

The value of imaging and clinical outcomes in a phase II clinical trial of a lysophosphatidic acid receptor antagonist in idiopathic pulmonary fibrosis

Grace Hyun J. Kim , Jonathan G. Goldin, Wendy Hayes, Andrea Oh, Benjamin Soule and Shuyan Du

Ther Adv Respir Dis

2021, Vol. 15: 1–11

DOI: 10.1177/
17534666211004238

© The Author(s), 2021.

Article reuse guidelines:
sagepub.com/journals-
permissions

Abstract

Background: Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive fibrotic lung disease characterized by worsening dyspnea and lung function and has a median survival of 2–3 years. Forced vital capacity (FVC) is the primary endpoint used most commonly in IPF clinical trials as it is the best surrogate for mortality. This study assessed quantitative scores from high-resolution computed tomography (HRCT) developed by machine learning as a secondary efficacy endpoint in a 26-week phase II study of BMS-986020 – an LPA₁ receptor antagonist – in patients with IPF.

Methods: HRCT scans from 96% (137/142) of randomized subjects were utilized. Quantitative lung fibrosis (QLF) scores were calculated from the HRCT images. QLF improvement was defined as $\geq 2\%$ reduction in QLF score from baseline to week 26.

Results: In the placebo arm, 5% of patients demonstrated an improvement in QLF score at week 26 compared with 15% and 27% of patients in the BMS-986020 600 mg once daily (QD) and twice daily (BID) arms, respectively [*versus* placebo: $p=0.08$ (600 mg QD); $p=0.0098$ (600 mg BID)]. Significant correlations were found between changes in QLF and changes in percent predicted FVC, diffusing capacity for carbon monoxide (DLCO), and shortness of breath at week 26 ($\rho=-0.41$, $\rho=-0.22$, and $\rho=0.27$, respectively; all $p < 0.01$).

Conclusions: This study demonstrated the utility of quantitative HRCT as an efficacy endpoint for IPF in a double-blind, placebo-controlled clinical trial setting.

The reviews of this paper are available via the supplemental material section.

Keywords: disease progression, interstitial, lung diseases, therapeutics, tomography

Received: 13 February 2021; revised manuscript accepted: 22 February 2021.

Introduction

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive fibrotic lung disease characterized by worsening dyspnea and lung function and has a median survival of 2–3 years from the time of diagnosis.^{1,2} Two anti-fibrotic drugs (pirfenidone and nintedanib) are approved for the treatment of IPF as each demonstrated significant impact on the slope of FVC decline. Although these therapies slow the rate of disease progression, they are often poorly tolerated,^{3,4} and many patients progress despite treatment. Given the limited options for this life-threatening disease,

there is a need for improved IPF therapies and measurable, reliable biomarkers to better monitor disease progression and the response to therapeutic intervention.⁵

Forced vital capacity (FVC) has been the primary endpoint used most commonly in IPF clinical trials.^{6,7} Additional measures of treatment efficacy include the patient reported outcome of dyspnea, measured by the California San Diego Shortness of Breath Questionnaire (UCSD-SOBQ),⁸ and diffusing capacity for carbon monoxide (DLCO) – a physiologic measure of oxygen uptake.^{1,9,10}

Correspondence to:

Shuyan Du
Bristol Myers Squibb,
Princeton, NJ, USA
shuyan.du@bms.com

Grace Hyun J. Kim
Department of
Radiological Sciences,
David-Geffen School of
Medicine, and Department
of Biostatistics, Fielding
School of Public Health,
University of California,
Los Angeles, CA, USA

Jonathan G. Goldin
Department of
Radiological Sciences,
David-Geffen School of
Medicine, University of
California, Los Angeles,
CA, USA

Wendy Hayes
Bristol Myers Squibb,
Princeton, NJ, USA

Andrea Oh
Department of Radiology,
National Jewish Health,
Denver, CO, USA

Benjamin Soule
Bristol Myers Squibb,
Princeton, NJ, USA

In addition, high resolution computed tomography (HRCT) plays a central role in the diagnosis and management of IPF.¹ The extent of radiologically detected lung fibrosis correlates with physiologic measures of lung function and can act as an important predictor of prognosis.¹¹ However, studies supporting the role of HRCT in evaluating fibrotic lung disease have used semi-quantitative visual scoring methods, which can be limited by interobserver variability.^{11–13} Computer-based quantitative HRCT evaluation has the potential for greater precision than visual scoring in the estimation of IPF extent and additionally allows for an assessment of changes that is more reproducible and objective.^{14,15} The quantitative lung fibrosis (QLF) score, an estimation of the extent of reticular patterns on a computed tomography (CT) scan using a machine learning algorithm, correlates with treatment outcomes and changes in physiological measurements in patients with IPF and scleroderma-related interstitial lung disease.^{16–18} Many computer-based algorithms have been developed with HRCT for IPF diagnosis and fibrosis quantification using machine learning and, more recently, deep learning, both of which are part of artificial intelligence (AI).^{15,19–21} However, few algorithms have been applied in a prospective, blinded clinical study. To enable robust AI development, it is critical to lock the algorithm, implement it into a system, then apply it in prospective studies.

The purpose of this study was to evaluate the utility of computer-based quantitative HRCT as an efficacy endpoint for assessing the change of extent of lung disease. Our approach was to examine the relationship between HRCT QLF score and select clinical and physiological variables of disease severity using data from a prospective, randomized, double-blind, placebo-controlled study of 26-week treatment with BMS-986020 – an anti-fibrotic drug candidate – in patients with IPF. This study demonstrated a statistically significant slowdown of the rate of FVC decline in patients with IPF treated with BMS-986020 600 mg twice daily (BID) compared with placebo.²²

Methods

Study design

NCT01766817 was a multicenter, randomized, double-blinded, placebo-controlled, phase II trial to examine the efficacy and safety of BMS-986020, a lysophosphatidic acid receptor 1 (LPA₁)

antagonist, (Bristol Myers Squibb, Princeton, NJ, USA) in patients with IPF.²² Eligible patients were blinded and randomized to receive BMS-986020 600 mg once daily (QD), BMS-986020 600 mg BID, or placebo orally for 26 weeks.

Patients

Eligible patients were adults between 40 and 90 years of age who had been diagnosed with usual interstitial pulmonary fibrosis (UIP) or IPF by the central multi-disciplinary team *via* HRCT or surgical lung biopsy (SLB) collected less than 6 years before randomization.

Eligible patients had symptoms consistent with IPF, greater extent of fibrotic changes (i.e., honeycombing and reticular changes) than emphysema on HRCT scan, post-bronchodilator percent predicted FVC between 45% and 90%, DLCO between 30% and 80%, no evidence of improvement in measures of IPF disease severity over the preceding year, ability to walk 150 m or more, and demonstrated an exertional decrease in oxygen saturation of 2% or greater.

Key exclusion criteria included (1) significant clinical worsening of IPF during the 30 day screening period, (2) forced expiratory volume in 1 second (FEV₁)/FVC ratio less than 0.8 after administration of a bronchodilator, and (3) absolute increase of 12% or greater and increase of 200 ml in FEV₁, FVC, or both after bronchodilator use.

Study endpoints

The primary endpoint was the rate of change in FVC (liters) over 26 weeks. Secondary and exploratory efficacy endpoints included change in QLF score on HRCT, change in percent predicted FVC, change in 6-minute walk test (6MWT) distance, change in DLCO, and change in patient reported dyspnea (UCSD-SOBQ) at week 26 compared with baseline. QLF score improvement from baseline was assessed by a predetermined smallest detectable change (SDC) value of 2%, which is the product of 0.72% and 2.77 ($= 2 \times \sqrt{2}$).¹⁸

Automated quantitative CT image analysis

All HRCT images were acquired while patients were in the prone position. Out of 142 randomized patients, HRCT scans from 137 patients (46/47 of

BMS-986020 600mg QD arm; 46/48 of BMS-986020 600mg BID arm, and 45/47 of placebo arm) were volumetric series and utilized in the QLF analysis. QLF scores were calculated using a machine learning technique and automated using five steps after the visually confirmed lung segmentation. The five steps were (a) denoise, (b) grid-sampling, (c) calculation of selected important texture features, (d) classification with support vector machine, and (e) production of a ratio of the classified fibrotic reticulation to the total grid sample in a percent scale.^{16,17} Of these 137 patients, an additional 7 were excluded due to HRCT scans with motion artifacts, inconsistent positioning, and/or improper Digital Imaging and Communications in Medicine (DICOM) formatting. The motion artifacts and inconsistent positioning in prone and supine images were confirmed by a central radiologist ($N=6$). Improper DICOM formatting was checked by standard technical quality control in the imaging analyses ($N=1$), where DICOM formatting is a necessary condition in quantification of CT images.

Of the 130 patients, 124 had paired HRCT images at baseline and at week 26. A thoracic radiologist reviewed the 124 sets of paired HRCT images and excluded an additional 6 sets with ≥ 700 cc in change of total lung capacity (TLC) volume and motion artifacts between baseline and week 26 scans.^{23,24} Thus, a total of 118 patients had paired HRCT scans ($n=39$ for BMS-986020 600mg QD; $n=41$ for BMS-986020 600mg BID; $n=38$ for placebo) that were utilized in this study to assess change in QLF between baseline and week 26.

As previously reported,^{25,26} an automated lung fibrosis classification system was used to compute the percentage of CT pixels representing QLF, quantitative honeycomb cysts (QHC), and quantitative ground-glass (QGG) in the whole lung. Quantitative interstitial lung disease (QILD) score was defined as the sum of QLF, QGG, and QHC. QLF (mL) was derived from the product of CT TLC volume and QLF (%); QILD (mL) was derived from the product of CT TLC volume and QILD (%).

Statistical analysis

Baseline characteristics were summarized by treatment arm. Changes from baseline to week 26 in QLF assessments for the treatment arms were compared with those for the placebo arm using t

tests or Wilcoxon rank tests depending upon meeting the normality assumption. A waterfall plot was used to show QLF changes at week 26 by treatment arm. The two-sample test of proportions was applied to measure the differences in QLF scores in each treatment arm compared with the placebo arm. The Holm–Bonferroni method was used to compare the proportion of responders in the three study arms by accounting for the family-wise error rate of 5% for multiple comparisons.²⁷ As part of the *post hoc* analyses, Spearman rank correlations were used to test for associations between pulmonary function (percent predicted FVC and DLCO), dyspnea (UCSD-SOBQ), and QLF scores at baseline and their change at week 26. All statistical analyses were performed using STATA V.14.0 (StataCorp LLC, College Station, TX, USA).

Results

Baseline characteristics

There were no statistically significant differences in the percent predicted FVC, DLCO, UCSD-SOBQ, or QLF scores across treatment arms at baseline (Table 1). Of note, mean QILD scores in both percentage scale and volume (mL) were numerically higher in the BMS-986020 600mg QD arm compared with the other two arms.

Treatment changes from baseline at week 26

Results of key clinical endpoints at week 26 are summarized in Table 2. Mean percent predicted FVC decreased by 5.78% in the placebo arm compared with 2.08% ($p=0.08$ versus placebo) and 2.71% ($p=0.03$ versus placebo) in the BMS-986020 600mg QD and BMS-986020 600mg BID arms, respectively.

Though no statistically significant differences were observed in the mean changes in DLCO and UCSD-SOBQ scores in the BMS-986020 treatment arms compared with placebo, a numeric increase of DLCO and a decrease in UCSD-SOBQ score were observed in the BMS-986020 600mg BID arm compared with the BMS-986020 600mg QD and placebo arms. There were no significant differences in the mean change in QLF and QILD scores in the BMS-986020 treatment arms compared with placebo, though there was a numerically smaller increase of QLF and QILD in the BMS-986020 600mg BID arm compared with the BMS-986020 600mg QD and placebo arms.

Table 1. Baseline characteristics of patients from study [ClinicalTrials.gov identifier: NCT01766817].

Parameter	BMS-986020 600 mg QD (n=46)	BMS-986020 600 mg BID (n=46)	Placebo (n=45)	Total (N=137)
Age, mean (SD), years	68.52 (6.65)	69.17 (8.36)	69.33 (7.56)	69.00 (7.55)
Sex, male, n (%)	32 (70)	34 (74)	33 (73)	99 (72)
Predicted FVC, mean (SD), %	69.95 (12.28)	68.59 (12.32)	69.22 (11.27)	69.25 (11.89)
Predicted DLCO, mean (SD), %	45.69 (16.92)	40.23 (12.62)	44.33 (12.77)	43.41 (14.34)
UCSD-SOBQ score, mean (SD)	45.76 (23.51)	42.52 (23.64)	37.42 (23.35)	41.93 (23.58)
QLF whole lung, mean (SD), %	19.00 (9.71)	17.90 (8.03)	17.36 (7.08)	18.09 (8.32)
QILD whole lung, mean (SD), %	40.54 (13.19)	37.82 (11.55)	37.07 (10.24)	38.48 (11.74)
QLF whole lung, mean (SD), mL ^a	703.18 (393.36)	623.93 (246.91)	622.91 (262.79)	650.21 (308.45)
QILD whole lung, mean (SD), mL ^b	1,491.37 (560.44)	1,349.32 (439.83)	1,348.52 (455.14)	1,396.75 (489.40)

^aQLF (mL) was derived from the product of CT TLC volume and QLF (%).
^bQILD (mL) was derived from the product of CT TLC volume and QILD (%).
 BID, twice daily; DLCO, diffusing capacity of the lungs for carbon monoxide; FVC, forced vital capacity; QD, once daily; QILD, quantitative interstitial lung disease; QLF, quantitative lung fibrosis; SD, standard deviation; TLC, total lung capacity; UCSD-SOBQ, University of California San Diego Shortness of Breath Questionnaire.

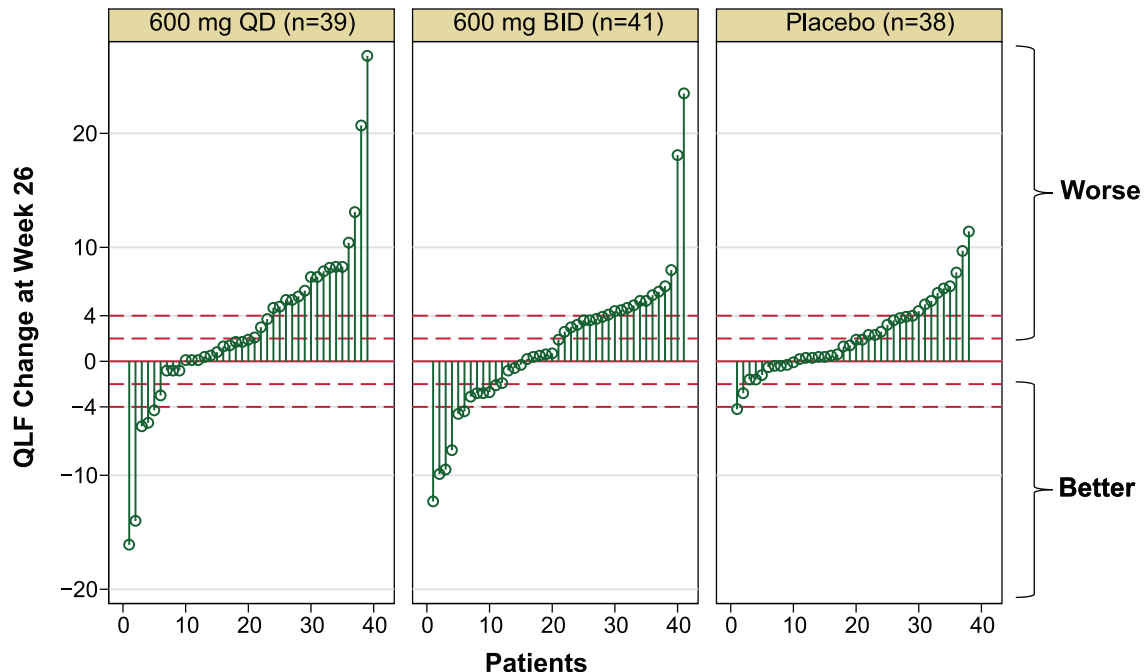
Table 2. Week 26 changes in clinical measurements and quantitative HRCT findings from baseline.

Parameter, mean (SD)	BMS-986020 600 mg QD (n=39)	BMS-986020 600 mg BID (n=41)	Placebo (n=38)
Predicted FVC, %	-2.08 (8.16)	-2.71 (9.81)	-5.78 (8.20)
Predicted DLCO, %	-4.34 (9.94)	2.90 (34.33)	-3.74 (18.57)
UCSD-SOBQ score	4.54 (17.24)	-2.54 (21.45)	3.63 (18.67)
QLF whole lung, % ^a	2.26 (6.62)	1.60 (6.49)	2.22 (3.37)
QILD whole lung, % ^b	2.77 (8.54)	1.85 (8.65)	2.77 (6.67)
QLF whole lung, mL ^a	57.99 (189.74)	40.35 (160.34)	59.11 (96.82)
QILD whole lung, mL ^b	73.46 (213.68)	31.41 (228.11)	55.66 (186.22)

n=38, *n*=41, and *n*=38 for DLCO; *n*=40, *n*=43, and *n*=40 for UCSD-SOBQ; for BMS-986020 600 mg QD, BMS-986020 600 mg BID, and placebo, respectively.
^aQLF (mL) was derived from the product of CT TLC volume and QLF (%).
^bQILD (mL) was derived from the product of CT TLC volume and QILD (%).
 BID, twice daily; DLCO, diffusing capacity of the lungs for carbon monoxide; FVC, forced vital capacity; HRCT, high-resolution computed tomography; QD, once daily; QILD, quantitative interstitial lung disease; QLF, quantitative lung fibrosis; SD, standard deviation; TLC, total lung capacity; UCSD-SOBQ, University of California San Diego Shortness of Breath Questionnaire.

Changes in QLF scores from baseline are illustrated in Figure 1 for each of the three treatment arms. The predetermined SDC value was 2%. In the placebo arm, 5% of patients demonstrated a QLF score

response, while 18% ($p=0.08$ versus placebo) and 27% ($p=0.0098$ versus placebo) of patients showed a QLF response in the BMS-986020 600 mg QD and BMS-986020 600 mg BID arms, respectively.



600 mg QD (n=39)	600 mg BID (n=41)	Placebo (n=38)
(Better, Same, Worse) at $\pm 2\%$	(Better, Same, Worse) at $\pm 2\%$	(Better, Same, Worse) at $\pm 2\%$
N: (6, 14, 19) \rightarrow (15%, 36%, 49%)	N: (11, 10, 20) \rightarrow (27%, 24%, 49%)	N: (2, 19, 17) \rightarrow (5%, 50%, 45%)
A trend and significant differences in improvement in QD and BID doses compared with placebo ($p=0.08$ and $p=0.0098$, respectively)		

Figure 1. Drop line plot of individual QLF scores at week 26 for the two BMS-986020 treatment groups and placebo group. N=number of patients within each of the three arms that had better, same, or worse QLF scores compared with baseline. Of note, two patients had a reduction of 5% in the BMS-986020 600 mg QD arm, thus only six drop lines can be seen, instead of seven. QD, once daily; QLF, quantitative lung fibrosis.

Table 3. Correlation between clinical measurements and quantitative HRCT findings at baseline.

Parameter, rho (p value)	% predicted FVC	% predicted DLCO ^a	UCSD-SOBQ total score
Predicted DLCO ^a , %	0.2528 ($p=0.0029$)		
UCSD-SOBQ score	-0.0543 ($p=0.53$)	-0.0385 ($p=0.65$)	
QLF whole lung, %	-0.3656 ($p<0.0001$)	-0.3816 ($p<0.0001$)	0.1049 ($p=0.29$)
QILD whole lung, %	-0.3328 ($p=0.0001$)	-0.2625 ($p=0.0019$)	0.09297 ($p=0.25$)
QLF whole lung, mL	-0.2699 ($p=0.0014$)	-0.3025 ($p=0.0003$)	0.0537 ($p=0.53$)
QILD whole lung, mL	-0.1737 ($p=0.0423$)	-0.1254 ($p=0.14$)	0.0583 ($p=0.50$)

^aQILD=QLF + QGG + QHC; QLF (mL) was derived from the product of CT TLC volume and QLF (%); Similarly, QILD (mL) was derived from the product of CT TLC volume and QILD (%).

DLCO, diffusing capacity of the lungs for carbon monoxide; FVC, forced vital capacity; HRCT, high-resolution computed tomography; QILD, quantitative interstitial lung disease; QLF, quantitative lung fibrosis; SD, standard deviation; TLC, total lung capacity; UCSD-SOBQ, University of California San Diego Shortness of Breath Questionnaire.

Table 4. Correlation between change in clinical measurements and HRCT at week 26.

Parameter, rho (<i>p</i> value)	Changes in % predicted FVC	Change in % predicted DLCO ^a	Change in UCSD-SOBQ total score
Changes in % predicted DLCO ^a	0.3659 (<i>p</i> =0.0001)		
Changes in UCSD-SOBQ total score	-0.1772 (<i>p</i> =0.0549)	-0.1217 (<i>p</i> =0.19)	
Changes in QLF whole lung, %	-0.4142 (<i>p</i> <0.0001)	-0.2166 (<i>p</i> =0.0196)	0.2680 (<i>p</i> =0.0034)
Changes in QILD whole lung, % ^b	-0.3775 (<i>p</i> <0.0001)	-0.2136 (<i>p</i> =0.0213)	0.1972 (<i>p</i> =0.0323)
Changes in QLF whole lung, mL	-0.3540 (<i>p</i> =0.0001)	-0.2022 (<i>p</i> =0.0295)	0.2723 (<i>p</i> =0.0029)
Changes in QILD whole lung, mL ^b	-0.2700 (<i>p</i> =0.0031)	-0.2069 (<i>p</i> =0.0258)	0.1491 (<i>p</i> =0.1072)

^a*n*=116 for DLCO.

^bQILD=(QLF + QGG + QHC).

QLF (mL) was derived from the product of CT TLC volume and QLF (%); Similarly, QILD (mL) was derived from the product of CT TLC volume and QILD (%). DLCO, diffusing capacity of the lungs for carbon monoxide; FVC, forced vital capacity; HRCT, high-resolution computed tomography; QILD, quantitative interstitial lung disease; QLF, quantitative lung fibrosis; UCSD-SOBQ, University of California San Diego Shortness of Breath Questionnaire.

Association of quantitative HRCT findings with clinical/physiologic measurements at baseline

A statistically significant inverse correlation was found between the quantitative HRCT lung fibrosis and ILD radiological scores (QLF and QILD) and pulmonary function measurements (percent predicted FVC and DLCO) at baseline (Table 3). Higher reticular or ILD patterns on HRCT indicated poorer lung function as indicated by pulmonary function measurements. No statistically significant correlations were found between the UCSD-SOBQ dyspnea score and pulmonary function measurements (FVC and DLCO) or quantitative lung fibrosis scores (QLF and QILD).

Correlations of quantitative HRCT findings with clinical/physiologic measurements of change at week 26

Statistically significant inverse correlations were found between changes in quantitative lung fibrosis scores (QLF and QILD) and pulmonary function measurements (percent predicted FVC and DLCO) at week 26 (Table 4). A statistically significant correlation was also observed between change in quantitative lung fibrosis scores (QLF and QILD) and the UCSD-SOBQ scores. The correlation between the change in UCSD-SOBQ scores and pulmonary function measurements (percent predicted FVC and DLCO) did not reach statistical significance (*p*=0.0549).

Representative HRCT images from baseline and week 26 from a patient in the BMS-986020 600 mg QD arm (Figure 2) show the associated changes in QLF, FVC, and UCSD-SOBQ scores and demonstrate overall improvement over time (Figure 2). In comparison, representative HRCT images from a patient in the placebo arm show worsening QLF, FVC, and UCSD-SOBQ scores (Figure 3).

Discussion

Automated quantitative analysis of HRCT images from a prospective clinical trial demonstrated that QLF and QILD can objectively measure changes in the extent of lung fibrosis and ILD radiological score as potential efficacy endpoints. The demonstration of an inverse correlation between QLF changes and pulmonary function tests (percent predicted FVC and DLCO) was consistent with other published studies.^{14,16,17,28–32} To our knowledge, this is the first prospective clinical study for which association of quantitative HRCT measures with changes in dyspnea (UCSD-SOBQ) has been demonstrated in patients with IPF.

Quantitative HRCT scores generated using machine learning have been previously developed and implemented as a potential tool for a clinical management or an outcome.¹⁵ Multiple clinical studies have demonstrated the utility of quantitative HRCT scores in IPF in retrospective and prospective trial settings.^{15–17,29–33} Retrospective analysis

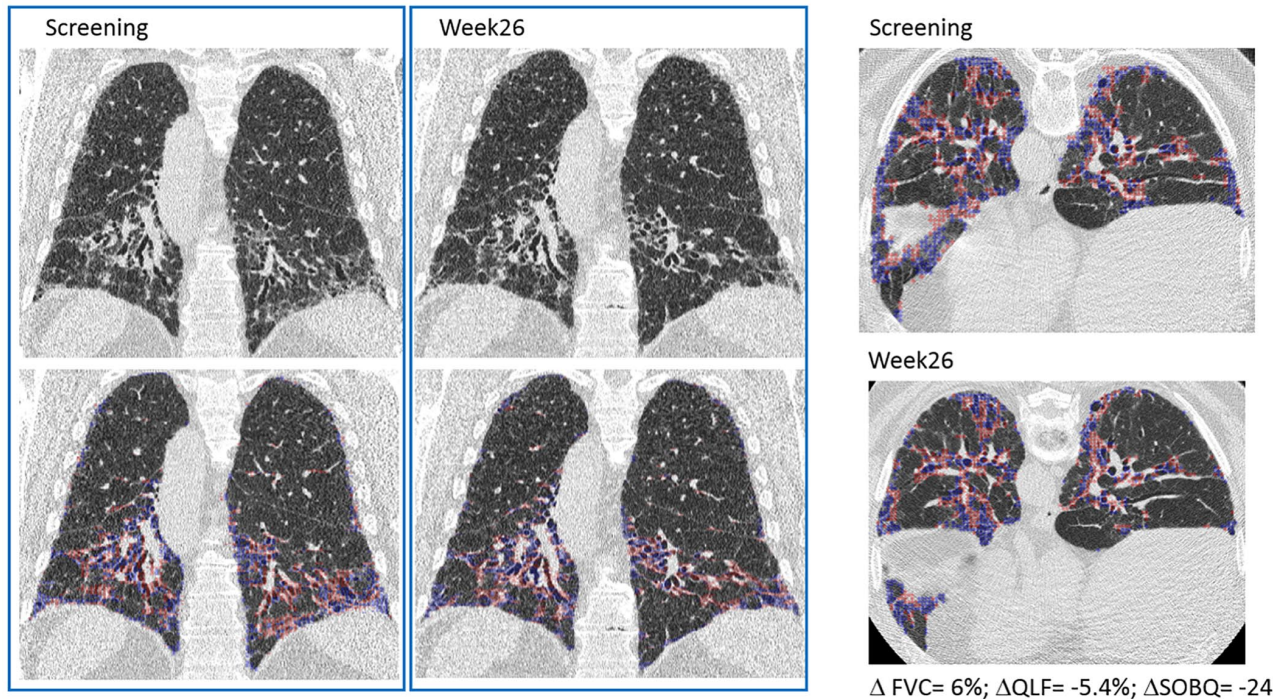


Figure 2. Representative coronal and axial HRCT images from a patient in the BMS-986020 QD arm at screening and week 26. Change over 26 weeks: FVC: 6%; QLF score: -5.4%, UCSD-SOBQ: -24. Classification overlay for QLF score in blue and red. FVC, forced vital capacity; HRCT, high resolution computed tomography; QD, once daily; QLF, quantitative lung fibrosis; UCSD-SOBQ, University of California San Diego Shortness of Breath Questionnaire.

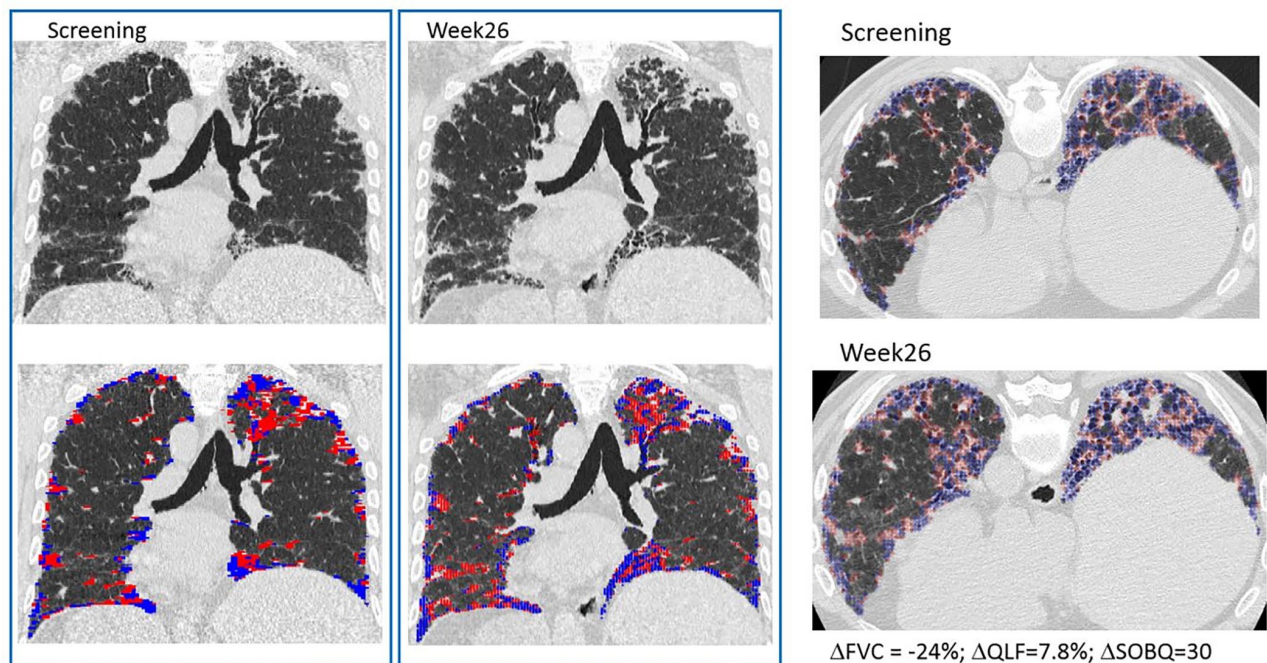


Figure 3. Representative coronal and axial HRCT images from a patient in the placebo arm at screening and week 26. Change over 26 weeks: FVC: -24%; QLF score: 7.8%; UCSD-SOBQ: 30. Classification overlay for QLF score is in blue and red. FVC, forced vital capacity; HRCT, high resolution computed tomography; QLF, quantitative lung fibrosis; UCSD-SOBQ, University of California San Diego Shortness of Breath Questionnaire.

showed that pirfenidone-treated patients with IPF had significantly lower increases in fibrotic reticulation patterns (F-patterns) on HRCT compared with age-matched controls, indicative of a treatment effect of pirfenidone in lung fibrosis.²⁹ An open-label study of an anti-connective tissue growth factor antibody treatment reported that mean change in HRCT QLF at month 6 is predictive of FVC change at month 12.¹⁶ HRCT vessel-related structure extent at baseline was reported as the strongest predictor of survival and FVC decline in observational studies.³⁰ A review of multiple methods of computer-based quantitative HRCT provides insight into the use of quantitative HRCT as a potential surrogate endpoint in clinical trials of IPF.¹⁵

One of the steps necessary to evaluate the accuracy of HRCT in assessing change in disease extent is to examine the relationship between prospectively-collected imaging data with the established FVC outcome in IPF patients.^{15,34} In our study, the observed improvement in the proportion of patients with HRCT QLF response in BMS-986020-treated arms *versus* the placebo arm is consistent with the previously reported improvement in the proportion of patients with no FVC decline in BMS-986020-treated arms *versus* the placebo arm.²² Of note, this study was a prospective, double-blind, placebo-controlled phase II clinical trial for patients with IPF in which 26-week treatment with BMS-986020 600 mg BID *versus* placebo significantly slowed the rate of decline in both FVC (liters) and percent predicted FVC.²² Analyses from this study demonstrated a significant inverse association between change in lung fibrosis and ILD radiological scores (QLF and QILD) and pulmonary function measurements (percent predicted FVC and percent predicted DLCO), suggesting that, as pulmonary function declines, there is a commensurate increase in lung fibrotic disease extent in radiology. The most robust association was observed between the QLF score and percent predicted FVC, which is the endpoint used most commonly in clinical trials for IPF. These study results support automated quantitative measurements of HRCT images as an efficacy endpoint to complement the pulmonary function measurement of percent predicted FVC in IPF clinical trials.

HRCT scores, percent predicted FVC, and percent predicted DLCO were not correlated with UCSD-SOBQ scores at baseline. Changes in HRCT scores were significantly associated with change in UCSD-SOBQ scores at week 26.

Changes in percent predicted FVC and percent predicted DLCO trended in the same direction with UCSD-SOBQ score changes. This observation may be explained by interpatient differences in the tolerance level for shortness of breath in symptom scores. However, the inpatient changes in lung function (percent predicted DLCO), lung structure (HRCT), and symptoms (SOBQ) were correlated. Worsening dyspnea was significantly correlated with decline in pulmonary function and deterioration in the extent of lung fibrosis. According to the United States Food and Drug Administration's patient-drug development initiative, there is a need to systematically gather the patient perspective about their condition and treatment options.³⁵ The results of this study that show a correlation between QLF and UCSD-SOBQ scores support the idea that, together with physiological assessments, quantitative HRCT could play an important role in the development and validation of patient-reported outcomes or performance outcomes (e.g., 6MWT) and biomarkers in patients with IPF.

This study had several limitations. First, the short 26-week follow-up time limited the sensitivity of HRCT to assess disease extent changes because study patients have a variable disease course comprised of a heterogeneously progressive overall decline in lung function, intermittent exacerbations, and variable response to BMS-986020 treatment. The imaging outcome of QLF was part of the secondary endpoint outcome in a clinical trial. The changes in radiological outcome is dependent upon the available and analyzable HRCT examinations in the trial. Second, although BMS-986020 was well tolerated in most study patients, hepatobiliary toxicity was observed in several patients, which led to the termination of BMS-986020 clinical development.²² Due to the study's early termination, a smaller than planned number of study patients completed the trial, which limited the statistical power to assess the effect of BMS-986020 on lung fibrosis. In addition, the sample size was further reduced because data from six study patients had to be excluded from the HRCT quantitative analysis due to volume artifacts, inconsistent positioning, and incomplete coverage of the lung in the field-of-view. This occurrence emphasizes the importance of standardization of the HRCT acquisition procedure in clinical trials to minimize quality control failure and maximize image data completion. A final limitation of this study was that it included

percent predicted FVC in the correlative analysis with HRCT only when FVC (liters) was the primary endpoint of the clinical trial. In future studies, it would be desirable to include systematically collected patient-report outcomes, reasons for missing pulmonary functional tests, and a performance measurement of 6MWT.

Conclusion

This study demonstrated the value of automated quantitative HRCT as an efficacy endpoint in a prospective, randomized, double-blind, placebo-controlled IPF clinical trial of BMS-986020 in assessing disease extent and changes over 26 weeks of treatment. There was significant association between pulmonary function (percent predicted FVC and DLCO) and lung fibrotic extent (QLF and QILD) on HRCT both at baseline and week 26. In addition, significant correlations were observed between the changes in lung fibrotic extent (QLF and QILD) and changes in pulmonary function (percent predicted FVC and DLCO), and between the changes in lung fibrotic extent and symptoms of shortness of breath (UCSD-SOBQ) with IPF. Taken together, these data suggest that quantitative HRCT is a valuable tool to assess disease extent and treatment efficacy in patients in IPF clinical trials.

Acknowledgements

The authors would like to gratefully acknowledge the patients who participated in the study, as well as their families and caregivers.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by Bristol Myers Squibb. The research materials were obtained as part of the phase II clinical trials [ClinicalTrials.gov identifier: NCT01766817].

Conflict of interest statement

The authors Wendy Hayes, Benjamin Soule, and Shuyan Du were employees of Bristol Myers Squibb at the time that the study was performed and they may hold company stock/stock options. Grace Hyun J. Kim has the issued patent (UC-2015-0324982-A1) of Transitional Changes in Quantitative Interstitial Lung Disease and is a research consultant at MedQIA LLC. Jonathan G Goldin is a founder of MedQIA LLC.

ORCID iD

Grace Hyun J. Kim  <https://orcid.org/0000-0003-1225-3489>

Supplemental material

The reviews of this paper are available via the supplemental material section.

References

1. 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.
2. Ley B, Collard HR and King TE Jr. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2011; 183: 431–440.
3. Carlos WG, Streck ME, Wang TS, *et al.* Treatment of idiopathic pulmonary fibrosis. *Ann Am Thorac Soc* 2016; 13: 115–117.
4. Margaritopoulos GA, Vasarmidi E and Antoniou KM. Pirfenidone in the treatment of idiopathic pulmonary fibrosis: an evidence-based review of its place in therapy. *Core Evid* 2016; 11: 11–22.
5. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001; 69: 89–95.
6. Raghu G, Collard HR, Anstrom KJ, *et al.* Idiopathic pulmonary fibrosis: clinically meaningful primary endpoints in phase 3 clinical trials. *Am J Respir Crit Care Med* 2012; 185: 1044–1048.
7. Wells AU. Forced vital capacity as a primary endpoint in idiopathic pulmonary fibrosis treatment trials: making a silk purse from a sow's ear. *Thorax* 2013; 68: 309–310.
8. Eakin EG, Resnikoff PM, Prewitt LM, *et al.* Validation of a new dyspnea measure: the UCSD Shortness of Breath Questionnaire. University of California, San Diego. *Chest* 1998; 113: 619–624.
9. du Bois RM, Weycker D, Albera C, *et al.* Ascertainment of individual risk of mortality for patients with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2011; 184: 459–466.
10. 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: 1382–1389.

11. Lynch DA, Godwin JD, Safrin S, *et al.* High-resolution computed tomography in idiopathic pulmonary fibrosis: diagnosis and prognosis. *Am J Respir Crit Care Med* 2005; 172: 488–493.
12. Goldin J, Elashoff R, Kim HJ, *et al.* Treatment of scleroderma-interstitial lung disease with cyclophosphamide is associated with less progressive fibrosis on serial thoracic high-resolution CT scan than placebo: findings from the scleroderma lung study. *Chest* 2009; 136: 1333–1340.
13. Watadani T, Sakai F, Johkoh T, *et al.* Interobserver variability in the CT assessment of honeycombing in the lungs. *Radiology* 2013; 266: 936–944.
14. Jacob J, Bartholmai BJ, Rajagopalan S, *et al.* Automated quantitative computed tomography versus visual computed tomography scoring in idiopathic pulmonary fibrosis: validation against pulmonary function. *J Thorac Imaging* 2016; 31: 304–311.
15. 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.
16. 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: 1481–1491.
17. 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.
18. 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: 2455–2465.
19. Walsh SLF, Calandriello L, Silva M, *et al.* Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. *Lancet Respir Med* 2018; 6: 837–845.
20. Christe A, Peters AA, Drakopoulos D, *et al.* Computer-aided diagnosis of pulmonary fibrosis using deep learning and CT images. *Invest Radiol* 2019; 54: 627–632.
21. Liu K, Li Q, Ma J, *et al.* Evaluating a fully automated pulmonary nodule detection approach and its impact on radiologist performance. *Radiol Artif Intell* 2019; 1: e180084.
22. Palmer SM, Snyder L, Todd JL, *et al.* Randomized, double-blind, placebo-controlled, phase 2 trial of BMS-986020, a lysophosphatidic acid receptor antagonist for the treatment of idiopathic pulmonary fibrosis. *Chest* 2018; 154: 1061–1069.
23. Chong D, Brown MS, Kim HJ, *et al.* Reproducibility of volume and densitometric measures of emphysema on repeat computed tomography with an interval of 1 week. *Eur Radiol* 2012; 22: 287–294.
24. 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.
25. Kim HJ, Li G, Gjertson D, *et al.* Classification of parenchymal abnormality in scleroderma lung using a novel approach to denoise images collected via a multicenter study. *Acad Radiol* 2008; 15: 1004–1016.
26. 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–S35.
27. Holm S. A simple sequentially rejective multiple test procedure. *Scand J Statist* 1979; 6: 70.
28. Kafaja S, Clements PJ, Wilhalme H, *et al.* Reliability and minimal clinically important differences of forced vital capacity: results from the Scleroderma Lung Studies (SLS-I and SLS-II). *Am J Respir Crit Care Med* 2018; 197: 644–652.
29. Iwasawa T, Ogura T, Sakai F, *et al.* CT analysis of the effect of pirfenidone in patients with idiopathic pulmonary fibrosis. *Eur J Radiol* 2014; 83: 32–38.
30. Jacob J, Bartholmai BJ, Rajagopalan S, *et al.* Predicting outcomes in idiopathic pulmonary fibrosis using automated computed tomographic analysis. *Am J Respir Crit Care Med* 2018; 198: 767–776.
31. Colombi D, Dinkel J, Weinheimer O, *et al.* Visual vs fully automatic histogram-based assessment of idiopathic pulmonary fibrosis (IPF) progression using sequential multidetector computed tomography (MDCT). *PLoS One* 2015; 10: e0130653.
32. Xu Y, van Beek EJ, Hwanjo Y, *et al.* Computer-aided classification of interstitial lung diseases via

- MDCT: 3D adaptive multiple feature method (3D AMFM). *Acad Radiol* 2006; 13: 969–978.
33. Humphries SM, Yagihashi K, Huckleberry J, *et al.* Idiopathic pulmonary fibrosis: data-driven textural analysis of extent of fibrosis at baseline and 15-month follow-up. *Radiology* 2017; 285: 270–278.
34. Wagner JA and Ball JR. Implications of the institute of medicine report: evaluation of biomarkers and surrogate endpoints in chronic disease. *Clin Pharmacol Ther* 2015; 98: 12–15.
35. U.S. Department of Health and Human Services. Patient-focused drug development: collecting comprehensive and representative input, <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-focused-drug-development-collecting-comprehensive-and-representative-input> (2020, accessed 17 March 2021).

Visit SAGE journals online
[journals.sagepub.com/
home/tar](https://journals.sagepub.com/home/tar)

 SAGE journals