Patient-Reported Outcomes in Idiopathic Pulmonary Fibrosis Research

Jeffrey J. Swigris, DO; and Diane Fairclough, DrPH

Patient-reported outcomes (PROs) include questionnaires or surveys that ask patients for their perceptions about things like symptoms they are experiencing or quality of life. For incurable, morbid, life-shortening conditions like idiopathic pulmonary fibrosis (IPF), PROs are particularly germane: They elucidate for clinicians and researchers what it is like for patients to live with such a disease, and they may detect important treatment effects not captured by other metrics (eg, pulmonary physiology). However, a relative paucity of research on PROs in IPF has left significant knowledge gaps in this area and contributed to the timidity investigators have about using PROs as prominent outcomes in IPF drug trials. Additional research on existing instruments is needed to establish or bolster their basic psychometric properties in IPF. When PROs are used as end points in therapeutic trials, analyzing PRO response data can be challenging, but these challenges can be overcome with a transparent, thoughtful, and sophisticated statistical approach. In this article, we discuss some of the basics of PRO assessment, existing knowledge gaps in IPF-related PRO research, and the potential usefulness of using PROs in IPF trials and conclude by offering specific recommendations for an approach to analyzing repeated-measures PRO data from IPF trials.

CHEST 2012; 142(2):291–297

Abbreviations: ATAQ-IPF = a tool to assess quality of life in idiopathic pulmonary fibrosis; BDI/TDI = baseline dyspnea index/transition dyspnea index; FDA = US Food and Drug Administration; HRQL = health-related quality of life; ICR = internal consistency reliability; IPF = idiopathic pulmonary fibrosis; LOCF = last observation carried forward; MANOVA = multivariate analysis of variance; PRO = patient-reported outcome; QOL = quality of life; SF-36 = Medical Outcomes Study 36-Item Instrument; SGRQ = St. George Respiratory Questionnaire; UCSD = University of California San Diego Shortness of Breath Questionnaire

Patient-reported outcomes (PROs) seek to ascertain information about patients' experiences from their own perspective. Administered as questionnaires or surveys and completed by patients themselves, PROs attempt to quantify a person's perception of things such as health status, symptoms, or quality of life (QOL). PROs can reveal information patients might not disclose spontaneously,¹ thus strengthening their voices in the often overwhelming realm of their disease.

In idiopathic pulmonary fibrosis (IPF), PROs used in either patient-perspectives research settings or ther-

apeutic trials have included dyspnea indexes, cough questionnaires, health status instruments, health-related quality-of-life (HRQL) measures, depression or anxiety questionnaires, and sleep surveys. Although other outcomes (eg, FVC) yield information on disease stage and provide prognostic information, only PROs clue us in to what patients with IPF are actually experiencing and how, in turn, that experience affects any number of life domains. At the core of PROs, and the thing that distinguishes them from other outcomes, is their incorporation of patients' values and judgments—the traits most influential in shaping patients' perceptions of what it is like to live with

Manuscript received October 11, 2011; revision accepted January 20, 2012.

Affiliations: From the Autoimmune Lung Center and Interstitial Lung Disease Program (Dr Swigris), National Jewish Health; and the Department of Biostatistics and Informatics (Dr Fairclough), Colorado Health Outcomes Program, University of Colorado Denver, Anschutz Medical Center, Denver, CO.

Funding/Support: Dr Swigris is supported in part by a Career Development Award from the National Institutes of Health [Grant K23 HL092227].

Correspondence to: Jeffrey J. Swigris, DO, Autoimmune Lung Center and Interstitial Lung Disease Program, National Jewish Health, 1400 Jackson St, Denver, CO 80206; e-mail: swigrisj@njc.org

© 2012 American College of Chest Physicians. Reproduction of this article is prohibited without written permission from the American College of Chest Physicians. See online for more details. DOI: 10.1378/chest.11-2602

IPF.^{1,2} None of the primary outcomes commonly used in IPF trials has anything to do with patient values or judgments: FVC is important to patients only in so far as it quantifies disease severity; it is the downstream effects of a low FVC (eg, activity limitation, dyspnea), rather than the FVC value itself, that influence patients' perceptions of their disease.

In some IPF studies, PRO scores have correlated with other outcomes (eg, FVC); in other studies (and quite often within an individual), PRO scores do not correlate with other outcome measures. Consider a 64-year-old woman with severe IPF who is perfectly content if she is able to sit at her computer and e-mail her friends. Although her FVC is only 30% of predicted for her age and height, and she requires supplemental oxygen 24 h per day, she is not necessarily bothered by these, because her perception is that she is still able to do the one thing she values most—e-mail her friends. An outsider, even a physician caring for this patient, might ask how she can be at ease, even content, with her current situation. Although her FVC reveals she has very severe IPF, her values and judgments shape her perception that things are not too bad. Other patients with IPF with the same FVC and oxygen requirements may perceive things quite differently. Because each patient formulates his or her own unique perceptions, and because those perceptions (and scores from PROs that capture them) may not parallel other outcome measures, PROs can identify differences between groups of patients with IPF undetected by physiologic or radiologic measures.

PROs IN IPF TRIALS

For patients with life-shortening diseases that have no curative therapies (like IPF), on the face of it, PROs would seem to be the ideal outcome measures.¹ Without medicinal therapies to prolong the lives of patients with IPF, perhaps the most important objective for practitioners is to identify ways to improve the QOL of patients with IPF—to manage and limit dyspnea, cough, and fatigue; to enable them to be as physically and socially active as possible; and to ease the frustration and anguish they endure living with IPF.

There are PROs available to assess these things. However, to date, investigators have not used a PRO as a primary end point, or as a component of a composite primary end point, in a multicenter therapeutic trial in IPF. This likely stems from several factors, including the following: the incorrect perception among IPF trial designers that PROs are “soft” end points and, even if met, would not be recognized as meaningful by the US Food and Drug Administra-

tion (FDA) in the drug-approval process; because of the relative paucity of research on PROs in IPF, investigators have not developed confidence in what PRO scores really mean; and until recently, IPF-specific PROs had not been developed.

NEW FDA GUIDELINES OF PROs

Recently, the FDA began formalizing recommendations for how a PRO might qualify (in their view for the drug indication approval process) as a valid, reliable outcome measure whose scores have “interpretable meaning” and one whose usefulness need not be reconfirmed when used again in the same target population.³ To qualify, the PRO (or the idea for one) is submitted to the FDA, which essentially vets the instrument if various criteria are viewed as acceptable, including the following: (1) the overall goal of the PRO, (2) the conceptual framework for the attribute to be measured, (3) the context of its use, (4) the developmental process, (5) various measurement properties (eg, reliability, construct validity), (6) score interpretation, and (7) language translation and cultural adaptation.

WHICH PROs SHOULD BE USED IN IPF TRIALS?

It depends. To our knowledge, there are no FDA-qualified PROs for IPF. We are not suggesting that only FDA-qualified PROs be used in IPF research, but if qualified, researchers could at least be confident that certain fundamental psychometric properties of the PRO have been met. For several PROs used in IPF trials, these properties are not known (Table 1),⁴⁻¹⁰ but must be before their scores can be interpreted with confidence.

As seen in Table 1, we do not know a lot about how existing PROs perform in IPF, at least not in longitudinal studies. At a minimum, to be deemed useful in IPF, and to yield scores in whose meaning we can be highly confident, internal consistency reliability (ICR) and test-retest for the PRO should be proved acceptable, and data to support its validity for use as a longitudinal outcome measure in IPF must be published. The ICR (ie, Cronbach α)—a measure of the interrelatedness of items composing an instrument—for a multidimensional instrument, used in research, should be >0.7 , and Streiner and Norman¹¹ recommend that it should not exceed 0.9. A test-retest (intraclass correlation) coefficient >0.7 is considered psychometrically acceptable.^{1,12-14} For comparison, in IPF, the test-retest correlation coefficient for FVC is 0.93.¹⁵

Besides ensuring that ICR and test-retest are acceptable, PRO users must confirm that the aggregate

Table 1—Psychometric Properties of Various PROs Used in Published IPF Trials

PRO	Internal Consistency in IPF	Test-Retest in IPF	Evidence to Support Ability to Detect Change in IPF	Evidence to Support Construct Validity for Use in IPF
Dyspnea				
BDI/TDI ^{4,7}	Unknown	Unknown	Some	Some
Borg ^{4,8}	Unknown	Unknown	None	None
UCSD ^{5,8,9}	Unknown	Unknown	None	Some
QOL/HRQL				
EQ5D ^{7,8}	Unknown	Unknown	None	None
SF-36 ^{4,6,8}	Domains vary, 0.67-0.90	Unknown	Some	Some
SGRQ ^{4,5,6,8-10}	Domains vary, 0.66-0.85	Unknown	Some	Some

BDI/TDI = baseline dyspnea index/transition dyspnea index; Borg = Borg dyspnea scale; EQ5D = EuroQOL 5 dimension instrument; IPF = idiopathic pulmonary fibrosis; PRO = patient-reported outcome; QOL/HRQL = quality of life/health-related quality of life; SF-36 = Medical Outcomes Study 36-Item Instrument; SGRQ = St. George Respiratory Questionnaire; UCSD = University of California San Diego Shortness of Breath Questionnaire.

of items from a PRO match the expected levels of the attribute in the target population.¹ For example, a PRO that contains only items focused on the ability to perform very strenuous physical activity (eg, carry a heavy load one city block, running a mile, digging a ditch) would likely be irrelevant to many, if not most, patients with IPF. These items connote a level of physical functioning far beyond that of an average patient with IPF and would be too difficult for the majority of the IPF population.¹⁶ Such a scale would not be able to differentiate patients with IPF on physical functioning, because responses would almost certainly be skewed toward the “easiest” of all items. This skew, caused by lack of congruence between item difficulty and respondent ability, can also lead to floor and ceiling effects and poor sensitivity to change over time.¹

Investigators wanting to use a PRO in IPF research will need to assess whether there are data to support its validity for use in IPF. Validation is the process of building understanding in the meaning of the scores of a PRO so they can be used to make accurate inferences about a person in the target population. It takes time and hypothesis testing, in multiple studies, to amass data to support validity of a PRO. As more studies that include a PRO are conducted, greater understanding of (and confidence in) the meaning of different scores from the PRO is gained. A mistake that has been perpetuated in the medical literature is that one cross-sectional correlation study “validates” an instrument.¹ Data from such studies, assuming correlations are significant and in the expected (hypothesized) directions, may support the so-called “concurrent validity” of the instrument, but more data are needed before the instrument could be used confidently to assess change in longitudinal studies.^{11,17,18} According to Nunnally,¹⁹ “one validates not a measurement instrument but rather some use to which the instrument is put.” For the majority of PROs in Table 1, there are numerous studies whose results

support concurrent validity²⁰⁻²²; for example, patients with IPF who score high on the St. George Respiratory Questionnaire (SGRQ) or low on the Medical Outcomes Study 36-Item Instrument (SF-36) (both indicating impairments in HRQL) tend to have significant dyspnea.

Although gaps remain in our knowledge of PROs in IPF, recently, a growing interest in PRO research in IPF has generated limited data that support the validity of the baseline dyspnea index/transition dyspnea index (BDI/TDI), University of California San Diego Shortness of Breath Questionnaire (UCSD), SGRQ, and SF-36 for use as longitudinal outcomes in IPF trials.²³⁻²⁵ Although the question of how many data (or studies) are needed to validate an instrument ignores the basic premise that validation is a process, when a PRO is being considered for use as an outcome measure in a trial, we believe the ability of the PRO to track change over time needs to have been well demonstrated in at least one previous study (eg, using data from a previously conducted therapeutic trial or a longitudinal observational study) in the target population.

If a PRO has acceptable psychometric properties (ICR, test-retest reliability, and data to support its validity for the intended purpose in the target population), then investigators will need to decide whether it adequately addresses all the domains (or aspects of an individual domain) they are interested in examining in their study. For example, if dyspnea is the outcome of interest, do the investigators want to know its severity, how burdensome or intrusive it is, how it impacts emotional health, or some other aspect? If they are most interested in the effect of dyspnea on emotional well-being (ie, whether it induces anxiety, frustration, fear, low mood), then they may wish to consider a dyspnea questionnaire other than the BDI/TDI or UCSD, because these instruments don’t fully tap that aspect of dyspnea. Assuming more than one psychometrically acceptable PRO is available to

give them the data they desire, investigators must review the candidate instruments and draw on their scientific curiosity, clinical experience, work done previously by other investigators, and perhaps even patient input to help shape their decision of which PRO(s) to use.

Because of lack of data to support the longitudinal validity of more than a handful of PROs, investigators should be prepared to choose from a limited number of candidate instruments (acceptable to use at this time), to perform the preliminary studies necessary to build a foundation of validity for an instrument, or to develop a new one. Although it is impossible for us to make recommendations about which PROs to use in a trial without knowing the specific goals of the trial, prior research has revealed that patients with IPF have impaired QOL, they cough, and they are dyspneic and fatigued. These are domains an investigator may wish to assess as outcomes in a trial. A generally accepted rule is that investigators use both a generic and a disease-specific instrument: The generic instrument allows comparisons to be made with other patient populations, and the disease-specific instrument is included in the hopes of maximizing the ability to capture change over time. Whatever instruments they choose, investigators should follow the same principles of trial design that guide them in their selection and formulation of efficacy end points. Importantly, hypotheses about PRO end points and data analysis plans should be developed and clearly stated *a priori*.¹

IPF-SPECIFIC INSTRUMENTS

We are aware of four IPF-specific PROs (all HRQL instruments) that have been developed (or are in the developmental phase), two of which have been published.^{26,27} The SGRQ-I is an IPF-specific variation of the original SGRQ, developed by applying Rasch analysis to SGRQ response data acquired in a multinational trial.^{4,27} In Rasch analysis, response data are fit to a mathematical model, the results of which place instrument items and respondents on the same scale. The analysis determines a difficulty for each item and a measure (or ability) for each respondent. An item's difficulty refers to how likely it is to be endorsed. Consider patients with IPF responding to an instrument that assesses the ability to perform physical activities: If response data fit the Rasch model, then an item that asks about the ability to run a marathon would have high difficulty (very low probability that many patients with IPF would endorse the ability to run a marathon), and an item that asks about the ability to walk from room to room would have lower difficulty (many, but not all, patients with IPF would endorse the ability to walk from one room to another in their home). A person's measure

(or ability) is determined solely by how he responds to the aggregate of items on the instrument and is encapsulated in that one number. In terms of our example, respondents with higher person measures have greater physical ability than respondents with lower person measures. The scale units are log-odds (logits); because they are linear, simple statistical tests can be run using them. One can see how useful Rasch can be when trying to quantify constructs more abstract than physical ability (eg, QOL). The SGRQ-I has 34 items that, like the original SGRQ, are divided into Symptoms, Activity, and Impact domains; however, SGRQ response options have been significantly modified for the SGRQ-I, so its scores cannot be derived directly from the original SGRQ. To our knowledge, data supporting the responsiveness of the SGRQ-I have not been published.

To our knowledge, a tool to assess quality of life in IPF (ATAQ-IPF) is the only published, IPF-specific, HRQL instrument developed by using the perspectives of patients with IPF to identify domains and items of interest.²⁶ The ATAQ-IPF is a multidimensional questionnaire consisting of 74 items that tap 13 domains, including breathlessness, cough, fatigue, independence, and emotional health. By including all domains of interest, using the ATAQ-IPF alleviates the need to administer additional questionnaires (eg, dyspnea or cough indexes).²⁶ Ongoing research aims to further build the validity of the ATAQ-IPF for use as an outcome measure in IPF trials.

PRO ADMINISTRATION TO SUBJECTS WITH IPF

Nonstatistical Issues

Prior to giving a PRO to a subject with IPF, it is important to describe what he or she will be completing and explain why he or she is being asked to do it. If the PRO is computerized, practice questions should be given first. PROs must be administered in a quiet, neutral, space where confidentiality is ensured and subjects can complete them without interruption. Print should be large enough to be read by someone with suboptimal vision. During trials, it is crucial that subjects with IPF complete PROs at the same time each visit, prior to pulmonary physiology testing or meeting with the investigator (knowing their test results or talking with the investigator could influence responses). Trial centers must be prepared to give patients with IPF supplemental oxygen and make certain that their oxygen supply does not run out while completing the PROs.

Statistical Issues

Missing Data: For longitudinal research in populations with significant morbidity, like IPF, one challenge

is missing data, specifically, selective dropout by the sickest patients (in whom death or disease progression precludes return to the study site for data collection). Investigators must anticipate this challenge and determine a priori how missing PRO data will be handled. Of course, implementing procedures to prevent missing data beforehand is the best solution, but in many studies missing some data is unavoidable.²⁸

There are several principles regarding the handling of missing data that investigators should consider: The first is to not create missing data “by design.” For example, stopping assessments on a subject who misses a single assessment creates missing data by design. It is advisable to continue assessments on subjects who stop an intervention, thus allowing a true intent-to-treat analysis. The second is to avoid analyzing with methods that do not use all available data, such as multivariate analysis of variance (MANOVA) or repeated univariate analyses at each assessment time point. The third principle is to recognize that, in most cases, dropout will result in missing data that cannot be ignored (as occurs when the sickest subjects drop out). Analyses using simple methods that erroneously treat such missing data as ignorable will make subjects appear stable over time when in fact they are declining.

As a first step in the analysis of PRO data, full reporting of the numbers of subjects with missing data, which could be accomplished via a plot for each treatment arm stratified by last assessment, should be performed (Fig 1). In addition to Figure 1, plots for each arm stratified by reason for drop-out/missing data (eg, lack of efficacy/disease progression, side effects of treatment, or other) would shed light on whether missing data are ignorable. When the curves overlay each other, missingness is independent of both the observed data and reason for dropout, and most methods of analysis will produce unbiased estimates of change. However, this is very unlikely for conditions like IPF, and more sophisticated analysis methods are required.

Recommended Analysis Methods: In the majority of large IPF trials, investigators have used MANOVA^{5,6,9,10,29} with imputation^{4-6,9,10,29} to perform statistical analyses. However, for analyzing longitudinal, repeated-measures data in which missing data are assumed non-ignorable, the most common recommendation is to use likelihood mixed effects (also called random effects, random-coefficient, or hierarchical) models, supplemented with sensitivity analyses, if there is a large proportion of missing data. The merits of mixed effects models include the following: (1) all available data can be used (negating the need for imputation or case-wise deletion due to missing data), (2) model

FIGURE 1. Graph of subjects in one arm of a trial grouped on number of follow-up visits. All subjects in the arm have complete data for the baseline visit. Subjects in group 1 have complete data for each of the three follow-up visits (3 mo, 6 mo, and 12 mo). Subjects in group 2 dropped out after the 6-mo visit; they have complete data for the 3-mo and 6-mo visits but no data for the 12-mo visit. Subjects in group 3 dropped out after the 3-mo visit; they have complete data for the 3-mo visit but no data for the 6-mo or 12-mo visits.

structure is extremely flexible and allows for time-varying covariates, and (3) there is great flexibility in structuring the covariance of the repeated measures. These beneficial aspects allow these models to yield information on how groups change over time, how interventions effect that change, and how patient characteristics might modify treatment effects.³⁰

When using simpler analytic methods like MANOVA, to avoid case-wise deletion imposed by these methods that limit analyses to only subjects with complete data, investigators often use imputation techniques for missing data. One frequently used technique is last observation carried forward (LOCF); that is, if data for a subject are missing (for whatever reason), the scores/values from the last study visit are carried forward and recorded as results for the missing scores/values. When LOCF is used incorrectly, there is danger of imputing overly favorable data for subjects with non-ignorable missing data (stemming from disease progression/death or treatment-related morbidity). Sensitivity analyses are always indicated when LOCF is used: Readers need to know that investigators are not making the arm with greater drop-out “look” better with their method for dealing with missing data. In Figure 1, although LOCF would seem appropriate for group 1, it would not be for groups 2 or 3, because the outcome is on

an upward trajectory for group 2 and a downward trajectory for group 3.

Summary of Knowledge Gaps and Recommendations

- For any PRO intended to be used as an outcome in IPF research, its basic psychometric properties, including ICR and test-retest interclass correlation coefficient, must be calculated and fall within the acceptable ranges. For PROs that have been used in IPF research, certain of these properties remain to be determined.
- If a PRO is to be used as an end point in longitudinal research (eg, a drug trial), data to support its longitudinal validity (ie, ability to capture underlying change when present) must have been generated in a research setting of similar design (eg, in a previously conducted therapeutic trial or longitudinal observational study). Validity data generated in cross-sectional studies (eg, correlations between PRO score and FVC) tell nothing of the ability of the PRO to capture change over time and, thus, do not contribute to determining its usefulness as an outcome measure in longitudinal research.
- Several studies have revealed that three of the most relevant physical domains in IPF are dyspnea, fatigue, and cough. As such, generating data to support the validity of PROs that assess these domains would seem to be a worthwhile endeavor. Including in drug trials those PROs that assess these domains (and have the backing validity data to support their intended use) is a good starting point.
- When PROs are used in therapeutic trials, investigators should analyze resultant data transparently by revealing how many data were missing and how the missing data were handled. We encourage the use of methods (such as mixed-effects models) that use all available data and handle missing data without introducing bias.
- PROs should not be the only end points in IPF trials; however, we are fast approaching the time when they can hold more prominent positions in the end point schemes (eg, as stand-alone primary end points or as components in composite end points) of therapeutic trials.

CONCLUSION

For IPF, PROs are important outcome measures; however, for a number of reasons, they have not held a prominent place as end points in IPF trials. There are gaps in our knowledge of PROs in IPF, including a lack of data on some of the most basic psycho-

metric properties of candidate instruments. A couple IPF-specific HRQL instruments have been developed, but data to support their validity as measures capable of assessing longitudinal change have yet to be published. Analyzing repeated-measures PRO data from IPF trials poses certain challenges, dealing with missing data being one of the more frequent and complex, many of which can be overcome by using a sophisticated approach that uses likelihood mixed effects models. With more research, IPF investigators will gain confidence in PROs, thus allowing those investigators to determine with greater certainty whether and how therapeutic interventions impact the lives of patients with IPF.

ACKNOWLEDGMENTS

Financial/nonfinancial disclosures: The authors have reported to *CHEST* that no potential conflicts of interest exist with any companies/organizations whose products or services may be discussed in this article.

Role of sponsors: The sponsor had no role in the design of the study, the collection and analysis of the data, or in the preparation of the manuscript.

REFERENCES

1. Revicki DA, Osoba D, Fairclough D, et al. Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. *Qual Life Res*. 2000;9(8):887-900.
2. Gill TM, Feinstein AR. A critical appraisal of the quality of quality-of-life measurements. *JAMA*. 1994;272(8):619-626.
3. Food and Drug Administration, Center for Drug Evaluation and Research. *Guidance for Industry: Qualification Process for Drug Development Tools*. Silver Spring, MD: US Department of Health and Human Services; 2010.
4. King TE Jr, Behr J, Brown KK, et al. BUILD-1: a randomized placebo-controlled trial of bosentan in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. 2008;177(1):75-81.
5. Raghu G, Brown KK, Bradford WZ, et al; Idiopathic Pulmonary Fibrosis Study Group. A placebo-controlled trial of interferon gamma-1b in patients with idiopathic pulmonary fibrosis. *N Engl J Med*. 2004;350(2):125-133.
6. Raghu G, Brown KK, Costabel U, et al. Treatment of idiopathic pulmonary fibrosis with etanercept: an exploratory, placebo-controlled trial. *Am J Respir Crit Care Med*. 2008;178(9):948-955.
7. King TE Jr, Brown KK, Raghu G, et al. BUILD-3: a randomized, controlled trial of bosentan in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. 2011;184(1):92-99.
8. Zisman DA, Schwarz M, Anstrom KJ, Collard HR, Flaherty KR, Hunninghake GW; Idiopathic Pulmonary Fibrosis Clinical Research Network. A controlled trial of sildenafil in advanced idiopathic pulmonary fibrosis. *N Engl J Med*. 2010;363(7):620-628.
9. King TE Jr, Albera C, Bradford WZ, et al; INSPIRE Study Group. Effect of interferon gamma-1b on survival in patients with idiopathic pulmonary fibrosis (INSPIRE): a multicentre, randomised, placebo-controlled trial. *Lancet*. 2009;374(9685):222-228.
10. Demedts M, Behr J, Buhl R, et al; IFIGENIA Study Group. High-dose acetylcysteine in idiopathic pulmonary fibrosis. *N Engl J Med*. 2005;353(21):2229-2242.

11. Streiner D, Norman G. Reliability. In: *Health Measurement Scales: a Practical Guide to Their Development and Use*. 4th ed. New York: Oxford University Press; 2008:104-126.
12. Hays RD, Anderson RT, Revicki DA. Assessing reliability and validity of measurement in clinical trials. In: Staquet M, Hays R, Fayers P, eds. *Quality of Life Assessment in Clinical Trials*. New York: Oxford University Press; 1998:169-182.
13. Nadorra RL, Landing BH. Pulmonary lesions in childhood onset systemic lupus erythematosus: analysis of 26 cases, and summary of literature. *Pediatr Pathol*. 1987;7(1):1-18.
14. Nunnally J, Bernstein J. The assessment of reliability. In: *Psychometric Theory*. 3rd ed. New York: McGraw-Hill; 1994: 248-292.
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(12):1382-1389.
16. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen, Denmark: Danish Institute of Educational Research; 1960.
17. Hays RD, Hadorn D. Responsiveness to change: an aspect of validity, not a separate dimension. *Qual Life Res*. 1992;1(1): 73-75.
18. Patrick DL, Chiang YP. Measurement of health outcomes in treatment effectiveness evaluations: conceptual and methodological challenges. *Med Care*. 2000;38(suppl 9): II14-II25.
19. Nunnally J. *Introduction to Psychological Measurement*. New York, NY: McGraw-Hill; 1970.
20. Swigris JJ, Gould MK, Wilson SR. Health-related quality of life among patients with idiopathic pulmonary fibrosis. *Chest*. 2005;127(1):284-294.
21. Swigris JJ, Kuschner WG, Jacobs SS, Wilson SR, Gould MK. Health-related quality of life in patients with idiopathic pulmonary fibrosis: a systematic review. *Thorax*. 2005;60(7): 588-594.
22. Nishiyama O, Taniguchi H, Kondoh Y, et al. Health-related quality of life in patients with idiopathic pulmonary fibrosis. What is the main contributing factor? *Respir Med*. 2005; 99(4):408-414.
23. 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(4): 459-466.
24. 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(9):1231-1237.
25. 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(2):296-304.
26. Swigris JJ, Wilson SR, Green KE, Sprunger DB, Brown KK, Wamboldt FS. Development of the ATAQ-IPF: a tool to assess quality of life in IPF. *Health Qual Life Outcomes*. 2010; 8:77.
27. Yorke J, Jones PW, Swigris JJ. Development and validity testing of an IPF-specific version of the St George's Respiratory Questionnaire. *Thorax*. 2010;65(10):921-926.
28. Bernhard J, Cella DF, Coates AS, et al. Missing quality of life data in cancer clinical trials: serious problems and challenges. *Stat Med*. 1998;17(5-7):517-532.
29. Noble PW, Albera C, Bradford WZ, et al; CAPACITY Study Group. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. *Lancet*. 2011; 377(9779):1760-1769.
30. Fairclough D. Models for longitudinal studies I. In: *Design and Analysis of Quality of Life Studies in Clinical Trials*. 2nd ed. New York, NY: CRC Press; 2010:53-81.