Patient-reported outcomes and patient-reported outcome measures in interstitial lung disease: where to go from here?

Meena Kalluri, Fabrizio Luppi, Ada Vancheri, Carlo Vancheri, Elisabetta Balestro, Francesco Varone, Nesrin Mogulkoc, Giulia Cacopardo, Elena Bargagli, Elisabetta Renzoni, et al.

Patient-reported outcomes and patient-reported outcome measures in interstitial lung disease: where to go from here?

Meena Kalluri [1,17], Fabrizio Luppi [2,17], Ada Vancheri[3], Carlo Vancheri[4], Elisabetta Balestro [5], Francesco Varone[6], Nesrin Mogulkoc[7], Giulia Cacopardo [8], Elena Bargagli[9], Elisabetta Renzoni[10], Sebastiano Torrisi[4], Mariarosaria Calvello[6], Alessandro Libra[3], Mauro Pavone[4], Francesco Bonella [11], Vincent Cottin [12,13], Claudia Valenzuela[14,17], Marlies Wijsenbeek[15,17] and Elisabeth Bendstrup [16,17]

[1]Division of Pulmonary Medicine, University of Alberta, Edmonton, AB, Canada. [2]Respiratory Diseases Unit, University of Milano-Bicocca. "S. Gerardo" Hospital, Monza, Italy. [3]Regional Referral Center for Rare Lung Diseases, University - Hospital "Policlinico G. Rodolico – San Marco", Catania, Italy. [4]Dept of Clinical and Experimental Medicine, Regional Referral Centre for Rare Lung Diseases, University - Hospital "Policlinico G. Rodolico – San Marco", University of Catania, Catania, Italy. [5]Dept of Cardiac, Thoracic, Vascular Sciences and Public Health, University of Padova, Padova, Italy. [6]UOC Pneumologia, Fondazione Policlinico A. Gemelli IRCCS, Rome, Italy. [7]Dept of Pulmonology, Ege University Hospiral, Bornova, Izmir, Turkey. [8]UOSD UTIR, A.R.N.A.S. Ospedali Civico Di Cristina Benfratelli, Palermo, Italy. [9]Respiratory Diseases Unit, Siena University, Siena, Italy. [10]Interstitial Lung Disease Unit, Royal Brompton Hospital, Imperial College, London, UK. [11]Pneumology Dept, Centre for Interstitial and Rare Lung Disease, Ruhrlandklinik University Hospital, University of Duisburg-Essen, Essen, Germany. [12]Dept of Respiratory Medicine, National Reference Coordinating Center for Rare Pulmonary Diseases, Louis Pradel Hospital, Hospices Civils de Lyon, Lyon, France. [13]University of Lyon, INRAE, IVPC, Lyon, France. [14]Pulmonology Dept, Hospital Universitario de la Princesa, Universidad Autonoma Madrid, Madrid, Spain. [15]Centre of excellence, Interstitial Lung Diseases and Sarcoidosis, Erasmus MC, University Medical Centre Rotterdam, Rotterdam, The Netherlands. [16]Dept of Respiratory Diseases and Allergy, Centre for Rare Lung Diseases, Aarhus University Hospital, Aarhus N, Denmark. [17]Shared first and last authorship.

Corresponding author: Meena Kalluri (kalluri@ualberta.ca)

Shareable abstract (@ERSpublications)
PROMs are essential tools for research and care in ILD and IPF. They report patient perceptions of the impact of disease and its treatments on whole-person wellbeing and can guide research to make care more patient-centred. https://bit.ly/3s7Y0a8

Cite this article as: Kalluri M, Luppi F, Vancheri A, *et al*. Patient-reported outcomes and patient-reported outcome measures in interstitial lung disease: where to go from here? *Eur Respir Rev* 2021; 30: 210026 [DOI: 10.1183/16000617.0026-2021].

Received: 6 Feb 2021
Accepted: 2 April 2021

Abstract

Patient-reported outcome measures (PROMs), tools to assess patient self-report of health status, are now increasingly used in research, care and policymaking. While there are two well-developed disease-specific PROMs for interstitial lung diseases (ILD) and idiopathic pulmonary fibrosis (IPF), many unmet and urgent needs remain. In December 2019, 64 international ILD experts convened in Erice, Italy to deliberate on many topics, including PROMs in ILD. This review summarises the history of PROMs in ILD, shortcomings of the existing tools, challenges of development, validation and implementation of their use in clinical trials, and the discussion held during the meeting. Development of disease-specific PROMs for ILD including IPF with robust methodology and validation in concordance with guidance from regulatory authorities have increased user confidence in PROMs. Minimal clinically important difference for bidirectional changes may need to be developed. Cross-cultural validation and linguistic adaptations are necessary in addition to robust psychometric properties for effective PROM use in multinational clinical trials. PROM burden of use should be reduced through appropriate use of digital technologies and computerised adaptive testing. Active patient engagement in all stages from development, testing, choosing and implementation of PROMs can help improve probability of success and further growth.

Introduction

Patient-reported outcomes (PROs) have become increasingly important in medicine and in the field of interstitial lung diseases (ILDs) including idiopathic pulmonary fibrosis (IPF) and other fibrosing ILDs. Many ILDs are not only irreversible, but also progressive in nature [1]. The debilitating symptoms and the knowledge of a poor prognosis heavily affects ILD patients' physical, mental and emotional wellbeing [2].

Antifibrotic therapies available in IPF were approved primarily because they slow down disease progression, but have not so far been demonstrated to have a significant impact on patient wellbeing [3]. In parallel with the need to identify interventions that can preserve or improve patient quality of life [4], there is a need to utilise and further develop objective measures of patients' own perceptions of their health-related quality of life (HRQoL) so they can be assessed appropriately [5]. A new PRO measure (PROM), the Living with Idiopathic Pulmonary Fibrosis questionnaire (L-IPF) has suggested a beneficial impact of antifibrotic treatment in non-IPF progressive fibrotic ILD [6]. Therefore, the use of such PROMs to study the impact of existing and novel antifibrotic medications on HRQoL are needed [4].

A PRO is a self-reported health outcome directly reported by the patient who experiences it without interpretation by an intermediary such as healthcare professionals, informal caregivers or their proxies, not obtainable in any other manner and represents the patients' perspective on their own symptoms, function, psychological problems, health status and overall HRQoL. These measures are extremely important as they provide a real understanding of the impact of disease and its treatments on health status from a patient's perspective. Objective outcome measures used in ILD clinics and research (for example, change in pulmonary function) that are considered important by physicians do not necessarily align with patient priorities. Qualitative studies suggest that the objective measures used in ILD clinics are viewed by patients as disconnected from their experience of the disease [7]. Inclusion of PROs in such situations draws clinical attention to what matters most to patients [8–11]. Thus, the use of PROs as research study end-points and clinical follow-up tools may allow inclusion of patient narratives that have been largely ignored by modern medicine [12]. PRO should be considered as complementary to traditional efficacy measures and are not meant to replace them.

The overarching purpose of PRO inclusion in clinical research is to ensure that research conducted eventually improves clinical care, patient experience and healthcare systems and is of high value to all stakeholders, including patients and society. As a result, the development of relevant PROs and their use in research and care has received encouragement from professional societies, regulatory bodies, health systems and payers and patient advocacy groups [13].

PROMs are dedicated tools to assess PROs. They may be collected *via* self-administered questionnaires or through patient interviews if the data obtained are reported directly. They are distinct from patient-reported experience measures, which report on patient experience of care, settings and processes [14]. PROs should not be confused with patient-centred outcomes that are directly relevant to patients, but may not be self-reported, *e.g.* end-of-life care outcomes, location of death, quality of dying and death, *etc.* [15] (figure 1).

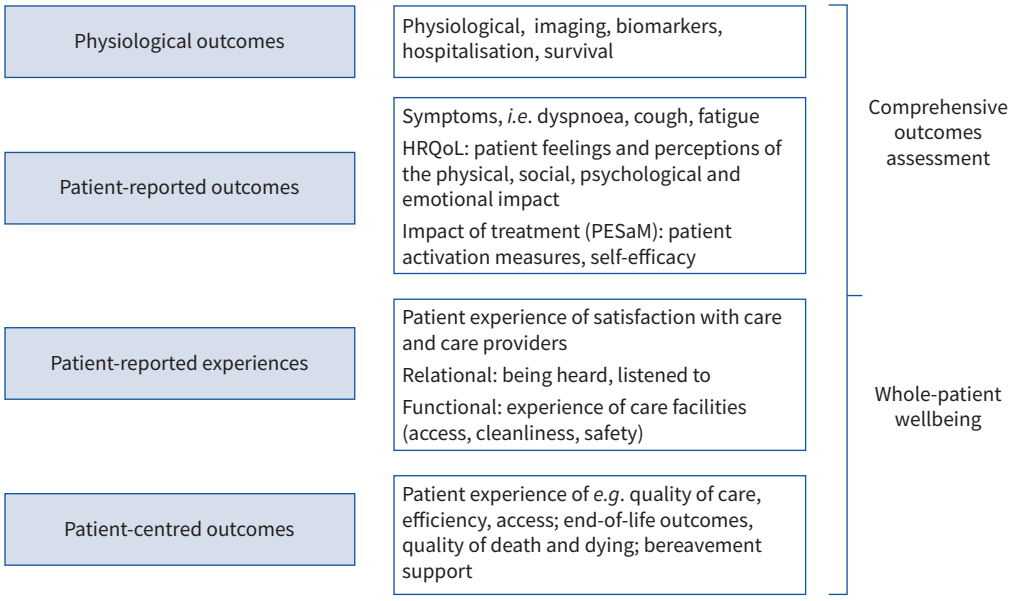

FIGURE 1: Comprehensive outcomes assessment in interstitial lung disease (ILD)/idiopathic pulmonary fibrosis/progressive fibrotic ILD. HRQoL: health-related quality of life; PESaM: patient experience and satisfaction with medication.

PROMs can be generic, symptom- or disease-specific measures. Generic PROMs are designed for use in a disease or general population sample and may exclude important aspects that are relevant to specific diseases. Symptom-specific measures are dedicated tools used to assess specific symptoms such as dyspnoea or cough. In contrast, condition-specific PROMs are designed to assess patient views across various domains within a particular health condition or diagnostic group such as cancer or chronic illness and are only suitable for use in that population. PROM use may improve patient satisfaction, clinician satisfaction, efficiency and communication, and can support joint decision-making [16–19]. PROMs can assist in monitoring symptoms, response to therapeutic interventions and serve as a screening tool for physical and emotional problems that require additional clinical interventions [20–23]. While most patients welcome their use in care, physicians often remain reluctant to use PROMs routinely because of lack of confidence in their value and the time needed for implementation [24].

In December 2019, a working group of 63 international ILD experts assembled for the third time at the Ettore Majorana Foundation and Centre for Scientific Culture in Erice, Italy, to discuss priority clinical and research issues in ILD. Participants represented 17 countries in Europe and North America. Among the topics of discussion during this 3-day meeting were the current PROMs for patients with IPF. In pro–con debate and round-table discussion formats, it was discussed whether PROMs must be incorporated as efficacy end-points in ILD clinical trials. Given the critical unmet need, the current experience and challenges with PROMs, this review was prepared by the primary participants of discussion on this issue (authors) and incorporating input from all other participants of the Erice ILD working group (collaborators) to catalogue needs, gaps and possible future directions to facilitate understanding and research in PROMs for ILD.

History of PROM use in ILD/IPF

Over the past decade, there has been an increasing recognition of the need to engage in patient-centred research and to use PROMs in the process [25–28]. The growing awareness that care must be comprehensive, holistic and benefit the whole person, and not just address disease pathobiology has propelled multiple stakeholders across various sectors to embrace PROMs. This requires tools that can measure concepts relevant to the wellbeing of the person as a whole and assess whether specific treatments benefit these domains of health (figure 1). It is equally important to recognise when the interventions could have a negative impact on HRQoL. Thus, traditional clinical research outcomes in ILD/IPF such as pulmonary function, while important, are not sufficient or comprehensive enough from the patient's perspective, and outcomes that include a focus of patient and carer perceptions are needed [7]. The assessment of how an intervention affects symptoms, function and HRQoL should be considered an equally important research outcome; for example, "Can I go out for a walk without feeling too breathless?" or "Could the side-effects of this therapy worsen my quality of life further?" instead of "Has my spirometry or walk distance improved?" Such aspects of disease experience can only be ascertained by asking patients directly or by using PROMs.

PROMs have been used for over a decade in IPF/ILD clinical drug trials, mostly as secondary end-points (table 1). Unfortunately, most of the trials failed to demonstrate significant effect on HRQoL, despite a benefit on lung function outcome for some interventions. This has multiple probably reasons, such as the nature of interventions aiming to modify fibrosis by slowing down disease progression, the use of PROMs not specifically developed for IPF/ILD, too short an observation time, inclusion of populations without sufficiently severe disease, and their use in multinational settings with PROMs failing to take into account the inherent sociocultural differences in expectation. PROMs have been used successfully used in pulmonary rehabilitation trials in ILD/IPF [46]. These were not included in our discussion as it was considered out of scope. A recent Cochrane systemic review of 21 such studies demonstrated meaningful improvement in dyspnoea and HRQoL as measured by the Chronic Respiratory Disease Questionnaire (CRQ) and the St George's Respiratory Questionnaire (SGRQ) for participants with ILD and for the subgroup of people with IPF.

Among the numerous available PROMs, SGRQ has been most frequently used; it was originally developed for COPD and asthma [47–49]. Although validated in IPF, it may not provide assessment of all relevant aspects of IPF/ILD [50]. Therefore, SGRQ was modified to create an IPF-specific version, SGRQ-I (SGRQ-IPF), and validated in IPF [51]. Other generic measures such as EuroQol five dimensions (EQ-5D), 36-item short form (SF-36) and University of California San Diego (UCSD) Shortness of Breath Questionnaire have also been used in IPF (table 2). Several of these PROMs (EQ-5D, SF-36, COPD Assessment Test) have been validated in the IPF population, resulting in confirmation of acceptable measurement properties [11, 52, 78, 79]. Importantly, none of these instruments were developed specifically for the ILD population and were created prior to release of PROM guidance documents from

TABLE 1 Summary of patient-reported outcome measures (PROMs) in idiopathic pulmonary fibrosis (IPF)/interstitial lung disease (ILD) pharmacotherapeutic clinical trials (randomised controlled trials only)

Study, clinicaltrials.gov identifier; year registered [reference]	Intervention	PROM	Used as primary or secondary end-point	PROM outcomes	
Shionogi, phase II [29]	Pirfenidone *versus* placebo	Hugh–Jones CRQ	Tertiary	NS NS	No difference in worsening of HRQoL
IFIGENIA phase III, NCT00639496; 2008 [30]	N-acetylcysteine *versus* placebo	SGRQ	Secondary	NS	No significant change in HRQoL
Etanercept phase II, NCT00063869; 2003 [31]	Etanercept *versus* placebo	SGRQ SF-36 MDI	Secondary	p<0.001 SF-36 physical p<0.001, mental p=0.09 MDI NR	Significant decline in HRQoL in both arms
INSPIRE phase II, NCT00075998; 2004 [32]	IFN-γ *versus* placebo	SGRQ UCSD-SOBQ	Secondary	NS NS	No difference in worsening of HRQoL and dyspnoea
STEP-IPF phase III, NCT00517933; 2007 [33]	Sildenafil *versus* placebo	SGRQ UCSD-SOBQ Borg SF-36 EQ-5D	Secondary	p=0.01 p=0.006 NS p=0.008 NS	Significant difference in HRQoL and dyspnoea favouring sildenafil
BUILD-3 phase III, NCT00391443; 2006 [34]	Bosentan *versus* placebo	SF-36 EQ-5D TDI	Secondary	NS NS NS	No differences in worsening of HRQoL and dyspnoea
CAPACITY 1 and 2 phase III, NCT00287716; 2006 [35]	Pirfenidone *versus* placebo	UCSD-SOBQ	Secondary	NS	Nonsignificant worsening in dyspnoea in both arms
BIBF-1120 phase II, NCT00514683; 2007 [36]	Nintedanib *versus* placebo	SGRQ	Secondary	p=0.007	Statistically significant improvement in HRQoL not meeting clinical significance
Ambrisentan phase II, NCT00768300; 2008 [37]	Ambrisentan *versus* placebo	SGRQ TDI SF-36	Secondary	NS NS NS	No difference in worsening of HRQoL and dyspnoea between the two arms
ASCEND phase III, NCT01366209; 2011 [38]	Pirfenidone *versus* placebo	UCSD-SOBQ	Secondary	NS	No difference in worsening dyspnoea between the two arms
PANTHER, NCT00650091; 2008 [39]	N-acetylcysteine, azathioprine, prednisone *versus* placebo	SF-36 EQ-5D SGRQ UCSD-SOBQ ICE-CAP	Secondary	p=0.03 NS NS NS p=0.01	No difference in worsening HRQoL and dyspnoea between the arms
INPULSIS phase III, NCT01335464, NCT01335477; 2011 [40]	Nintedanib	SGRQ CASA-Q	Secondary	INPULSIS 1 NS INPULSIS 2 p=0.02 CASA-Q NR	No difference in worsening HRQoL in pooled data *versus* placebo
AmbOx phase IV, NCT02286063; 2014 [41]	Oxygen	K-BILD UCSD-SOBQ SGRQ HADS	Primary Secondary Secondary Secondary	p<0.0001 p<0.0001 p=0.018 NS	Statistically significant improvement in HRQoL and dyspnoea with ambulatory oxygen use in patients with ILD with isolated exertional hypoxia. Did not meet minimum clinical significance
INSTAGE phase III, NCT02802345; 2016 [42]	Nintedanib Sildenafil	SGRQ UCSD-SOBQ	Primary Secondary	NS NS	No difference in change from baseline in the total score (worsening) on the SGRQ at week 12. No difference in worsening of dyspnoea from baseline between the two arms

Continued

TABLE 1 Continued

Study, clinicaltrials.gov identifier; year registered [reference]	Intervention	PROM	Used as primary or secondary end-point		PROM outcomes
INBUILD phase III, NCT02999178; 2016 [43]	Nintedanib	K-BILD	Secondary	NS	No difference in worsening of HRQoL between the two arms
Unclassifiable FILD phase II, NCT03099187; 2017 [44]	Pirfenidone	UCSD-SOBQ LCQ Cough VAS SGRQ	Secondary	NS NS NS NS	No difference in worsening of HRQoL and symptoms between arms
Advanced IPF phase II B, NCT02951429; 2016 [45]	Pirfenidone, sildenafil	SGRQ UCSD-SOBQ	Secondary Secondary	NS NS	No difference in worsening of HRQoL and dyspnoea between the two arms

CRQ: Chronic Respiratory Disease Questionnaire; NS: nonsignificant; HRQoL: health-related quality of life; SGRQ: St George's Respiratory Questionnaire; SF-36: 36-item short form; MDI: Mahler Dyspnoea Index; NR: not reported; IFN: interferon; UCSD-SOBQ: University of San Diego Shortness of Breath Questionnaire; EQ-5D: EuroQol five dimensions; TDI: Transitional Dyspnoea Index; ICE-CAP: Investigating Choice Experiences Capability instrument; CASA-Q: Cough and Sputum Assessment Questionnaire; K-BILD: King's Brief Interstitial Lung Disease Questionnaire; HADS: Hospital Anxiety and Depression Score; LCQ: Leicester Cough Questionnaire; VAS: visual analogue scale.

regulatory authorities such as the US Food and Drug Administration (FDA) [80] and the European Medicines Agency (EMA) [81]. In light of the increasing importance of PROMs and their use in clinical drug trials, these regulatory authorities created guidance documents to facilitate proper development, validation, testing and use of these tools in research settings. The King's Brief Interstitial Lung Disease Questionnaire (K-BILD) is the first ILD-specific PROM that was developed explicitly for this population [61]. It has been validated for use in IPF in subsequent studies [82]. This was a key step in the right direction as demonstrated by its effective use in a recent clinical pharmacotherapeutic trial by Visca *et al*. [41]. In a randomised controlled trial (RCT) to assess the effect of ambulatory oxygen on quality of life using PROMs as primary outcomes, clinically significant results were noted for the first time in an ILD drug trial [41]. In contrast to previous studies, this trial used K-BILD, an ILD-specific PROM and studied a treatment that directly targeted symptoms and quality of life. This suggests that the use of disease-specific PROMs that are validated rigorously and chosen thoughtfully may be more efficacious and may help move the field forward. L-IPF is a recently validated PROM developed specifically for IPF and concordant with FDA guidance [66]. L-IPF assesses symptoms, disease impact and HRQoL, but it only indirectly assesses mood [6]. In addition, it does not assess therapy-related effects that can adversely affect HRQoL and longitudinal validity remains to be established. L-IPF was used as a secondary outcome measure in INBUILD, an RCT investigating the effects of nintedanib on progressive fibrotic ILD and showed significant results published as an abstract [6]. As more high-quality PROMs are developed and tested using rigorous methodologies, we hope they will be increasingly used in clinical trials to assess the impact of interventions on HRQoL.

PROM development

PROM development can be a long, arduous and costly process. Figure 2 describes the various steps in its development as outlined by FDA draft guidance and the EMA reflective paper. The first step involves the development of a preliminary instrument with a pool of items by creating a conceptual framework or model, based on literature and expert review. The content pool undergoes further refinement with inclusion of new items generated based on interviews with patients and further expert input. Item response theory and Rasch measurement theory are the preferred methods for item development and testing [83]. Response options, recall period and tool format are determined. The item bank then undergoes further refinement and reduction, again based on interviews with patients. This is done through cognitive interviewing or testing that involves "verbal probing" to determine the readability, acceptability of the format and structure of the instrument, interpretation of items and formulation of response [84]. After cognitive testing with patients, a pilot draft is further optimised to create the final instrument and subsequently tested. Once the tool has been developed in this fashion, cross-sectional and longitudinal validation is performed (figure 2). Thereafter, the questionnaires may be translated into various languages for wider use and undergo related appropriate validation steps. The chosen PROMs must also have appropriate cultural adaptations to account

https://doi.org/10.1183/16000617.0026-2021

TABLE 2 Characteristic and properties of patient-reported outcome measures (PROMs) used in idiopathic pulmonary fibrosis (IPF)/interstitial lung disease (ILD) clinical trials

	Domain of PROM	Source population	FDA/EMA guidance concordant development	Cross-sectional validation	Longitudinal validation MCID	MCID	Number of items Response time	References
SGRQ	HRQoL 3 domains: symptoms (frequency and severity); activity (effects of breathlessness on mobility and activity); impact (psychosocial impact of disease)	COPD, asthma	No	Yes (at least in CTD-ILDs)	Yes Yes	Total 7 (4–8) Symptoms 8 Activity 5 Impact 7	Paper or electronic 50 items, 10 min Recall periods: symptoms: past 4 weeks	[48, 50, 52–54]
SGRQ-I	HRQoL 3 domains: symptoms (frequency and severity); activity (effects of breathlessness on mobility and activity); impact (psychosocial impact of disease)	COPD	No	Yes	Yes	Improvement/decline total 3.9/4.9 symptoms 9.0/8.1 activities 9.8/10.4 impact 5.4/5.4	Paper or electronic 34 items 10–15 min Recall periods: symptoms: past 4 weeks	[51, 55, 56]
EQ-5D	HRQoL 5 domains: assesses impact of disease on mobility, self-care, usual activities, pain/discomfort and anxiety/depression	COPD	No	No	No Yes	0.028 (range 0.017–0.033)	Paper or electronic 3 levels: 15 items 5 levels: 25 items 8 min Recall period: immediate situation	[57–59]
CRQ	HRQoL 4 domains: assesses dyspnoea "individualised" (impact on 5 activities chosen by patients including chores and social function); fatigue (severity and frequency); emotional functioning (satisfaction with life and frequency of impairment); and mastery (frequency of pain *versus* confidence in self-care)	COPD	No	No	No No	Not available	Paper or electronic 20 items 15–25 min (administered) 8–10 min (self) Recall period: past 4 weeks	[60]
K-BILD	HRQoL 3 domains: psychological (frequency of impairment); breathlessness and activities (frequency of impact on activities); and chest symptoms (frequency of air hunger, wheezing and tightness)	ILD, some IPF	Yes	Yes	Yes Yes	Improvement/decline Total score 4.7/2.7 Breathlessness and activities 3.6/3.6 Chest symptoms 7/6 Psychological 4.8/3.5	Paper or electronic 15 items 5–7 minutes Recall period: past 2 weeks	[51, 55, 61–63]

Continued

https://doi.org/10.1183/16000617.0026-2021

TABLE 2 Continued

	Domain of PROM	Source population	FDA/EMA guidance concordant development	Cross-sectional validation	Longitudinal validation MCID	MCID	Number of items Response time	References
SF-36	HRQoL 8 domains: physical functioning (frequency of limitation); physical role limitations (degree of limitation); bodily pain; general health perceptions; energy/vitality (loss of energy or presence of fatigue); social functioning (limitations); emotional role limitations (difficulties with work and daily activities); mental health (presence of depression/nervousness)		No	Yes	Yes Yes	PCS 5; MCS 7; MID range 7–14	Paper or electronic 36 items <10 min Recall period: past 1 week	[64, 65]
L-IPF	HRQoL 2 modules: symptoms (3 domains-shortness of breath, cough, energy) and impact	IPF	Yes	Yes	No No	Not available	Not available 35 items Recall periods: symptoms: past 24 h impact: past 1 week	[66]
IPF-PROM	HRQoL 4 domains: physical experience of breathlessness; psychological experience of breathlessness; emotional wellbeing; and energy level	IPF	Yes	Yes	No No	Not available	Not available 12 items	[67, 68]
UCSD-SOBQ	Dyspnoea: assesses dyspnoea severity with activities (21 items) and limitations from dyspnoea and related fear and harm (3 items)	COPD	No	Yes	Yes No	5 units	Paper or electronic 24 items 5–10 min Recall period: immediate situation	[69–71]
Dyspnoea-12	Dyspnoea 2 domains: physical (breathlessness severity) and affective, incorporating both physical and affective aspects, and does not depend on activity limitation	COPD, ILD	No	Yes	No No	Not available	Paper 12 items 5–10 min Recall period: immediate situation	[72, 73]
LCQ	Cough 3 domains: physical, psychological and social (frequency of impairment ranging from all to none of the time)	COPD and chronic cough	No	No	No No	Not available	Paper or electronic 19 items <5 min Recall period: past 2 weeks	[74, 75]
CASA-Q	Cough 4 domains: cough symptoms; cough impact; sputum symptoms; and sputum impact (assesses frequency ranging from never to a lot)	COPD	No	Yes	No	Not available	Paper 20 items Recall period: past 1 week	[76, 77]

FDA: US Food and Drug Administration; EMA: European Medicines Agency; MCID: minimal clinically important difference; SGRQ: St George's Respiratory Questionnaire; SGRQ-I: SGRQ IPF-specific version; EQ-5D: EuroQol five dimensions; CRQ: Chronic Respiratory Disease Questionnaire; K-BILD: King's Brief Interstitial Lung Disease Questionnaire; SF-36: 36-item short form; L-IPF: Living with Idiopathic Pulmonary Fibrosis; UCSD-SOBQ: University of California San Diego Shortness of Breath Questionnaire; LCQ: Leicester Cough Questionnaire; CASA-Q: Cough and Sputum Assessment Questionnaire; HRQoL: health-related quality of life; CTD-ILD: connective tissue disease-associated ILD; PCS: physical component summary; MCS: mental health component summary; MID: minimally important difference.

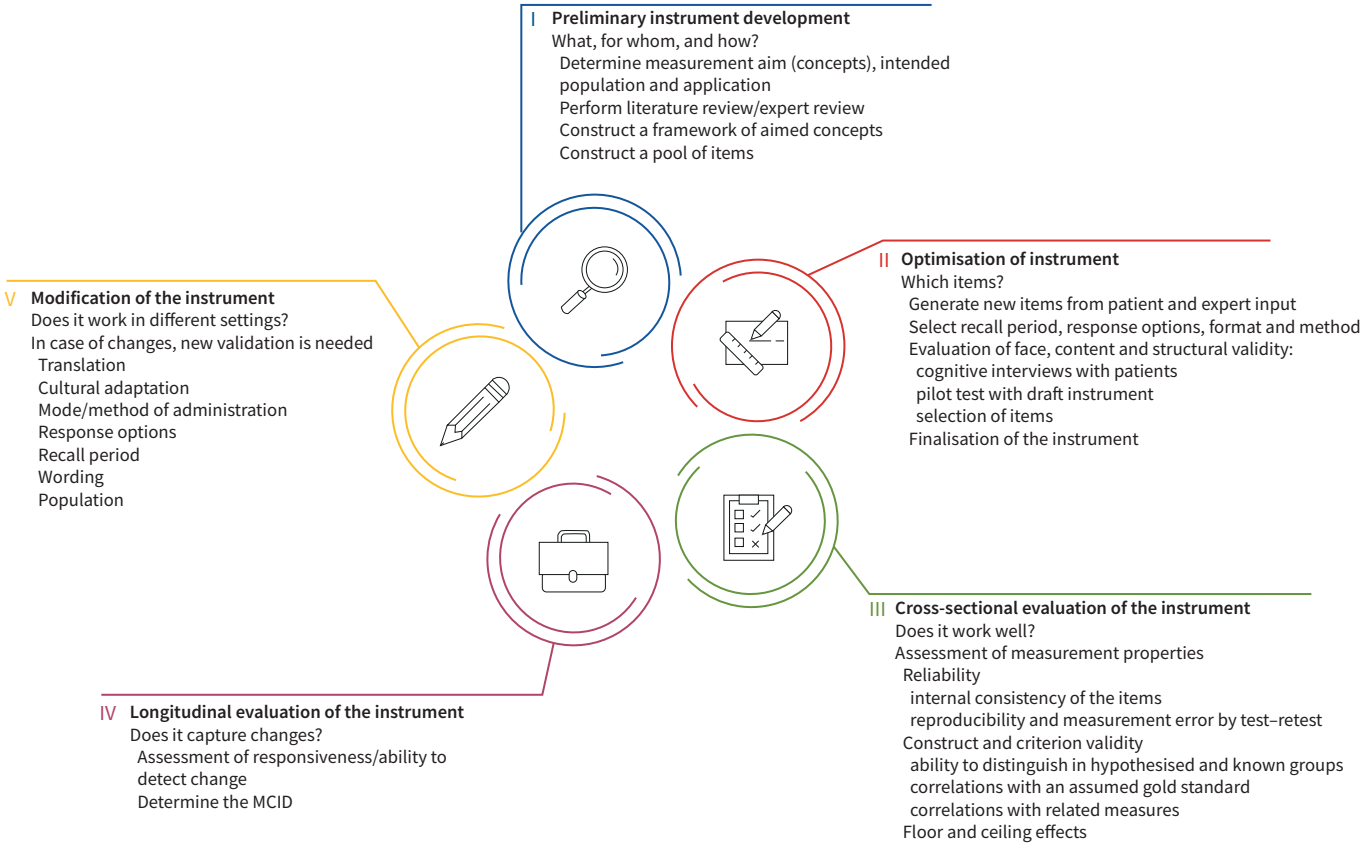

FIGURE 2: Development of patient-reported outcome measures (PROMs). MCID: minimal clinically important difference. Adapted from US Food and Drink Administration guidance on the development of a PROM instrument for drug development to support labelling claims [80].

for varying patient perceptions governed by social and cultural conditions. For PROMs to be useful in diverse, multicultural, multilingual populations of rare diseases and clinical trials, they must be adapted to target populations.

Challenges of PROMs: development, validation and implementation

Problems in development

As noted earlier, PRO end-points mostly fail as (predominantly secondary) end-points in clinical trials aimed at disease modification in IPF [85]. Table 2 reviews characteristics and measurement properties of PROMs commonly used in ILD that can help inform the reasons behind this disappointing result. As outlined, all PROMs except K-BILD and L-IPF were developed in non-ILD populations. The lack of disease-specific rigorously developed tools has precluded confident use of PROMs in ILD research [85]. Furthermore, none of the symptom-specific PROMs, as listed in table 2, were originally developed for IPF [70, 86–88]. Although cough is a common symptom of many ILDs, questionnaires have not been developed to assess cough severity or its impacts specifically for patients with ILD [89]. The Leicester Cough Questionnaire and the Cough Quality of Life Questionnaire, developed in patients with chronic cough, are currently used in IPF-related cough [90, 91]. Similarly, there are no ILD-specific questionnaires to assess fatigue, a common and debilitating symptom for many patients with ILD. However, the Fatigue Assessment Scale, developed and validated in sarcoidosis, has been used in IPF [92, 93].

Another concern with PROMs is that they are likely to be more precise in capturing patient perceptions and their variation over time in homogeneous target populations, such as patients within a set range of disease severity [94]. This may limit their generalisability to large, multicentre and multinational clinical trials with widespread geographic, cultural and socioeconomic differences between participating patient populations [95]. Furthermore, if a wide range of disease severity is represented in a therapeutic trial, the realistic patient goals, expectations and perceptions may further vary greatly. This is a well-known challenge in many rare diseases with heterogeneous populations and variable disease courses. This was

observed in the INPULSIS trials, where a significant reduction in loss of HRQoL as measured by PROMs (including SGRQ) was seen in the treatment group compared to placebo in patients with more severe IPF, but not in those with less advanced disease, suggesting that the PROMs used may have a lower sensitivity to change in patients with less severe IPF [96].

Other explanations for the low sensitivity of PROMs in ILD/IPF trials could be related to the confounding effect of differences in language, practices and expectations between people from different countries and cultures [97]. The importance of linguistic validation has increasingly been acknowledged with PROMs intended for use in a heterogeneous population (see also later). All these issues may lead to limitations in generalisability of PROMs.

Cross-sectional validity and reliability

The validity of PROMs can be assessed in a variety of ways, including the demonstration of moderate-to-high correlation with existing measures that address the same concept (convergent validity), low correlation with measures that assess other concepts (divergent validity) and responsiveness to meaningful change. Reliability refers to the consistency of a measure. The main types of consistency are over time (test–retest reliability) and across items (internal consistency) [98–100]. Internal consistency determines the consistency of responses from an individual across the items on a multiple-item measure [100]. The FDA guidance emphasised test–retest reliability as the most important reliability type, while internal consistency may be used in the absence of information on the test–retest reliability [100].

Evaluating the measurement properties of PROMs in rare diseases can be challenging because of the small number of patients in addition to heterogeneous populations. For example, the COPD-specific SGRQ, used to assess HRQoL in IPF [47], is self-reported, and has 50 items split into three domains: “symptoms”, “activity” and “impacts” [101]. Validation in IPF [50] concluded that the internal consistency of the activity and impacts domains and the total score from the SGRQ was excellent, whereas the consistency of the symptoms domain was moderate, probably due to different symptom characteristics (*e.g.* wheeze and sputum production are not common in patients with IPF/ILD). An important point to note is that for most PROMs, as part of validation, the PROM data are usually compared to other quality-of-life tools and lung function data. Comparing newer ILD tools to SGRQ, not specific to IPF, in whole or single domains is less than ideal. In the same way, many HRQoL and dyspnoea PROMs are validated against lung function data or other tools that in turn use lung function for validation. However, a study by Grønseth *et al.* [102] showed that only 13% of dyspnoea variance can be explained by the usual confounders, including lung function, suggesting that they are not ideal comparators for validation. In addition, their study highlighted significant variation in dyspnoea prevalence across 15 countries, probably associated with differences in patient perception of symptoms. Therefore, these data call into question the cross-cultural validity of many such instruments, not just limited to dyspnoea scales, but also the HRQoL tools that include dyspnoea as part of their domains [102]. K-BILD, although an ILD-specific PROM, has an important limitation in the lack of some health issues relevant to ILD, such as cough, that are not included in the final version of the K-BILD [61]. The process of development, testing and validation of K-BILD was concordant with FDA indications, but unfortunately led to the exclusion of cough by the Rasch theory [80]. This highlights the particular challenge of including a large heterogeneous derivative cohort to cover all relevant aspects of disease that can vary across severity.

Longitudinal responsiveness

Longitudinal studies are performed to ensure reliability (stable and consistent results), reproducibility (to obtain similar results under similar conditions) and responsiveness of PROMs (the ability to detect changes over time). Such robust psychometric properties are a pre-condition for PROM use and tools are generally considered valid if supported by this type of statistical evidence. For research use, PROMs must not only be valid and reliable, but also have minimal clinically important difference (MCID) determination, *i.e.* measure the smallest change in outcome that an individual patient would identify as important. Access to validated MCIDs for PROMs in ILD/IPF will facilitate understanding of the clinical relevance and magnitude of the effect of a specific intervention from a patient’s perspective, allow sample size calculations and help define expected end-points in clinical trials [103]. In addition, the context of MCID use and interpretation is an important factor to consider. For example, a recent study by Prior *et al.* [55] suggested that the MCID for determining improvement in HRQoL may be different from that for worsening as detected by both SGRQ-I and K-BILD in IPF. MCID for improvement differed from deterioration for both SGRQ-I total score (3.9 and 4.9, respectively) and K-BILD total score (4.7 and 2.7, respectively). The MCID of an instrument in different disease stages may be variable. In IPF, quality of life seems to worsen abruptly in the last years before death [104], and it may therefore be that PROMs have divergent outcomes in more advanced disease stages compared to early disease.

Cultural differences, differing expectations and language issues may also influence MCID. A systematic review suggested that a wide range of MCID values may exist for the same questionnaire and the difference between significance and clinical importance needs to be clarified. Therefore, MCID determined in one cohort of IPF patients may not be applicable to multicentre trials that enrol participants with wide-ranging sociocultural backgrounds with varying understanding of health and disease and related expectations. Consensus is needed for appropriate usage of MCID-determining equations [105]. Typically, a combination of anchor- and distribution-based methods are applied in this process, where current health status and change over time from a patient's point of view are used as anchors.

Specifically considering PROM measurement in ILD/IPF, longitudinal responsiveness was documented for some of the most used PROMs, including SGRQ, SGRQ-I and K-BILD [51, 62, 82]. However, a growing interest in PRO research in IPF has generated limited data that also support the validity of the baseline dyspnoea index/transition dyspnoea index, UCSD Shortness of Breath Questionnaire, SGRQ, SF-36, five-level EQ-5D and EQ-VAS [57] for use as longitudinal outcomes in IPF trials [25].

Sociocultural diversity in the different cohorts engaged in development and testing of PROMs are necessary to ensure required adaptations, appropriate sensitivity and accuracy in reflecting patient perceptions. For a PROM to be usable, it should embrace the intended patient population. For example, symptoms perceived as side-effects in Western cultures (for instance diarrhoea due to antifibrotic treatment) may in other cultures be seen as a desirable elimination of disease-related "badness". Thus, PROMs should be validated with respect to cultural background in addition to the usual psychometric properties.

PROM burden of use

The administrative burden of responding to PRO questionnaires, their collection, storage, interpretation and use is another relevant issue. The length of the questionnaire may impose a challenge even for the healthier patients; sicker patients with advanced disease might find that responding to PRO questionnaires is even more burdensome [106]. Potential factors related to response burden include the length and/or formatting of the questionnaire or interview, issues with literacy level and issues related to the mode of administration (*e.g.* paper-, telephone- or web-based surveys). In addition, there may be issues related to sensitive content of items that participants may be unwilling to answer, or patients' perception that an interviewer expects a specific response.

Many of the available IPF/ILD questionnaires are not practical for clinical use, mainly due to the time taken to complete the questionnaires. Time is the main barrier reported by clinicians to implementing questionnaires in their routine care, probably because many PROMs are extensive and mostly developed for research use [107]; shorter and more reliable measures can improve response rates [24]. However, longer questionnaires provide more possibilities to capture all aspects that are deemed important for an adequate evaluation of several aspects of the disease. There is an urgent need to make PROMs more efficient to aid implementation and compliance while maintaining its integrity. The use of computerised adaptive testing offers a potential solution by decreasing the question burden on the patients [108]. Data science tells us that the scores of PROMs can be accurately predicted from fewer questions if the correct questions are asked [109]. Computerised adaptive testing aims to identify the subset of questions from the full instrument to ask each patient based on his/her previous responses. This can reduce patient burden without compromising the precision of measuring their perceptions [110]. Therefore, computerised adaptive testing methodology for disease-specific PROMs must be studied in ILD. Striking the right balance between comprehensiveness, utility and length is challenging and may be tailored towards different aims, *i.e.* research or clinical practice. This requires frequent monitoring and reassessment of processes. Finally, successful implementation of PROMs requires support for healthcare practitioners and substantial resource investment for training the staff that could make practical data collection challenging.

Recommendations

PROMs are increasingly emphasised in ILD as we transition from disease-centred to patient-centred, and from volume-based to value-based healthcare. Patient-focused drug development also requires increasing use of PROMs. In the case of IPF/ILD, where a patient's quality of life is significantly impaired to begin with, assessments of any intervention must include patient perspectives on overall impact on function, symptoms and quality of life in addition to impact on traditional biomedical outcomes. PROMs must be measured and reported in trials, even if the interventions are not expected to improve PROs, to ensure that patient perspectives on tolerability, impact and burden of therapy are fully assessed. The use of such measures will allow clinicians and researchers alike to understand the total impact of disease and therapy on PROs and shed light on issues of tolerability and acceptability of therapy, which cannot be discerned

otherwise. Given the number of therapeutic agents in development with varying targets and probable diverse adverse effects that can negatively affect HRQoL, such measures should be included in clinical drug trials.

We highlight needs and recommendations for PROM use in IPF/ILD in table 3. The quality of a PROM and its ability to provide dependable results is essential for its meaningful use. ILD-specific PROMs should be prioritised and validated in both cross-sectional and longitudinal settings where different measurement properties are studied. The development and validation of instruments such as K-BILD, SGRQ-I and L-IPF will help move the field forward, and wherever possible, efforts should be made to use ILD-specific PROMs over PROMS such as SGRQ that were originally developed for other diseases. Validation can be perceived as an iterative process and should serve to improve the robustness and generalisability by assessing the PROM instruments in different populations, cultures and disease settings. In addition to robust psychometric properties, we need culturally sensitive, specific PROM tools responsive to varying disease severity, including end of life. Varying cultural, social and psychological factors influence patient perceptions; hence, what is meaningful and expected in one geographical area may differ significantly from another, underlining the need for testing across different populations. Cultural adaptation is often overlooked when developing and translating PROMs and this can have significant impact on quality of data, including attrition and missing data. There are many recommendations for adapting PROMs to ensure equivalence and validity across various cultural groups [111]. Pooling of data from international studies to assess cultural validity of psychometric properties of a PROM or to establish suitable MCIDs can be considered. Determination of MCID for both improvement and deterioration is desirable for appropriate interpretations and reporting of patients' HRQoL.

When deciding on PROM use in a clinical trial, the concept of interest measured by the PROMs should meet the aim of the research study. For example, in ILD drug trials, the investigators may want to comprehensively assess impact of intervention on all domains relevant to patients in addition to measuring traditional biomedical markers. This necessitates the use of PROMs that are disease-specific and validated HRQoL instruments (*e.g.* K-BILD, L-IPF) instead of using instruments such as SGRQ that were developed primarily for COPD populations. Another concept of interest can be patient self-efficacy; CRQ may be a good choice in this case, as it measures mastery of symptom control as a domain. Similarly, Dyspnoea-12 can be helpful to assess physical and affective responses to breathlessness without depending on activity levels, which are the focus in many dyspnoea instruments. Recall periods of PROMs are variable, and this should be kept in mind when choosing instruments. PROM use in drug trials poses another challenge. All drug interventions are likely to cause adverse effects with varying impacts on HRQoL; these may not be captured by the disease-specific PROMs and additional instruments may be needed to assess impact of these adverse effects. For example, diarrhoea or skin rash, which as adverse effects can impair HRQoL, are not measured by ILD/IPF-specific HRQoL instruments, but can be meaningfully reported from the patient perspective. Specific PROMs to capture adverse effects, such as the Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events [112], were developed and tested for use in cancer

TABLE 3 Urgent needs and recommendations for patient-reported outcome measures (PROMs) in idiopathic pulmonary fibrosis/interstitial lung disease (ILD)

	Recommendations
Choice of PROM	ILD-specific PROMs should be prioritised The concept of interest in the PROM should meet the aim of the clinical trial PROMs should be selected according to the population of interest PROMs to measure impact of adverse effect/harm from drug interventions Measure patient experience of therapy burden Development based on IRT or Rasch measurement theory
PROM validation	Language validation Cultural validation Cross-sectional and longitudinal validation (MCID) MCID determination for improvement and deterioration
Operability	PROMs should be simple and short PROMs should be linguistically adapted to the population of interest PROM could utilise computerised adaptive testing
Collection of data in real time	Use a device/device of patients' choice Use new technology, including audio and touch screen options

IRT: item response theory; MCID: minimal clinically important difference.

drug trials. Such tools can help assess impact/harm of cancer therapies as PROs. Similar tools for ILD/IPF are needed. Reporting of adverse effect frequency, severity and distribution alone is not enough, but the degree of interference with HRQoL should also be considered. The need for measurement of both harm and benefit from the patient perspective should be considered when selecting PROMs in drug trials. The burden of therapy may need to be measured if the administration is complex (parenteral route or number of pills, *etc*.). Instruments such as patient experience and satisfaction with medication and other patient-reported experience measures may be useful in this regard. The selection of PROMs that measure symptoms of interest such as dyspnoea, cough or fatigue may also be considered, as they are more relevant to patients and it is important to document positive or negative impact of interventions on symptoms. It should be selected according to the population of interest (*e.g.* early *versus* advanced disease) and be available in several languages, with cultural validity as described earlier. Results of PROM may vary in different populations. For instance, patients in clinical trials are highly selected and differences between trial and non-trial populations may very well exist. This must be acknowledged in the interpretation of results. A PROM should consider whether the population is initiating treatment, on treatment or living with or beyond the disease [113]. In addition, appropriate selection requires that users are familiar with PROMs, their properties, MCID, attendant limitations and the need for contextual interpretation. Researchers must select the appropriate MCID for the PROM of interest, as the unit change for improvement and deterioration may be different. One way to improve PROMs is to involve patients more, not only as consultants, but also as active partners in the development and implementation of PROMs. Patient involvement can optimise the relevance of a PROM and support patient-centred research and care [114]. Patients can advise on using acceptable language to make the completion of an instrument easier, and to ensure that it is not burdensome for well or unwell patients [113]. To improve the integration of the patient voice throughout the development, implementation and dissemination of a PROM, patients need to be part of the process from the beginning [115]. By including patients from the outset, this can enable a truly collaborative approach with patients, patient advocates and, in some instances, caregivers [115, 116].

PROMs should be practically feasible and operational, and easy to implement for patients and the healthcare staff administering them. New digital tools enable patients to complete questionnaires online with direct data transmission either from home or at the clinic, reducing some of the burden. Use of digital platforms can ensure consistent administration, storage and interpretation of data. The imposed response burden on patients and healthcare providers must be carefully weighed against the clinical and research need for comprehensive tools. PROMs with the minimum possible number of items that can provide a broad assessment of relevant domains are needed to reduce this burden. With increasing confidence in well-developed PROM tools and their greater use in clinical studies, we can look forward to exciting times in medicine where patients and their perceptions are truly at the centre of care and research.

Provenance: Submitted article, peer reviewed

Acknowledgements: 3rd International Summit for ILD (ISILD), Erice collaborators listed below. Carlo Albera, Goksel Altinisik, Kjetil Ask, Elisabetta Balestro, Elena Bargagli, Elisabeth Bendstrup, Marialuisa Bocchino, Francesco Bonella, Martina Bonifazi, Giulia Cacopardo, Maria Calvello, Diego M. Castillo, Nazia Chaudhuri, Ulrich Costabel, Vincent Cottin, Bruno Crestani, Manuela Funke-Chambour, Jack Gauldie, Peter M. George, Johannes C. Grutters, Sergio Harari, Richard G. Jenkins, Kerri A. Johannson, Mark G. Jones, Meena Kalluri, Michael P. Keane, Maria A. Kokosi, Michael Kreuter, Donato Lacedonia, Brett Ley, Alessandro Libra, Fabrizio Luppi, Toby M. Maher, George A. Margaritopoulos, Fernando J. Martinez, Jelle Miedema, Nesrin Mogulkoc, Maria Molina-Molina, Philip L. Molyneaux, Julie Morisset, Stefano Palmucci, Mauro Pavone, Ganesh R. Raghu, Elisabetta A. Renzoni, Luca Richeldi, Gianluca Sambataro, Alfredo Sebastiani, Paolo Spagnolo, Giulia Maria Stella, Martina Sterclova, Irina Strambu, Sara Tomassetti, Sebastiano Torrisi, Jacopo Simonetti, Haluk Turktas, Argyrios Tzouvelekis, Claudia Valenzuela, Ada Vancheri, Carlo Vancheri, Francesco Varone, Patrizio Vitulo, Athol U. Wells, Marlies S. Wijsenbeek, Wim A. Wuyts, and London North West University Hospital Healthcare Trust (London, UK).

The authors would like to thank M. Wapenaar for the development of figure 2 (Centre of excellence, Interstitial Lung Diseases and Sarcoidosis, Erasmus MC, University Medical Centre Rotterdam, Rotterdam, The Netherlands).

Conflict of interest: M. Kalluri reports grants and personal fees from Boehringer Ingelheim, and Roche, outside the submitted work. F. Luppi reports grants and lecture fees from Roche and lecture fees from Boehringer Ingelheim. A. Vancheri has nothing to disclose. C. Vancheri reports grants and personal fees from Roche and Boehringer Ingelheim, outside the submitted work. E. Balestro reports personal fees from Roche and Boehringer Ingelheim, outside the submitted work. F. Varone reports unrestricted grants, consultancy and lecture fees from Roche and Boehringer Ingelheim. N. Mogulkoc has nothing to disclose. G. Cacopardo has nothing to disclose. E. Bargagli has nothing to disclose. E. Renzoni reports grants from Boeringher Ingelheim, and lecture fees from Boeringher

Ingelheim and Roche, outside the submitted work. S. Torrisi reports personal fees from Hoffman La-Roche and Boehringer Ingelheim, outside the submitted work. M. Calvello has nothing to disclose. A. Libra has nothing to disclose. M. Pavone has nothing to disclose. F. Bonella reports grants and personal fees from Boehringer Ingelheim and Roche, and consultancy fees from BMS, Galapagos, GSK and Savara, outside the submitted work. V. Cottin reports personal fees and non-financial support from Actelion and Roche/Promedior; grants, personal fees and non-financial support from Boehringer Ingelheim; and personal fees from Bayer/MSD, Novartis, Sanofi, Celgene/BMS, Galapagos, Galecto, Shionogi, AstraZeneca and Fibrogen, outside the submitted work. C. Valenzuela reports personal fees from Boehringer Ingelheim, F. Hoffmann-La Roche, Galapagos and BMS, outside the submitted work. M. Wijsenbeek reports grants and other funding from Boehringer Ingelheim and Hoffman-La Roche, and other funding from Respivant, Galapagos, Novartis and Savara, outside the submitted work. All fees and grants were paid to her institution. E. Bendstrup reports grants and personal fees from Boehringer Ingelheim and Hoffmann-La Roche, and personal fees from Galapagos and AstraZeneca, outside the submitted work.

References

1 Wijsenbeek M, Cottin V. Spectrum of fibrotic lung diseases. *N Engl J Med* 2020; 383: 958–968.
2 Swigris JJ, Stewart AL, Gould MK, *et al.* Patients' perspectives on how idiopathic pulmonary fibrosis affects the quality of their lives. *Health Qual Life Outcomes* 2005; 3: 61.
3 van Manen MJG, Geelhoed JJM, Tak NC, *et al.* Optimizing quality of life in patients with idiopathic pulmonary fibrosis. *Ther Adv Respir Dis* 2017; 11: 157–169.
4 International Alliance of Patients' Organizations (IAPO). What is Patient-Centred Healthcare? A Review of Definitions and Principles. 2007, London, IAPO.
5 Ferrara G, Luppi F, Birring SS, *et al.* Best supportive care for idiopathic pulmonary fibrosis: current gaps and future directions. *Eur Respir Rev* 2018; 27: 170076.
6 Swigris J, Richeldi L, Wisjenbeek M, *et al.* Effects of nintedanib on dyspnea, cough and quality of life in patients with progressive fibrosing interstitial lung diseases: findings from the INBUILD trial. *Am J Respir Crit Care Med* 2020; 201: A2754.
7 Sampson C, Gill BH, Harrison NK, *et al.* The care needs of patients with idiopathic pulmonary fibrosis and their carers (CaNoPy): results of a qualitative study. *BMC Pulm Med* 2015; 15: 155.
8 Maher TM, Molina-Molina M, Russell A-M, *et al.* Unmet needs in the treatment of idiopathic pulmonary fibrosis – insights from patient chart review in five European countries. *BMC Pulm Med* 2017; 17: 124.
9 Moor CC, Wijsenbeek MS, Balestro E, *et al.* Gaps in care of patients living with pulmonary fibrosis: a joint patient and expert statement on the results of a Europe-wide survey. *ERJ Open Res* 2019; 5: 00124–2019.
10 Bonella F, Wijsenbeek M, Molina-Molina M, *et al.* European IPF Patient Charter: unmet needs and a call to action for healthcare policymakers. *Eur Respir J* 2016; 47: 597–606.
11 Kalluri M, Luppi F, Ferrara G. What patients with idiopathic pulmonary fibrosis and caregivers want: filling the gaps with patient reported outcomes and experience measures. *Am J Med* 2020; 133: 281–289.
12 Kalitzkus V, Matthiessen PF. Narrative-based medicine: potential, pitfalls, and practice. *Perm J* 2009; 13: 80–86.
13 Acquadro C, Berzon R, Dubois D, *et al.* Incorporating the patient's perspective into drug development and communication: an *ad hoc* task force report of the Patient-Reported Outcomes (PRO) Harmonization Group meeting at the Food and Drug Administration, February 16, 2001. *Value Health* 2003; 6: 522–531.
14 Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. *BJA Education* 2017; 17: 137–144.
15 Jayadevappa R. Patient-centered outcomes research and patient-centered care for older adults: a perspective. *Gerontol Geriatr Med* 2017; 3: 2333721417700759.
16 Field J, Holmes MM, Newell D. PROMs data: can it be used to make decisions for individual patients? A narrative review. *Patient Relat Outcome Meas* 2019; 10: 233–241.
17 Black N. Patient reported outcome measures could help transform healthcare. *BMJ* 2013; 346: f167.
18 Nelson EC, Eftimovska E, Lind C, *et al.* Patient reported outcome measures in practice. *BMJ* 2015; 350: g7818.
19 Santana M-J, Feeny D. Framework to assess the effects of using patient-reported outcome measures in chronic care management. *Qual Life Res* 2014; 23: 1505–1513.
20 Hjollund NHI, Larsen LP, Biering K, *et al.* Use of patient-reported outcome (PRO) measures at group and patient levels: experiences from the generic integrated PRO system, WestChronic. *Interact J Med Res* 2014; 3: e5.
21 Basch E, Dueck AC. Patient-reported outcome measurement in drug discovery: a tool to improve accuracy and completeness of efficacy and safety data. *Expert Opin Drug Discov* 2016; 11: 753–758.
22 Rotenstein LS, Huckman RS, Wagle NW. Making patients and doctors happier – the potential of patient-reported outcomes. *N Engl J Med* 2017; 377: 1309–1312.
23 Jensen RE, Rothrock NE, DeWitt EM, *et al.* The role of technical advances in the adoption and integration of patient-reported outcomes in clinical care. *Med Care* 2015; 53: 153–159.
24 Weldring T, Smith SMS. Patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs). *Health Serv Insights* 2013; 6: 61–68.

25 Swigris JJ, Fairclough D. Patient-reported outcomes in idiopathic pulmonary fibrosis research. *Chest* 2012; 142: 291–297.
26 Russell A-M, Sprangers MA, Wibberley S, *et al.* The need for patient-centred clinical research in idiopathic pulmonary fibrosis. *BMC Med* 2015; 13: 240.
27 Nathan SD, Meyer KC. IPF clinical trial design and endpoints. *Curr Opin Pulm Med* 2014; 20: 463–471.
28 Raghu G, Collard HR, Anstrom KJ, *et al.* Idiopathic pulmonary fibrosis: clinically meaningful primary endpoints in phase 3 clinical trials. *Am J Respir Crit Care Med* 2012; 185: 1044–1048.
29 Azuma A, Nukiwa T, Tsuboi E, *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.
30 Demedts M, Behr J, Buhl R, *et al.* High-dose acetylcysteine in idiopathic pulmonary fibrosis. *N Engl J Med* 2005; 353: 2229–2242.
31 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: 948–955.
32 King TE, Albera C, Bradford WZ, *et al.* Effect of interferon gamma-1b on survival in patients with idiopathic pulmonary fibrosis (INSPIRE): a multicentre, randomised, placebo-controlled trial. *Lancet* 2009; 374: 222–228.
33 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.
34 King TE, 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: 92–99.
35 Noble PW, Albera C, Bradford WZ, *et al.* Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. *Lancet* 2011; 377: 1760–1769.
36 Richeldi L, Costabel U, Selman M, *et al.* Efficacy of a tyrosine kinase inhibitor in idiopathic pulmonary fibrosis. *N Engl J Med* 2011; 365: 1079–1087.
37 Raghu G, Behr J, Brown KK, *et al.* Treatment of idiopathic pulmonary fibrosis with ambrisentan: a parallel, randomised trial. *Ann Intern Med* 2013; 158: 641–649.
38 King TE, 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.
39 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.
40 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.
41 Visca D, Mori L, Tsipouri V, *et al.* Effect of ambulatory oxygen on quality of life for patients with fibrotic lung disease (AmbOx): a prospective, open-label, mixed-method, crossover randomised controlled trial. *Lancet Respir Med* 2018; 6: 759–770.
42 Kolb M, Raghu G, Wells AU, *et al.* Nintedanib plus sildenafil in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 2018; 379: 1722–1731.
43 Flaherty KR, Wells AU, Cottin V, *et al.* Nintedanib in progressive fibrosing interstitial lung diseases. *N Engl J Med* 2019; 381: 1718–1727. .
44 Maher TM, Corte TJ, Fischer A, *et al.* Pirfenidone in patients with unclassifiable progressive fibrosing interstitial lung disease: a double-blind, randomised, placebo-controlled, phase 2 trial. *Lancet Respir Med* 2020; 8: 147–157.
45 Behr J, Nathan SD, Wuyts WA, *et al.* Efficacy and safety of sildenafil added to pirfenidone in patients with advanced idiopathic pulmonary fibrosis and risk of pulmonary hypertension: a double-blind, randomised, placebo-controlled, phase 2b trial. *Lancet Respir Med* 2021; 9: 85–95.
46 Dowman L, Hill CJ, May A, *et al.* Pulmonary rehabilitation for interstitial lung disease. *Cochrane Database Syst Rev* 2021; 2: CD006322.
47 Jones PW, Quirk FH, Baveystock CM. The St George's Respiratory Questionnaire. *Respir Med* 1991; 85: 25–31.
48 Jones PW, Quirk FH, Baveystock CM, *et al.* A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire. *Am Rev Respir Dis* 1992; 145: 1321–1327.
49 Barr JT, Schumacher GE, Freeman S, *et al.* American translation, modification, and validation of the St. George's Respiratory Questionnaire. *Clin Ther* 2000; 22: 1121–1145.
50 Swigris JJ, Esser D, Conoscenti CS, *et al.* The psychometric properties of the St George's Respiratory Questionnaire (SGRQ) in patients with idiopathic pulmonary fibrosis: a literature review. *Health Qual Life Outcomes* 2014; 12: 124.
51 Prior TS, Hoyer N, Shaker SB, *et al.* Validation of the IPF-specific version of St. George's Respiratory Questionnaire. *Respir Res* 2019; 20: 199.
52 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.
53 Suzuki A, Kondoh Y, Swigris JJ, *et al.* Performance of the St George's Respiratory Questionnaire in patients with connective tissue disease-associated interstitial lung disease. *Respirology* 2018; 23: 851–859.

54 Welling JBA, Hartman JE, Ten Hacken NHT, *et al.* The minimal important difference for the St George's Respiratory Questionnaire in patients with severe COPD. *Eur Respir J* 2015; 46: 1598–1604.
55 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.
56 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: 921–926.
57 Tsai APY, Hur SA, Wong A, *et al.* Minimum important difference of the EQ-5D-5L and EQ-VAS in fibrotic interstitial lung disease. *Thorax* 2021; 76: 37–43.
58 Dyer MT, Goldsmith KA, Sharples LS, *et al.* A review of health utilities using the EQ-5D in studies of cardiovascular disease. *Health Qual Life Outcomes* 2010; 8: 13.
59 Bae E, Choi S-E, Lee H, *et al.* Validity of EQ-5D utility index and minimal clinically important difference estimation among patients with chronic obstructive pulmonary disease. *BMC Pulm Med* 2020; 20: 73.
60 Guyatt GH, Berman LB, Townsend M, *et al.* A measure of quality of life for clinical trials in chronic lung disease. *Thorax* 1987; 42: 773–778,
61 Patel AS, Siegert RJ, Brignall K, *et al.* The development and validation of the King's Brief Interstitial Lung Disease (K-BILD) health status questionnaire. *Thorax* 2012; 67: 804–810.
62 Sinha A, Patel AS, Siegert RJ, *et al.* The King's Brief Interstitial Lung Disease (KBILD) questionnaire: an updated minimal clinically important difference. *BMJ Open Respir Res* 2019; 6: e000363.
63 Nolan CM, Birring SS, Maddocks M, *et al.* King's Brief Interstitial Lung Disease questionnaire: responsiveness and minimum clinically important difference. *Eur Respir J* 2019; 54: 1900281.
64 Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992; 30: 473–83.
65 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.
66 Swigris JJ, Andrae DA, Churney T, *et al.* Development and initial validation analyses of the Living with Idiopathic Pulmonary Fibrosis (L-IPF) questionnaire. *Am J Respir Crit Care Med* 2020; 202: 1689–1697.
67 Russell A-M, Wickremasinghe M, Saketkoo LA, *et al.* Preliminary testing of the Idiopathic Pulmonary Fibrosis Patient Reported Outcome Measure (IPF PRoM). *Am J Respir Crit Care Med* 2018; 197: A7707.
68 Russell A-M, Jones G, Saketkoo L, *et al.* Development and preliminary testing of the Idiopathic Pulmonary Fibrosis Patient Reported Outcome Measure (IPF-PRoM): UK and Ireland Multi-Centre Study. *Am J Respir Crit Care Med* 2017; 195: A1031.
69 Eakin EG, Resnikoff PM, Prewitt LM, *et al.* Validation of a new dyspnea measure: the UCSD Shortness of Breath Questionnaire. *Chest* 1998; 113: 619–624.
70 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.
71 Kupferberg DH, Kaplan RM, Slymen DJ, *et al.* Minimal clinically important difference for the UCSD Shortness of Breath Questionnaire. *J Cardiopulm Rehabil* 2005; 25: 370–377.
72 Yorke J, Moosavi SH, Shuldham C, *et al.* Quantification of dyspnoea using descriptors: development and initial testing of the Dyspnoea-12. *Thorax* 2010; 65: 21–26.
73 Yorke J, Swigris J, Russell A-M, *et al.* Dyspnea-12 is a valid and reliable measure of breathlessness in patients with interstitial lung disease. *Chest* 2011; 139: 159–164.
74 Birring SS, Prudon B, Carr AJ, *et al.* Development of a symptom specific health status measure for patients with chronic cough: Leicester Cough Questionnaire (LCQ). *Thorax* 2003; 58: 339–343.
75 Berkhof FF, Boom LN, ten Hertog NE, *et al.* The validity and precision of the Leicester Cough Questionnaire in COPD patients with chronic cough. *Health Qual Life Outcomes* 2012; 10: 4.
76 Crawford B, Monz B, Hohlfeld J, *et al.* Development and validation of a cough and sputum assessment questionnaire. *Respir Med* 2008; 102: 1545–1555.
77 Gries KS, Esser D, Wiklund I. Content validity of CASA-Q cough domains and UCSD-SOBQ for use in patients with idiopathic pulmonary fibrosis. *Glob J Health Sci* 2013; 5: 131–141.
78 Grufstedt HK, Shaker SB, Konradsen H. Validation of the COPD Assessment Test (CAT) in patients with idiopathic pulmonary fibrosis. *Eur Clin Respir J* 2018; 5: 1530028.
79 Szentes BL, Kreuter M, Bahmer T, *et al.* Quality of life assessment in interstitial lung diseases: a comparison of the disease-specific K-BILD with the generic EQ-5D-5L. *Respir Res* 2018; 19: 101.
80 U.S. Food and Drug Administration. Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009. www.fda.gov/media/77832/download
81 European Medicines Agency. Reflection Paper on the Regulatory Guidance for the Use of Health Related Quality of Life (HRQL) Measures in the Evaluation of Medicinal Products. 2005. www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-regulatory-guidance-use-healthrelated-quality-life-hrql-measures-evaluation_en.pdf
82 Prior TS, Hilberg O, Shaker SB, *et al.* Validation of the King's Brief Interstitial Lung Disease questionnaire in idiopathic pulmonary fibrosis. *BMC Pulm Med* 2019; 19: 255.

83 Slade A, Isa F, Kyte D, *et al.* Patient reported outcome measures in rare diseases: a narrative review. *Orphanet J Rare Dis* 2018; 13: 61.
84 Willis GB. Cognitive Interviewing: A Tool for Improving Questionnaire Design. 1st Edn. Thousand Oaks: Sage Publications, 2004.
85 Wijsenbeek M, van Manen M, Bonella F. New insights on patient-reported outcome measures in idiopathic pulmonary fibrosis: only PROMises? *Curr Opin Pulm Med* 2016; 22: 434–441.
86 Nishiyama O, Taniguchi H, Kondoh Y, *et al.* A simple assessment of dyspnoea as a prognostic indicator in idiopathic pulmonary fibrosis. *Eur Respir J* 2010; 36: 1067–1072.
87 Mahler DA, Witek TJ. The MCID of the transition dyspnea index is a total score of one unit. *COPD* 2005; 2: 99–103.
88 Borg GAV. Psychophysical bases of perceived exertion. *Med Sci Sports Exerc* 1982; 14: 377–381.
89 Moor CC, Heukels P, Kool M, *et al.* Integrating patient perspectives into personalized medicine in idiopathic pulmonary fibrosis. *Front Med* 2017; 4: 226.
90 Lechtzin N, Hilliard ME, Horton MR. Validation of the Cough Quality-of-Life Questionnaire in patients with idiopathic pulmonary fibrosis. *Chest* 2013; 143: 1745–1749.
91 Key AL, Holt K, Hamilton A, *et al.* Objective cough frequency in idiopathic pulmonary fibrosis. *Cough* 2010; 6: 4.
92 Michielsen HJ, De Vries J, Van Heck GL. Psychometric qualities of a brief self-rated fatigue measure: the Fatigue Assessment Scale. *J Psychosom Res* 2003; 54: 345–352.
93 Kølner-Augustson L, Prior TS, Skivild V, *et al.* Fatigue in idiopathic pulmonary fibrosis measured by the Fatigue Assessment Scale during antifibrotic treatment. *Eur Clin Respir J* 2020; 8: 1853658.
94 Kreuter M, Swigris J, Pittrow D, *et al.* Health related quality of life in patients with idiopathic pulmonary fibrosis in clinical practice: INSIGHTS-IPF registry. *Respir Res* 2017; 18: 139.
95 Bristowe K, Murtagh FEM, Clift P, *et al.* The development and cognitive testing of the positive outcomes HIV PROM: a brief novel patient-reported outcome measure for adults living with HIV. *Health Qual Life Outcomes* 2020; 18: 214.
96 Kreuter M, Wuyts WA, Wijsenbeek M, *et al.* Health-related quality of life and symptoms in patients with IPF treated with nintedanib: analyses of patient-reported outcomes from the INPULSIS® trials. *Respir Res* 2020; 21: 36.
97 Neale J, Vitoratou S, Lennon P, *et al.* Development and early validation of a patient-reported outcome measure to assess sleep amongst people experiencing problems with alcohol or other drugs. *Sleep* 2018; 41: zsy013.
98 Cronbach L. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; 16: 297–334.
99 Cronbach L. Test “reliability”: its meaning and determination. *Psychometrika* 1947; 12: 1–16.
100 Bottomley A, Jones D, Claassens L. Patient-reported outcomes: assessment and current perspectives of the guidelines of the Food and Drug Administration and the reflection paper of the European Medicines Agency. *Eur J Cancer* 2009; 45: 347–353.
101 Olson AL, Brown K, Swigris J. Understanding and optimizing health-related quality of life and physical functional capacity in idiopathic pulmonary fibrosis. *Patient Relat Outcome Meas* 2016; 7: 29–35.
102 Grønseth R, Vollmer WM, Hardie JA, *et al.* Predictors of dyspnoea prevalence: results from the BOLD study. *Eur Respir J* 2014; 43: 1610–1620.
103 Rebelo P, Oliveira A, Paixão C, *et al.* Minimal clinically important differences for patient-reported outcome measures of cough and sputum in patients with COPD. *Int J Chron Obstruct Pulmon Dis* 2020; 15: 201–212.
104 Rajala K, Lehto JT, Sutinen E, *et al.* Marked deterioration in the quality of life of patients with idiopathic pulmonary fibrosis during the last two years of life. *BMC Pulm Med* 2018; 18: 172.
105 Mouelhi Y, Jouve E, Castelli C, *et al.* How is the minimal clinically important difference established in health-related quality of life instruments? Review of anchors and methods. *Health Qual Life Outcomes* 2020; 18: 136.
106 Atkinson TM, Schwartz CE, Goldstein L, *et al.* Perceptions of response burden associated with completion of patient-reported outcome assessments in oncology. *Value Health* 2019; 22: 225–230.
107 Kocks JWH, Blom CMG, Kasteleyn MJ, *et al.* Feasibility and applicability of the paper and electronic COPD assessment test (CAT) and the clinical COPD questionnaire (CCQ) in primary care: a clinimetric study. *NPJ Prim Care Respir Med* 2017; 27: 20.
108 Hsueh I-P, Chen J-H, Wang C-H, *et al.* Development of a computerized adaptive test for assessing activities of daily living in outpatients with stroke. *Phys Ther* 2013; 93: 681–693.
109 Weiss DJ. Improving measurement quality and efficiency with adaptive testing. *Appl Psychol Meas* 1982; 6: 473–492.
110 Chien T-W, Wang W-C, Huang S-Y, *et al.* A web-based computerized adaptive testing (CAT) to assess patient perception in hospitalization. *J Med Internet Res* 2011; 13: e61.
111 Beaton DE, Bombardier C, Guillemin F, *et al.* Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000; 25: 3186–3191.
112 Dueck AC, Mendoza TR, Mitchell SA, *et al.* Validity and reliability of the US National Cancer Institute’s Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *JAMA Oncol* 2015; 1: 1051–1059.

113 Retzer A, Keeley T, Ahmed K, *et al.* Evaluation of patient-reported outcome protocol content and reporting in UK cancer clinical trials: the EPiC study qualitative protocol. *BMJ Open* 2018; 8: e017282.
114 Staniszewska S, Haywood KL, Brett J, *et al.* Patient and public involvement in patient-reported outcome measures: evolution not revolution. *Patient* 2012; 5: 79–87.
115 de Wit MPT, Kvien TK, Gossec L. Patient participation as an integral part of patient-reported outcomes development ensures the representation of the patient voice: a case study from the field of rheumatology. *RMD Open* 2015; 1: e000129.
116 Wilson H, Dashiell-Aje E, Anatchkova M, *et al.* Beyond study participants: a framework for engaging patients in the selection or development of clinical outcome assessments for evaluating the benefits of treatment in medical product development. *Qual Life Res* 2018; 27: 5–16.