

Forced Vital Capacity in Patients with Idiopathic Pulmonary Fibrosis

Test Properties and Minimal Clinically Important Difference

Roland M. du Bois¹, Derek Weycker², Carlo Albers³, Williamson Z. Bradford⁴, Ulrich Costabel⁵, Alex Kartashov², Talmadge E. King, Jr.⁶, Lisa Lancaster⁷, Paul W. Noble⁸, Steven A. Sahn⁹, Michiel Thomeer¹⁰, Dominique Valeyre¹¹, and Athol U. Wells¹²

¹Imperial College, London, United Kingdom; ²Policy Analysis Inc., Brookline, Massachusetts; ³Department of Clinical and Biological Sciences, University of Turin, Turin, Italy; ⁴InterMune Inc., Brisbane, California; ⁵Ruhrlandklinik and Medical Faculty, University of Duisburg/Essen, Essen, Germany; ⁶University of California, San Francisco, San Francisco, California; ⁷Vanderbilt University Medical Center, Nashville, Tennessee; ⁸Duke University School of Medicine, Durham, North Carolina; ⁹Medical University of South Carolina, Charleston, South Carolina; ¹⁰University Hospital Gasthuisberg, Leuven, Belgium; ¹¹Assistance Publique-Hôpitaux de Paris, Hospital Avicenne, Bobigny, France; and ¹²Royal Brompton Hospital, London, United Kingdom

Rationale: Forced vital capacity (FVC) is an established measure of pulmonary function in idiopathic pulmonary fibrosis (IPF). Evidence regarding its measurement properties and minimal clinically important difference (MCID) in this population is limited.

Objectives: To assess the reliability, validity, and responsiveness of FVC and estimate the MCID in patients with IPF.

Methods: The study population included all 1,156 randomized patients in two clinical trials of IFN- γ 1b. FVC and other measures of functional status were measured at screening or baseline and 24-week intervals thereafter. Reliability was assessed based on two proximal measures of FVC, validity was assessed based on correlations between FVC and other measures of functional status, and responsiveness was assessed based on the relationship between 24-week changes in FVC and other measures of functional status. Distribution-based and anchor-based methods were used to estimate the MCID.

Measurements and Main Results: Correlation of percent-predicted FVC between measurements (mean interval, 18 d) was high ($r = 0.93$; $P < 0.001$). Correlations between FVC and other parameters were generally weak, with the strongest observed correlation between FVC and carbon monoxide diffusing capacity ($r = 0.38$; $P < 0.001$). Correlations between change in FVC and changes in other parameters were slightly stronger (range, $r = 0.16$ – 0.37 ; $P < 0.001$). Importantly, 1-year risk of death was more than twofold higher ($P < 0.001$) in patients with a 24-week decline in FVC between 5% and 10%. The estimated MCID was 2–6%.

Conclusions: FVC is a reliable, valid, and responsive measure of clinical status in patients with IPF, and a decline of 2–6%, although small, represents a clinically important difference.

Keywords: interstitial lung disease; idiopathic pulmonary fibrosis; forced vital capacity; minimal clinically important difference; mortality prediction

Idiopathic pulmonary fibrosis (IPF) is a progressive, life-threatening, interstitial lung disease of unknown etiology. It is characterized anatomically by scarring of the lungs and symptomatically by exertional dyspnea. IPF is the most common diffuse fibrosing lung disease and one of the most lethal, with a median survival of only 2 to 3 years after diagnosis (1–3).

(Received in original form May 12, 2011; accepted in final form August 30, 2011)

Supported by InterMune Inc., Brisbane, CA.

Correspondence and requests for reprints should be addressed to Roland M. du Bois, M.D., Imperial College, 1B Manresa Road, London SW3 6LR, UK. E-mail: ron@du-bois.co.uk

Am J Respir Crit Care Med Vol 184, pp 1382–1389, 2011

Copyright © 2011 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.201105-0840OC on September 22, 2011

Internet address: www.atsjournals.org

AT A GLANCE COMMENTARY

Scientific Knowledge on the Subject

Forced vital capacity (FVC) is a widely used measure of disease status and a common endpoint in clinical trials in patients with idiopathic pulmonary fibrosis. However, the measurement properties of FVC in this population have not been systematically evaluated.

What This Study Adds to the Field

Our findings demonstrate that FVC is a reliable, valid, and responsive measure of disease status in patients with idiopathic pulmonary fibrosis and suggest that the minimal clinically important difference for percent-predicted FVC is between 2% and 6%.

Forced vital capacity (FVC) has been a standard spirometric measure of pulmonary function in IPF for many decades. Longitudinal change in serial measures of lung volume (either FVC or vital capacity) is a widely accepted reflection of disease progression in patients with IPF and a commonly used primary endpoint in therapeutic studies in IPF (4–8). Moreover, several studies have identified change in percent-predicted FVC as an independent predictor of mortality in patients with IPF (9–14). Despite its widespread usage and clear prognostic utility, the measurement properties of percent-predicted FVC in IPF have not been formally examined. Furthermore, the minimal clinically important difference (MCID) in percent-predicted FVC among patients with IPF is currently unknown.

The MCID is the smallest difference in a measure that may be perceived to be important, either beneficial or harmful, and that would lead a clinician to consider a change in a patient's therapy (15, 16). MCID is a clinically important concept, because it may assist with the interpretation of the significance of observed changes in a measure and may influence the perceived efficacy of an intervention. The MCID also may have implications for the design of clinical trials in terms of the selection of primary and secondary endpoints and the determination of sample size (17, 18).

In the present study, we used data from two of the largest clinical trials to date in patients with IPF to assess the reliability, validity, and responsiveness of FVC and estimate the MCID in patients with this disease. This study has been presented in part at the 2010 European Respiratory Society Annual Congress (19).

METHODS

Study Population

The study population comprised all randomized subjects ($n = 1,156$) in two placebo-controlled clinical trials of IFN- γ 1b (Protocols GIPF-001 [$n = 330$] and GIPF-007 [$n = 826$]) irrespective of treatment assignment (placebo [$n = 443$] or IFN- γ 1b [$n = 713$]), given that both of these studies were negative. The designs of these trials are described in detail elsewhere (20, 21). Briefly, eligible patients were required to have a high-resolution computed tomography scan showing features consistent with protocol-defined criteria for either a definite or probable diagnosis of IPF. Surgical lung biopsy was required to confirm a suspected diagnosis in all patients with a clinical and radiographic diagnosis of probable IPF, and all patients less than age 50 years, regardless of the degree of certainty associated with the clinical and radiographic diagnoses.

Data Collection

Data collected in the aforementioned trials included patient demographic and clinical characteristics (age, sex, duration of disease [from date of IPF diagnosis], smoking status, and use of supplemental oxygen); physiologic assessments (percent-predicted FVC, percent-predicted carbon monoxide diffusing capacity [DL_{CO}], and resting alveolar-arterial oxygen pressure at ambient temperature [A-a gradient]); measures of functional status (6-minute-walk distance [6MWD], in Protocol GIPF-007), dyspnea (University of California at San Diego Shortness of Breath Questionnaire [UCSD-SOBQ]), and health-related quality of life (HRQL; St. George's Respiratory Questionnaire [SGRQ] and Medical Outcomes Study 36-item short-form [SF-36], the latter in Protocol GIPF-001); hospitalizations; and vital status. Total score on the UCSD-SOBQ ranges from 0–120, with higher scores reflecting worse dyspnea (22). The SGRQ comprises three respiratory-specific domains (symptoms, activity, and impacts) and each domain ranges from 0–100, with an increasing score indicating worsening HRQL (23).

Information on percent-predicted FVC, percent-predicted DL_{CO} , A-a gradient (Protocol GIPF-001), 6MWD (Protocol GIPF-007), dyspnea, and HRQL was collected at the screening or baseline visit and periodically (i.e., every 12 wks in Protocol GIPF-001 and every 24 wks in Protocol GIPF-007) thereafter through the end of the trial. Resting A-a gradient was assessed at the screening visit and every 48 weeks thereafter in Protocol GIPF-007. Hospitalizations and vital status were tracked throughout the conduct of the trials.

Statistical Analyses

Reliability. Reliability was assessed based on the stability of percent-predicted FVC values between the study screening visit (Day –28 to Day –1) and the baseline visit (Day 1). The intraclass correlation coefficient was used to assess the strength of the relationship between these assessments; a value of 0.80 or greater was assumed to represent “good” reliability. Analyses were conducted using observed data for all subjects with FVC values from both the screening and baseline visits in Protocol GIPF-001 (screening data on percent-predicted FVC were not collected in Protocol GIPF-007).

Validity. Criterion validity was assessed based on relationships between percent-predicted FVC and the following parameters at the same visit: percent-predicted DL_{CO} , A-a gradient, 6MWD, UCSD-SOBQ, SGRQ, and SF-36. Distribution-independent (Spearman) correlation coefficients were used to assess the strength of these relationships. Strength of correlation was designated as follows: greater than 0.5, large; 0.5–0.3, moderate; 0.3–0.1, small; and less than 0.1, trivial (24).

Construct validity was assessed by comparing mean percent-predicted FVC values across subgroups of patients presumed to have different levels of physiologic function, defined on the basis of percent-predicted DL_{CO} , A-a gradient, 6MWD, UCSD-SOBQ, SGRQ, and SF-36 at the same visit. Patients were stratified into subgroups based on the quintiles of the corresponding distributions. One-way analysis of variance was used for statistical comparisons.

Responsiveness. Responsiveness was assessed using Spearman correlation coefficients between changes (i.e., screening and baseline to Week

24) in percent-predicted FVC and changes over the same period in percent-predicted DL_{CO} , A-a gradient, 6MWD, UCSD-SOBQ, SGRQ, and SF-36. The relationship between mean changes in percent-predicted FVC and changes in other measures (stratified into quintiles) was examined using analysis of variance. Analyses were conducted using observed data for the full study population.

Responsiveness was also evaluated by examining the relationship between change in percent-predicted FVC over 24 weeks and 1-year risk of death using a Cox proportional hazards model. Change in percent-predicted FVC was evaluated over the 24-week periods immediately preceding the Week 24 and Week 72 trial visits, respectively, and was defined categorically (i.e., absolute change in percent-predicted FVC $\leq -10\%$, -5% to -9% , $> -5\%$) based on prior research (9–14). All deaths occurring over the 48-week period after the Week 24 and Week 72 trial visits, respectively, were included in these analyses; subjects who were lost to follow-up and those who underwent lung transplant during follow-up were censored on the corresponding date. Potential confounding from the inclusion of subjects receiving active drug during follow-up was evaluated by including a term in the model for treatment assignment and an interaction term for treatment assignment and change in percent-predicted FVC; percent-predicted FVC at baseline (i.e., baseline trial visit and Week 48 trial visit, respectively) was also included as a model covariate. The assumption of proportional hazards was evaluated using published methods (25).

Minimal clinically important difference. Both distribution-based and anchor-based methods were used to estimate the MCID. Distribution-based methods included the SEM and effect size. SEM was calculated for percent-predicted FVC by multiplying the estimated standard deviation at baseline for all randomized subjects by the square root of one minus the estimated reliability coefficient (26, 27). One SEM was defined to be the MCID. Because the SEM is sample-independent, corresponding estimates are considered to be bidirectional in nature. Analyses were repeated using estimates of standard deviation for percent-predicted FVC at Weeks 24, 48, 72, and 96, respectively.

Effect size was calculated by dividing the difference in mean percent-predicted FVC at baseline and Week 48 by the standard deviation of this measure at baseline. A change in value corresponding to a “small” effect size, defined as 0.2 to less than 0.5, per conventional benchmarks, was considered to approximate the MCID (17, 24, 28, 29). One-third of the estimated standard deviation has also been suggested as an approximation of MCID (30). Analyses were repeated evaluating the difference in mean percent-predicted FVC over alternative 48-week intervals (i.e., Weeks 24–72 and Weeks 48–96) and intervals of different durations (i.e., Weeks 0–24, Weeks 0–72, and Weeks 0–96).

Anchor-based methods included the patient-referencing and criterion-referencing approaches. The patient-referencing approach compared the mean change in percent-predicted FVC between baseline and Week 48 across subgroups of patients defined on the basis of their global rating of change in health status at Week 48. Determination of change in global health status was based on the following question from the SF-36: “Compared with one year ago, how would you rate your health in general now?” Possible responses to this question were as follows: (1) “much better,” (2) “somewhat better,” (3) “same,” (4) “somewhat worse,” and (5) “much worse.” The second and fourth responses (i.e., somewhat better and somewhat worse, respectively) were considered to represent minimal but clinically important changes; in these analyses, the valence of change in percent-predicted FVC was reversed for patients reporting worse health (31). Mean changes in percent-predicted FVC were examined on both an unadjusted and an adjusted basis. FVC mean change values were adjusted for the mean change in the group reporting no change in global health status to account for potential recall bias in patient responses, and thus to “normalize” change scores relative to the group of patients who rated their health status as “the same” as 1 year ago. Data for all randomized subjects in Protocol GIPF-001 were used in these analyses; analyses were repeated evaluating change in percent-predicted FVC between Weeks 12 and 60, and Weeks 24 and 72, respectively.

The criterion-referencing approach involved estimation of differences in percent-predicted FVC at baseline between patients who did and did not experience selected health events (specifically, all-cause hospitalization, death, and the composite endpoint, hospitalization or death) during the subsequent 48-week period. An independent samples t test

was used for statistical comparisons. Analyses were conducted using observed data from all randomized subjects who were not lost to follow-up during the 48-week period.

RESULTS

Patient Characteristics

A total of 1,156 patients were randomized either to placebo (n = 443) or IFN- γ 1b (n = 713) in the two clinical trials (Protocols GIPF-001 [n = 330] and GIPF-007 [n = 826]) (Table 1). Mean (\pm SD) age at study entry was 65 (\pm 8) years and 70% of patients were male. The mean value for percent-predicted FVC at baseline was 70.1% (12.8) and the interquartile range was 60–78.7%. There was substantial variation across study subjects in measures of physiologic function, functional status, dyspnea, and HRQL.

Reliability

Percent-predicted FVC seems to have good reliability in patients with IPF. The intraclass correlation coefficient between percent-predicted FVC values (n = 91) at the screening and baseline visits (mean number of days between visits, 18) was 0.93 (P < 0.001). There were no apparent differences in correlation coefficients for percent-predicted FVC values based on age or sex (data not shown).

TABLE 1. BASELINE CHARACTERISTICS

Characteristic	Value*
Age, yr	
Mean (SD)	65.3 (8.1)
Median (IQR)	66 (60–72)
Sex, n (%)	
Male	812 (70.2)
Female	344 (29.8)
FVC, % predicted	
Mean (SD)	70.1 (12.8)
Median (IQR)	68 (60–78.7)
D _{LCO} , % predicted	
Mean (SD)	44.4 (10.8)
Median (IQR)	43.9 (37–50.8)
Resting AaPo ₂ , mm Hg	
Mean (SD)	20.8 (11.1)
Median (IQR)	20.4 (13–28.2)
6MWT distance, m	
Mean (SD)	392.4 (108.5)
Median (IQR)	395 (328–462)
UCSD SOBQ score	
Mean (SD)	37.9 (23.5)
Median (IQR)	35 (19.6–55)
SGRQ score	
Mean (SD)	43.8 (17.7)
Median (IQR)	43.6 (31–56.2)
SF-36 MCS	
Mean (SD)	50.6 (10.2)
Median (IQR)	52.1 (43–59.3)
SF-36 PCS	
Mean (SD)	34.5 (8.6)
Median (IQR)	33.2 (28.2–40.4)

Definition of abbreviations: 6MWT = 6-minute-walk test; AaPo₂ = alveolar-arterial oxygen gradient; D_{LCO} = carbon monoxide diffusion capacity; IQR = interquartile range; SF-36 MCS = Medical Outcomes Study 36-item short form (mental component summary); SF-36 PCS = Medical Outcomes Study 36-item short form (physical component summary); SGRQ = St. George's Respiratory Questionnaire; UCSD SOBQ = University of California San Diego Shortness of Breath Questionnaire.

* n = 1,156 except for D_{LCO} (1,152); AaPo₂ (1,139); 6MWT distance (822); UCSD SOBQ (1,120); SGRQ (1,060); SF-36 MCS (307); and SF-36 PCS (307).

TABLE 2. CORRELATION BETWEEN PERCENT-PREDICTED FVC AND OTHER MEASURES

Variable	N	Coefficient*	P Value†
D _{LCO} , % predicted	1,152	0.38	<0.001
Resting AaPo ₂ , mm Hg	1,139	–0.17	<0.001
6MWT distance, m	822	0.12	<0.001
UCSD SOBQ score	1,120	–0.17	<0.001
SGRQ score	1,060	–0.16	<0.001
SF-36 MCS	307	0.00	0.944
SF-36 PCS	307	0.14	0.015

Definition of abbreviations: 6MWT = 6-minute-walk test; AaPo₂ = alveolar-arterial oxygen gradient; D_{LCO} = carbon monoxide diffusion capacity; SF-36 MCS = Medical Outcomes Study 36-item short form (mental component summary); SF-36 PCS = Medical Outcomes Study 36-item short form (physical component summary); SGRQ = St. George's Respiratory Questionnaire; UCSD SOBQ = University of California San Diego Shortness of Breath Questionnaire.

* Spearman correlation coefficient.

† One-way analysis of variance.

Validity

The relationship between percent-predicted FVC and other measures of disease status is summarized in Tables 2 and 3. Correlations between percent-predicted FVC and measures of gas exchange, functional status, dyspnea, and HRQL were in the expected direction but generally weak (absolute values of all coefficients, <0.17, except for percent-predicted D_{LCO} [r = 0.38; P < 0.001]) (Table 2). Mean values for percent-predicted FVC were generally significantly lower for patients with poorer levels of gas exchange, functional status, dyspnea, and HRQL; percent-predicted FVC did not vary across quintiles of patients defined on the basis of the SF-36 (Table 3).

Responsiveness

Correlations between changes in percent-predicted FVC and changes in measures of physiologic function, functional status, dyspnea, and HRQL were in the expected direction, but moderately to weakly correlated (absolute values of all coefficients, <0.37) (Table 4). Decline in percent-predicted FVC was consistently greater for patients with larger declines in levels of physiologic function, functional status, and HRQL (Table 5). Findings were largely unchanged when focusing on each of the respective treatment groups from the clinical trials and when assessing changes over a different period of time (i.e., 48 wk).

Change in percent-predicted FVC over 24 weeks was highly predictive of death over the subsequent 1-year period. Risk of death was nearly fivefold higher (hazard ratio, 4.78; 95% confidence interval [CI], 3.12–7.33; P < 0.001) for patients with absolute declines in percent-predicted FVC of greater than or equal to 10% (e.g., a decline from 70–60%), and more than twofold higher (hazard ratio, 2.14; 95% CI, 1.43–3.20; P < 0.001) for those with absolute declines between 5% and 10%, compared with patients who experienced declines in percent-predicted FVC of less than 5% (Table 6). Treatment assignment and the interaction term for treatment assignment and change in percent-predicted FVC (defined continuously) were not found to be important predictors of death, and the proportional hazards assumption for change in percent-predicted FVC was not violated.

Minimal Clinically Important Difference

The estimated SEM for percent-predicted FVC, and the corresponding MCID, was 3.4 (95% CI, 3.2–3.5) (Table 7). Using observed data from postbaseline visits, the estimated SEM ranged from 3.8–4.3. The estimated effect size for percent-predicted FVC was 0.27 (95% CI, 0.23–0.31), based on a difference in mean baseline and mean Week 48 values of 3.4%;

TABLE 3. MEAN (SD) PERCENT-PREDICTED FVC BY QUINTILES OF OTHER MEASURES

Variable	N	% Predicted FVC*	P Value†
D_{LCO}, % predicted			
<36.00	221	63.1 (10.5)	<0.001
≥36.00 to <41.02	239	67.4 (10.9)	<0.001
≥41.02 to <46.44	231	69.2 (11.3)	<0.001
≥46.44 to <52.73	230	73.1 (12.6)	<0.001
≥52.73	231	77.3 (13.8)	—
Resting AaPo₂, mm Hg			
<10.89	228	72.4 (12.2)	<0.001
≥10.89 to <17.50	227	71.1 (12.3)	<0.001
≥17.50 to <23.50	228	71.3 (12.6)	<0.001
≥23.50 to <29.58	229	69.3 (13.5)	0.009
≥29.58	227	66.2 (13.4)	—
6MWT distance, m			
<306	163	70.4 (12.6)	<0.001
≥306 to <369	165	71.4 (11.9)	0.010
≥369 to <419	162	72.5 (11.8)	0.064
≥419 to <480	165	73.1 (13.1)	0.169
≥480	167	75 (13.6)	—
UCSD SOBQ score			
<15.83	224	74 (13.5)	<0.001
≥15.83 to <27.37	224	70.4 (12.6)	0.002
≥27.37 to <42.00	214	70.4 (11.6)	0.003
≥42.00 to <59.00	223	68.6 (12.3)	0.124
≥59.00	235	66.8 (12.5)	—
SGRQ score			
<28.09	212	73.1 (12.8)	<0.001
≥28.09 to <38.11	212	70.4 (13.2)	0.003
≥38.11 to <48.18	212	69.7 (12)	0.015
≥48.18 to <59.60	211	68.6 (12.8)	0.139
≥59.60	213	66.8 (11.5)	—
SF-36 MCS			
<40.98	61	63.7 (11.2)	0.870
≥40.98 to <48.66	61	63.9 (11.8)	0.935
≥48.66 to <55.82	62	63.6 (10.9)	0.808
≥55.82 to <60.42	62	63.8 (10.4)	0.890
≥60.42	61	64 (10.9)	—
SF-36 PCS			
<26.83	61	63.2 (9.9)	0.410
≥26.83 to <31.22	62	61.3 (10.3)	0.075
≥31.22 to <36.22	61	62.6 (11.8)	0.251
≥36.22 to <42.23	62	67 (11.4)	0.267
≥42.23	61	64.8 (10.8)	—

Definition of abbreviations: 6MWT = 6-minute-walk test; AaPo₂ = alveolar-arterial oxygen gradient; D_{LCO} = carbon monoxide diffusion capacity; SF-36 MCS = Medical Outcomes Study 36-item short form (mental component summary); SF-36 PCS = Medical Outcomes Study 36-item short form (physical component summary); SGRQ = St. George’s Respiratory Questionnaire; UCSD SOBQ = University of California San Diego Shortness of Breath Questionnaire.

* Data are presented as mean (SD).

† One-way analysis of variance.

according to Cohen’s criteria (24), such an effect should be considered “small.” Similar results were obtained when focusing on alternative time intervals, with effect sizes ranging from 0.15–0.46 based on differences in mean values for percent-predicted FVC ranging from 2–6%. One-third of the estimated standard deviation at baseline yielded a figure of 4.2; figures based on data from other visits ranged from 4.2–4.9.

Patients who reported that their global health status was “much better” at Week 48 versus baseline experienced an (unadjusted) absolute increase in percent-predicted FVC of 2.3%, on average; patients reporting that their global health status was “somewhat better,” “same,” “somewhat worse,” or “much worse” experienced decreases in percent-predicted FVC (–2.1%, –2.8%, –6.5%, and –6.1%, respectively). On an adjusted basis, mean change in percent-predicted FVC for the subgroup who reported that their health status was “somewhat better” or “somewhat worse,” which was considered to

TABLE 4. CORRELATION BETWEEN CHANGE IN PERCENT-PREDICTED FVC AND CHANGE IN OTHER MEASURES

Variable	N	Coefficient*	P Value†
D _{LCO} , % predicted	1,047	0.29	<0.001
Resting AaPo ₂ , mm Hg	290	–0.37	<0.001
6MWT distance, m	762	0.22	<0.001
UCSD SOBQ score	1,001	–0.25	<0.001
SGRQ score	897	–0.32	<0.001
SF-36 MCS	246	0.16	0.012
SF-36 PCS	246	0.26	<0.001

Definition of abbreviations: 6MWT = 6-minute-walk test; AaPo₂ = alveolar-arterial oxygen gradient; D_{LCO} = carbon monoxide diffusion capacity; SF-36 MCS = Medical Outcomes Study 36-item short form (mental component summary); SF-36 PCS = Medical Outcomes Study 36-item short form (physical component summary); SGRQ = St. George’s Respiratory Questionnaire; UCSD SOBQ = University of California San Diego Shortness of Breath Questionnaire.

* Spearman correlation coefficient.

† One-way analysis of variance.

represent the MCID, was 2.2%. In analyses using change in percent-predicted FVC between Weeks 12 and 60 and change between Weeks 24 and 72, respectively, corresponding estimates of MCID were 2.7% and 4.3%.

Criterion-referencing analyses demonstrated that values for percent-predicted FVC at baseline were significantly different for patients who died versus those who did not during the 48-week follow-up period. Mean (± SD) baseline percent-predicted FVC in patients who died during the 48-week follow-up period was 64.9% (± 11.2), compared with 70.7% (± 12.8) in those who survived. Based on this difference, the estimated MCID was 5.8% (P < 0.001).

DISCUSSION

FVC is a widely used measure of pulmonary function and disease status in patients with IPF. Despite its ubiquitous use in clinical practice and therapeutic clinical trials, FVC test performance characteristics in patients with IPF have not been formally evaluated. Additionally, the MCID for percent-predicted FVC in this patient population is currently unknown. The present study assessed reliability, validity, and responsiveness of FVC and estimated the MCID in a large cohort of patients with a confident IPF diagnosis and mild to moderate impairment in measures of baseline physiologic function. Taken collectively, the data suggest that percent-predicted FVC is a robust measure of clinical status in patients with IPF.

Reliability, as assessed by the stability of two proximal measures of FVC, was found to be good, with an overall interclass correlation coefficient of 0.93. Comparison of baseline percent-predicted FVC values with selected measures of physiologic function, dyspnea, and HRQL revealed generally weak correlations; however, comparison of mean values for percent-predicted FVC across quintiles defined on the basis of other measures demonstrated consistently lower values for patients with poorer levels of performance on these measures. Responsiveness, as measured by the correlation between change in FVC and changes in other parameters, was slightly stronger, with coefficients consistently in the range of 0.16–0.37. Additionally, change in percent-predicted FVC was found to be highly predictive of mortality. Although the objective of this analysis (i.e., to evaluate responsiveness) was different from that of prior research (i.e., to evaluate independent effect), our findings are consistent with observations from previous studies (9–14). Specifically, a decline in percent-predicted FVC greater than or equal to 10% at 24 weeks was associated with a nearly fivefold increase in the risk of mortality over the subsequent year, whereas a decline of

TABLE 5. MEAN (SD) CHANGE IN PERCENT-PREDICTED FVC BY QUINTILES OF CHANGE IN OTHER MEASURES

Variable	N	% Predicted FVC*	P Value†
D_{lCO}, % predicted			
< -8.00	208	-6 (7.2)	<0.001
≥ -8.00 to < -4.17	212	-2.6 (6.3)	<0.001
≥ -4.17 to < -1.00	203	-1.5 (6.4)	<0.001
≥ -1.00 to < 2.84	215	-0.8 (5.6)	<0.001
≥ 2.84	209	1.4 (6.6)	—
Resting AaPo₂, mm Hg			
< -6.52	59	1 (5.4)	<0.001
≥ -6.52 to < -1.44	57	-0.4 (7)	<0.001
≥ -1.44 to < 3.20	57	-0.8 (4.4)	<0.001
≥ 3.20 to < 7.46	58	-3.8 (6)	0.242
≥ 7.46	59	-5 (5.6)	—
6MWT distance, m			
< -72.00	151	-4.7 (7.5)	<0.001
≥ -72.00 to < -23.00	152	-2.3 (6.5)	0.011
≥ -23.00 to < 1.00	148	-1.6 (6.8)	0.091
≥ 1.00 to < 32.00	157	-0.9 (7.6)	0.395
≥ 32.00	154	-0.2 (6.8)	—
UCSD SOBQ score			
< -9.00	189	0.1 (6.6)	<0.001
≥ -9.00 to < -1.00	208	-1.3 (6.6)	<0.001
≥ -1.00 to < 5.00	203	-1.4 (5.6)	<0.001
≥ 5.00 to < 14.82	203	-2.1 (6)	<0.001
≥ 14.82	198	-4.9 (7.5)	—
SGRQ score			
< -8.61	181	0.9 (7.4)	<0.001
≥ -8.61 to < -2.59	179	-1.1 (6.3)	<0.001
≥ -2.59 to < 2.65	178	-1.3 (6.5)	<0.001
≥ 2.65 to < 10.02	180	-2.9 (6)	<0.001
≥ 10.02	179	-5.3 (7.4)	—
SF-36 MCS			
< -7.26	48	-2.3 (8.1)	0.334
≥ -7.26 to < -1.96	50	-2.1 (6.1)	0.409
≥ -1.96 to < 2.05	50	-1.9 (6.1)	0.485
≥ 2.05 to < 6.98	50	-0.9 (4.8)	0.911
≥ 6.98	48	-1 (5.6)	—
SF-36 PCS			
< -5.90	48	-3.7 (5.8)	<0.001
≥ -5.90 to < -1.87	50	-3.2 (5.6)	<0.001
≥ -1.87 to < 1.28	49	-0.9 (6.6)	0.059
≥ 1.28 to < 6.21	50	-1.7 (5.6)	0.010
≥ 6.21	49	1.4 (6.2)	—

Definition of abbreviations: 6MWT = 6-minute-walk test; AaPo₂ = alveolar-arterial oxygen gradient; D_{lCO} = carbon monoxide diffusion capacity; SF-36 MCS = Medical Outcomes Study 36-item short form (mental component summary); SF-36 PCS = Medical Outcomes Study 36-item short form (physical component summary); SGRQ = St. George's Respiratory Questionnaire; UCSD SOBQ = University of California San Diego Shortness of Breath Questionnaire.

* Data are presented as mean (SD).

† One-way analysis of variance.

5–10% conferred a more than twofold increase in the risk of 1-year mortality. This latter finding is particularly noteworthy, because it supports recent data suggesting that changes in percent-predicted FVC that were previously regarded as evidence of clinically stable disease are medically relevant and worthy of further clinical evaluation (13).

Consistent with the recommendations of Yost and Eton (30), we used a number of alternative methods to estimate the MCID for percent-predicted FVC, including distribution-based and anchor-based approaches, and we report a range of estimates of the MCID. Our results suggest that the MCID for percent-predicted FVC lies between 2% and 6% based on these various alternative methods of estimation. We note, however, that the upper limit of this range may not appropriately reflect the smallest difference that would be clinically important to patients with IPF because it was estimated using an anchor-based method that used death as a health event, which is clearly not minimally

TABLE 6. COX PROPORTIONAL HAZARDS MODEL

	Patient Visits (n)	Deaths (n)	1-Year Risk of Death	
			HR (95% CI)	P Value
ΔFVC, % predicted				
≤ -10	166	39	4.78 (3.12–7.33)	<0.001
-5 to -10	373	45	2.14 (1.43–3.20)	<0.001
> -5	1,316	56		
FVC, % predicted				
≤ 50	203	42	7.44 (3.28–16.87)	<0.001
51 to 65	691	65	4.09 (1.87–8.98)	<0.001
66 to 79	594	26	1.97 (0.85–4.55)	0.111
≥ 80	374	7		

Definition of abbreviations: ΔFVC = 24-week absolute change in percent-predicted FVC (e.g., change from 70–65% = 5% absolute change); CI = confidence interval; HR = hazard ratio.

important in nature. We also note that the lower limit of this range may not reflect the MCID because a value lower than 3% was observed in only one of the many different analyses used, and this estimate was generated using data on a relatively small sample of patients (n = 114) from only one of the two clinical trials. In addition, measurement error may limit the use of small MCID values in the assessment of an individual patient; for this reason, it has been suggested that the high end of the range should be used to assess the clinical relevance of changes in an individual patient, whereas the lower limit of the range should be used to assess change at the population level (30). Importantly, our estimates of MCID were largely unchanged (maximum value, 5%) when using data from other visits and over intervals of different durations. This finding, coupled with the similarity of the estimates derived by the various alternative analytic approaches, demonstrates a high degree of internal consistency. Moreover, our findings are remarkably consistent with those of a recent study by Zappala and coworkers (13), which evaluated the prognostic significance of marginal declines in percent-predicted FVC in 84 patients with biopsy-proved IPF. A *post hoc* analysis of mortality risk at various thresholds of change in percent-predicted FVC at 6 months suggested that the optimal threshold value was between -3% and -4%.

Several study limitations should be noted. First, reliability of percent-predicted FVC was assessed using data from a relatively small subgroup of patients (n = 91) for whom both screening and baseline values were available. The analysis was limited to these data because the mean interval between the screening and baseline visits was only 18 days, compared with the relatively long intervals between postbaseline visits in GIPF-001 and GIPF-007 (12 and 24 wk, respectively). This allowed us to minimize potential confounding from changes in disease status and other factors. Of note, this subgroup was not clinically or statistically different in terms of their baseline characteristics from other patients enrolled in GIPF-001. Second, although the analyses of responsiveness ideally would have been limited to patients randomized to the placebo arm in the clinical trials, we concluded based on the absence of evidence for any treatment effect that the enhanced power of the study to characterize the relationship between changes in percent-predicted FVC and changes in other measures of disease status justified the inclusion of all randomized patients. Analyses of responsiveness and all other analyses (reliability, validity, and MCID) were robust when focusing on different populations (e.g., all subjects vs. placebo subjects) and when assessing changes over different periods of time (e.g., 24 vs. 48 wk). Third, although the SEM is sample-independent and the corresponding estimates of MCID should therefore be considered bidirectional in nature, findings from the patient-referencing approach suggest that perceived gains and losses

TABLE 7. ESTIMATION OF THE MINIMAL CLINICALLY IMPORTANT DIFFERENCE IN PERCENT-PREDICTED FVC IN PATIENTS WITH IPF

Standard Error of Measurement					
	N	Mean	SD	Correlation	SEM (95% CI)
% Predicted FVC	1,156	70.1	12.8	0.93	3.4 (3.2–3.5)
Effect Size					
	N	Baseline	Follow-up*	Difference	Effect Size
Mean % predicted FVC (SD)	984	70.8 (12.7)	67.4 (15.2)	–3.4 (8.2)	0.27
Patient Referencing					
Change in Percent-predicted FVC*					
	N	Unadjusted Mean (SD)		Adjusted Mean (SD) [†]	
Global change in health status [‡]					
Much better	25	2.3 (7.3)		5.1 (7.3)	
Somewhat better	55	–2.1 (6.4)		0.7 (6.4)	
Same	96	–2.8 (5.8)		0 (5.8)	
Somewhat worse	59	–6.5 (6.6)		–3.7 (6.6)	
Much worse	14	–6.1 (9.5)		–3.3 (9.5)	
Somewhat better or worse	114	2.3 (7.7)		2.2 (6.6)	
Criterion Referencing [§]					
	N	Percent-predicted FVC		P Value [¶]	
Hospitalization				0.226	
No	929	70 (12.9)			
Yes	213	71.1 (12.2)			
Difference		1.2			
Death				<0.001	
No	1,037	70.7 (12.8)			
Yes	105	64.9 (11.2)			
Difference		5.8			
Hospitalization or death				0.265	
No	849	70.5 (12.9)			
Yes	293	69.5 (12.4)			
Difference		1			

Definition of abbreviation: CI = confidence interval.

* Assessed at the Week 48 study visit.

† Percent-predicted FVC values adjusted by mean value for group reporting no change in health status.

‡ Based on the SF-36 question: “Compared to one year ago, how would you rate your health in general now?”

Somewhat better and somewhat worse are considered to represent minimal but clinically important changes.

§ Comparison of baseline percent-predicted FVC between patients who did and did not experience selected health events during the subsequent 48-week period.

|| Data are presented as mean (SD).

¶ Independent samples *t* test.

may not be equivalent vis-à-vis change in FVC; small sample sizes in the latter analyses precluded a more thorough investigation of this issue. Fourth, in our criterion-referencing approach to estimating the MCID, we used hospitalization and death from any cause as health events. Clearly, these events are not minimally important in nature; as a result, estimates based on this approach may not appropriately reflect the smallest differences that would be clinically important to patients with IPF. These estimates may thus be viewed as conservative, although they were largely consistent with estimates yielded by the other approaches. Fifth, the clinical trials from which data were used for this study enrolled a group of patients with mild-to-moderate impairment of pulmonary function at baseline; patients who were too ill or considered at high risk for dying during the course of the trial were excluded from the trial populations. It is thus unknown whether the measurement properties and MCID for percent-predicted FVC would be comparable in a population of patients with more severe IPF. Finally, for each of the various analyses, we used all

observed data rather than limiting the study population to patients with complete information on all measures at all times or using an imputation algorithm to estimate missing values; thus, the size of the population and mix of patients may have varied across analyses. In addition, although the exclusion of patients in some analyses and informative censoring caused by patients who were transplanted or lost to follow-up in other analyses (i.e., responsiveness, MCID based on effect size and anchor-based methods) is a potential concern, the percentage of patients who were transplanted or lost to follow-up was low. For example, from the trial baseline visit to the Week 48 visit (the follow-up period used in some of the previously noted analyses), only 31 (2.7%) and 32 (2.7%) patients, respectively, were transplanted or lost to follow-up, and this subgroup of patients was largely comparable with the rest of the population in terms of baseline characteristics. As a result, the potential for any such bias is small.

The findings, combined with those of a previous study evaluating the performance characteristics of the 6-minute-walk test in

patients with IPF (32), provide two validated measures of clinical status that can be used to assess the clinical efficacy of experimental therapies in future trials in patients with IPF. Moreover, the observation that 24-week change in FVC and 6MWD is predictive of 1-year mortality despite relatively weak correlations between these and other measures of disease status suggests that both tests measure clinically important domains of the disease process that are not captured by other measures. Finally, the establishment of a MCID for both percent-predicted FVC and 6MWD provides important benchmarks for assessing the clinical significance of longitudinal changes in these parameters in patients with IPF.

Importantly, the findings of this study do not militate against current clinical practice, in which changes in both FVC and DL_{CO} are used to assess disease status. We believe, however, that our findings provide robust reassurance about the performance characteristics of the FVC measurement and provide the clinician with an anchor around which to decide if any given change is clinically meaningful. In this regard, we believe that the MCID will be one of several factors that the clinician may use in the assessment of disease progression and will be of value to clinicians and researchers.

In conclusion, our results demonstrate that percent-predicted FVC is a reliable, valid, and responsive measure of clinical status in patients with IPF, and that a decline in percent-predicted FVC of 2 to 6% represents a small but clinically important difference. Collectively, these findings demonstrate that FVC is a clinically useful measure of disease status and a valid endpoint for clinical trials in patients with IPF.

Author Disclosures are available with the text of this article at www.atsjournals.org.

Acknowledgment: This study was funded by InterMune, Inc. The authors thank Kenneth Glasscock for medical writing and editorial assistance and the participating staff members and patients at all study centers.

References

- King TE Jr, Tooze JA, Schwarz MI, Brown K, Cherniack RM. Predicting survival in idiopathic pulmonary fibrosis. Scoring system and survival model. *Am J Respir Crit Care Med* 2001;164:1171–1181.
- Latsi PI, du Bois RM, Nicholson AG, Colby TV, Bisirtzoglou D, Nikolakopoulou A, Veeraraghavan S, Hansell DM, Wells AU. Fibrotic idiopathic interstitial pneumonia: the prognostic value of longitudinal functional trends. *Am J Respir Crit Care Med* 2003;168:531–537.
- Ley B, Collard HR, King TE Jr. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2011;183:431–440.
- American Thoracic Society, European Respiratory Society. Idiopathic pulmonary fibrosis: diagnosis and treatment. International consensus statement. *Am J Respir Crit Care Med* 2000;161:646–664.
- Azuma A, Nukiwa T, Tsuboi E, Suga M, Abe S, Nakata K, Taguchi Y, Nagai S, Itoh H, Ohi M, et al. Double-blind, placebo-controlled, trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2005;171:1040–1047.
- Demedts M, Behr J, Buhl R, Costabel U, Dekhuijzen R, Jansen HM, MacNee W, Thomeer M, Wallaert B, Laurent F, et al. High-dose acetylcysteine in idiopathic pulmonary fibrosis. *N Engl J Med* 2005;353:2229–2242.
- Taniguchi H, Ebina M, Kondoh Y, Azuma A, Ogura T, Taguchi Y, Suga M, Takahashi H, Nakata K, Sato A, et al. Pirfenidone in idiopathic pulmonary fibrosis: a phase III clinical trial in Japan. *Eur Respir J* 2010;35:821–829.
- Noble PW, Albera C, Bradford WZ, Costabel U, Glassberg MK, Kardatzke D, King TE Jr, Lancaster L, Sahn SA, Swarcberg J, et al. The CAPACITY Program: two randomised, double-blind, placebo controlled trials of pirfenidone in patients with idiopathic pulmonary fibrosis. *Lancet* 2011;377:1760–1769.
- Collard HR, King TE Jr, Bartelson BB, Vourlekis JS, Schwarz MI, Brown KK. Changes in clinical and physiologic variables predict survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2003;168:538–542.
- Flaherty KR, Mumford JA, Murray S, Kazerooni EA, Gross BH, Colby TV, Travis WD, Flint A, Toews GB, Lynch JP, et al. Prognostic implications of physiologic and radiographic changes in idiopathic interstitial pneumonia. *Am J Respir Crit Care Med* 2003;168:543–548.
- Jegal Y, Kim DS, Shim TS, Lim CM, Do Lee S, Koh Y, Kim WS, Kim WD, Lee JS, Travis WD, et al. Physiology is a stronger predictor of survival than pathology in fibrotic interstitial pneumonia. *Am J Respir Crit Care Med* 2005;171:639–644.
- King TE Jr, Safrin S, Starko KM, Brown KK, Noble PW, Raghu G, Schwartz DA. Analyses of efficacy end points in a controlled trial of interferon gamma-1b for idiopathic pulmonary fibrosis. *Chest* 2005;127:171–177.
- Zappala CJ, Latsi PI, Nicholson AG, Colby TV, Cramer D, Renzoni EA, Hansell DM, du Bois RM, Wells AU. Marginal decline in FVC is associated with a poor outcome in idiopathic pulmonary fibrosis. *Eur Respir J* 2009;35:830–836.
- du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, Lancaster L, Noble PW, Raghu G, Sahn SA, et al. Ascertainment of individual risk of mortality for patients with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2011;184:459–466.
- Guyatt GH, Osoba D, Wu AW, Wyrich KW, Norman GR. Methods to explain the clinical significance of health measures. *Mayo Clin Proc* 2002;77:371–383.
- Holland AE, Hill CJ, Conron M, Munro P, McDonald CF. Small changes in six-minute walk distance are important in diffuse parenchymal lung disease. *Respir Med* 2009;103:1430–1435.
- Man-Son-Hing M, Laupacis A, O'Rourke K, Molnar FJ, Mahon J, Chan KB, Wells G. Determination of the clinical importance of study results. *J Gen Intern Med* 2002;17:469–476.
- Schunemann HJ, Guyatt GH. Commentary—goodbye M(C)ID! hello MID, where do you come from? *Health Serv Res* 2005;40:593–597.
- du Bois RM, Albera C, Bradford WZ, Costabel U, Kartashov A, Noble PW, Swarcberg J, Thomeer M, Valeyre D, Weycker D, et al. Percent predicted forced vital capacity is a reliable, valid, and responsive measure of clinical status in patients with idiopathic pulmonary fibrosis [abstract]. European Respiratory Society Annual Congress 2010. Barcelona, Spain, September 18–22. A3632.
- Raghu G, Brown KK, Bradford WZ, Starko K, Noble PW, Schwartz DA, King TE Jr. A placebo-controlled trial of interferon gamma-1b in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 2004;350:125–133.
- King TE Jr, Albera C, Bradford WZ, Costabel U, Hormel P, Lancaster L, Noble PW, Sahn SA, Swarcberg J, Thomeer M, et al. Effect of interferon gamma-1b on survival in patients with idiopathic pulmonary fibrosis (INSPIRE): a multicentre, randomised, placebo-controlled trial. *Lancet* 2009;374:222–228.
- Eakin EG, Resnikoff PM, Prewitt LM, Ries AL, Kaplan RM. Validation of a new dyspnoea measure: the UCSD shortness of breath questionnaire. University of California, San Diego. *Chest* 1998;113:619–624.
- Chang JA, Curtis JR, Patrick DL, Raghu G. Assessment of health-related quality of life in patients with interstitial lung disease. *Chest* 1999;116:1175–1182.
- Cohen J. Statistical power analysis for behavioral sciences, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
- Allison PD. Survival analysis using the SAS System: a practical guide. Cary, NC: SAS Institute Inc., 1995. pp. 292.
- Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care* 1999;37:469–478.
- Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol* 1999;52:861–873.

28. Ries AL. Minimally clinically important difference for the UCSD Shortness of Breath Questionnaire, Borg Scale, and Visual Analog Scale. *COPD* 2005;2:105–110.
29. Nichol MB, Epstein JD. Separating gains and losses in health when calculating the minimum important difference for mapped utility values. *Qual Life Res* 2008;17:955–961.
30. Yost KJ, Eton DT. Combining distribution- and anchor-based approaches to determine minimally important differences. *Eval Health Prof* 2005;28:172–191.
31. Walters SJ, Brazier JE. What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. *Health Qual Life Outcomes* 2003;1:4.
32. du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, Lancaster L, Noble PW, Sahn SA, Szwarzberg J, *et al.* Six-minute walk test in idiopathic pulmonary fibrosis: test validation and minimal clinically important difference. *Am J Respir Crit Care Med* 2011;183:1231–1237.