An updated approach to determine minimal clinically important differences in idiopathic pulmonary fibrosis

Mohleen Kang ⓘ, Srihari Veeraraghavan, Greg S. Martin and Jordan A. Kempker ⓘ

Emory University School of Medicine, Division of Pulmonary, Allergy, Critical Care and Sleep Medicine, Atlanta, GA, USA.

Corresponding author: Mohleen Kang (mkang30@emory.edu)

Abstract

Introduction Current medications for idiopathic pulmonary fibrosis (IPF) have not been shown to impact patient-reported outcome measures (PROMs), highlighting the need for accurate minimal clinically important difference (MCID) values. Recently published consensus standards for MCID studies support using anchor-based over distribution-based methods. The aim of this study was to estimate MCID values for worsening in IPF using only an anchor-based approach.

Methods We conducted secondary analyses of three randomised controlled trials with different inclusion criteria and follow-up intervals. The health transition question in the 36-Item Short-Form Health Survey (SF-36) questionnaire was used as the anchor. We used receiver operating curves to assess responsiveness between the anchor and 10 variables (four physiological measures and six PROMs). We used an anchor-based method to determine the MCID values of variables that met the responsiveness criteria (area under the curve ⩾0.70).

Results 6-min walk distance (6MWD), the St George's Respiratory Questionnaire (SGRQ), physical component score (PCS) of SF-36 and University of California, San Diego, Shortness of Breath Questionnaire (UCSD SOBQ) met the responsiveness criteria. The MCID value for 6MWD was −75 m; the MCID value for SF-36 PCS was −7 points; the MCID value for SGRQ was 11 points; and the MCID value for the UCSD SOBQ was 11 points.

Conclusions The MCID estimates of 6MWD, SGRQ, SF-36 and UCSD SOBQ using only anchor-based methods were considerably higher compared to previously proposed values. A single MCID value may not be applicable across all classes of disease severity or durations of follow-up time.

Introduction

Idiopathic pulmonary fibrosis (IPF) is a chronic fibrosing lung disease that is progressive and has a median survival of 2–3 years after diagnosis [1]. The disease progression is associated with increased symptom burden and is punctuated by episodic acute exacerbations that can lead to hospitalisation and acute respiratory failure. Mortality and hospitalisation are meaningful but challenging primary end-points in IPF, as they require large sample sizes and long follow-up periods [2]. Therefore, measures of lung function such as forced vital capacity (FVC) are more feasible end-points for drug trials. There are currently two pharmacological treatment options, pirfenidone and nintedanib, which have been shown to decrease the rate of annual decline of FVC [3–6]. However, neither of these medications has shown an impact on patient-reported outcome measures (PROMs) as measured by the St George's Respiratory Questionnaire (SGRQ) or the University of California, San Diego Shortness of Breath Questionnaire (UCSD SOBQ). This raises an important issue as to what minimal clinically important difference (MCID) in outcome measures such as FVC and SGRQ would be associated with clinically meaningful change in patients.

MCID is a threshold value for a change in a measure considered meaningful by the patient and which, per JAESCHKE *et al.* [7], who first defined the concept in 1989, "would mandate, in the absence of troublesome side-effects and excessive cost, a change in patient management". MCID is often used in trial design to estimate effect size for sample size calculation and in evaluating the clinical importance of trial results. For instance, a statistically significant difference in a primary end-point such as FVC between the treatment and control groups may not be clinically important for patients if it falls below the MCID value of that primary end-point. MCID values have traditionally been determined by three different methods: anchor-based, distribution-based and expert opinion. The anchor-based methods estimate MCID as the quantity of change in a measure that is associated with patient's report of minimal improvement or worsening, *i.e.* the anchor. Distribution-based methods use statistical methods to determine the minimal change that can be detected beyond statistical error without incorporating patient input. Expert opinion incorporates formal or informal clinician judgements as the MCID value. While there is no gold-standard methodology to determine MCID values, proposed tools and consensus approaches support anchor-based over distribution-based methods [8, 9].

Among the 10 articles that have studied MCID values of various measures in IPF, there are some limitations [10–19]. Nine out of the 10 studies utilised distribution-based methods to calculate MCID [10–17, 19]. Distribution-based methods do not incorporate patient input and, therefore, may not necessarily reflect patient-centred differences [20, 21]. Additionally, while these studies also used anchor-based methods, some of the studies used mortality and or hospitalisation as anchors, which, while clinically important to patients, may determine "maximal" rather than "minimal" important changes [14–16]. Similarly, physiological measures, such as FVC, do not incorporate patient input about change and may be less than ideal when used as sole anchors in a study [10–13, 18]. The overall aim of this exploratory study is to estimate the MCID values of various physiological measures and PROMs in three different IPF cohorts using only an anchor-based approach consistent with the core criteria of the Minimally Important Difference Credibility Assessment Tool (©2018, McMaster University) developed for evaluating anchor-based MCID studies [9]. We hypothesised that for a chronic progressive lung disease like IPF, most patients would either be unchanged or worsened at the end of the specified follow-up period. Therefore, we calculated MCID values associated with patient worsening only.

Methods

Data sources

We conducted secondary analyses of data from three randomised controlled trials: Sildenafil Trial of Exercise Performance in Idiopathic Pulmonary Fibrosis (STEP-IPF), AntiCoagulant Effectiveness in Idiopathic Pulmonary Fibrosis (ACE-IPF) and Prednisone, Azathioprine, and N-Acetylcysteine: A Study That Evaluates Response in Idiopathic Pulmonary Fibrosis (PANTHER-IPF) [22–24]. These three IPFnet trials were conducted by the same clinical trials group and around the same time period with similar diagnostic and adjudication process [25]. Data from these trials was obtained from the National Heart, Lung, and Blood Institute *via* the Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC) programme. While each of these trials enrolled patients with IPF, each had different inclusion criteria and study duration: 1) the STEP-IPF trial followed patients with severe lung function impairment for 24 weeks; 2) ACE-IPF followed patients with a progressive phenotype for 48 weeks; and 3) the PANTHER-IPF trial followed patients with mild to moderate impairment for 60 weeks (supplementary table S1). Given that the three studies had different inclusion and exclusion criteria and different follow-up time periods, three separate analyses following the same procedures were conducted for each. We used both placebo and treatment arm patients in our analysis.

Study measures

For our anchor, we selected the health transition question (SF2) in the 36-Item Short-Form Health Survey (SF-36). SF2 asks the patients to rate their health on a 5-point Likert scale in response to the following question: "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) "about the same", 4) "somewhat worse" and 5) "much worse" [26]. The SF2 is a general question that has been used in MCID determination in other studies and meets the requirements of patient-reported anchor proposed by the Minimally Important Difference Credibility Assessment Tool [9, 15]. Since it is not specific for a domain such as dyspnoea or physical function, SF2 is a suitable anchor for all the measures of interest in the analysis. In addition, it was available for all three studies and for all follow-up intervals. Data for other possible anchors, such as one of the PROMs or subdomains of PROMs, were not available for all study cohorts. We analysed patients with complete SF2 data at the end of the respective study follow-up time period.

The physiological measures included in our analysis were FVC, total lung capacity (TLC), diffusing capacity of the lung for carbon monoxide (D_{LCO}) and 6-min walk distance (6MWD). We evaluated both absolute change in percentage of predicted FVC and FVC in litres (L) separately. Additionally, we analysed relative change in FVC in L which was expressed as a percentage. For D_{LCO} we evaluated absolute difference in percentage of predicted D_{LCO} and D_{LCO} measured as mL·min^{-1}·mmHg^{-1}. The STEP-IPF dataset obtained from BioLINCC did not include percentage of predicted values for FVC and D_{LCO}. We used National Health and Nutrition Examination Survey spirometry reference values to compute percentage of predicted values for FVC for the STEP-IPF cohort [27]. Percentage of predicted values for D_{LCO} were not computed for the STEP-IPF cohort. For TLC, the absolute difference in TLC in L was analysed in the ACE-IPF and PANTHER-IPF cohorts. The TLC values were not available in the STEP-IPF dataset. For 6MWD, we analysed absolute difference in 6MWD in metres.

The PROMs we examined included Borg dyspnoea scale, SF-36 physical and mental component scores, EuroQol score index and visual analogue scores, SGRQ, UCSD SOBQ and Investigating Choice Experiences for the Preferences of Older People Capability Instruments for Adults (ICECAP) questionnaire. The STEP-IPF data set did not include total scores for SGRQ, SF-36 physical and mental components, UCSD SOBQ, EuroQoL index and visual analogue scale or ICECAP questionnaire. We calculated the total scores for UCSD SOBQ and the EuroQol index and visual analogue scale (using the SAS code provided by the EuroQol Group). We were unable to compute total scores for SGRQ, SF-36 and ICECAP in the STEP-IPF, cohort due to missing components.

Statistical analysis

All analyses were conducted using SAS (version 9.4; SAS Institute, Cary, NC, USA) and IBM SPSS Statistics (version 26; SPSS, Chicago, IL, USA). All analyses were conducted using observed cases. If patients had missing data at follow-up, then those patients were not included in the MCID analysis. We initially performed descriptive univariate analyses for each patient measure retaining all outliers in the analysis. We calculated mean change between follow-up and baseline (score difference) of each measure for patients in each of the categories in the SF2 question.

For MCID calculation we followed a step-wise approach detailed in the supplementary material. Briefly, we assessed responsiveness of each measure with SF2 by using receiver operating curve analysis. Only those measures that met the criteria for responsiveness, *i.e.* area under the curve (AUC) ⩾0.70, were selected for MCID estimation. We determined the score difference of the measures from baseline to follow-up in patients who answered "somewhat worse" in response to SF2 as the MCID.

Results

Baseline characteristics

140 patients had follow-up data at 24 weeks in the STEP-IPF cohort; 111 patients had follow-up data at 48 weeks in the ACE-IPF cohort; and 228 patients had follow-up data at 60 weeks in the PANTHER-IPF trial. Participants from all three cohorts were predominantly male (71–81%) and white (92–96%). The STEP-IPF cohort had a mean±SD age of 68.47±9.11 years with mean±SD FVC of 58.52±15.50% pred and mean±SD D_{LCO} of 7.92±2.12 mL·min^{-1}·mmHg^{-1} (supplementary table S2). The ACE-IPF cohort had a mean±SD age of 66.65±7.49 years with a mean±SD FVC of 61.94±15.19% pred and mean±SD D_{LCO} of 36.16±12.90% pred (supplementary table S3). The PANTHER-IPF cohort had mean±SD age of 67.05±8.32 years with a mean±SD FVC of 73.81±15.05% pred and D_{LCO} of 46.18±11.36% pred (supplementary table S4).

Response to anchor SF2

In the STEP-IPF cohort, 110 (78.6%) out of the 140 patients were either in the "about the same" or in the "somewhat worse" category according to SF2 response at follow-up (supplementary table S5). 6MWD was the only measure in the STEP-IPF cohort that met the responsiveness criteria (AUC ⩾0.70) for further MCID estimation (table 1). The AUC for other physiological measures and PROMs in the STEP-IPF ranged from 0.55 to 0.68 (supplementary table S5). In the ACE-IPF cohort, 98 (88.3%) out of the 111 patients with follow-up data at 48 weeks answered "about the same" or "somewhat worse" in response to the SF2 question at 48 weeks (supplementary table S6). None of the physiological measures or the PROMs in the ACE-IPF cohort met the pre-specified responsiveness criteria for further MCID determination with AUC ranging from 0.53 to 0.61 (supplementary table S6). In the PANTHER-IPF cohort, 175 (76.8%) out of 228 patients answered "about the same" or "somewhat worse" in response to the SF2 question (supplementary table S7). In the PANTHER-IPF cohort, the physical component score (PCS) of the SF-36 questionnaire, the total SGRQ and UCSD SOBQ scores were the only measures that met criteria for next

TABLE 1 Change in measures that met responsiveness criteria over 24 weeks by health transition question (SF2) categorical responses in patients with idiopathic pulmonary fibrosis in the Sildenafil Trial of Exercise Performance in Idiopathic Pulmonary Fibrosis (STEP-IPF) trial

	SF2 response at 24-week follow-up					AUC (95% CI)[¶]
	Much better	Somewhat better	About the same	Somewhat worse[#]	Much worse	
Subjects n	7	12	46	58	9	
Physiological measures						
Δ6MWD m	−3.71±42.33	−35.08±144.51	−32.29±95.46	−74.89±69.32	−163.00±122.93	0.72 (0.61–0.83)

Data are presented as mean±SD, unless otherwise stated. AUC: area under the curve; Δ: absolute change over 24 weeks; 6MWD: 6-min walk distance. [#]: the mean change in scores in the "somewhat worse" category is the minimal clinically important difference value for measures that meet the responsiveness criteria; [¶]: receiver operating curve comparing dichotomous health transition question (SF2) response ("about the same" *versus* "somewhat worse") with mean change score of variable.

stage of MCID calculation (table 2). The AUC for other physiological measures and PROMs in the PANTHER-IPF ranged from 0.47 to 0.69 (supplementary table S7).

Anchor-based MCID values for worsening

The following measures did not meet responsiveness criteria (AUC ⩾0.70) in any of the cohorts: FVC, TLC, D_{LCO}, Borg dyspnoea score, SF-36 mental component score, EuroQol score index and visual analogue scores, and ICECAP scores. Therefore, no MCID values were determined for these measures. 6MWD met the responsiveness criteria only in the STEP-IPF cohort; therefore, the MCID values for 6MWD was determined only at 24 weeks. SGRQ, SF-36 PCS and UCSD SOBQ met the responsiveness criteria only in the PANTHER-IPF cohort; therefore, MCIDs were determined only at 60 weeks' interval for these measures. The mean change from baseline to follow-up (24 weeks for 6MWD and 60 weeks for the other three measures) in patients who answered "somewhat worse" in response to SF2 was selected as the MCID. MCID value for 6MWD was −74.89 m (95% CI −93.11– −56.66 m) over 24 weeks; MCID value for SF-36 PCS over 60 weeks was −6.79 points (95% CI −8.66– −4.92 points); MCID value for total SGRQ score over 60 weeks was 10.95 points (95% CI 7.81–14.1 points); and MCID value for the total UCSD SOBQ score over 60 weeks was 11.38 points (95% CI 7.83–14.93 points) (table 3).

Discussion

This is the first study in IPF to conduct a comprehensive exploratory analysis of multiple physiological measures and PROMs in three different cohorts using only an anchor-based approach consistent with recently proposed standards in the MCID literature, and it demonstrates several key points [9]. First, the MCID estimates of 6MWD, SGRQ, SF-36 and UCSD SOBQ were higher than previously calculated point estimates. These previous studies not only used different methodology, but in most instances, conducted their analyses on patients with different baseline disease severity and with different follow-up intervals, which makes direct comparison difficult. Second, in our analysis, no one measure met responsiveness

TABLE 2 Change in measures that met responsiveness criteria over 60 weeks by health transition question (SF2) categorical responses in patients with idiopathic pulmonary fibrosis in the Prednisone, Azathioprine, and N-Acetylcysteine: A Study That Evaluates Response in Idiopathic Pulmonary Fibrosis (PANTHER-IPF) trial

	SF2 response at 60-week follow-up					AUC (95% CI)[¶]
	Much better	Somewhat better	About the same	Somewhat worse[#]	Much worse	
Subjects n	11	32	101	74	10	
Physiological measures						
ΔSF-36 physical component score[+]	8.00±11.57	−0.92±6.26	−0.06±5.72	−6.79±8.07	−5.29±9.26	0.75 (0.67–0.83)
Subjective measures						
ΔTotal SGRQ score[§]	−13.70±14.55	−0.97±9.60	1.35±9.18	10.95±13.19	25.86±17.46	0.71 (0.63–0.79)
ΔUCSD SOBQ total score	−4.45±11.79	3.19±9.60	0.59±12.81	11.38±15.31	34.90±23.65	0.72 (0.65–0.80)

Data are presented as mean±SD, unless otherwise stated. AUC: area under the curve; Δ: absolute change over 60 weeks; SF-36: 36-Item Short-Form Health Survey; SGRQ: St George's Respiratory Questionnaire; UCSD SOBQ: University of California, San Diego Shortness of Breath Questionnaire. [#]: the mean change in scores in the "somewhat worse" category is the minimal clinically important difference value for measures that meet the responsiveness criteria; [¶]: receiver operating curve comparing dichotomous health transition question (SF2) response ("about the same" *versus* "somewhat worse") with mean change score of variable; [+]: missing n=5 (0 "much better", 1 "somewhat better", 4 "about the same", 0 "somewhat worse", 0 "much worse"); [§]: missing n=13 (0 "much better", 2 "somewhat better", 8 "about the same", 3 "somewhat worse", 0 "much worse").

TABLE 3 Anchor-based estimates of minimal clinically important difference (MCID) for worsening in idiopathic pulmonary fibrosis from current study and comparison with MCID estimates from previous studies

	Current MCID estimate		Previous MCID estimate[#]	
	Time period[¶]	MCID[+] (95% CI)	Time period [ref.]	MCID
6MWD m	24 weeks	−74.89 (−93.11– −56.66)	48 weeks [14]	24–45[§]
			48 weeks [16]	21.7 –37.0[§]
			52 weeks [11]	28
SF-36 physical component score	60 weeks	−6.79 (−8.66– −4.92)	26 weeks [13]	3
			No specified time period [10]	5
Total SGRQ score	60 weeks	10.95 (7.81–14.1)	26 weeks [13]	7 (5–10)[##]
			52 weeks [12]	4–5
UCSD SOBQ total score	60 weeks	11.38 (7.83–14.93)	24 weeks [19]	8 (5–11)[##]

6MWD: 6-min walk distance; SF-36: 36-Item Short-Form Health Survey; SGRQ: St George's Respiratory Questionnaire; UCSD SOBQ: University of California, San Diego Shortness of Breath Questionnaire. [#]: previous studies have estimated MCID values for worsening and improvement using both anchor-based and distribution-based methods; [¶]: 6MWD met responsiveness criteria only in the Sildenafil Trial of Exercise Performance in Idiopathic Pulmonary Fibrosis (STEP-IPF) cohort, while the SF-36 physical component score, SGRQ and UCSD SOBQ met criteria in only the Prednisone, Azathioprine, and N-Acetylcysteine: A Study That Evaluates Response in Idiopathic Pulmonary Fibrosis (PANTHER-IPF) cohort; therefore, time period for MCID is 24 weeks or 60 weeks; [+]: MCID for worsening only, since it was calculated as mean change in "somewhat worse" health transition question (SF2) group from baseline to follow-up; [§]: range; [##]: point estimate (range).

criteria in more than one cohort. Third, the variable FVC, the primary end point in major trials, did not meet responsiveness criteria in any of the three cohorts. This variation in responsiveness of outcome measures may be due to random chance, different duration of follow-up compared to the anchor, study procedures or bias; or some combination of them all. Our findings demonstrate the complexities of MCID calculation, which has large implications for trial design and evaluation.

Our study's results must be understood in the context of its limitations. First, the different time periods for the three trial cohorts limited an analysis of a combined cohort and thereby restricted the sample size for the analysis. In addition, the lack of similar follow-up time limited our ability to validate MCID values from one cohort in another cohort. Further studies in other trial cohorts using a similar anchor-based approach are needed to verify and validate the results of our study. Second, we used a single anchor, SF2, for our analysis. While general transition rating questions such as SF2 have been widely used as anchors, the results of our analysis should be confirmed with other anchors [9, 20]. Additionally, SF2 asks patient to recall their general health over the past year which makes it prone to recall bias, and using this to anchor changes over other time periods may not be ideal. Third, the anchor-based method used in our analysis is prone to regression to the mean [28, 29]. There are no clear guidelines on which anchor-based methods to use in estimating MCID [9]. Additional consensus recommendations on the most accurate and precise anchor-based methods are warranted and would lead to further standardisation of the MCID calculation. Finally, our study assessed responsiveness and estimated MCID, but did not assess the validity or psychometric properties of these measures. Previous studies have evaluated convergent validity and some psychometric properties of 6MWD, SF-36, SGRQ and UCSD SOBQ in IPF [11–13, 19, 30, 31]. Even with these limitations, the MCID values estimated in our analysis represent some significant methodological strengths over prior IPF work.

We used a systematic approach consistent with the core criteria proposed by the recently published Minimally Important Difference Credibility Assessment Tool to evaluate MCID studies, with one notable methodologic exception [9]. Most MCID studies and the above-mentioned credibility instrument propose using a correlation coefficient (usually $\geqslant 0.3$ or 0.5) to assess responsiveness of the change in the measure with the anchor [28, 29]. This approach is suitable for diseases such as chronic pain where patients are expected to be categorised somewhat evenly into the 5-point Likert scale categories of an anchor like SF2. However, in a chronic progressive disease like IPF, most patients may fall into only two of the five anchor categories, as was the case in our analysis. Therefore, using correlation coefficient may not accurately identify variables that are responsive to the anchor. Given the imbalance in categories, which was seen in all three cohorts in our study, the receiver operating curve analysis with AUC $\geqslant 0.70$ was used to assess responsiveness of variables to a dichotomous anchor [32, 33].

Compared to previous MCID studies in IPF, we did not use distribution-based methods in our calculation. Since distribution-based methods do not take into account a patient's report of their health, they essentially report the minimal detectable change. However, minimal detectable change and MCID are two different

concepts, as illustrated by DE VET and TERWEE [21]. Previous MCID studies in IPF have used distribution-based methods along with anchor-based methods and have reported lower point estimates when compared to our calculated values (table 3). The MCID estimate for 6MWD at 75 m in our analysis is much higher compared to previously reported values ranging from 21.7–45 m [11, 14, 16]. The estimate for SF-36 PCS of 7 points is also higher when compared to previous values of 3 points and 5 points [10, 13]. While the difference in baseline disease severity and follow-up intervals in some of the previous studies makes direct comparison difficult, in certain cases our MCID values fall within the reported ranges of previous studies even if they are higher than the point estimates. For instance, only one study thus far has determined MCID estimates of total UCSD SOBQ scores and used the STEP-IPF cohort for their analysis [19]. They reported an MCID estimate of 8 points for both improvement and worsening with a range of 5–11 over 24 weeks using SGRQ's activity domain for anchor-based method along with distribution-based methods [19]. The UCSD SOBQ score did not meet responsiveness criteria in our analysis of STEP-IPF cohort, but our reported anchor-based MCID values for UCSD SOBQ at 11.38 points over a 60-week time period using mild to moderate disease patients of the PANTHER-IPF trial is close to the reported range of 5–11 points in the previous study.

Similarly, an earlier study reported an MCID of SGRQ as 7 points with a range of 5–10 points using both anchor-based and distribution-based methods in IPF patients with mild to moderate severity [13]. In our analysis, we estimated higher MCID of SGRQ of 10.95 over 60-week time period for worsening using a similar mild to moderate category of patients, which again is within the range of the previous study, but higher than the reported point estimate. However, another more recent study estimated MCID for SGRQ in IPF using mild to moderate severity patients over 52 weeks and proposed a threshold of 4–5 points for both improvement and worsening using both distribution and anchor-based methods and is much lower than our estimate [12]. Further research is needed to study the impact of MCID methodology, disease severity and follow-up interval on MCID estimation and there are efforts underway to study some of these relationships in other diseases such as asthma [34]. A study of MCID of three questionnaires including SGRQ in COPD patients found stable MCID values over different follow-up intervals ranging from 3 weeks to 12 months [35]. A large real-world dataset of IPF patients, such as the newer patient registries, with patients of varying disease severity and multiple follow-up measurements at set intervals may be useful for standardised MCID research of physiological measures and PROMs, provided they have appropriate anchors for MCID estimation [36, 37].

Conclusions

Our study highlights the fact the anchor-based MCID estimates of 6MWD, SGRQ, SF-36 and UCSD SOBQ in our study were considerably higher when compared to point estimates from previously proposed values. Further research is needed to assess MCID values of various physiologic measures and PROMs in IPF using a more current and standardised approach in different patient cohorts over different time periods to better design and evaluate clinical trials. There is further need to establish MCID of newer physiologic measures such as home spirometry and actigraphy [38, 39]. In addition, PROMs designed specifically for IPF patients are needed to better capture the patient experience in clinical trials, since PROMs such as SGRQ were developed for patients with obstructive diseases. The newly proposed Living with Idiopathic Pulmonary Fibrosis questionnaire is one such endeavour to better incorporate the patient experience [40]. With these advances, future intervention trials in IPF may be better poised to accurately evaluate patient quality of life.

References

1   Raghu G, Richeldi L. Current approaches to the management of idiopathic pulmonary fibrosis. *Respir Med* 2017; 129: 24–30.

2   Collard HR, Brown KK, Martinez FJ, *et al.* Study design implications of death and hospitalization as end points in idiopathic pulmonary fibrosis. *Chest* 2014; 146: 1256–1262.

3   Raghu G, Rochwerg B, Zhang Y, *et al.* An official ATS/ERS/JRS/ALAT clinical practice guideline: treatment of idiopathic pulmonary fibrosis. An update of the 2011 clinical practice guideline. *Am J Respir Crit Care Med* 2015; 192: e3–e19.

4   Noble PW, Albera C, Bradford WZ, *et al.* Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. *Lancet* 2011; 377: 1760–1769.

5   Richeldi L, du Bois RM, Raghu G, *et al.* Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med* 2014; 370: 2071–2082.

6   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: 2083–2092. .

7   Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; 10: 407–415.

8   McGlothlin AE, Lewis RJ. Minimal clinically important difference. *JAMA* 2014; 312: 1342–1343.

9   Devji T, Carrasco-Labra A, Qasim A, *et al.* Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. *BMJ* 2020: m1714.

10   Witt S, Krauss E, Barbero MAN, *et al.* Psychometric properties and minimal important differences of SF-36 in idiopathic pulmonary fibrosis. *Respir Res* 2019; 20: 47.

11   Swigris JJ, Wamboldt FS, Behr J, *et al.* The 6 minute walk in idiopathic pulmonary fibrosis: longitudinal changes and minimum important difference. *Thorax* 2010; 65: 173–177.

12   Swigris JJ, Wilson H, Esser D, *et al.* Psychometric properties of the St George's Respiratory Questionnaire in patients with idiopathic pulmonary fibrosis: insights from the INPULSIS trials. *BMJ Open Respir Res* 2018; 5: e000278.

13   Swigris JJ, Brown KK, Behr J, *et al.* The SF-36 and SGRQ: validity and first look at minimum important differences in IPF. *Respir Med* 2010; 104: 296–304.

14   du Bois RM, Weycker D, Albera C, *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.

15   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.

16   Nathan SD, du Bois RM, Albera C, *et al.* Validation of test performance characteristics and minimal clinically important difference of the 6-minute walk test in patients with idiopathic pulmonary fibrosis. *Respir Med* 2015; 109: 914–922.

17   Nolan CM, Delogu V, Maddocks M, *et al.* Validity, responsiveness and minimum clinically important difference of the incremental shuttle walk in idiopathic pulmonary fibrosis: a prospective study. *Thorax* 2017; 73: 680–682.

18   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.

19   Swigris JJ, Han M, Vij R, *et al.* The UCSD shortness of breath questionnaire has longitudinal construct validity in idiopathic pulmonary fibrosis. *Respir Med* 2012; 106: 1447–1455.

20   Terwee CB, Roorda LD, Dekker J, *et al.* Mind the MIC: large variation among populations and methods. *J Clin Epidemiol* 2010; 63: 524–534.

21   de Vet HC, Terwee CB. The minimal detectable change should not replace the minimal important difference. *J Clin Epidemiol* 2010; 63: 804–805.

22   Noth I, Anstrom KJ, Calvert SB, *et al.* A placebo-controlled randomized trial of warfarin in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2012; 186: 88–95.

23   Idiopathic Pulmonary Fibrosis Clinical Research Network, Raghu G, Anstrom KJ, *et al.* Prednisone, azathioprine, and N-acetylcysteine for pulmonary fibrosis. *N Engl J Med* 2012; 366: 1968–1977.

24   Idiopathic Pulmonary Fibrosis Clinical Research Network, Zisman DA, Schwarz M, *et al.* A controlled trial of sildenafil in advanced idiopathic pulmonary fibrosis. *N Engl J Med* 2010; 363: 620–628.

25 de Andrade J, Schwarz M, Collard HR, *et al.* The Idiopathic Pulmonary Fibrosis Clinical Research Network (IPFnet): diagnostic and adjudication processes. *Chest* 2015; 148: 1034–1042.

26 Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992; 30: 473–483.

27 Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general US population. *Am J Respir Crit Care Med* 1999; 159: 179–187.

28 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: 128–136.

29 Malec JF, Ketchum JM. A standard method for determining the minimal clinically important difference for rehabilitation measures. *Arch Phys Med Rehabil* 2020; 101: 1090–1094.

30 Swigris JJ, Esser D, Wilson H, *et al.* Psychometric properties of the St George's Respiratory Questionnaire in patients with idiopathic pulmonary fibrosis. *Eur Respir J* 2017; 49: 1601788.

31 Swigris JJ, Streiner DL, Brown KK, *et al.* Assessing exertional dyspnea in patients with idiopathic pulmonary fibrosis. *Respir Med* 2014; 108: 181–188.

32 de Vet HC, Terwee CB, Mokkink LB, *et al.* Measurement in Medicine: A Practical Guide. Cambridge, Cambridge University Press, 2011.

33 Prinsen CAC, Mokkink LB, Bouter LM, *et al.* COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018; 27: 1147–1157.

34 Lanario J, Jones R, Hyland M, *et al.* Is the minimally clinically important difference (MCID) fit for purpose? A planned study using the SAQ. *Eur Respir J* 2020; 56: Suppl. 64, 2241.

35 Alma H, de Jong C, Jelusic D, *et al.* The minimal clinically important difference for COPD health status tools measured with global ratings of change during different time periods. *Eur Respir J* 2016; 48: Suppl. 60, PA873.

36 Guenther A, Krauss E, Tello S, *et al.* The European IPF registry (eurIPFreg): baseline characteristics and survival of patients with idiopathic pulmonary fibrosis. *Respir Res* 2018; 19: 141.

37 Wang BR, Edwards R, Freiheit EA, *et al.* The pulmonary fibrosis foundation patient registry: rationale, design, and methods. *Ann Am Thorac Soc* 2020; 17: 1620–1628.

38 Russell AM, Adamali H, Molyneaux PL, *et al.* Daily home spirometry: an effective tool for detecting progression in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. 2016; 194: 989–997.

39 Nathan SD, Flaherty KR, Glassberg MK, *et al.* A randomized, double-blind, placebo-controlled study of pulsed, inhaled nitric oxide in subjects at risk of pulmonary hypertension associated with pulmonary fibrosis. *Chest* 2020; 158: 637–645.

40 Swigris JJ, Andrae DA, Churney T, *et al.* Development and initial validation analyses of the Living with Idiopathic Pulmonary Fibrosis questionnaire. *Am J Respir Crit Care Med* 2020; 202: 1689–1697.