BMJ Open

Minimum clinically important difference in Quantitative Lung Fibrosis score associated with all-cause mortality in idiopathic pulmonary fibrosis: subanalysis from two phase II trials of pamrevlumab

Grace Hyun Kim [ID],[1] Xueping Zhang,[2] Matthew S Brown,[1] Lona Poole,[2] Jonathan Goldin[1]

[1]University of California Los Angeles David Geffen School of Medicine, Los Angeles, California, USA
[2]FibroGen Inc, San Francisco, California, USA

Correspondence to
Dr Grace Hyun Kim;
GraceKim@mednet.ucla.edu

ABSTRACT

Objectives Idiopathic pulmonary fibrosis (IPF) is a progressive interstitial lung disease. Chest high-resolution CT (HRCT) is instrumental in IPF management, and the Quantitative Lung Fibrosis (QLF) score is a computer-assisted metric for quantifying lung disease using HRCT. This study aimed to assess the change in QLF score associated with a minimum clinically important difference (MCID) of IPF symptoms and physiological lung function, and also determine the MCID of QLF change associated with all-cause mortality to serve as an imaging biomarker to confirm disease progression and response to therapy.

Design and study setting We conducted post hoc analyses of prospective data from two IPF phase II studies of pamrevlumab, a fully human monoclonal antibody that binds to and inhibits connective tissue growth factor activity.

Participants Overall, 152 patients with follow-up visits after week 24.

Methods We used the anchor-based Jaeschke's method to estimate the MCID of the QLF score that corresponded with the already established MCID of St. George's Respiratory Questionnaire (SGRQ) and percent-predicted forced vital capacity (ppFVC). We also conducted a Cox regression analysis to establish a sensitive and robust MCID of the QLF score in predicting all-cause mortality.

Results QLF changes of 4.4% and 3.6% corresponded to the established MCID of a 5-point increase in SGRQ and a 3.4% reduction in ppFVC, respectively. QLF changes of 1% (HR=4.98, p=0.05), 2% (HR=4.04, p=0.041), 20 mL (HR=6.37, p=0.024) and 22 mL (HR=6.38, p=0.024) predicted mortality.

Conclusion A conservative metric of 2% can be used as the MCID of QLF for predicting all-cause mortality. This may be considered in IPF trials in which the degree of structural fibrosis assessed via HRCT is an endpoint. The MCID of SGRQ and FVC corresponds with a greater amount of QLF and may reflect that a greater amount of change in fibrosis is required before there is functional change.

Trial registration number NCT01262001, NCT01890265.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This study demonstrates the utilisation of an anchor-based approach and an early prediction of mortality in estimating a minimum clinically important difference (MCID).

⇒ This study estimates the MCID for extensive pulmonary fibrosis in idiopathic pulmonary fibrosis using high-resolution CT (HRCT) as an imaging biomarker based on two clinical trials, in which subjects underwent HRCT scans according to clinical protocols—reducing the potential bias compared with observational data.

⇒ The limitation of this study is that the MCID estimation was based on post hoc research of using the existing data from clinical trials.

INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is a rare, progressive interstitial lung disease that includes symptoms of cough, worsening of dyspnoea and progressive lung injury and scarring. Together, these symptoms limit physical activity and reduce patient health-related quality of life (HRQOL).[1–3] There is no cure for IPF,[3] and its prognosis is very poor. Median survival is estimated to be no more than 2–5 years after diagnosis.[4] Two approved antifibrotic drugs (pirfenidone and nintedanib) significantly reduce the rate of lung-function decline in IPF.[5–7] However, individual responses to treatment are variable and unpredictable, and HRQOL does not improve.[6 7] Identifying individual small, detectable and clinically meaningful changes in patient-level correlation will be beneficial for both physicians and patients in making informed decisions for available antifibrotic

treatments and for ongoing novel therapeutic discovery in clinical trials.

Pamrevlumab is a fully human monoclonal antibody that binds to and inhibits the activity of connective tissue growth factor.[8–10] Two phase II studies, one open-label and the other placebo-controlled intravenous administration of pamrevlumab, demonstrated slowing the rate of lung-function decline, progression of lung fibrosis evident on CT, and a trend towards improved HRQOL. Adverse events were generally mild.[9 10] However, a recent phase III trial of pamrevlumab for IPF (ZEPHYRUS-1) did not meet its primary endpoint of absolute change in forced vital capacity (FVC) from baseline to week 48.[11 12] Its companion study (ZEPHYRUS-2) was terminated.[13]

IPF treatment options are limited, and improved monitoring and a sensitive metric for assessing therapeutic efficacy are needed. Radiologically detected lung fibrosis correlates with physiological lung function and symptomatic changes in IPF, and early and sensitive imaging biomarkers are needed to confirm disease progression or worsening of FVC and response to therapy as quickly as possible to optimise drug development and patient care.[2 3 14–17]

FVC is the most common measure for assessing treatment efficacy in IPF.[18 19] HRQOL and other patient-reported outcome (PRO) measures are also important endpoints for evaluating disease progression and treatment efficacy including St. George's Respiratory Questionnaire (SGRQ).[3 20] Use of chest high-resolution CT (HRCT) is expanding and is instrumental in the diagnosis and management of IPF.[1] Computer-assisted methods for quantifying lung disease on HRCT calculate textural features derived from image data and classify different patterns of interstitial lung diseases based on machine learning algorithms.[21–23] Computational quantitative scoring systems that analyse HRCT images have been used as imaging biomarkers in IPF clinical trials to assess the degree and progression of structural lung fibrosis.

Of these, the Quantitative Lung Fibrosis (QLF) score has demonstrated high reproducibility[22–24] (figure 1). QLF is associated with more prospective validation than other quantitative CT techniques and has been used in recent clinical trials of IPF.[21 25] In two phase II IPF trials of pamrevlumab, significant correlations were observed between QLF changes and the changes in percent-predicted FVC (ppFVC) (ranging from −0.51 to −0.64), as well as with changes in PRO, SGRQ (ranging from 0.27 to 0.30).[9 10 26] Furthermore, QLF changes of <2% were associated with better long-term survival than changes ≥2% for patients with interstitial lung disease in scleroderma.[27]

The minimum clinically important difference (MCID) is an important standard for determining meaningful changes related to a clinical intervention or measurement tool[28] and represents the smallest detectable and beneficial change.[29] Both distribution-based methods (using variations from repeated measures) and anchor-based approaches (relying on established MCIDs from other relevant clinical variables) are used to determine MCIDs.[30] The MCID of QLF changes in an IPF cohort has not been evaluated. For clinically meaningful validation, an MCID of the QLF threshold should provide a tool for both identifying an effective treatment and detecting a difference in mortality over time. This is especially important in IPF, which is a progressive disease with a considerably shorter median survival than other chronic lung diseases.[31]

Our aim is to assess the change in QLF score associated with MCID of a PRO measure, SGRQ, and a key measure in lung function physiology, FVC, using the anchor-based Jaeschke's method, and to determine the MCID of QLF change based on its association with all-cause mortality through a post hoc analysis of prospective data from two phase II studies of pamrevlumab, exploring the potential of QLF as an imaging marker to confirm disease progression and response to therapy.
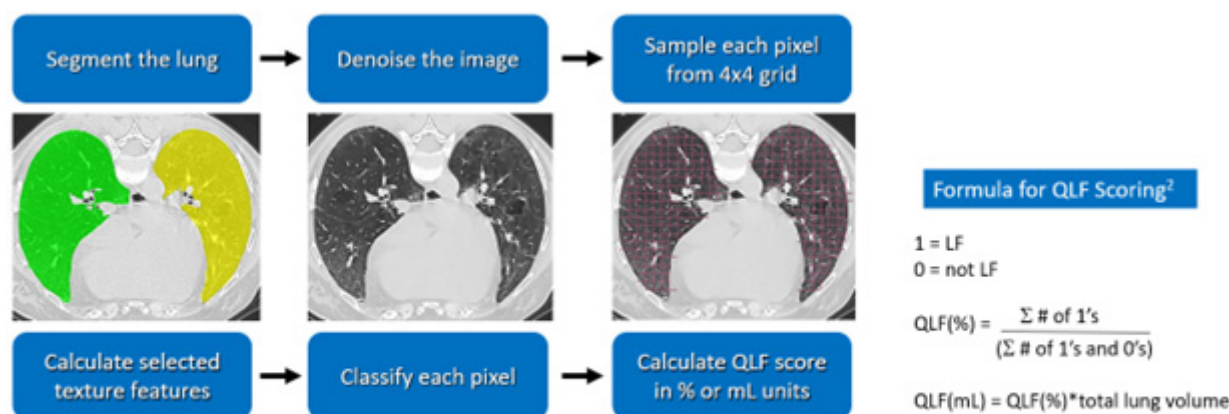

Figure 1 Use of HRCT to calculate QLF. QLF is a specific method (UCLA patent) that uses image normalisation (denoising) to minimise cross-site variability within images, resulting in decomposed CT images prior to texture calculation. HRCT, high-resolution CT; LF, lung fibrosis; QLF, Quantitative Lung Fibrosis; UCLA, University of California at Los Angeles.

METHODS
Patients
This was a secondary analysis of Study 049[9] and the phase II PRAISE study.[10] The two study populations were pooled to include a total of 190 patients with IPF. Eligibility criteria for the two studies were similar.[9 10] Study 049 (NCT01262001), conducted between March 2011 and December 2012 at 18 centres in the USA, was a single-arm, open-label study.[9] Pamrevlumab was administered every 3 weeks for 45 weeks: cohort 1 received 15 mg/kg and cohort 2 received 30 mg/kg.[9] PRAISE (Study 067 (NCT01890265)), conducted between August 2013 and July 2017 at 39 centres throughout North America, Australia, Africa and Europe, was a double-blind, placebo-controlled study.[10] Patients were randomised to receive placebo or pamrevlumab 30 mg/kg every 3 weeks for 45 weeks.[10]

Local ethics committees/institutional review boards (ECs/IRBs) approved the protocol for each site, and all patients provided written informed consent before enrolment (Study 049: Aspire IRB00004587; Study 067 (PRAISE): Quorum (now Advarra) 00023875; both studies WIRB (now WCG IRB) IRB00000533).

Of the 190 patients, 155 had follow-up visits after week 24 and data from week 48 visits, including the primary outcome measure of FVC. For both studies, pulmonary function tests, including spirometry, were performed at baseline and every 12 weeks thereafter, and HRCT was performed at baseline and every 24 weeks. SGRQ was completed at baseline and weeks 24 and 48. Mortality data were collected for the lengths of the respective studies.

Patient and public involvement
No patients or members of the public were involved in the design, conduct, reporting or dissemination plans of this study.

Outcomes
QLF scores were estimated from standardised non-contrast thin-section volumetric HRCT scans using an established radiomic texture-based quantification algorithm. QLF uses image normalisation (denoising) to minimise cross-site variability within images prior to texture calculation.[22] QLF was measured as extent (%) and volume (mL). Online supplemental figure 1 provides an example of QLF extent (%) and volume (mL) on HRCT and overlaid images for a patient with IPF. QLF measures the amount of reticulation with architectural distortion in the lung. Scores range from 0% to 100% for extent of fibrosis and from 0 mL to total lung capacity for volume of fibrosis. Greater scores represent increased fibrosis.[10 21] For this analysis, we considered 24-week or 48-week changes in QLF in the whole lung, which were calculated from the QLF scores of baseline HRCT.

Estimation of an MCID
We used the anchor-based Jaeschke's method with predefined criteria for establishing the MCID of QLF that corresponded with the established MCIDs of SGRQ and ppFVC. We used a landmark Cox proportional hazards regression analysis using all-cause mortality as an anchor by applying several thresholds of 24-week QLF changes. Patients did not have follow-up visits if they died or received a lung transplant.

The SGRQ is a self-administered questionnaire that assesses HRQOL in respiratory diseases. SGRQ total score ranges from 0 to 100, and greater scores indicate deterioration in HRQOL, and in this study, SGRQ was used to represent severity of symptoms. The MCID of the SGRQ was assumed to be ±5 points.[32] Changes in FVC are often used as primary endpoints in trials of respiratory diseases. The ppFVC is an estimate of lung function, with greater percentages indicating better function. The MCID of ppFVC was assumed to be ±3.4%.[33] In this study, SGRQ was used to represent the severity of symptoms and lung function, respectively.

Changes in longitudinal QLF scores were initially correlated with established MCID changes in SGRQ and ppFVC. The anchor-based Jaeschke's method was used to estimate the MCID of QLF scores from these changes in SGRQ and ppFVC from baseline at weeks 24 and 48. Jaeschke's method describes the mean change in the measurement of interest for patients who experience a change in an anchor.[31] Multiple anchors were chosen to obtain robust, unbiased estimates of the MCID.[34] Another anchor-based Cox proportional hazards regression was used for all-cause mortality, in which duration of survival or time to death was used as an anchor. A preliminary threshold was derived from a previous reproducibility study[35] and 6-month change observed in a clinical trial.[21] The reproducibility coefficient of QLF score was estimated to be approximately 0.4% ($\approx 2.77 \times 0.14 = \sqrt{2} \times 1.96 \times 0.14$),[35 36] and the mean of 6-month change was 0.98% for extent QLF and 21.7 mL for volume QLF from a nintedanib arm.[21] Thresholds increased incrementally as extent changes of 1%, 2%, 3% and 4% and volume changes of 20, 22, 24 and 26 mL.[21 35] Covariates of age and ppFVC at baseline were adjusted in the regression analysis. (Of note, the covariates of sex and percent-predicted diffusing capacity of lungs for carbon monoxide (ppDLCO) were not used in Cox regression due to the imbalanced distribution of sex in the multiple thresholds and the collinearity among ppDLCO, ppFVC and QLF.) Continuous scale and multiple thresholds of QLF scores were compared with test differences in mortality risk. In addition, the MCID from each anchor (SGRQ and ppFVC) was tested in a Cox regression model as a threshold.

Summary statistics are reported for demographics and clinical variables. Continuous variables are reported as mean and SD, and categorical variables are reported as frequencies and percentages.

RESULTS
There were no notable differences in demographics or baseline characteristics between the cohorts (table 1,

Table 1 Demographics and baseline clinical characteristics

Variable	Category	Combined group (Study 049+Study 067) (n=190)	Study 049 (n=89)	Study 067 (n=101)
Age, years	Mean (SD)	68.1 (7.06)	67.9 (7.04)	68.2 (7.11)
Sex, n (%)	Male	145 (76.3)	71 (79.8)	74 (73.3)
	Female	45 (23.7)	18 (20.2)	27 (26.7)
Race, n (%)	American Indian or Alaska Native	1 (0.5)	1 (1.1)	0
	Asian	8 (4.2)	0	8 (7.9)
	Black or African American	1 (0.5)	1 (1.1)	0
	White	172 (90.5)	87 (97.8)	85 (84.2)
	Other	8 (4.2)	0	8 (7.9)
Weight, kg	n	190	89	101
	Mean (SD)	86.40 (16.46)	89.50 (14.82)	83.67 (17.39)
Time since first IPF diagnosis, years	n	189	89	100
	Mean (SD)	1.60 (1.35)	1.91 (1.54)	1.32 (1.09)
Tobacco smoking status, n (%)	Current	3 (1.6)	0	3 (3.0)
	Former	60 (31.6)	0	60 (59.4)
	Never	38 (20.0)	0	38 (37.6)
	Missing	89 (46.8)	89 (100.0)	0
Lung function				
FVC, L	n	190	89	101
	Mean (SD)	2.79 (0.78)	2.73 (0.80)	2.84 (0.76)
FVC, % predicted	n	190	89	101
	Mean (SD)	70.44 (13.42)	66.79 (14.68)	73.66 (11.33)
DLCO, % predicted	n	190	89	101
	Mean (SD)	51.38 (13.72)	48.85 (13.32)	53.61 (13.74)
FEV_1/FVC ratio	n	190	89	101
	Mean (SD)	0.81 (0.06)	0.82 (0.06)	0.80 (0.06)
HRCT*				
Whole lung volume, L	n	185	89	96
	Mean (SD)	3.95 (1.01)	3.81 (0.94)	4.07 (1.05)
QLF, %	n	185	89	96
	Mean (SD)	16.8 (10.33)	19.6 (11.20)	14.3 (8.76)
QLF, mL	n	185	89	96
	Mean (SD)	622.74 (367.67)	701.13 (395.62)	550.07 (325.19)
GAP†				
GAP score	n	190	89	101
	Mean (SD)	3.8 (1.33)	4.1 (1.33)	3.6 (1.30)
Symptom score				
SGRQ	n	186	88	98
	Mean (SD)	44.04 (17.87)	46.24 (14.78)	42.05 (20.11)
UCSD-SOBQ	n	135	89	46
	Mean (SD)	33.10 (20.32)	32.13 (18.68)	34.98 (23.28)

*Computer-assisted QLF scores were derived from volumetric scans of the whole lung.
†Gender-Age-Pulmonary (GAP) function score. Median and ranges were reported in online supplemental table 1.
DLCO, diffusing capacity for carbon monoxide; FEV_1, forced expiratory volume in 1 second; FVC, forced vital capacity; GAP, gender, age, physiology; HRCT, high-resolution CT; IPF, idiopathic pulmonary fibrosis; QLF, Quantitative Lung Fibrosis score; SGRQ, St. George's Respiratory Questionnaire; UCSD-SOBQ, University of California San Diego, Shortness of Breath Questionnaire.

Table 2 Relationship of QLF extent (%) and volume (mL) with the anchor-based MCID of SGRQ and FVC using Jaeschke's method

Anchor		Adjusted corresponding QLF values of thresholds compared with stable condition*			
		Week 24		**Week 48**	
		SGRQ change			
SGRQ	Whole lung	Improved† (N=48)	Worsened† (N=41)	Improved† (N=41)	Worsened† (N=42)
	QLF, %	≤0.5	≥4.4‡	≤ −2.8‡	≥2.3
	95% CI	(−1.5, 2.5)	(2.3, 6.5)	(−5.4, −0.2)	(−0.3, 4.9)
	QLF, mL	≤10	≥91‡	≤ −54	≥81‡
	95% CI	(−38, 58)	(41, 141)	(−118, 9)	(18, 144)
		ppFVC change			
		Improved† (N=20)	Worsened† (N=49)	Improved† (N=20)	Worsened† (N=53)
ppFVC	QLF, %	≤−2.0	≥3.6‡	≤−2.1	≥4.3‡
	95% CI	(−4.5, 0.6)	(1.7, 5.4)	(−5.1, 0.8)	(2.1, 6.4)
	QLF, mL	≤−56	≥65‡	≤−65	≥108‡
	95% CI	(−118, 7)	(20, 110)	(−143, 13)	(51, 164)

*Shifted by the mean changes at ±5 points SGRQ and ±3.4% ppFVC; stated with two significant figures.
†Worsened: SGRQ change ≥5 points or ppFVC change ≤3.4%; stable: SGRQ change between −5 and 5 points or ppFVC change between −3.4% and 3.4%; improved: SGRQ change ≤−5 points or ppFVC change ≥3.4%.
‡Two 95% CIs of the difference between the improved and stable groups and the difference between worse and stable groups are not overwrapped. 95% CIs were obtained from a one-way ANOVA model in the model. Comparisons are obtained from the same model using the least squares means (LSMEANS) statement.
ANOVA, analysis of variance; FVC, forced vital capacity; MCID, minimum clinically important difference; ppFVC, percent-predicted forced vital capacity; QLF, Quantitative Lung Fibrosis score; SGRQ, St. George's Respiratory Questionnaire.

online supplemental table 1). The median (±IQR) length of the follow-up period was 337 (±504) days. Because changes in QLF outcomes for all-cause mortality were derived from week 24 data and the screening HRCT scan, the median observed survival was relatively short. In total, 185 available screening HRCT scans, 33 patients had no available survival analyses after week 24 because they discontinued prior to week 24 visits (n=19), or they died prior to week 24 (n=13) or did not undergo scan (n=1) (online supplemental figure 2).

Anchor-based analyses assessed the relationship between QLF and the established MCIDs for SGRQ, ppFVC (table 2) or both (online supplemental table 2). Thresholds of QLF changes (extent and volume) at week 24 were 4.4% and 91 mL for symptomatic worsening, respectively, when applying Jaeschke's method using SGRQ as an anchor, and 3.6% and 65 mL for worsening lung function, respectively, when using ppFVC as an anchor. For improved condition of symptom by SGRQ or lung function by ppFVC, the thresholds of QLF changes were ≤0.5% and 10 mL and −2.0% and −56 mL, respectively. At week 48 for worsened condition, the thresholds of QLF changes were 2.3% and 81 mL by SGRQ, and 4.3% and 108 mL by ppFVC, respectively, and for improved condition, the QLF changes were −2.8% and −54 mL by SGRQ, and −2.1% and −65 mL by ppFVC, respectively.

Agreement between the two components of symptoms and lung function and their corresponding changes in QLF score are reported in online supplemental table 2.

Similar percentages of patients experienced concordance in changes at weeks 24 and 48. If changes in SGRQ and FVC were both considered worsened, the mean (SD) changes in extent of QLF were 8.1% (8.27) at week 24 and 7.9% (7.37) at week 48. If changes were both considered stable, the mean (SD) changes in QLF were 0.5% (3.75) at week 24 and 2.5% (4.31) at week 48. If changes were both considered improved, the mean (SD) changes in QLF were −1.6% (3.71) at week 24 and −3.4% (6.50) at week 48. Overall, the concordant changes for worsening, stability and improvement followed expected directional changes in QLF scores.

IPF is a complex disease that includes dynamic changes in lung symptoms, function and structure over time. These changes do not always progress at the same rate, and most patients experienced discordance in symptoms and lung function. The discordant changes when one parameter worsened were associated with the mean QLF changes of approximately 2%; discordant changes with one parameter improved were associated with QLF changes of approximately within ±1% (online supplemental table 2).

The relationship of QLF changes to each of the MCIDs of SGRQ and ppFVC is provided in online supplemental table 3. Unadjusted analyses demonstrated that mean changes in MCIDs were greater (eg, 4.88%, 95% CI (3.46%, 6.30%) vs 4.39%, 95% CI (2.97%, 5.81%) when applying SGRQ as an anchor at week 24) compared with the stable group (eg, 0.49%, 95% CI (−0.31%, 2.32%) vs

Table 3 Cox proportional hazards model using all-cause mortality as an anchor with QLF with cut-off of week 24 changes

QLF changes in whole lung	Total patients (N=152)	Total deaths (n=11) (%)	HR (95% CI)	P value
Continuous*, %			1.20 (1.08, 1.34)	0.001
Continuous*, mL			1.01 (1.01, 1.02)	<0.001
Week 24 QLF changes in whole lung	**Total patients meeting listed threshold, n (%)**	**Total deaths meeting listed threshold, n (%)**	**HR (95% CI)**	**P value**
≥0%	118 (78)	9 (82)	1.48 (0.31, 7.02)	0.623
≥0 mL	99 (65)	9 (82)	2.57 (0.55, 12.1)	0.232
Thresholds QLF, %				
≥1%	101 (66)	9 (82)	4.98 (1.00, 24.8)	0.050
≥2%	87 (57)	8 (73)	4.04 (1.06, 15.4)	0.041
≥3%	66 (43)	6 (55)	4.69 (1.24, 17.7)	0.023
≥4%	50 (33)	5 (45)	4.52 (1.18, 17.3)	0.028
QLF volume				
≥20 mL	84 (55)	8 (73)	6.37 (1.27, 32.0)	0.024
≥22 mL	83 (55)	8 (73)	6.38 (1.27, 32.0)	0.024
≥24 mL	81 (53)	8 (73)	6.83 (1.37, 34.2)	0.019
≥26 mL	80 (53)	8 (73)	7.63 (1.51, 38.6)	0.014

Note: Mean change from previously modelled measurement variation of QLF of 0.4% (2.77×2×SD).
*The proportional hazards assumption is being examined and met using Schoenfield residuals.
QLF, Quantitative Lung Fibrosis score.

0.00%, 95% CI (−1.15%, 1.15%) when applying SGRQ as an anchor) and trended in the correct direction. The effect sizes of QLF changes (both extent and volume) were approximately 1 or slightly greater for patient groups with worsened SGRQ scores and worsened ppFVC (ie, 1.08 for extent and 1.03 for volume when applying SGRQ as an anchor, 1.02 for extent and 0.93 for volume when applying ppFVC as an anchor, respectively), indicating a strong relationship between QLF and the anchor parameter. The effect sizes for worsened conditions were greater than those observed for patients with not-worsened conditions (ie, 0.93, 0.33, −0.17 for worsening, stable and better in the effect size of QLF volume change, respectively, when applying ppFVC as an anchor).

A Cox proportional hazards regression model of QLF changes is presented in table 3. Changes ranging from 1% to 4% and from 20 to 26 mL were associated with statistically significant differences in all-cause mortality. A minimum threshold of a 1% change in QLF at week 24 was associated with an increased risk of death. A twofold to fivefold increased risk of death was observed for patients with sizeable changes in QLF (ie, changes greater than the thresholds established by the single-variable anchor-based analyses). Online supplemental figure 3A,B presents patients with IPF with QLF changes of 1% (25 mL) and 2% (46 mL), respectively, at week 24. We used an MCID derived from the increase in SGRQ and the decrease in ppFVC, which indicates worsening in IPF, from table 2. After applying these thresholds, estimated MCID from QLF changes at week 24, using all-cause mortality, Cox proportional hazards model revealed a fourfold to ninefold increased risk of death for patients with sizeable changes in QLF at week 24 (table 4). Significant differences were observed from 1% to 4% changes (20–26 mL for QLF volume) (table 3), whereas 4.4% or 3.6% changes (or 91 or 65 mL) for volume changes were derived from the changes of QLF corresponding to the anchors of SGRQ and ppFVC, respectively (table 4).

DISCUSSION

This study established the MCID for change in QLF score in the setting of IPF as it relates to all-cause mortality. A minimum threshold of change in QLF of 1% or 20 mL at week 24 was associated with an increased risk of death for patients with IPF. To include changes observed with SGRQ and FVC, a conservative estimate of 2% can be adopted as the MCID of QLF, based on the week 24 mean QLF changes when patients experienced worsening of either SGRQ or ppFVC. An increased risk of death was also associated with sizeable QLF changes using Jaeschke's method with anchors of IPF symptoms (SGRQ) and lung function (ppFVC) to determine MCID.

Changes in QLF were consistent with changes in SGRQ and ppFVC, and mean QLF changes coincided with both symptom and lung-function changes. Changes in QLF and SGRQ were positively correlated, and changes in QLF and ppFVC were inversely correlated.[9 10 26] This indicates that, generally, a responder in IPF clinical trial was associated with a reduction in QLF, a reduction in SGRQ

Table 4 Anchor-based Cox proportional hazards model with an MCID derived from SGRQ and ppFVC as anchors

	Total (N=155)	Total deaths (n=11)	Risk of death, HR (95% CI)	P value
QLF changes in whole lung by SGRQ using ±5 as anchor				
≥4.4%	27 (18%)	4 (36%)	6.17 (1.63, 23.4)	0.007
≥91 mL	47 (31%)	6 (55%)	7.35 (1.98, 27.3)	0.003
QLF changes in whole lung by ppFVC with ±3.4% as anchor				
≥3.6%	41 (27%)	4 (36%)	4.33 (1.14, 16.5)	0.031
≥65 mL	60 (39%)	7 (64%)	8.89 (2.13, 37.0)	0.003

Note: The anchor-based values, marked with bold fonts, are results from table 2.
. MCID, minimum clinically important difference; ppFVC, percent-predicted forced vital capacity; QLF, Quantitative Lung Fibrosis score; SGRQ, St. George's Respiratory Questionnaire.

and an increase in ppFVC. The threshold of QLF change for improved symptoms was close to zero (ie, 0.5%) at week 24 compared with the reduction (−2.0%) in MCID for improved lung function, but at week 48, further reduced in both symptomatic changes (ie, −2.8%) and functional changes (−2.1%) (see table 2 for the details). In contrast, the threshold of QLF change for worsened symptoms had a greater magnitude at week 24 (ie, 4.4%) than the threshold for worsened lung function (ie, 3.6%), but at week 48, functional changes (ie, 4.3%) were greater than symptomatic changes (ie, 2.3%) (see table 2 for the details). This suggests that symptomatic changes (measured by SGRQ) improve slower or more inconsistently than functional changes (FVC) for improvement but were sensitive in worsening faster. This is likely a result, in part, of the effects of limited symptom recall and the subjective nature of HRQOL.

Quantitative HRCT tools for measuring pulmonary fibrosis are critical in therapeutic development in ILD to confirm efficacy or evaluate the safety of an experimental drug.[18 21 25] Assessments of QLF change have mostly served as secondary or exploratory quantitative imaging outcomes to estimate changes in lung fibrosis in clinical trials.[11 13 25] QLF score based on HRCT images is traceable and can visualise regions of fibrosis. The incremental changes in QLF extent (%) or volume (mL) highlight the structural worsening in IPF that is associated with decreased FVC.[10] The incremental worsening in ppFVC from week 24 and week 48 confirms that FVC is a reasonably reliable assessment in IPF and supports its use as the primary efficacy endpoint in clinical trials.[19] A role of quantitative tools in the future can be expanded in patient care using a digital AI platform, when a trial is approved with a positive outcome from a primary endpoint or a secondary endpoint of an imaging outcome.

In this post hoc analysis of prospective IPF clinical trial data, an anchor-based method was used to estimate the threshold of QLF change associated with established MCIDs of SGRQ and ppFVC, and a sensitivity-based method was used to establish the MCID of QLF from all-cause mortality. Sensitivity-based methods for estimating MCID ideally require a baseline variable or characteristic to reliably quantify disease severity. For HRCT-based

QLF, this requires repeat HRCT scans in a coffee break type (ie, approximately 15 min)[37 38] of experiment for patients with IPF, but this method poses ethical challenges because of the unnecessary risk of radiation exposure. The variability can be estimated statistically,[25] but we recognise this as a limitation of the analysis. Because repeated HRCT scans were not usually available, an anchor-based approach, which relies on the variability of anchored measurements, was selected in this study. Anchor-based methods require measurements of longitudinal change for the tool of interest and other anchor measurements that already have an established MCID. The estimates of MCIDs from other anchored measurements are likely overestimated because of the nature of additive variabilities. MCID estimates are approximations, and the recommended approach is to use multiple anchors to define a range of MCID estimates.[39] Both function and PROs are relevant when interpreting the outcomes of clinical trials.

The MCIDs of the relationships between QLF and SGRQ (PRO) or FVC (lung function) in this study were less sensitive than the MCID of QLF using all-cause mortality as an anchor. The degree of QLF change needed to attain a meaningful, absolute change was 0.5–4.4% for SGRQ as an anchor, 2.0–4.3% for FVC as an anchor and −1.6%–8.1% for concordance of SGRQ and FVC in non-stable change. This suggests that, for a meaningful change in PROs and function derived from Jaeschke's method, a greater amount of QLF change is needed in the evaluation of QLF with SGRQ or FVC than when using all-cause mortality as an anchor with multiple QLF thresholds. This reflects the progressive nature of IPF, which contrasts with other chronic lung diseases, and the fact that the observations from the follow-up visits are based on stable or worsening disease. This is similar to findings by Kon et al, who reported that the MCID of an assessment tool for chronic obstructive pulmonary disease from a receiver operating characteristic analysis was smaller than the MCID from an anchor-based approach.[40] Further, multivariable regression modelling, such as the anchor-based Jaeschke's method, that combine both clinical and subjective (eg, PRO) parameters to quantify changes in the outcome of interest (eg, QLF) offer less-biased estimates of MCIDs

than distribution-based methods, which only assess the statistical significance of a change.[41]

The QLF changes of 1% and 2% at week 24 presented in online supplemental figure 3A,B, respectively, correspond to 25 and 46 mL changes. The structural changes are visualised both in the right side (online supplemental figure 3A) and left side of the lung (online supplemental figure 3B). These QLF changes are similar to the absolute mean changes of 2% and 56 mL (table 2) when FVC was used as an anchor in our analysis. When retesting the MCID of the QLF score derived from the MCID using SGRQ and ppFVC as anchors, the HR ranged from 4.33 to 8.89, which is similar to the HRs of 4.98 and 4.04 for the relatively smaller MCIDs of 1% and 2%, respectively. The QLF scores associated with the MCIDs of SGRQ and ppFVC were greater than the QLF score associated with survival, which suggests that these biomarkers require greater structural disease progression before they can detect meaningful change. This also suggests that QLF may be more sensitive than either ppFVC or SGRQ as a trial endpoint.[42]

There were 11 deaths in this study that occurred after week 24. Because QLF changes were derived from week 24 and the screening HRCT scans, the median survivals were relatively short. Week 48 data were omitted from the survival analysis because most changes in QLF were observed within 24 weeks. In addition, including data beyond week 24 had the potential to skew the results by including patients with mild to moderate IPF who were more likely to remain alive at 1 year. QLF change as a volume is a suitable clinical trial endpoint, as noted by the high HR in table 4. Change in QLF extent can provide a normalised measurement regardless of the volume differences between patients of different sex or height. Finally, the MCID of QLF is an early biomarker of change in lung fibrosis, so 48 weeks of data, which is often used for the primary endpoints in IPF clinical trials, are not needed to determine the value of the MCID of the QLF score and its clinical applicability for predicting early change.

The multiple thresholds of QLF changes can be found in other studies. In an independent cohort of patients who received nintedanib (n=42), an antifibrotic approved drug, mean absolute changes in QLF were 0.98% and a 21.7 mL increase at month 6, and 1.4% and a 27.6 mL increase at month 12. In the placebo arm of the same trial, the changes were 1.33% and 37.3 mL at month 6 and 2.2% and 67.0 mL at month 12. Negative correlations were observed between change in QLF score and change in FVC at month 6, supporting the findings of the QLF score.[21] In a retrospective analysis of approximately 200 patients with IPF, a 4% change in QLF score for the most severe lobe and for the whole lung at 6 months was associated with a threefold to fivefold increased risk of clinical progression.[43] Further, a placebo-controlled phase II trial of 137 patients with IPF reported significant correlations between QLF and ppFVC changes, as well as other symptoms of IPF, where most subjects in the placebo arm were within ±2% changes at week 24.[25] Overall, QLF measured on HRCT, where the most of 6-month mean changes range from 1% to 4%, has proven to be useful as an efficacy endpoint in clinical trial settings.

This study has several limitations. First, it was a post hoc rather than an a priori analysis of data from two phase II clinical trials. Thus, due to the nature of phase II studies, mortality was based on a short follow-up period. In addition, the allocation of treatment arms and study locations was different between the studies. Specifically, Study 049 was a single-arm study, whereas Study 067 was a randomised study with one-to-one allocation of placebo and pamrevlumab. Further, Study 049 took place only in the USA, and Study 067 involved patients worldwide. Subcohorts of patients who received pamrevlumab or other treatments were not analysed separately here for purposes of simplification. Additionally, this study analysed the usefulness of QLF change for predicting mortality risk only over a short period of time. Second, we used MCID derived from a distribution-based approach for ppFVC and the symmetric changes in ±5 points SGRQ, which is close to 4.9 changes reported by Prior *et al* for deterioration, where most IPF subjects feel worsening or stable in their symptoms. We used ±5 where the subject-level change of SGRQ is an integer change, and MCID of SGRQ is considered to be around 4–5 points.[32 44 45] Third, we did not estimate MCID using a distribution-based approach because the extra radiation exposure required for patients to estimate the MCID was not well-justified. Fourth, a single quantitative HRCT score for IPF was applied. The estimated MCID may not be generalisable to other available quantitative scores. Fifth, caution should be applied when applying the estimated MCID for the observational or registry studies, in which HRCT scans are not performed routinely. In this study, HRCT scans were scheduled and performed as part of clinical trials. Lastly, phase III studies did not show the efficacy of pamrevlumab.[11–13] This study used survival as the primary endpoint to assess MCID.

We believe our analyses begin the evidence-generation process of using multiple thresholds for validation of a biomarker.[46] The MCID of QLF in IPF has demonstrated clinical validity. The estimated MCID of 2% may be considered for associating changes in mortality, lung function and patient symptoms in ongoing and future trials of IPF, where the metric can be normalised to the volume of QLF changes for both sexes. The greater MCID of QLF using the MCID of SGRQ and ppFVC may suggest that structural changes precede functional changes. The change of QLF volume is a sensitive measurement that can be considered in applying an imaging outcome as a potential efficacy endpoint when the extent of structural fibrosis is assessed via HRCT.

ORCID iD
Grace Hyun Kim http://orcid.org/0000-0003-1225-3489

REFERENCES

1. Raghu G, Remy-Jardin M, Richeldi L, *et al*. 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* 2022;205:e18–47.
2. Richeldi L, Costabel U, Selman M, *et al*. Efficacy of a tyrosine kinase inhibitor in idiopathic pulmonary fibrosis. *N Engl J Med* 2011;365:1079–87.
3. Swigris JJ, Esser D, Conoscenti CS, *et al*. The psychometric properties of the St George's Respiratory Questionnaire (SGRQ) in patients with idiopathic pulmonary fibrosis: a literature review. *Health Qual Life Outcomes* 2014;12:124.
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* 2011;183:788–824.
5. Behr J, Prasse A, Wirtz H, *et al*. Survival and course of lung function in the presence or absence of antifibrotic treatment in patients with idiopathic pulmonary fibrosis: long-term results of the INSIGHTS-IPF registry. *Eur Respir J* 2020;56:1902279.
6. Richeldi L, du Bois RM, Raghu G, *et al*. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med* 2014;370:2071–82.
7. King TE Jr, Bradford WZ, Castro-Bernardini S, *et al*. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 2014;370:2083–92.
8. Saito S, Deskin B, Rehan M, *et al*. Novel mediators of idiopathic pulmonary fibrosis. *Clin Sci* 2022;136:1229–40.
9. Raghu G, Scholand MB, de Andrade J, *et al*. FG-3019 anti-connective tissue growth factor monoclonal antibody: results of an open-label clinical trial in idiopathic pulmonary fibrosis. *Eur Respir J* 2016;47:1481–91.
10. Richeldi L, Fernández Pérez ER, Costabel U, *et al*. Pamrevlumab, an anti-connective tissue growth factor therapy, for idiopathic pulmonary fibrosis (PRAISE): a phase 2, randomised, double-blind, placebo-controlled trial. *Lancet Respir Med* 2020;8:25–33.
11. ZEPHYRUS-1: evaluation of efficacy and safety of pamrevlumab in patients with idiopathic pulmonary fibrosis, Available: https://clinicaltrials.gov/ct2/show/NCT03955146
12. Raghu G, Richeldi L, Fernández Pérez ER, *et al*. Pamrevlumab for idiopathic pulmonary fibrosis: The ZEPHYRUS-1 randomized clinical trial. *JAMA* 2024;332:380–9.
13. ZEPHYRUS-2: Efficacy and safety study of pamrevlumab in participants with idiopathic pulmonary fibrosis (IPF), Available: https://www.clinicaltrials.gov/ct2/show/NCT04419558
14. Zisman DA, Schwarz M, Anstrom KJ, *et al*. A controlled trial of sildenafil in advanced idiopathic pulmonary fibrosis. *N Engl J Med* 2010;363:620–8.
15. 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. *The Lancet* 2009;374:222–8.
16. Peng S, Li Z, Kang J, *et al*. Cross-sectional and longitudinal construct validity of the Saint George's respiratory questionnaire in patients with IPF. *Respirology* 2008;13:871–9.
17. Raghu G, Behr J, Brown KK, *et al*. Treatment of idiopathic pulmonary fibrosis with ambrisentan: a parallel, randomized trial. *Ann Intern Med* 2013;158:641–9.
18. Raghu G, Ghazipura M, Fleming TR, *et al*. Meaningful endpoints for idiopathic pulmonary fibrosis (ipf) clinical trials: emphasis on 'Feels, functions, survives. Report of a Collaborative Discussion in a Symposium with Direct Engagement from Representatives of Patients, Investigators, the National Institutes of Health, a Patient Advocacy Organization, and a Regulatory Agency.647–69.
19. Wells AU. Forced vital capacity as a primary end point in idiopathic pulmonary fibrosis treatment trials: making a silk purse from a sow's ear. *Thorax* 2013;68:309–10.
20. Nolan CM, Birring SS, Maddocks M, *et al*. King's brief interstitial lung disease questionnaire: responsiveness and minimum clinically important difference. *Eur Respir J* 2019;54:1900281.
21. Lancaster L, Goldin J, Trampisch M, *et al*. Effects of nintedanib on quantitative lung fibrosis score in idiopathic pulmonary fibrosis. *Open Respir Med J* 2020;14:22–31.
22. Kim HG, Tashkin DP, Clements PJ, *et al*. A computer-aided diagnosis system for quantitative scoring of extent of lung fibrosis in scleroderma patients. *Clin Exp Rheumatol* 2010;28:S26–35.
23. Wu X, Kim GH, Salisbury ML, *et al*. Computed tomographic biomarkers in idiopathic pulmonary fibrosis. the future of quantitative analysis. *Am J Respir Crit Care Med* 2019;199:12–21.
24. Kim HJ, Brown MS, Chong D, *et al*. Comparison of the quantitative CT imaging biomarkers of idiopathic pulmonary fibrosis at baseline

and early change with an interval of 7 months. *Acad Radiol* 2015;22:70–80.

25 Kim GHJ, Goldin JG, Hayes W, *et al*. The value of imaging and clinical outcomes in a phase II clinical trial of a lysophosphatidic acid receptor antagonist in idiopathic pulmonary fibrosis. *Ther Adv Respir Dis* 2021;15:17534666211004238.

26 Kouchakji E, Goldin J, Kim G, *et al*. PRAISE ipf study correlation between quantitative lung fibrosis (qlf) measured by quantitative high resolution computed tomography (qhrct) and quality of life (qol) assessed by st. george's respiratory questionnaire (sgrq). ERS International Congress 2019 abstracts; September 28, 2019:A1321.

27 Volkmann ER, Tashkin DP, Roth MD, *et al*. Early radiographic progression of scleroderma: lung disease predicts long-term mortality. *Chest* 2022;161:1310–9.

28 Cook CE. Clinimetrics corner: The Minimal Clinically Important Change Score (MCID): a necessary pretense. *J Man Manip Ther* 2008;16:E82–3.

29 Copay AG, Subach BR, Glassman SD, *et al*. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J* 2007;7:541–6.

30 Hays RD, Woolley JM. The concept of clinically meaningful difference in health-related quality-of-life research. How meaningful is it? *Pharmacoeconomics* 2000;18:419–23.

31 Kang M, Marts L, Kempker JA, *et al*. Minimal clinically important difference in idiopathic pulmonary fibrosis. *Breathe (Sheff)* 2021;17:200345.

32 Prior TS, Hoyer N, Hilberg O, *et al*. Responsiveness and minimal clinically important difference of SGRQ-I and K-BILD in idiopathic pulmonary fibrosis. *Respir Res* 2020;21:91.

33 du Bois RM, Weycker D, Albera C, *et al*. Forced vital capacity in patients with idiopathic pulmonary fibrosis: test properties and minimal clinically important difference. *Am J Respir Crit Care Med* 2011;184:1382–9.

34 Kafaja S, Clements PJ, Wilhalme H, *et al*. Reliability and minimal clinically important differences of FVC. *Am J Respir Crit Care Med* 2018;197:644–52.

35 Kim HJ, Brown MS, Elashoff R, *et al*. Quantitative texture-based assessment of one-year changes in fibrotic reticular patterns on HRCT in scleroderma lung disease treated with oral cyclophosphamide. *Eur Radiol* 2011;21:2455–65.

36 Obuchowski NA, Reeves AP, Huang EP, *et al*. Quantitative imaging biomarkers: a review of statistical methods for computer algorithm comparisons. *Stat Methods Med Res* 2015;24:68–106.

37 Ye S, Lim JY, Huang W. Statistical considerations for repeatability and reproducibility of quantitative imaging biomarkers. *BJR Open* 2022;4:20210083.

38 Oxnard GR, Zhao B, Sima CS, *et al*. Variability of lung tumor measurements on repeat computed tomography scans taken within 15 minutes. *J Clin Oncol* 2011;29:3114–9.

39 Revicki D, Hays RD, Cella D, *et al*. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008;61:102–9.

40 Kon SSC, Canavan JL, Jones SE, *et al*. Minimum clinically important difference for the COPD assessment test: a prospective analysis. *Lancet Respir Med* 2014;2:195–203.

41 Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. *J Clin Epidemiol* 2017;82:128–36.

42 Goldin J, Kim G, Lu P, *et al*. QLF change is more sensitive than fvc change in predicting progression free survival (pfs) in ipf trials. ERS International Congress 2017 abstracts; September 2017:A1958.

43 Kim GHJ, Weigt SS, Belperio JA, *et al*. Prediction of idiopathic pulmonary fibrosis progression using early quantitative changes on CT imaging for a short term of clinical 18-24-month follow-ups. *Eur Radiol* 2020;30:726–34.

44 Jones PW, Quirk FH, Baveystock CM. The St George's respiratory questionnaire. *Respir Med* 1991;85 Suppl B:25–31.

45 Jones PW. St. George's Respiratory Questionnaire: MCID. *COPD: J Chronic Obstructive Pulmonary Disease* 2005;2:75–9.

46 Food and Drug Administration, National Institutes of Health. BEST (Biomarkers, EndpointS, and other Tools) resource, Available: https://www.ncbi.nlm.nih.gov/books/NBK326791/?report=reader