with idiopathic pulmonary fibrosis. *Eur Respir J* 52(2):1702593
- Nalysnyk L, Cid-Ruzafa J, Rotella P, Esser D (2012) Incidence and prevalence of idiopathic pulmonary fibrosis: review of the literature. *Eur Respir Rev* 21:355–361
- Raghu G, Remy-Jardin M, Richeldi L et al (2022) Idiopathic pulmonary fibrosis (an update) and progressive pulmonary fibrosis in adults: an Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *Am J Respir Crit Care Med* 205:e18–e47
- Hutchinson J, Fogarty A, Hubbard R, McKeever T (2015) Global incidence and mortality of idiopathic pulmonary fibrosis: a systematic review. *Eur Respir J* 46:795–806
- Rivera-Ortega P, Molina-Molina M (2019) Interstitial lung diseases in developing countries. *Ann Global Health* 85(1):4
- Wakwaya Y, Brown KK (2019) Idiopathic pulmonary fibrosis: epidemiology, diagnosis and outcomes. *Am J Med Sci* 357:359–369
- Salciccioli JD, Marshall DC, Goodall R et al (2022) Interstitial lung disease incidence and mortality in the UK and the European Union: an observational study, 2001–2017. *ERJ Open Res* 8(3):00058–2022
- Costabel U, Inoue Y, Richeldi L et al (2016) Efficacy of nintedanib in idiopathic pulmonary fibrosis across prespecified subgroups in INPULSIS. *Am J Respir Crit Care Med* 193:178–185
- Martusewicz-Boros M, Gorska K (2020) Nintedanib - efficacy, safety and practical aspects of treatment for patients with idiopathic pulmonary fibrosis. *Adv Respir Med* 88:599–607
- Aono Y, Nakamura Y, Kono M et al (2020) Prognostic significance of forced vital capacity decline prior to and following anti-fibrotic therapy in idiopathic pulmonary fibrosis. *Ther Adv Respir Dis* 14:1753466620953783.
- Durheim MT, Bendstrup E, Carlson L et al (2021) Outcomes of patients with advanced idiopathic pulmonary fibrosis treated with nintedanib or pirfenidone in a real-world multicentre cohort. *Respirology* 26:982–988
- Mooney J, Reddy SR, Chang E, Broder MS, Gokhale S, Corral M (2021) Antifibrotic therapies reduce mortality and hospitalization among Medicare beneficiaries with idiopathic pulmonary fibrosis. *J Manag Care Spec Pharm* 27:1724–1733
- Raghu G, Collard HR, Egan JJ et al (2011) 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
- Paterniti MO, Bi Y, Rekić D, Wang Y, Karimi-Shah BA, Chowdhury BA (2017) Acute exacerbation and decline in forced vital capacity are associated with increased mortality in idiopathic pulmonary fibrosis. *Ann Am Thorac Soc* 14:1395–1402
- Durheim MT, Collard HR, Roberts RS et al (2015) Association of hospital admission and forced vital capacity endpoints with survival in patients with idiopathic pulmonary fibrosis: analysis of a pooled cohort from three clinical trials. *Lancet Respir Med* 3:388–396
- Zurkova M, Kriegova E, Kolek V et al (2019) Effect of pirfenidone on lung function decline and survival: 5-yr experience from a real-life IPF cohort from the Czech EMPIRE registry. *Respir Res* 20(1):16
- Pellegrino R, Viegi G, Brusasco V et al (2005) Interpretative strategies for lung function tests. *Eur Respir J* 26:948–968
- Cottin V, Cordier J-F (2009) The syndrome of combined pulmonary fibrosis and emphysema. *Chest* 136(1):1–2
- Cottin V (2013) The impact of emphysema in pulmonary fibrosis. *Eur Respir Rev* 22:153–157
- Hansell DM, Goldin JG, King TE, Lynch DA, Richeldi L, Wells AU (2015) CT staging and monitoring of fibrotic interstitial lung diseases in clinical practice and treatment trials: a Position Paper from the Fleischner Society. *Lancet Respir Med* 3:483–496
- Behr J (2017) Disease progression in idiopathic pulmonary fibrosis FVC is not enough. *Am J Respir Crit Care Med* 196:1094–1095
- Yoon HY, Kim TH, Seo JB et al (2019) Effects of emphysema on physiological and prognostic characteristics of lung function in idiopathic pulmonary fibrosis. *Respirology* 24:55–62
- Schmidt SL, Nambiar AM, Tayob N et al (2011) Pulmonary function measures predict mortality differently in IPF versus combined pulmonary fibrosis and emphysema. *Eur Respir J* 38:176–183
- Watadani T, Sakai F, Johkoh T et al (2013) Interobserver variability in the CT assessment of honeycombing in the lungs. *Radiology* 266:936–944
- Widell J, Liden M (2020) Interobserver variability in high-resolution CT of the lungs. *Eur J Radiol Open* 7:100228
- Walsh SLF (2018) Imaging biomarkers and staging in IPF. *Curr Opin Pulm Med* 24:445–452
- Chen A, Karwoski RA, Gierada DS, Bartholmai BJ, Koo CW (2020) Quantitative CT analysis of diffuse lung disease. *Radiographics* 40:28–43
- Jankharia B, Angirish B (2021) Computer-aided quantitative analysis in interstitial lung diseases—a pictorial review using CALIPER. *Lung India* 38:161–167
- Jacob J, Bartholmai BJ, Rajagopalan S et al (2018) Predicting outcomes in idiopathic pulmonary fibrosis using automated computed tomographic analysis. *Am J Respir Crit Care Med* 198:767–776
- Rashed BM, Popescu N, Ieee (2021) Machine learning techniques for medical image processing. 9th IEEE international conference on e-health and bioengineering (EHB), Grigore T Popa Univ Med

- & Pharmacy, Electr Network. <https://doi.org/10.1109/EHB52898.2021.9657673>
31. Erickson BJ, Korfiatis P, Akkus Z, Kline TL (2017) Machine learning for medical imaging(1). *Radiographics* 37:505–515
 32. Bartholmai BJ, Raghunath S, Karwoski RA et al (2013) Quantitative computed tomography imaging of interstitial lung diseases. *J Thorac Imaging* 28:298–307
 33. Newell JD Jr, Tschirren J, Peterson S, Beinlich M, Sieren J (2019) Quantitative CT of interstitial lung disease. *Semin Roentgenol* 54:73–79
 34. Robbie H, Daccord C, Chua F, Devaraj A (2017) Evaluating disease severity in idiopathic pulmonary fibrosis. *Eur Respir J* 26(145):170051
 35. Koo CW, Larson NB, Parris-Skeete CT et al (2021) Prospective machine learning CT quantitative evaluation of idiopathic pulmonary fibrosis in patients undergoing anti-fibrotic treatment using low- and ultra-low-dose CT. *Clin Radiol*. <https://doi.org/10.1016/j.crad.2021.11.006>
 36. Sverzellati N, Silva M, Seletti V et al (2020) Stratification of long-term outcome in stable idiopathic pulmonary fibrosis by combining longitudinal computed tomography and forced vital capacity. *Eur Radiol* 30:2669–2679
 37. Romei C, Tavanti LM, Taliani A et al (2020) Automated computed tomography analysis in the assessment of idiopathic pulmonary fibrosis severity and progression. *Eur J Radiol* 124:108852
 38. Jacob J, Bartholmai BJ, Rajagopalan S et al (2016) Automated quantitative computed tomography versus visual computed tomography scoring in idiopathic pulmonary fibrosis: validation against pulmonary function. *J Thorac Imaging* 31:304–311
 39. Jacob J, Bartholmai BJ, Rajagopalan S et al (2018) Serial automated quantitative CT analysis in idiopathic pulmonary fibrosis: functional correlations and comparison with changes in visual CT scores. *Eur Radiol* 28:1318–1327
 40. Jacob J, Bartholmai BJ, Van Moorsel CH et al (2018) Longitudinal prediction of outcome in idiopathic pulmonary fibrosis using automated CT analysis. *Am J Respir Crit Care Med* 197:A5926
 41. Maldonado F, Moua T, Rajagopalan S et al (2014) Automated quantification of radiological patterns predicts survival in idiopathic pulmonary fibrosis. *Eur Respir J* 43:204–212
 42. Crews MS, Bartholmai BJ, Adegunsoye A et al (2020) Automated CT analysis of major forms of interstitial lung disease. *J Clin Med* 9(11):3776
 43. Jacob J, Bartholmai BJ, Rajagopalan S et al (2017) Mortality prediction in idiopathic pulmonary fibrosis: evaluation of computer-based CT analysis with conventional severity measures. *Eur Respir J* 49(1):1601011
 44. Mogulkoc N, Brutsche MH, Bishop PW et al (2001) Pulmonary function in idiopathic pulmonary fibrosis and referral for lung transplantation. *Am J Respir Crit Care Med* 164:103–108
 45. Li X, Peng S, Wei L, Li Z (2014) Relevance analysis of clinical and lung function parameters changing and prognosis of idiopathic pulmonary fibrosis. *Int J Clin Exp Med* 7:4759–4769
 46. Latsi PI, du Bois RM, Nicholson AG et al (2003) Fibrotic idiopathic interstitial pneumonia: the prognostic value of longitudinal functional trends. *Am J Respir Crit Care Med* 168:531–537
 47. Collard HR, King TE Jr, Bartelson BB, Vourlekis JS, Schwarz MI, Brown KK (2003) Changes in clinical and physiologic variables predict survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 168:538–542
 48. Ley B, Ryerson CJ, Vittinghoff E et al (2012) A multidimensional index and staging system for idiopathic pulmonary fibrosis. *Ann Intern Med* 156:684–U658
 49. Fisher M, Nathan SD, Hill C et al (2017) Predicting life expectancy for pirfenidone in idiopathic pulmonary fibrosis. *J Manag Care Spec Pharm* 23:S17–s24
 50. Procter AJ, Jacob J (2019) Visual vs. computer-based computed tomography analysis for the identification of functional patterns in interstitial lung diseases. *Curr Opin Pulm Med* 25:426–433
 51. Salisbury ML, Xia M, Zhou Y et al (2016) Idiopathic pulmonary fibrosis: Gender-Age-Physiology index stage for predicting future lung function decline. *Chest* 149:491–498
 52. Lynch DA, Sverzellati N, Travis WD et al (2018) Diagnostic criteria for idiopathic pulmonary fibrosis: a Fleischner Society White Paper. *Lancet Respir Med* 6:138–153

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.