The measurement properties of tests and tools used in cystic fibrosis studies: a systematic review

Charlie McLeod [1,2,3], Jamie Wood [4], Allison Tong [5,6], André Schultz [7,8], Richard Norman [9], Sherie Smith [10], Christopher C. Blyth [1,2,3,11], Steve Webb [12,13], Alan R. Smyth [10] and Thomas L. Snelling [5,14]

[1]Wesfarmers Centre for Vaccines and Infectious Diseases, Telethon Kids Institute, Nedlands, Australia. [2]Division of Paediatrics, Faculty of Medicine, University of Western Australia, Nedlands, Australia. [3]Infectious Diseases Dept, Perth Children's Hospital, Nedlands, Australia. [4]Physiotherapy Dept, Sir Charles Gairdner Hospital, Nedlands, Australia. [5]Sydney School of Public Health, The University of Sydney, Sydney, Australia. [6]Centre for Kidney Research, The Children's Hospital at Westmead, Westmead, Australia. [7]Wal-yan Respiratory Research Centre, Telethon Kids Institute, University of Western Australia, Nedlands, Australia. [8]Dept of Respiratory and Sleep Medicine, Perth Children's Hospital, Nedlands, Australia. [9]School of Population Health, Curtin University, Bentley, Australia. [10]Evidence Based Child Health Group, University of Nottingham, Queens Medical Centre, Nottingham, UK. [11]Pathwest Laboratory Medicine WA, QEII Medical Centre, Nedlands, Australia. [12]St John of God Hospital, Subiaco, Australia. [13]School of Population Health and Preventive Medicine, Monash University, Melbourne, Australia. [14]Menzies School of Health Research, Royal Darwin Hospital Campus, Tiwi, Australia.

Corresponding author: Charlie McLeod (charlie.mcleod@health.wa.gov.au)

Shareable abstract (@ERSpublications)
A diverse range of tests and tools were found, these varied with respect to their measurement properties. There was inconsistency in the selection of tests and tools to measure the same/similar outcomes across studies. Consensus is required. https://bit.ly/3nw2uoE

Cite this article as: McLeod C, Wood J, Tong A, *et al*. The measurement properties of tests and tools used in cystic fibrosis studies: a systematic review. *Eur Respir Rev* 2021; 30: 200354 [DOI: 10.1183/16000617.0354-2020].

This article has supplementary material available from err.ersjournals.com

Received: 3 Nov 2020
Accepted: 26 Dec 2020

Abstract

There is no consensus on how best to measure responses to interventions among children and adults with cystic fibrosis (CF). We have systematically reviewed and summarised the characteristics and measurement properties of tests and tools that have been used to capture outcomes in studies among people with CF, including their reliability, validity and responsiveness. This review is intended to guide researchers when selecting tests or tools for measuring treatment effects in CF trials. A consensus set of these tests and tools could improve consistency in how outcomes are captured and thereby facilitate comparisons and synthesis of evidence across studies.

Introduction

Research is conducted to generate evidence of the efficacy and safety of interventions to inform best clinical practice and thereby improve outcomes for patients. When designing studies, it is necessary to establish which outcomes must be evaluated to meet the study objectives and how these outcomes will be measured and analysed as end-points [1]. Tests or tools may be required for outcome measurement. To improve consistency and facilitate synthesis of evidence across studies, there is a need to establish a consensus set of these tests and tools for measuring outcomes in studies in people with cystic fibrosis (CF). These must be responsive to changes in the outcome of interest and capture outcomes with sufficient validity, reliability and precision [2, 3]. This is necessary so that results can be interpreted with confidence and be used to support the translation of evidence into practice.

There is no existing consensus on the selection of tests or tools for measuring outcomes. Selection is challenging for a number of reasons. First, tests and tools may lack appropriation validation, and hence their quality might be uncertain. Secondly, criteria to facilitate interpretation of the results of the test or tool may not exist. Thirdly, logistic or economic constraints may restrict the use of some. Finally, although initiatives to improve and standardise the use of patient-reported outcome measures (PROMs) have been

established [4, 5], there is no standardised approach for evaluating and selecting optimal tests and tools more generally in clinical research.

As a first step towards developing a consensus set of tests and tools for measuring outcomes in CF studies, we aimed to evaluate and summarise the characteristics and properties of tests and tools that have been used in previous CF studies. In the interim, we hope this review will be used by clinicians, people with CF, researchers and policy makers to guide optimal selection of these tests and tools, and to encourage validation or development of new tests or tools for measurement where required.

Methods

Search strategy and selection criteria

This was a PROSPERO registered systematic review (CRD42020151785). The search strategy is provided in the supplementary material. Medline, Embase and the Cochrane database were searched from inception until July 2020. Outcome measures proposed for evaluation in the Clinicaltrials.gov registry were also evaluated.

Inclusion criteria were reports of randomised controlled trials, observational studies, conference abstracts and reviews written in English, evaluating one or more measurement properties of a test or tool used to measure health outcomes in studies among people with CF. Original studies were sought to provide additional information about the characteristics or measurement properties of the tests and tools where necessary. Registered trials without published results proposing evaluation of one or more measurement properties of novel tests or tools were also included. Exclusion criteria were tests or tools developed for diagnostic purposes or used for evaluation of microbiological outcomes, or validation studies written in languages other than English. Tests and tools used in people with CF but validated in non-CF populations were beyond the scope of this review.

Titles and abstracts were screened independently by two reviewers (C. McLeod and J. Wood). Potentially eligible studies were downloaded to Endnote and duplicates removed. Full text articles were retrieved and eligibility confirmed by consensus of the reviewers. A third reviewer (T.L. Snelling) was used to confirm eligibility where consensus could not be achieved. Relevant data were extracted by C. McLeod and recorded in an Excel database and cross-checked by J. Wood.

The following characteristics of selected tests or tools were recorded: the outcome construct measured; the target population; mode of administration of the test or tool; recall period (if relevant); time taken to perform the test; the range of possible scores; and the ease of use (feasibility). The following properties of measurement were critically appraised: 1) validity, including content validity, construct validity (including convergent and discriminant performance of the test, the structural validity and cross-cultural validity) and criterion validity (including concurrent and predictive validity); 2) reliability, including the test–retest and inter-/intra-rate reliability, internal consistency and measurement error; 3) responsiveness; and 4) the minimal clinically important difference (MCID). Definitions for these measurement properties were based on those provided by the Consensus-based Standards for the selection of health Measurement INstruments initiative (COSMIN); these are presented in supplementary appendix 1.

Definitions, abbreviations and citations

Quality of life (QoL) tools were broadly defined as those which capture an individuals' perception of their life satisfaction relative to their goals in the context of their culture and value systems, and not those that capture QoL based solely on the health status of the individual *per se* (health-related quality of life; HRQoL) [6]. Disease-specific QoL tools were defined as those developed for measuring QoL in people with CF, whereas generic QoL tools were defined as those originally developed for use in other populations that have also been applied in studies involving people with CF.

A full list of abbreviations and their meanings used throughout this manuscript and supplementary materials are alphabetically listed in supplementary appendix 2. References for information presented in the tables throughout this manuscript are provided in the supplementary materials.

Results

The review process is depicted in figure 1. 118 studies evaluating the measurement properties of 74 tests and tools used in studies among people with CF identified from Medline, Embase or Clinical Trials met the inclusion criteria [7–119]; a summary of these studies is provided in table S3. This review included three registered studies proposing validation of tests or tools used in people with CF with unpublished

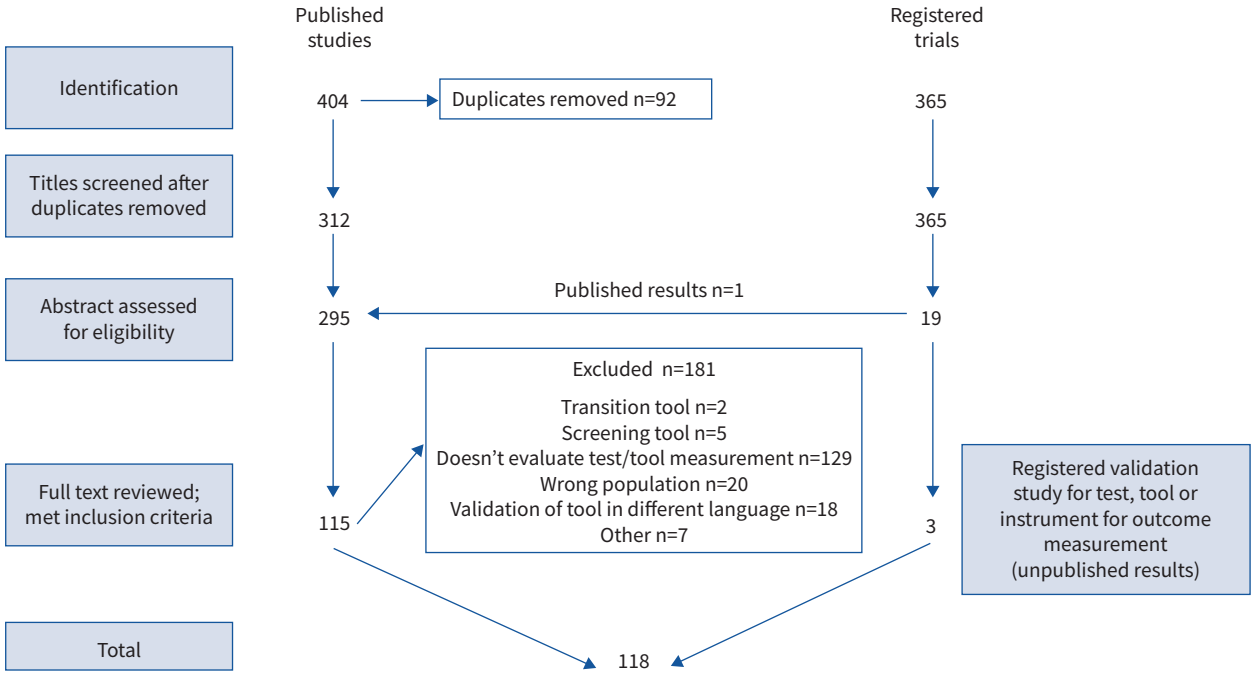

FIGURE 1 Search strategy flow chart.

results [120]. Nine source articles describing the characteristics or measurement properties of tests or tools were also included [121–129].

Characteristics of tests and tools

Tests or tools were categorised as PROMs capturing QoL or other patient-reported outcomes, clinical scores, radiological scores or tests capturing functional exercise performance, CF transmembrane conductance regulator (CFTR) function or sputum characteristics.

QoL tools

17 generic QoL tools evaluated in CF populations and seven CF-specific QoL were identified. Characteristics of these generic and CF-specific QoL tools are detailed in table S5 and table 1, respectively.

The Cystic Fibrosis Questionnaire-Revised (CFQ-R; original version 2003) has been the most widely used QoL tool reported in CF studies and is available in 34 languages [130]. It is endorsed for use in clinical trials by the US Food and Drug Administration (FDA) and European Medicines Agency [131, 132]. There are five versions available; these are described in table 1.

One QoL questionnaire for use by carers of people with CF was identified, the Carer QoL in CF questionnaire (CQOLCF); this is a 35-item questionnaire designed to evaluate how providing care for someone with CF impacts on a carers' physical, emotional and social functioning and family [69].

Tools capturing patient-reported outcomes (excluding QoL)

Six questionnaires designed to evaluate self-reported levels of physical activity were identified [86, 101]. Two questionnaires capturing body image for use in people aged ≥14 years [116] and one tool measuring dietary intake in children aged between 7 months and 12 years (table S5) were also found [40]. One PROM has been used to evaluate the impact of CF on stigma, disclosure, public attitudes and negative self-image among adults with CF and their carers [96]. A separate PROM originally developed for use in people with asthma has been used to capture work productivity and activity in people with CF [92].

12 clinical scores calculated from outcome data reported by people with CF were identified; three were developed for use in CF pulmonary exacerbations [12, 106], three captured respiratory symptoms [22, 74,

TABLE 1 Characteristics of quality of life (QoL) tools developed for use in people with cystic fibrosis (CF)

Test or tool	Description	Constructs(s)	Target population	Administration	Recall period	Range of scores	Feasibility/ cultural validity
CFIQ	40 items; 5–6 min to complete	Activity limitation (physical, social, leisure), school/ work limitations, vulnerability/lack of control, emotional impact, treatment burden and future outlook	Children and adults with CF and their carers; interview templates for children aged 6–11 years, adolescents, adults >12 years and carers of children aged 0–18 years	Paper	36 items 7 day recall, the remainder "current status"	5- or 7-point verbal rating scale	Largely developed in Caucasian population; further validation required
CFQoL	QoL and symptom scoring tool; 52 items over nine domains; 15–20 min to complete	Two symptom scales (chest and emotional) and seven QoL domains: physical functioning (10), social functioning (4), treatment issues (3), future concerns (6), interpersonal relationships (10), body image (3), career issues (4)	Adults and adolescents	Paper	14 days	Each response 6-point Likert scale; total score 0–100	Time consuming
CFQoL scale (single item)	VAS: how has CF affected your QoL in the last 2 weeks? Couple of mins to complete	Single QoL question	Adults	Paper	14 days	0–10	Quick Validated in population who had not had pulmonary exacerbation for past 6 weeks
CFQ-R	QoL and symptom scoring tool; four scales: 1) CFQ-R >14 years (44 items over 9 domains); 2) CFQ-R child (8 domains, 35 items), interviewer-administered 6–13 years and self-report 12–13 years; 3) CFQ-R parent: 44 items, 10 min; 4) preschool 3–6 years, 28 items, 15 min to complete	Activity limitation (physical, social, leisure), school/ work limitations, vulnerability, lack of control, emotional impact, treatment burden and future outlook	CFQ-R >14 years; CFQ-R child: interviewer-administered 6–13 years and self-report 12–13 years CFQ-R parent: proxy report for children 6–13 years Separate are not proxy report for children 4–60 months	Paper/ electronic	14 days	4-point Likert scale; total score 0–100	Most widely used HRQoL questionnaire in CF; translated into 34 languages; EMA/FDA supports use in clinical trials

Continued

TABLE 1 Continued

Test or tool	Description	Constructs(s)	Target population	Administration	Recall period	Range of scores	Feasibility/ cultural validity
eCF-QUEST	Electronic, 3 domains, 4 items	Global measure (40 items), gastrointestinal (5 items) and general health (2 items)	Adults	Paper/ electronic	NR	NR	NR
DISABKIDS-CF	6 items, 2 min to complete	Impact and treatment dimensions	Exclusively for use in children and adolescents 8–17 years; self-report and proxy version (carer)	Paper	Each dimension 0–100	5-point Likert scale; scores 0–100% (higher score=higher QoL)	English and Spanish versions
FLZ-CF	Healthy and general patient population; 18 items over 8 domains, 9-item weighted scale, 5 min to complete	9 items: cough/ dyspnoea, abdominal, sleep, eating, therapy routine, adherence, understanding by others, being needed by others, disadvantage	Adolescents >15 years and adults	Paper	28 days	0–100 (high score=high satisfaction)	Screening test
CQOLCF	35 item carer QoL instrument, <10 minutes	Physical, emotional, family and social functioning	Carers of people with CF	Paper	NR	Each response 5-point Likert scale	NR

CFIQ: CF impact questionnaire; CFQoL: CF QoL; CFQ-R: revised CF questionnaire; eCF-QUEST: electronic version of the CFQoL evaluative self-administered test; FLZ-CF: Questions of Life Satisfaction; CQOLCF: caregiver QoL CF scale; VAS: visual analogue scale; HRQoL: health-related QoL; EMA: European Medicines Association; FDA: US Food and Drug Administration; NR: not reported.

85, 101, 115], three characterised pain [90, 133], two quantified abdominal symptoms [28, 31, 41] and one has been used to evaluate physical and psychological symptom burden [105, 106] (table 2). Of these, the CFRSD-CRISS (chronic respiratory infection symptom score) has been the most widely used in CF studies and is available in 38 languages; it evaluates eight respiratory symptoms and is validated for use in people with CF aged $\geq$12 year [44].

Clinical scores

The modified Schwachman scale (first described in 1964) was developed as a longitudinal clinical assessment tool. It includes activity levels, chest findings, cough, growth, nutrition, the character of stool and radiological changes; lung function is not included in this measure [26]. A test developed by Radine *et al.* [45] measuring nocturnal cough over two consecutive nights was found to be safe and feasible.

Three prognostic scoring tools were found. The most recent, in 2004, was a 5-year survival prediction tool [63], which was designed to guide eligibility for lung transplant; survival is predicted based on age, sex, forced expiratory volume in 1 s (FEV_1 % pred), weight for age z-score, pancreatic function, presence of diabetes, infection with *Staphylococcus aureus* or *Burkholderi cepacia* and the annual number of pulmonary exacerbations. The second, the modified Huang score (first described in 1997) [26], was developed for use as a prognostic and longitudinal assessment tool for those with terminal disease and captures clinical features (including lung function), radiological features and complications of disease. The oldest tool, first described in 1973, is the National Institutes of Health score (NIH) [26, 108], which was developed for people aged 5–30 years. It predicts the probability of death within 3 years based on lung function, chest radiograph (CXR) changes, and physiological and psychological features.

TABLE 2 Scoring tools incorporating patient-reported outcome

Test or tool	Description	Constructs(s)	Target population	Administration	Recall period	Range of scores	Feasibility/cultural validity
Pulmonary exacerbations							
AWESCORE	5 domains, each with 2 items	Respiratory (cough and sputum), physical (energy and exercise), nutritional (appetite and weight), psychological (mood and anxiety) and general health "wellness" and sleep	Adults	Paper	NR	0–100	NR
CFRSD/ CFRSD-CRISS	Symptom score: respiratory and emotional items (respiratory only in short version); developed for pulmonary exacerbation	8 respiratory symptoms, 4 emotional items and 4 other items (or short-version CFRSD-CRISS; 8 respiratory items: difficulty breathing, fever, tired, mucus, chills/sweats, cough, mucus, chest tightness, wheezing)	>12 years	Paper/ electronic	Daily	3–4 point Likert scale, total score 0–100	Available in 38 languages
Symptom score system	4 items; pulmonary exacerbation assessment	Cough, sputum volume and viscosity, breathlessness, fatigue	Adults	Paper	Daily	4-point Likert scale	NR
Respiratory							
ReS-CF	4-item questionnaire; <1 min to complete	Self-reported VAS for respiratory symptoms, cough, chest congestion and sputum	Adults	Paper	NR	Each VAS scored separately 0–10 (worst)	Screening tool: respiratory symptoms
SOBQ	Developed for PEX assessment; 17 items (13 respiratory and 4 CF-related impacts)	0–6 years and 7–11 years		NR	NR	Not applicable	
SOBQ	24 items; patients with COPD, CF and lung transplant recipients	Measures SOB while performing activities of daily living	Adults	Paper	NR	5-point Likert scale for each response; scores 0–120 (worst)	NR
Pain							
BPI	Severity and impact of pain on daily functions in people with chronic diseases; short: 5 min, long: 10 min to complete	7 domains: general activity, mood, walking ability, normal work (including housework), relationships with others, sleep, enjoyment in life	Adults	Paper	Daily	NR	Psychometrically and linguistically validated in 24 languages
DPAQ-CF	7 items	Frequency, duration, intensity, location and coping response to pain	Adolescents and adults	Paper/ electronic	Daily	5-point Likert scale for each response; total score 0–10	NR

Continued

TABLE 2 Continued

Test or tool	Description	Constructs(s)	Target population	Administration	Recall period	Range of scores	Feasibility/cultural validity
MPI	52 items, 3 domains, 12 subscales; 15–30 min to complete	Pain experiences, responses of others to the patient's communicated pain, the extent to which patients participate in activities of daily living	18–64 years and >65 years versions	Paper	NR	NR	Available in 6 languages
Abdominal							
CF-Abd score	28-item PROM for assessment of gastrointestinal involvement	Abdominal pain, appetite, bowel movement and symptom-related QoL	>6 years	Paper	14 days	NR	NR
Gastrointestinal symptom tracker	PROM for assessment of gastrointestinal and nutritional issues	4 domains; abdominal pain, stools, eating challenges and adherence	Adolescents and adults	Electronic (iPad)	10–14 days	0–100 (worst)	Easy to administer and complete
Symptoms involving multiple systems							
MSAS-CF	Physical and psychological symptom burden	QoL: respiratory (6 items), psychological burden (5 items) and gastrointestinal (4 items)	Adults	Paper	7 days	Each symptom 4- or 5-point Likert scale	Previously validated in people with cancer, heart disease, HIV and critical illness

AWESCORE: Alfred Wellness Score; CFRSD: cystic fibrosis (CF) respiratory symptom diary; CRISS: chronic respiratory infection symptom score; ReS-CF: respiratory symptoms in CF tool; SOBQ: Shortness of Breath Questionnaire; BPI: brief pain inventory; DPAQ-CF : Daily Pain Assessment Questionnaire in CF; MPI: multidimensional pain inventory; CF-Abd: CF abdominal; MSAS: Memorial Symptom Assessment Scale; NR: not reported; VAS: visual analogue scale; PEX: pulmonary exacerbation of CF; SOB: shortness of breath; PROM: patient-reported outcome measures; QoL: quality of life.

Lung function tests

Tests used to measure lung function for which measurement properties were described include spirometry, raised volume rapid thoracic compression (RVRTC), impulse oscillometry and lung clearance index (LCI). Characteristics of these tests are reported in table S5.

Spirometry has been the most frequently used lung function test in CF studies, and the measure of lung function most commonly reported has been FEV_1 [30]. This has been variously measured as the crude volume (in litres) or as the percentage predicted volume for age and height, or z-score for age, sex, height and ethnicity; within-individual changes in the FEV_1 have been reported as either the absolute change, or change relative to the baseline measure [54, 72].

Imaging scoring tools

Four CXR scoring systems used to quantify the degree of structural lung disease were identified. The oldest, the Chrispin–Norman score in 1974, is based on chest configuration and the presence or absence of different types of "shadows" [123]. First described in 1979, the Brasfield or Birmingham score (scored between 3 and 25) aims to capture radiographic evidence of air trapping, bronchial wall thickening, bronchiectasis, atelectasis and general severity [73, 122]. From 1993, the Wisconsin score (0–100) has been used to evaluate six attributes including hyperinflation, peribronchial thickening, bronchiectasis, opacities and atelectasis [54, 129]. The Brasfield scoring system has been reported to be easier to use and quicker to perform than the Wisconsin score [73]. The Northern score (introduced in 1994) is calculated based on the presence of linear, cystic or confluent opacities in each lung quadrant rated on a four-point Likert scale (normal to very severe) on a single film [124].

Three computed tomography (CT) scoring tools were identified: the Brody score (I and II), originally developed in 1999, and the CF-CT score (introduced in 2011), which is based on the Brody II score and

aims to improve standardisation of the latter [111]. These tools have been used in people aged >5 years. In 2014, the Perth Rotterdam Annotated Grid Morphometric Analysis method (PRAGMA-CT) score was developed for application in children and infants [102] and takes an experienced person ∼30 min to calculate per CT scan [111].

Scoring tools based on quantitative magnetic resonance imaging (MRI) are still in development [54].

Functional exercise performance

The most frequently studied field exercise test performed in people with CF is the 6-min walk test (6MWT) [36]. Characteristics of tests used to measure functional exercise performance are summarised in table 3.

CFTR function

Characteristics of tests used to directly (*e.g.* sweat chloride tests) or indirectly (*e.g.* nasal potential difference tests) measure CFTR function are summarised in supplementary table S5.

Sputum tests

Rheometry tests which capture the characteristics of sputum, and tests used to capture markers of inflammation in sputum, are summarised in table S5.

Measurement properties of tests, tools or instruments

QoL tools

The measurement properties of generic QoL tools based on their evaluation in CF populations are detailed in table S6, and the properties of CF-specific QoL tools are summarised in table 4.

The development of the CFQ-R involved people with CF, and it has been shown to be reliable, with sound content (including face) validity. The tool has been shown to have good internal consistency for all constructs examined, including parent proxy report of physical, eating and respiratory subscales (α=0.73–0.86), but not for treatment burden (r=0.44). The CFQ-R score correlates with FEV_1 and body mass index, and discriminates different degrees of disease severity [130], but a correlation with mortality has not been reported. Based on clinician judgement, a change of 0.8 units in the CFQ-R score was considered the MCID in the context of treatment for pulmonary exacerbations [134].

Patient-reported symptoms and function

The measurement properties of patient-reported symptoms and function are summarised in table 5.

Clinical scores

A validation study that evaluated nocturnal cough as an outcome found people with CF coughed more than healthy subjects ($p<0.001$); the reliability for repeated measurements was higher when cough epochs were scored (multiple coughs with <2 s between individual coughs) compared to discrete coughs (internal consistency coefficient (ICC) 0.75 *versus* 0.49, respectively) [45].

The interobserver reliability of the modified Schwachman score captured as Pearson's r coefficient was 0.71, 0.64 and 0.85 for the history, examination and growth domains, respectively [26]; the correlation was 0.92 with the NIH score and 0.67 with FEV_1.

The internal consistency of the modified Huang score was reported to be α=0.6 (except the domain relating to complications). The correlation of this score with FEV_1 % pred in moderate (score 35–60, r^2=0.3) and severe disease (<35 points, r^2=0.43) was greater than in asymptomatic or mild disease [26]. The NIH score was found to be significantly lower in the 5 years before death compared to CF controls who did not die ($p<0.001$); those with a score between 61 and 70 had a 25% chance of dying within 3 years. The internal consistency of this score was reported to be α=0.81 and the inter-rater (Pearson's r) score was 0.90; this was predominantly attributed to the robustness of the pulmonary domain on subscale analysis [26].

Lung function tests

Low FEV_1 was shown to correlate with death, with a relative risk of death within 2 years of 2.0 (95% CI 1.9–2.2) for each 10% reduction in FEV_1 below the predicted value after adjustment for age and sex [126]. Among people with the same FEV_1, the risk of death was more than double for females compared to males (RR 2.2 (95% CI 1.6–3.1)). FEV_1 was also shown to correlate with QoL; a 5% change in FEV_1 was associated with a change in CFQ-R score from 0.5 to 2.3 points [125].

TABLE 3 Functional measures of exercise capacity

Test or tool	Description	Construct(s)	Target population	Administration	Recall period	Range of scores	Feasibility/ cultural validity
iSTEP	Externally paced test; speed increases every 2 min Expired gas analysis	Expired gas analysis	Younger, fitter patients	Performance test	10 min	Variable	Portable, standardised and easy to administer field exercise test
MSWT	15-level modification of ISWT Office based walk/run test	Peak oxygen uptake	Children	Performance test; standard protocol	20 min	NA	10 m required, used in younger and fitter patients Excludes those with *i.v.* lines or those requiring oxygen support
$Power_{STS}$	1-min sit-to-stand power index	Quadriceps power	Moderate–severe CF	Performance test; standard protocol	1 min	N/A	Quick and easy to perform
Triple hop distance	Starting at one end of a tape, asked to hop three times consecutively on dominant leg, trying to cover as much distance as possible	Lower extremity power	Older children and adults	Performance test	NR	Distance recorded in cm	NR
Vertical jump test	90-cm^2 mat connected to a timer next to a wall; time off mat converted to a vertical jump (cm) using a controlled (90 degree) and uncontrolled knee angle	Power and posture	Older children and adults	Performance test	NR	Vertical distance recorded in cm	NR
3MST	Submaximal stress test (distance covered in m)	Externally paced test (metronome paced at 12 beats·min^{-1}) step up and down a 6-inch step for 3 min	Good choice in severe CLD	Performance test; standard protocol	3 min	NA	Requires the least amount of space; 6-inch step required
6MWT	Submaximal stress test (distance covered in m)	Standard protocol; distance walked within 6 min (enough O_2 to maintain saturations >90%)	Validated in 7–23 years Good choice in severe CLD	Performance test; standard protocol	6 min	NA	Most frequently studied exercise test in CF; easy but requires 30 m
30 s or 1-min-STS	Cardiorespiratory response during a 30-s or 1-min STS test (chair height 40 cm without armrest; full knee extension); as many repetitions as possible in 1 min	Exercise capacity	Moderate–severe CF	Performance test; standard protocol	1 min	Total number of full repetitions in 30 s or 1 min	Quick and easy to perform

iSTEP: incremental field step test; MSWT: modified shuttle walk test; $Power_{STS}$: 1-min sit-to-stand power index; 3MST: 3-min sit-to-stand test; 6MWT: 6-min walk test; STS: sit-to-stand test; ISWT: incremental shuttle walk test; N/A: not applicable; CF: cystic fibrosis; CLD: chronic lung disease.

The RVRTC test demonstrated good test–retest reliability with a coefficient of variation reported to be 2–6%; it differentiated people with CF from healthy controls, including among children aged <6 years [91]. Parameters were shown to improve in children aged between 4 months and 1 year, raising the possibility that lung damage may be reversible during this time [91]. However, RVRTC testing has not been appropriately standardised and consequently has not yet been recommended by authoritative bodies such as the European CF Society Clinical Trial Network as a primary outcome measure for use in CF studies [32].

Measurement of LCI has been found to be reliable, valid and responsive during treatment of pulmonary exacerbations and for monitoring disease progression [23, 37, 89, 110]. Measurements obtained by N_2 washout and by SF_6 were comparable (limits of agreement −2.5 to 1.2) [23]. These tests were found to

https://doi.org/10.1183/16000617.0354-2020

TABLE 4 Measurement properties of quality of life (QoL) tools specific for people with cystic fibrosis (CF)

Test or tool	Content validity	Convergent validity	Discriminant validity	Concurrent	Predictive	Intra- or inter-rater and test-retest	Internal consistency	Measurement error	Responsiveness	Comments/ MCID
CFIQ	Demonstrated; people >6 years with CF and carers used in construction	NR	NR	NR	NR	NR	NR	NR	NR	Requires further validation; content validity established
CFQoL	Easy to understand/ complete; people with CF involved in construction	Correlation of emotional scores with SF-36 r=0.64; p<0.001	Chest and emotional scores distinguished between severity of chronic lung disease (FEV_1 % pred >70, 40–70 or <40)	Chest score correlation with FEV_1 not tested	NR	Test–retest r_s=0.74–0.94 (p<0.01) Robust after 7–10 days; 0.9 for emotional scores and 0.93 for respiratory	Cronbach's α=0.3	NR	Chest symptom scores increased during pulmonary exacerbation treatment Chest and emotional score responsive over a 2-week application period in hospital (47–70.3, p=0.006) *versus* home groups (49.7–68.8, p=0.03)	NR

Continued

https://doi.org/10.1183/16000617.0354-2020

TABLE 4 Continued

Test or tool	Content validity	Convergent validity	Discriminant validity	Concurrent	Predictive	Intra- or inter-rater and test-retest	Internal consistency	Measurement error	Responsiveness	Comments/ MCID
CFQ-R	Patients involved in testing clarity of items Preschool version: children able to understand and answer questions	Correlation between CFQ-R and SF-36 on physical (r=0.81, p<0.01), health perceptions/ general health (r=0.79, p<0.01), vitality (r=0.84, p<0.01), role/ role-physical (r=0.73, p<0.01), emotional functioning/ mental health (r=0.74, p<0.01) and social (r=0.57, p<0.01) domains Strong convergence between child and parent proxy reports, although children generally reported better HRQoL than parents	CFQ-R: no significant difference between age groups (6–11 years, 12–13 years *versus* ≥14 years) for all domains except treatment between 6- to 11-year-olds and ≥14-year-olds Significant association between CFTR genotype and CFQ-R scores (K=9.34, p<0.01) Strong parent–child agreement found for scales measuring respiratory symptoms, but children reported more fatigue and difficulty running/walking	Respiratory score established using FEV_1; correlation with FEV_1 % pred r=0.42, p-value NR; correlation with number of intravenous antibiotic courses r=−0.27, p-value NR	NR	Acceptable	Cronbach's α=0.6–0.76 with the exception of treatment burden (α=0.44) Parent proxy report for CFQ-R physical, eating and respiratory subscales α=0.73–0.86	NR	Based on clinician judgement, a moderate change=0.5 units and an important change=0.8 units pre- & post-exacerbation	NR

Continued

TABLE 4 Continued

Test or tool	Content validity	Convergent validity	Discriminant validity	Concurrent	Predictive	Intra- or inter-rater and test-retest	Internal consistency	Measurement error	Responsiveness	Comments/ MCID
CF-QUEST	NR	NR	NR	r_s was 0.951 for the total CF-QUEST score, 0.929 for gastrointestinal module and 0.941 for GHQ module for paper/ electronic versions	NR	NR	NR	NR	NR	Excellent correlation and agreement of electronic version with its validated paper counterpart Patient preference tended towards electronic version
DISABKIDS-CFM	NR	Convergent validity with KINDL-R established; r=0.6	NR	NR	NR	NR	Cronbach's α=0.55 (p=0.011) for the impact dimension and 0.480 (p=0.02) for the treatment dimension	NR	NR	NR

Continued

https://doi.org/10.1183/16000617.0354-2020

TABLE 4 Continued

Test or tool	Content validity	Convergent validity	Discriminant validity	Concurrent	Predictive	Intra- or inter-rater and test-retest	Internal consistency	Measurement error	Responsiveness	Comments/ MCID
FLZ-CF	NR	Pearson's correlation r=0.75 with the generic satisfaction with health scale of the FLZ-CF, r=0.3 with FEV_1 % pred and r=−0.26 with daily time for home therapy Leisure time/ hobbies, physical condition, ability to relax, energy for life and satisfaction with health r_s >0.5 with positive mood and ability to relax and SF-36 physical functioning, general health, vitality, social function and mental health	The scale discriminated significantly between patients with and without need for assistance with daily life and between patients with and without a partner	Physical condition/fitness and FEV_1 % pred r_s=0.66	NR	NR	Cronbach's α=0.82–0.89	NR	NR	Reliable and valid Targets general healthy and general patient population Short enough to be used as a screening instrument

Continued

https://doi.org/10.1183/16000617.0354-2020

https://doi.org/10.1183/16000617.0354-2020

TABLE 4 Continued

Test or tool	Content validity	Convergent validity	Discriminant validity	Concurrent	Predictive	Intra- or inter-rater and test-retest	Internal consistency	Measurement error	Responsiveness	Comments/ MCID
Single item CFQoL questionnaire	NR	Most of the CFQoL variables were moderately correlated (r=0.38–0.61, p<0.001) with the single item scale weakly correlated with body image (r=0.25), p<0.01 Higher scores correlated negatively with frequency of hospital admissions in the previous year (r=−0.39, p<0.001)	Ability to distinguish adult CF patients with lower compared to higher CFQoL scores	Single-item scale correlation with FEV_1 r=0.21	NR	ICC 0.78 (95% CI 0.59–0.88)	NR	NR	NR	Acceptable, valid and repeatable measurement tool that can be easily used
CQOLCF	NR	Correlation with mental health r=0.634, emotional distress r= −0.687 and physical health r=0.049	NR	NR	NR	NR	Cronbach's α=0.909	NR	NR	Appears to be valid, reliable and internally consistent scale

MCID: minimal clinically important difference; CFIQ: CF Impact Questionnaire; CFQoL: CF QoL Questionnaire; CFQ-R: revised CF Questionnaire; CF-QUEST: electronic version of the CFQoL evaluative self-administered test; FLZ-CF: Questions of Life Satisfaction; CQOLCF: Caregiver QoL CF scale; NR: not reported; SF-36: Short-Form-36 Item Questionnaire; HRQoL: health-related QoL; CFTR: CF transmembrane regulator; FEV_1: forced expiratory volume in 1 s; GHQ: General Health Questionnaire; KINDL-R: Child QoL Questionnaire-Revised; ICC: internal consistency coefficient; rs: Spearman's correlation coefficient.

https://doi.org/10.1183/16000617.0354-2020

TABLE 5 Measurement properties of scoring tools based on outcomes reported by people with cystic fibrosis (CF)

Test or tool	Content validity	Convergent validity	Discriminant validity	Concurrent	Predictive	Intra or inter-rater and test–retest	Internal consistency	Measurement error	Responsiveness	Comments/ MCID
Pulmonary exacerbations										
AWESCORE	NR	NR	NR	Correlation of total AWESCORE and CFQ-R scores: r=0.632 (p=0.003)	NR	Pearson's correlation coefficient 0.854, p<0.0005	NR	NR	For exacerbation, score 47.5 (SD 11.2) at start of treatment and 21.6 (SD 15.6) at end of treatment (100=highest symptom severity)	11 points Mean change of −16.5 (95% CI −13.2 to −19.7 for exacerbation reported) No MCID for emotional score
CFRSD/ CFRSD-CRISS	Involved people with CF in testing clarity of items	Step-rate significantly higher in those who did NOT experience difficulty breathing, cough, tightness or feeling tired (respiratory items) or feeling worried, cranky or frustrated (emotional items)	Respiratory scores distinguished between moderate/ severe and mild/severe disease; emotional scores distinguished between mild/ severe disease	Respiratory and emotional score established using daily step count (not FEV_1)	NR	ICC 0.79 for respiratory scale using a 1-day interval	Cronbach's α for CFRSD-CRISS was 0.77	Test–retest reliability after 7–10 days; 0.9 for the emotional and 0.93 for the chest score	Total score been demonstrated to improve over 2 weeks' *i.v.* treatment	No MCID suggested on the basis of statistical analysis, but MCID >1 after 2 weeks of *i.v.* ABX suggested based on experience with COPD patients

Continued

TABLE 5 Continued

Test or tool	Content validity	Convergent validity	Discriminant validity	Concurrent	Predictive	Intra or inter-rater and test–retest	Internal consistency	Measurement error	Responsiveness	Comments/ MCID
Symptom score	Patients not involved in construction	All 4 items correlated with each other r>0.38; p<0.001	NR	Total score correlation with FEV_1: r=−0.41, (p<0.0001) and respiratory score on CFQ-R: r=−0.62 (p<0.001) and CFQ-R: r=−0.47 (p<0.001)	NR	NR	NR	NR	NR	
Respiratory										
Borg Dyspnoea Scale	NR	NR	NR	NR	NR	ICC=0.933	NR	NR	Mean change in score −3.1 with mean effect size 1.3 from baseline to 4 weeks	Appears to be valid, reliable and responsive in CF For those reporting improvement, scores changed −2.9 overall, −3.5 for cough, −3.5 for congestion and −3.1 for sputum domains
ReS-CF	NR	NR	NR	Correlation between ReS-CF and CFQ-R; r_s=−0.5 (p<0.001)	NR	ICCs for 4 scores >0.7	NR	NR	NR	

Continued

https://doi.org/10.1183/16000617.0354-2020

TABLE 5 Continued

Test or tool	Content validity	Convergent validity	Discriminant validity	Concurrent	Predictive	Intra or inter-rater and test–retest	Internal consistency	Measurement error	Responsiveness	Comments/ MCID
SOBQ	NR	SOBQ scores correlated negatively with physiological measures of disease severity (FVC % pred: r=−0.36, p<0.05 and FEV_1 % pred: r=−0.5, p<0.01) Scores correlated positively with Borg scale ratings of perceived breathlessness after 6MWT and QWB (r= −0.41, p<0.01)	NR	NR	NR	NR	α=0.96	NR	NR	MCID: 5 unit change

Continued

https://doi.org/10.1183/16000617.0354-2020

TABLE 5 Continued

Test or tool	Content validity	Convergent validity	Discriminant validity	Concurrent	Predictive	Intra or inter-rater and test–retest	Internal consistency	Measurement error	Responsiveness	Comments/ MCID
Pain										
BPI	NR	Correlation of BPI pain interference and airway clearance therapy (p=0.002), coughing and breathing (p<0.012), pain prevalence and CFQoL physical function (p=0.01), CFQoL treatment (p=0.03), CFQoL work/ school (p=0.02), CFQoL social (p=0.013) and CFQoL emotional scale (p=0.017) Pain intensity also correlated with CFQoL physical function, CFQoL treatment and CFQoL school/ work (p⩽0.01)	NR	Correlation of BPI pain prevalence and sleep quality (p=0.045), sleep disturbance (p<0.001), daytime dysfunction (p=0.001) and sleep interference and global BPI score rho-0.56, p<0.0001 OR 1.27 (p=0.012) of impaired sleep quality in those with pain	BPI pain severity correlated with risk of exacerbations (OR 1.65, p=0.04) for exacerbations with higher pain intensity and OR of 2.28 (p=0.008) of death with higher pain intensity	NR	NR	NR	NR	

Continued

https://doi.org/10.1183/16000617.0354-2020

https://doi.org/10.1183/16000617.0354-2020

TABLE 5 Continued

Test or tool	Content validity	Convergent validity	Discriminant validity	Concurrent	Predictive	Intra or inter-rater and test–retest	Internal consistency	Measurement error	Responsiveness	Comments/ MCID
DPAQ-CF	NR	CFQ-R social function (r=0.269, p<0.01), CFQ-R treatment burden (r=0.269, p<0.01), CFQ-R respiratory symptoms (r=0.241, p<0.05), HADS-depression scale (r=0.29, p<0.01) and HADS-anxiety (r=0.29, p<0.01) DPAQ-CF pain intensity correlated with CFQ- treatment burden and respiratory symptoms (p<0.01) DPAQ-CF pain duration correlated with CFQ-R treatment burden and respiratory symptoms (p<0.01)	NR	Correlation of pain and ppFEV$_1$ R=0.239, P<0.05	NR	NR	NR	NR	NR	

Continued

https://doi.org/10.1183/16000617.0354-2020

TABLE 5 Continued

Test or tool	Content validity	Convergent validity	Discriminant validity	Concurrent	Predictive	Intra or inter-rater and test–retest	Internal consistency	Measurement error	Responsiveness	Comments/ MCID
MPI	NR	Correlation of BPI pain severity and Shwachman scale history scale; r=0.24 (p=0.04) and BPI pain interference and total Shwachman score r=0.2 (p=0.09)	NR	NR	NR	NR	NR	NR	NR	
Abdominal										
CF-Abd score	NR	NR	Differentiated patients with CF and healthy controls with large effect size (17.3+1.1 *versus* 8.0 +0.7 points; p<0.001; Cohen's d=0.85)	NR	NR	ICC 0.932 (95% CI 0.874–0.963)	Cronbach's α=0.7–0.92	NR	NR	
Gastrointestinal symptom tracker	NR	Nutritional status is related to more stable lung function and fewer exacerbations	NR	NR	NR	Reliability established based on test–retest and internal consistency (unspecified)	NR	NR	NR	

Continued

https://doi.org/10.1183/16000617.0354-2020

TABLE 5 Continued

Test or tool	Content validity	Convergent validity	Discriminant validity	Concurrent	Predictive	Intra or inter-rater and test–retest	Internal consistency	Measurement error	Responsiveness	Comments/ MCID
Symptom and impact score										
MSAS-CF	Developed in accordance with COSMIN recommendations; patients not consulted	MSAS-Resp: correlation with CFQ-R (r=−0.60, p<0.05) and CFQoL (r=−0.7, p<0.05) MSAS-Psych: emotional scale good correlation with CFQ-R (r=−0.69, p<0.05) Poor correlation with MSAS-GI: strongest with weight (CFQ-R r=−0.49, p<0.05) Subscales moderately correlated with symptoms on CFQ-R and CFQoL	Respiratory, gastrointestinal and psychiatric scores were higher in patients with low FEV_1 <40% pred (p<0.05)	Correlation with CFQ-R respiratory score (r=−0.6) and CFQoL chest score (r=−0.7, p<0.05) and CFQ-R emotional functioning score (r=−0.69, p<0.05). Weak correlation with CFQ-R digestive score (r=−0.19, p<0.05)	NR	NR	α 0.74–0.86 High in all domains; MSAS-Physical α 0.92, MSAS-Psych α 0.95, MSAS-Global α 0.82	NR	NR	General tool; not specific for exacerbations; originally developed for an oncology population

MCID: minimal clinically important difference; AWESCORE: Alfred Wellness Score; CFRSD: CF respiratory symptom diary; CRISS: chronic respiratory infection symptom score; ReS-CF: respiratory symptoms in CF tool; SOBQ: Shortness of Breath Questionnaire; BPI: brief pain inventory; DPAQ-CF: Daily Pain Assessment Questionnaire in CF; MPI: multiple pain inventory; CF-Abd: CF abdominal; MSAS: Memorial Symptom Assessment Scale; NR: not reported; FEV_1: forced expiratory volume in 1 s; FVC: forced vital capacity; ICC: internal consistency coefficient; ABX: antibiotics ;6MWT: 6-min walk test; QWB: Quality of Well-Being Questionnaire; HADS: Hospital Anxiety and Depression Scale; Resp: respiratory; GI: gastrointestinal; CFQ-R: CF Questionnaire-revised; COSMIN: Consensus-based Standards for the selection of health Measurement INstruments initiative.

discriminate between people with CF and healthy controls, as well as those at different disease stages based on age, infection and structural abnormalities identified on high-resolution CT imaging or MRI [110]. A correlation with clinical outcomes has not been established.

Imaging scoring tools

While the Brasfield and Wisconsin CXR scores performed similarly and both have been found to be reproducible (intra-observer agreement r=0.86–0.99 and 0.78–0.96, respectively) and reliable (inter-rater agreement 0.76–0.90 and 0.74–0.97, respectively), they appear to be insensitive to early disease [122]. The correlation between these scores was reported as r=0.86, p<0.0001. Both scores correlated with lung function (FEV_1 and forced vital capacity, all p<0.001) [73]. The correlation of scores with FEV_1 was highest for the Northern score (r=−0.82) compared to the Brasfield (r=0.81) or Crispin–Norman scoring methods (r=−0.83) [124].

CT scoring tools have been found to have higher sensitivity for detecting lung disease progression than FEV_1 % pred. The test–retest reliability based on the intraclass correlation coefficient of the PRAGMA-CF score was shown to be >0.9 for percentage disease, 0.85 for percentage bronchiectasis and 0.96 for percentage air trapping; the intra-observer reliability was >0.90 for bronchiectasis, air trapping and percentage disease [50].

The test–retest reliability of a semi-quantitative MRI score was r^2=0.76 (p=0.0047) [54] and correlation with FEV_1 was r=0.81 (p=0.0023) [54].

Functional exercise performance

A summary of the measurement properties of tests used to capture functional exercise capacity is provided in table S6.

Many tests capturing functional exercise performance were compared to cardiopulmonary exercise testing (CPET), which has historically been viewed as the gold standard for assessing exercise capacity according to Von Berg *et al.* [57]. Rand *et al.* [47] found that the incremental field step test had acceptable concurrent validity compared to CPET in children for measuring peak oxygen uptake, minute ventilation, heart rate, change in oxygen saturation and CO_2 ventilation and perceived exertion [47].

Submaximal exercise tests included the 6MWT, 3-min step test (3MST), modified shuttle walk test (MSWT) and 30-s or 1-min sit-to-stand test [57]. Good concurrent validity of the MSWT with maximum oxygen capacity on CPET has been reported; however, results for concurrent validity were inconsistent for the 6MWT and 3MST. The ability of the 6MWT to predict pre-transplant survival was variable [36]. A reduction of 50 m or more in the modified shuttle test was associated with a hazard ratio of death or lung transplant within 1 year in adults with CF of 1.91 (95% CI 1.09–3.35, p<0.024) [20]. Convergent validity of 3MST and MSWT with FEV_1 (r=0.61, p=0.002) was found [36], but this was variable for the 6MWT.

CFTR function tests

Intestinal current measurement and nasal potential difference (NPD) tests, which directly measure CFTR function, were strongly correlated and have been found to distinguish people with CF from healthy controls (k=0.83 *versus* k=0.33, respectively, p<0.001) [68]. Changes in NPD have been reported over 14 days in trials of the CFTR function-modifying drug ivacaftor. Some evidence for the reliability of intestinal organoid volume has been found, but evidence to support its validity has not [68]. Some evidence for the validity and reliability of indirect measures of gastrointestinal CFTR function such as intestinal pH, faecal calprotectin and faecal elastase-1 has been found; however, these data are not described in detail in the review included in our study (table S6).

Sputum tests

Tests characterising sputum rheology, including viscoelasticity and solid content properties, demonstrated poor to fair test–retest reliability with ICCs ranging from 0.22 to 0.42 (with wide confidence intervals) [46, 103]. Reproducibility of biomarkers in the sputum such as total cell count, neutrophils, tumour necrosis factor-α, interleukin-8 and neutrophil elastase was demonstrated in one study [128] as follows: ICC=0.76, 0.82, 0.93, 0.82 and 0.74, respectively; however, there was marked between-patient variability [103, 128].

Measurement error

The systematic and random error of a patient's score not attributable to true changes in the construct that was measured was poorly reported across all studies (table 4, table 5 and table S6).

Discussion

While the measurement properties of PROMs evaluating HRQoL in CF studies have been previously evaluated [135], this is the first effort to systematically review evidence of the measurement properties of all tests and tools used in CF studies. A diverse range of tests and tools were identified which vary with respect to their reliability, responsiveness and validity. There was inconsistency in the use of tests and tools to measure the same or similar outcomes across studies. This highlights the need to establish consensus over which outcomes should be measured in CF studies and how they should be measured; this has been recommended by the COSMIN initiative group [136]. Compared to older tools, many recently developed tools incorporate self-reported outcomes by patients (*e.g.* CFRSD-CRISS, CFQ-R, CF Impact Questionnaire and CF Quality of Life (CFQoL)) and have involved people with CF in their development, consistent with the recommendation made by the US FDA in 2017 [137].

Evidence to support the reliability of spirometry testing was found; this has also been substantiated in other populations, such as in people with other chronic lung disease [138]. Poor FEV_1 is strongly correlated with death, progression to lung transplant (most transplant recipients have a FEV_1 <30% pred) [139] and reduced QoL [110] in people with CF and is also associated with a greater risk of hospitalisation, pulmonary exacerbations and colonisation with *Pseudomonas aeruginosa* [140]. Compared to crude or percentage predicted FEV_1 values, z-scores have been proposed as a less biased and more accurate measure for defining meaningful changes in lung function since they take into account sex and ethnicity in addition to age and height; this approach has been endorsed by the Global Lung Initiative since 2013 [141]. This, however, has not yet been universally adopted as the preferred measure for capturing lung function in CF studies. Consensus regarding the MCID for FEV_1 was not identified in this review, but MCIDs have been proposed. In the TRAFFIC and TRANSPORT phase 3 trials, which evaluated lumacaftor–ivacaftor *versus* placebo for people homozygous for the Phe508del CFTR mutation [142, 143], a mean relative difference of 3.3% (2.3–4.4, p<0.0001) and 2.8% (1.7–3.8, p<0.0001) was found in those with baseline FEV_1 ⩾40% pred and baseline FEV_1 <40%, respectively. It was proposed that this represents a clinically significant improvement since the annual rate of decline of FEV_1 % pred has been estimated to be 1.92% per annum for people with CF aged 1824 years (n=2793) and 1.45% for those aged >25 years [144].

While FEV_1 has been shown to be reproducible and repeatable in children aged ≥6 years and adults, its variability is affected by the person's age and the severity of their underlying lung disease [110]. In the early stages of CF disease, FEV_1 often remains within the normal range, while in severe lung disease FEV_1 is significantly compromised and unlikely to demonstrate variability [89]. LCI testing represents an alternative test for children aged <6 years who are incapable of performing spirometry. Since measurement is dependent on body size, the relative rather than the absolute change is considered more appropriate, at least before 6 years of age [145]. LCI has been shown to correlate strongly with structural abnormalities detected on high-resolution CT and abnormal preschool LCI is associated with spirometry deficits performed within 3 years from baseline in school-age children [146]. However, further standardisation and evaluation of the relationship of LCI with morbidity and mortality is warranted.

Evidence of the reliability, responsiveness and validity of two commonly used QoL tools, the CFQ-R and the CFQoL, as well as the CFRSD-CRISS symptom scoring tool has been reported previously and has been substantiated by this review. The content validity (including face validity) of these tools is strengthened by involving people with CF in their development [26]. The CFQ-R has been shown to correlate moderately with FEV_1 % pred [97].

There have been significant advances in treatment and long-term health outcomes for people with CF in recent decades, which raises a concern about the current content validity of some of the outcome scoring tools developed in the second half of the 20th century, many of which did not involve people with CF in their development [43]. Many of these have not undergone sufficient validation and consequently have not been recommended for use in clinical practice or in research.

The use of imaging modalities and scoring tools in CF has evolved with time; however, considerable variability exists between treatment centres, for example whether to use CXR or CT for longitudinal disease monitoring. An important limitation of CXR imaging is its poor sensitivity for detecting structural lung changes in early disease and progression in those with established disease [73]. This modality, however, is still used for monitoring disease progression in some treatment centres, and it has an established role in identifying pathology in the context of an acute clinical deterioration, such as consolidation or pneumothorax. Extensive collaboration has occurred within the CF community to standardise CT and MRI radiological scores, especially in young children, to enable quantification of the degree of structural lung damage. While CT is currently the most sensitive method for detecting structural

airways disease [147], MRI shows promise because it delivers non-ionising radiation and allows assessment of functional aspects of the lung such as perfusion, pulmonary haemodynamics and ventilation [111]. It may be possible to automate imaging scoring algorithms in the future, which may improve the efficiency and reliability of results. However, further assessment of the validity and reproducibility of MRI scoring tools is required, and the extent to which imaging scores predict clinical outcomes of significance requires further elucidation, including in children [111].

The strengths of this review include the use of a systematic approach to identify studies by two independent reviewers. There were four major limitations. First, tests and tools used in practice in people with CF that have been validated in non-CF populations (*e.g.* generic scores capturing abdominal symptoms) were considered beyond the scope of this review. Secondly, details about the systematic error (bias) and random error (noise) for each of the tests and tools (*i.e.* variation beyond that attributable to the outcome of interest) have been poorly described in the literature. Measurement error is an important source of bias; this information is necessary to appraise the quality of tests and tools and should be an important factor influencing selection. Thirdly, medical devices used to capture outcomes were beyond the scope of this review (such as weighing scales or stadiometers used to capture anthropometric outcomes). Finally, given the large scope of this review, an exhaustive critique of the measurement properties of individual tests and tools was not feasible.

Conclusions

This systematic review highlights the diversity of tests and tools which have been used for outcome measurement in CF studies and their variable characteristics and properties. While there have been concerted efforts within the CF research community to improve and standardise these tests and tools, further work is needed, particularly to optimise tools for outcome measurement in young children and those with mild or severe disease. A consensus set of tests and tools for measurement in CF studies is needed; this should be developed together with people with CF and other relevant stakeholders. This would likely improve the consistency of reporting and measurement of similar outcomes, allowing comparison and synthesis of evidence across studies and improving the value of the research that is conducted.

Provenance: Submitted article, peer reviewed

Acknowledgements: This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors. C. McLeod is supported by an NHMRC post-graduate scholarship (GNT1150996), and top-up grants from the Wesfarmers Centre and the Perth Children's Hospital Foundation (9772 and 9757). T.L. Snelling is supported by a Career Development Fellowship from the National Health and Medical Research Council (GNT1111657). C.C. Blyth is also supported by the National Health and Medical Research Council (1111596/1173163).

Author contributions: C. McLeod was responsible for the study conceptualisation and overall methodology. C. McLeod and J. Wood were responsible for data curation. T.L. Snelling, C. McLeod, S. Smith, S. Webb, C.C. Blyth, A. Tong, J. Wood, A. Schultz and A.R. Smyth elaborated the study protocol. C. McLeod drafted the manuscript. All authors were involved in the interpretation of data and revision of the manuscript. All authors approved the final manuscript.

Conflict of interest: C. McLeod has nothing to disclose. J. Wood has nothing to disclose. A. Tong has nothing to disclose. A. Schultz reports personal fees from Vertex Pharmaceuticals, outside the submitted work. R. Norman has nothing to disclose. S. Smith has nothing to disclose. C.C. Blyth has nothing to disclose. S. Webb has nothing to disclose. A.R Smyth reports grants from Vertex, speaker honoraria and expenses from TEVA and Novartis, and personal fees from Vertex, outside the submitted work. In addition, A.R. Smyth has a patent "Alkyl quinolones as biomarkers of *Pseudomonas aeruginosa* infection and uses thereof" issued. T.L. Snelling has nothing to disclose.

Support statement: Funding was received from Perth Children's Hospital Foundation (grants 9772 and 9757), National Health and Medical Research Council (GNT1111657) and Wesfarmers Centre for Vaccines and Infectious Diseases. Funding information for this article has been deposited with the Crossref Funder Registry.

References

1 Curtis L, Hernandez A, Weinfurt K. Choosing and Specifying Outcomes and Endpoints. Section 1, Introduction. *In:* NIH Collaboratory Living Textbook of Pragmatic Clinical Trials. Bethesda, MD, NIH, 2020.

2 Stevens S. On the theory of scales of measurement. *Science* 1946; 103: 677–680.

3 Eisenstein E, Anstrom K, Zozus M, *et al.* Choosing and Specifying Outcomes and Endpoints. Section 5, Inpatient endpoints in pragmatic clinical trials. *In:* NIH Collaboratory Living Textbook of Pragmatic Clinical Trials. Bethesda, MD, NIH, 2020

4 Mokkink LB, Terwee CB, Patrick DL, *et al.* The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010; 63: 737–745.

5 Fitzpatrick R DC, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess* 1998; 2: i–iv, 1–74.

6 World Health Organization. WHOQOL: Measuring Quality of Life 2020. www.who.int/healthinfo/survey/whoqol-qualityoflife/en/ Date last updated: 12 March 2021; date last accessed: 8 August 2020.

7 Anagnostopoulou P, Yammine S, Schmidt A, *et al.* False normal Lung Clearance Index in infants with cystic fibrosis due to software algorithms. *Pediatr Pulmonol* 2015; 50: 970–977.

8 Archangelidi O, Abbott J, Bryon M, *et al.* Quality of life in patients with CF using three online research questionnaires: a feasibility study. *Pediatr Pulmonol* 2019; 54: Suppl. 2, 419.

9 Bayfield KJ, McGovern M, Simpson AJ, *et al.* Reliability of measurements using Innocor breath by breath analyser during a maximal exercise test in cystic fibrosis patients. *Thorax* 2014; 69: A167–A168.

10 Brody AS, Kosorok MR, Li Z, *et al.* Reproducibility of a scoring system for computed tomography scanning in cystic fibrosis. *J Thorac Imaging* 2006; 21: 14–21.

11 Bullinger M, Petersen C, Muehlan H, *et al.* A European approach to measuring quality of life in children with chronic conditions: the DISABKIDS experience. *Psychother Psychosom* 2013; 82: 14.

12 Button BM, Wilson LM, Finlayson F, *et al.* Alfred wellness score: effects of pulmonary exacerbations on adult CF patient-reported outcomes-stability and validity. *Pediatr Pulmonol* 2019; 54: Suppl. 2, 355.

13 Cebrian M, Ferrero J, Ansotegui E, *et al.* Comparative study of three quality of life instruments in adolescents and adults with cystic fibrosis. *Eur Respir J* 2012; 40: Suppl. 56, P3371.

14 Chaplin JE, Koopman HM, Schmidt S, *et al.* DISABKIDS Smiley questionnaire: the TAKE 6 assisted health-related quality of life measure for 4 to 7-year-olds. *Clin Psychol Psychother* 2008; 15: 173–180.

15 Cheney JL, Saddington H, Shevill E, *et al.* Development of the pictorial preschool cystic fibrosis questionnaire-revised: child and parent report. *Pediatr Pulmonol* 2014; 49: Suppl. 38, 435–436.

16 Chetta A, Pisi G, Zanini A, *et al.* Six-minute walking test in cystic fibrosis adults with mild to moderate lung disease: comparison to healthy subjects. *Respir Med* 2001; 95: 986–991.

17 Connett G, Staab D, Hubert D, *et al.* The development and use of a pancreatic exocrine insufficiency questionnaire to assess symptoms and their impacts in cystic fibrosis. *J Cyst Fibros* 2019; 18: Suppl. 1, S136.

18 Daftary A, Acton J, Heubi J, *et al.* Fecal elastase-1: utility in pancreatic function in cystic fibrosis. *J Cyst Fibros* 2006; 5: 71–76.

19 De Boeck K, Kent L, Davies J, *et al.* CFTR biomarkers: time for promotion to surrogate end-point? *Eur Respir J* 2013; 41: 203–216.

20 Doeleman W, Burghard M, Twisk J, *et al.* The Modified Shuttle Test to predict survival in cystic fibrosis. *J Cyst Fibros* 2019; 18: Suppl. 1, S4–S5.

21 Doeleman WR, Hulzebos E. The minimal clinically important difference for survival in patients with cystic fibrosis. *J Cyst Fibros* 2018; 17: Suppl. 3, S98–SS9.

22 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.

23 Gibney KA, Stanojevic S, Salazar J, *et al.* Validation of a nitrogen (N2) based multiple breath washout technology in healthy children and children with cystic fibrosis. *Pediatr Pulmonol* 2012; 47: 308.

24 Gonska T, Ip W, Avolio J, *et al.* Validation of sweat gland potential difference measurements as a practical *in vivo* method of assessing CFTR function. *Pediatr Pulmonol* 2011; 46: 287.

25 Goss CH, Dellon EP, Lymp JF, *et al.* Advancing patient reported outcomes in children with cystic fibrosis. *Pediatr Pulmonol* 2011; 46: 294–295.

26 Hafen GM, Ranganathan SC, Robertson CF, *et al.* Clinical scoring systems in cystic fibrosis. *Pediatr Pulmonol* 2006; 41: 602–617.

27 Jardine J, Glinianaia SV, McConachie H, *et al.* Self-reported quality of life of young children with conditions from early infancy: a systematic review. *Pediatrics* 2014; 134: e1129–e1148.

28 Jaudszus A, Zeman E, Michl R, *et al.* Comparing abdominal symptoms in cystic fibrosis patients and healthy controls with a novel multimodal questionnaire (CF-Abd Score). *J Cyst Fibros* 2018; 17: Suppl. 3, S6.

29 Lowman JD, Moore K, Peeples A, *et al.* Reliability of musculoskeletal outcome measures. *Pediatr Pulmonol* 2010; 33: 387.

30 Madan Kumar H, Prestridge AL. To believe or not to believe first pulmonary function test in cystic fibrosis patients. *Pediatr Pulmonol* 2011; 46: 341–342.

31 Mainz JG, Tabori H, Lorenz M, *et al.* Validity and reliability of a novel multimodal questionnaire for the assessment of abdominal symptoms in people with cystic fibrosis (CFAbd-Score). *J Cyst Fibros* 2019; 18: Suppl. 1, S33.

32 Matouk E, Ghezzo RH, Gruber J, *et al.* Internal consistency reliability and predictive validity of a modified N. Huang clinical scoring system in adult cystic fibrosis patients. *Eur Respir J* 1997; 10: 2004–2013.
33 Matouk E, Ghezzo RH, Gruber J, *et al.* Construct and longitudinal validity of a modified Huang clinical scoring system in adult cystic fibrosis patients. *Eur Respir J* 1999; 13: 552–559.
34 Mayersohn GS, Ramos A, Kim R, *et al.* Parent and child differences in psychosocial symptom reporting. *Pediatr Pulmonol* 2016; 51: Suppl. 45, 469.
35 Modi AC, Quittner AL. Validation of a disease-specific measure of health-related quality of life for children with cystic fibrosis. *J Pediatr Psychol* 2003; 28: 535–545.
36 Moran FC, Rodda J. Systematic review of field exercise testing in cystic fibrosis – reliability and validity. *Pediatr Pulmonol* 2011; 46: 354.
37 Nyilas S, Bigler A, Yammine S, *et al.* Alternate gas washout indices: assessment of ventilation inhomogeneity in mild to moderate pediatric cystic fibrosis lung disease. *Pediatr Pulmonol* 2018; 53: 1485–1491.
38 Palermo TM, Long AC, Lewandowski AS, *et al.* Evidence-based assessment of health-related quality of life and functional impairment in pediatric psychology. *J Pediatr Psychol* 2008; 33: 983–996.
39 Parekh MH, Ipatova A, Dashputre AA, *et al.* Quality of life in adult patients with cystic fibrosis – a systematic review. *Value Health* 2015; 18: A288.
40 Powers SW, Patton SR, Henry R, *et al.* A tool to individualize nutritional care for children with cystic fibrosis: reliability, validity, and utility of the CF Individualized NuTritional Assessment of Kids Eating (CF INTAKE). *Child Health Care* 2005; 34: 113–131.
41 Quittner AL, Alpern AN, Wolin D, *et al.* Psychometric analyses of a new GI symptom tracker for cystic fibrosis. *Pediatr Pulmonol* 2017; 52: Suppl. 47, 453.
42 Quittner AL, Biesen J, Marciel K, *et al.* Health-related quality of life in adolescents with cystic fibrosis: associations with age, gender, nutrition and pulmonary function. *Am J Respir Crit Care Med* 2020; 201: A1176.
43 Quittner AL, Marciel KK, Kimberg CI. Content validity of the cystic fibrosis questionnaire-revised (CFQ-R). *Pediatr Pulmonol* 2011; 46: 419–420.
44 Quittner AL, Morris AM, Wainwright C, *et al.* Development of the preschool cystic fibrosis questionnaire-revised (CFQ-R). *Qual Life Res* 2015; 24: Suppl. 1, 41.
45 Radine A, Werner C, Raidt J, *et al.* Comparison of nocturnal cough analysis in healthy subjects and in patients with cystic fibrosis and primary ciliary dyskinesia: a prospective observational study. *Respiration* 2019; 97: 60–69.
46 Radtke T, Boni L, Bohnacker P, *et al.* The many ways sputum flows – dealing with high within-subject variability in cystic fibrosis sputum rheology. *Respir Physiol Neurobiol* 2018; 254: 36–39.
47 Rand S, Prasad SA, Main E. New incremental field step-test (iStep) is valid and feasible in measuring near maximal exercise performance in children with cystic fibrosis. *Physiotherapy* 2015; 101: eS931–eS932.
48 Rosenow T, Kuo W, De Bruijne M, *et al.* A new gold standard for assessing CT in early CF lung disease? *Eur Respir J* 2014; 44: Suppl. 58, 3446.
49 Rosenow T, Tiddens H, De Bruijne M, *et al.* Quantitation of chest CT abnormalities in early life CF: back to the basics. *Pediatr Pulmonol* 2013; 48: 340.
50 Rosenow T, Tiddens HA, Oudraad M, *et al.* PRAGMA: further support for use as a quantitative CT outcome measure. *Pediatr Pulmonol* 2014; 49: 295.
51 Sehgal S, Small B, Highland KB. Activity monitors in pulmonary disease. *Respir Med* 2019; 151: 81–95.
52 Sheppard E, Chang K, Cotton J, *et al.* Functional tests of leg muscle strength and power in adults with cystic fibrosis. *Respir Care* 2019; 64: 40–47.
53 Sherman AC SS, Reddy R, O'Brien C, *et al.* Finding meaning in illness: development and initial validation of a measure of sense-seeking and sense-making. *Psychosom Med* 2013; 3: A133.
54 Taylor-Cousar JL, Biederer J, Puderbach M, *et al.* Correlation of a semi-quantitative lung magnetic resonance imaging score with baseline FEV1 in cystic fibrosis. *Am J Respir Crit Care Med* 2011; 183: A11118.
55 Toucheque M, Etienne A. Assessment using e-health technologies in pediatric cystic fibrosis: developing a CF-specific module for the quality of life systemic inventory for children (QLSI-C)-iPad version. *Pediatr Pulmonol* 2014; 49: 441.
56 Tullis E, Cain E, Griffin K, *et al.* Validation of an electronic version of the Cystic Fibrosis Quality of life Evaluative Self-administered Test (eCF-QUEST). *J Cyst Fibros* 2019; 18: Suppl. 1, S179.
57 Von Berg K. Field exercise tests: which ones, and what can they tell you? *Pediatr Pulmonol* 2016; 51: Suppl. 45, 187–188.
58 Ward N, Stiller K, Rowe H, *et al.* Assessment of cough in cystic fibrosis. *Respirology* 2016; 21: Suppl. 2, 168.
59 Willgoss TG, Trigg A, Meysner S, *et al.* Understanding the suitability of cystic fibrosis-specific clinical outcome assessments for clinical trials and to support medical product labeling. *Value Health* 2015; 18: A286.
60 Williamson N, Janssen-van Solingen G, Arbuckle R, *et al.* Psychometric validation of a patient-reported outcome questionnaire in patients with pancreatic exocrine insufficiency. *Pancreatology* 2017; 17: Suppl. 3, S19–S20.

61 Wilschanski M, Yaakov Y, Omari I, *et al.* Comparison of nasal potential difference and intestinal current measurements as surrogate markers for CFTR function. *J Pediatr Gastroenterol Nutr* 2016; 63: e92–e97.

62 Ahrens RC, Standaert TA, Launspach J, *et al.* Use of nasal potential difference and sweat chloride as outcome measures in multicenter clinical trials in subjects with cystic fibrosis. *Pediatr Pulmonol* 2002; 33: 142–150.

63 Aigner C JP, Seebacher G, Mazhar S, *et al.* Cystic fibrosis and lung transplantation – determination of the survival benefit. *Wien Klin Wochenschr* 2004; 116: 318–321.

64 Alpern AN, Brumback LC, Ratjen F, *et al.* Initial evaluation of the Parent Cystic Fibrosis Questionnaire-Revised (CFQ-R) in infants and young children. *J Cyst Fibros* 2015; 14: 403–411.

65 Andrade Lima C, Dornelas de Andrade A, Campos SL, *et al.* Six-minute walk test as a determinant of the functional capacity of children and adolescents with cystic fibrosis: a systematic review. *Respir Med* 2018; 137: 83–88.

66 Balfour L, Amstrong M, Holly C, *et al.* Development and psychometric validation of a cystic fibrosis knowledge scale. *Respirology* 2014; 19: 1209–1214.

67 Bell SC, Mainz JG, MacGregor G, *et al.* Patient-reported outcomes in patients with cystic fibrosis with a G551D mutation on ivacaftor treatment: results from a cross-sectional study. *BMC Pulm Med* 2019; 19: 146.

68 Bodewes FA, Verkade HJ, Taminiau JA, *et al.* Cystic fibrosis and the role of gastrointestinal outcome measures in the new era of therapeutic CFTR modulation. *J Cyst Fibros* 2015; 14: 169–177.

69 Boling W, Macrina DM, Clancy JP. The Caregiver Quality of Life Cystic Fibrosis (CQOLCF) scale: modification and validation of an instrument to measure quality of life in cystic fibrosis family caregivers. *Qual Life Res* 2003; 12: 1119–1126.

70 Boon M, Claes I, Havermans T, *et al.* Assessing gastro-intestinal related quality of life in cystic fibrosis: validation of PedsQL GI in children and their parents. *PLoS One* 2019; 14: e0225004.

71 Bradley J, Dempster M, Wallace E, *et al.* The adaptations of a quality of life questionnaire for routine use in clinical practice: the Chronic Respiratory Disease Questionnaire in cystic fibrosis (CRDQ). *Qual Life Res* 1999; 8: 65–71.

72 Buchs C, Coutier L, Vrielynck S, *et al.* An impulse oscillometry system is less efficient than spirometry in tracking lung function improvements after intravenous antibiotic therapy in pediatric patients with cystic fibrosis. *Pediatr Pulmonol* 2015; 50: 1073–1081.

73 Cleveland RH, Stamoulis C, Sawicki G, *et al.* Brasfield and Wisconsin scoring systems have equal value as outcome assessment tools of cystic fibrosis lung disease. *Pediatr Radiol* 2014; 44: 529–534.

74 Cullen DL, Rodak B. Clinical utility of measures of breathlessness. *Respir Care* 2002; 47: 986–993.

75 dos Santos DM, Deon KC, Bullinger M, *et al.* Validity of the DISABKIDS-Cystic Fibrosis Module for Brazilian children and adolescents. *Rev Lat Am Enfermagem* 2014; 22: 819–825.

76 Driscoll KA, Modi AC, Filigno SS, *et al.* Quality of life in children with CF: psychometrics and relations with stress and mealtime behaviors. *Pediatr Pulmonol* 2015; 50: 560–567.

77 Edwards TC, Emerson J, Genatossio A, *et al.* Initial development and pilot testing of observer-reported outcomes (ObsROs) for children with cystic fibrosis ages 0–11 years. *J Cyst Fibros* 2018; 17: 680–686.

78 Eidt-Koch D, Mittendorf T, Greiner W. Cross-sectional validity of the EQ-5D-Y as a generic health outcome instrument in children and adolescents with cystic fibrosis in Germany. *BMC Pediatr* 2009; 9: 55.

79 Gee L, Abbott J, Conway SP, *et al.* Development of a disease specific health related quality of life measure for adults and adolescents with cystic fibrosis. *Thorax* 2000; 55: 946–954.

80 Gee L, Abbott J, Conway SP, *et al.* Validation of the SF-36 for the assessment of quality of life in adolescents and adults with cystic fibrosis. *J Cyst Fibros* 2002; 1: 137–145.

81 Goldbeck L, Schmitz TG. Comparison of three generic questionnaires measuring quality of life in adolescents and adults with cystic fibrosis: the 36-item short form health survey, the quality of life profile for chronic diseases, and the questions on life satisfaction. *Qual Life Res* 2001; 10: 23–36.

82 Goldbeck L, Schmitz TG, Henrich G, *et al.* Questions on life satisfaction for adolescents and adults with cystic fibrosis: development of a disease-specific questionnaire. *Chest* 2003; 123: 42–48.

83 Gruet M, Brisswalter J, Mely L, *et al.* Use of the peak heart rate reached during six-minute walk test to predict individualized training intensity in patients with cystic fibrosis: validity and reliability. *Arch Phys Med Rehabil* 2010; 91: 602–607.

84 Gruet M, Peyre-Tartaruga LA, Mely L, *et al.* The 1-minute sit-to-stand test in adults with cystic fibrosis: correlations with cardiopulmonary exercise test, 6-minute walk test, and quadriceps strength. *Respir Care* 2016; 61: 1620–1628.

85 Hoffman BM, Stonerock GL, Smith PJ, *et al.* Development and psychometric properties of the Pulmonary-specific Quality-of-Life Scale in lung transplant patients. *J Heart Lung Transplant* 2015; 34: 1058–1065.

86 Hulzebos E, Dadema T, Takken T. Measurement of physical activity in patients with cystic fibrosis: a systematic review. *Expert Rev Respir Med* 2013; 7: 647–653.

87 Johnson CD, Williamson N, Janssen-van Solingen G, *et al.* Psychometric evaluation of a patient-reported outcome measure in pancreatic exocrine insufficiency (PEI). *Pancreatology* 2019; 19: 182–190.

88 Kaplan RM, Anderson JP, Wu AW, *et al.* The Quality of Well-being Scale. Applications in AIDS, cystic fibrosis, and arthritis. *Med Care* 1989; 27: Suppl 3, S27–S43.

89 Kent L, Reix P, Innes JA, *et al.* Lung clearance index: evidence for use in clinical trials in cystic fibrosis. *J Cyst Fibros* 2014; 13: 123–138.

90 Lee AL, Rawlings S, Bennett KA, *et al.* Pain and its clinical associations in individuals with cystic fibrosis: a systematic review. *Chron Respir Dis* 2016; 13: 102–117.

91 Matecki S, Kent L, de Boeck K, *et al.* Is the raised volume rapid thoracic compression technique ready for use in clinical trials in infants with cystic fibrosis? *J Cyst Fibros* 2016; 15: 10–20.

92 McCarrier KP, Hassan M, Hodgkins P, *et al.* The Cystic Fibrosis Impact Questionnaire: qualitative development and cognitive evaluation of a new patient-reported outcome instrument to assess the life impacts of cystic fibrosis. *J Patient Rep Outcomes* 2020; 4: 36.

93 Mesbahi M, Shteinberg M, Wilschanski M, *et al.* Changes of CFTR functional measurements and clinical improvements in cystic fibrosis patients with non p.Gly551Asp gating mutations treated with ivacaftor. *J Cyst Fibros* 2017; 16: 45–48.

94 Munzenberger PJ, Van Wagnen CA, Abdulhamid I, *et al.* Quality of life as a treatment outcome in patients with cystic fibrosis. *Pharmacotherapy* 1999; 19: 393–398.

95 Padilla A, Olveira G, Olveira C, *et al.* [Validity and reliability of the St George's Respiratory Questionnaire in adults with cystic fibrosis]. *Arch Bronconeumol* 2007; 43: 205–211.

96 Pakhale S, Armstrong M, Holly C, *et al.* Assessment of stigma in patients with cystic fibrosis. *BMC Pulm Med* 2014; 14: 76.

97 Quittner AL, Buu A, Messer MA, *et al.* Development and validation of The Cystic Fibrosis Questionnaire in the United States: a health-related quality-of-life measure for cystic fibrosis. *Chest* 2005; 128: 2347–2354.

98 Quittner AL, Sawicki GS, McMullen A, *et al.* Erratum to: Psychometric evaluation of the Cystic Fibrosis Questionnaire-Revised in a national, US sample. *Qual Life Res* 2012; 21: 1279–1290.

99 Quon BS, Patrick DL, Edwards TC, *et al.* Feasibility of using pedometers to measure daily step counts in cystic fibrosis and an assessment of its responsiveness to changes in health state. *J Cyst Fibros* 2012; 11: 216–222.

100 Radtke T, Hebestreit H, Puhan MA, *et al.* The 1-min sit-to-stand test in cystic fibrosis – insights into cardiorespiratory responses. *J Cyst Fibros* 2017; 16: 744–751.

101 Radtke T, Puhan MA, Hebestreit H, *et al.* The 1-min sit-to-stand test – a simple functional capacity test in cystic fibrosis? *J Cyst Fibros* 2016; 15: 223–226.

102 Rosenow T, Oudraad MC, Murray CP, *et al.* PRAGMA-CF. A quantitative structural lung disease computed tomography outcome in young children with cystic fibrosis. *Am J Respir Crit Care Med* 2015; 191: 1158–1165.

103 Sagel SD, Chmiel JF, Konstan MW. Sputum biomarkers of inflammation in cystic fibrosis lung disease. *Proc Am Thorac Soc* 2007; 4: 406–417.

104 Salek MS, Jones S, Rezaie M, *et al.* Do patient-reported outcomes have a role in the management of patients with cystic fibrosis? *Front Pharmacol* 2012; 3: 38.

105 Sawicki GS, Sellers DE, Robinson WM. Self-reported physical and psychological symptom burden in adults with cystic fibrosis. *J Pain Symptom Manage* 2008; 35: 372–380.

106 Schmid-Mohler G, Caress AL, Spirig R, *et al.* Patient-reported outcome measures for symptom perception during a cystic fibrosis exacerbation. *Respir Care* 2018; 63: 353–366.

107 Simeoni MC, Schmidt S, Muehlan H, *et al.* Field testing of a European quality of life instrument for children and adolescents with chronic conditions: the 37-item DISABKIDS Chronic Generic Module. *Qual Life Res* 2007; 16: 881–893.

108 Sockrider MM, Swank PR, Seilheimer DK, *et al.* Measuring clinical status in cystic fibrosis: internal validity and reliability of a modified NIH score. *Pediatr Pulmonol* 1994; 17: 86–96.

109 Solomon GM, Liu B, Sermet-Gaudelus I, *et al.* A multiple reader scoring system for nasal potential difference parameters. *J Cyst Fibros* 2017; 16: 573–578.

110 Stanojevic S, Ratjen F. Physiologic endpoints for clinical studies for cystic fibrosis. *J Cyst Fibros* 2016; 15: 416–423.

111 Szczesniak R, Turkovic L, Andrinopoulou ER, *et al.* Chest imaging in cystic fibrosis studies: what counts, and can be counted? *J Cyst Fibros* 2017; 16: 175–185.

112 Tluczek A, Becker T, Grieve A, *et al.* Health-related quality of life in children and adolescents with cystic fibrosis: convergent validity with parent-reports and objective measures of pulmonary health. *J Develop Behav Pediatr* 2013; 34: 252–261.

113 Usatin D, Yen EH, McDonald C, *et al.* Differences between WHO and CDC early growth measurements in the assessment of Cystic Fibrosis clinical outcomes. *J Cyst Fibros* 2017; 16: 503–509.

114 VanDevanter DR, Heltshe SL, Spahr J, *et al.* Rationalizing endpoints for prospective studies of pulmonary exacerbation treatment response in cystic fibrosis. *J Cyst Fibros* 2017; 16: 607–615.

115 Ward N, Stiller K, Rowe H, *et al.* The psychometric properties of the Leicester Cough Questionnaire and Respiratory Symptoms in CF tool in cystic fibrosis: a preliminary study. *J Cyst Fibros* 2017; 16: 425–432.

116 Wenninger K, Weiss C, Wahn U, *et al.* Body image in cystic fibrosis – development of a brief diagnostic scale. *J Behav Med* 2003; 26: 81–94.

117 Yohannes AM, Dodd M, Morris J, *et al.* Reliability and validity of a single item measure of quality of life scale for adult patients with cystic fibrosis. *Health Qual Life Outcomes* 2011; 9: 105.

118 Young NL, Varni JW, Snider L, *et al.* The Internet is valid and reliable for child-report: an example using the Activities Scale for Kids (ASK) and the Pediatric Quality of Life Inventory (PedsQL). *J Clin Epidemiol* 2009; 62: 314–320.

119 Gruet M. The 1-minute sit-to-stand test in adults with cystic fibrosis: correlations with cardiopulmonary exercise test, 6-minute walk test, and quadriceps strength. *Respir Care* 2016; 61: 1620–1628.

120 U.S. National Library of Medicine. Clinical Trials Register. www.clinicaltrials.gov/ Date last accessed: 2 October 2020.

121 Aigner C, Jaksch P, Seebacher G, *et al.* Cystic fibrosis and lung transplantation – determination of the survival benefit. *Wien Klin Wochenschr* 2004; 116: 318–321.

122 Brasfield D, Hicks G, Soong S, *et al.* The chest roentgenogram in cystic fibrosis: a new scoring system. *Pediatrics* 1979; 63: 24–29.

123 Chrispin AR, Norman AP. The systematic evaluation of the chest radiograph in cystic fibrosis. *Pediatr Radiol* 1974; 2: 101–105.

124 Conway SP, Pond MN, Bowler I, *et al.* The chest radiograph in cystic fibrosis: a new scoring system compared with the Chrispin-Norman and Brasfield scores. *Thorax* 1994; 49: 860–862.

125 Dill EJP, Dawson RP, Sellers DEP, *et al.* Longitudinal trends in health-related quality of life in adults with cystic fibrosis. *Chest* 2013; 144: 981–989.

126 Kerem E, Reisman J, Corey M, *et al.* Prediction of mortality in patients with cystic fibrosis. *N Engl J Med* 1992; 326: 1187–1191.

127 Liou TG, Adler FR, FitzSimmons SC, *et al.* Predictive 5-year survivorship model of cystic fibrosis. *Am J Epidemiol* 2001; 153: 345–352.

128 Smountas AA, Lands LC, Mohammed SR, *et al.* Induced sputum in cystic fibrosis: within-week reproducibility of inflammatory markers. *Clin Biochem* 2004; 37: 1031–1036.

129 Weatherly MR, Palmer CG, Peters ME, *et al.* Wisconsin cystic fibrosis chest radiograph scoring system. *Pediatrics (Evanston)* 1993; 91: 488.

130 American Thoracic Society. Cystic Fibrosis Questionnaire (CFQ) and Cystic Fibrosis Questionnaire Revised (CFQ-R). https://qol.thoracic.org/sections/instruments/ae/pages/cfq-cfq-r.html Date last updated: April 2002; date last accessed: 29 September 2020.

131 U.S. Food and Drug Administration. Table of surrogate endpoints that were the basis of drug approval or licensure. www.fda.gov/drugs/development-resources/table-surrogate-endpoints-were-basis-drug-approval-or-licensure Date last updated: 20 August 2020; date last accessed: 27 September 2020.

132 European Medicines Agency. Report of the workshop on endpoints for cystic fibrosis clinical trials. www.ema.europa.eu/en/documents/report/report-workshop-endpoints-cystic-fibrosis-clinical-trials_en.pdf Date last updated: 29 November 2012; date last accessed: 13 October 2020.

133 Blackwell LS, Quittner AL. Daily pain in adolescents with CF: effects on adherence, psychological symptoms, and health-related quality of life. *Pediatr Pulmonol* 2015; 50: 244–251.

134 Henry B, Aussage P, Grosskopf C, *et al.* Development of the Cystic Fibrosis Questionnaire (CFQ) for assessing quality of life in pediatric and adult patients. *Qual Life Res* 2003; 12: 63–76.

135 Ratnayake I, Ahern S, Ruseckaite R. A systematic review of patient-reported outcome measures (PROMs) in cystic fibrosis. *BMJ Open* 2020; 10: e033867.

136 COSMIN. www.cosmin.nl/

137 US Department of Health and Human Services. Multiple Endpoints for Clinical Trials: Guidance for Industry. www.fda.gov/media/102657/download Date last updated: January 2017; date last accessed: 15 October 2020.

138 Graham BL, Steenbruggen I, Miller MR, *et al.* Standardization of Spirometry 2019 Update. An Official American Thoracic Society and European Respiratory Society Technical Statement. *Am J Respir Crit Care Med* 2019; 200: e70–e88.

139 Cystic Fibrosis Foundation Patient Registry. 2018 Annual Data Report. Bethesda, MD, National Institutes of Health, 2018.

140 Szczesniak R, Heltshe SL, Stanojevic S, *et al.* Use of FEV1 in cystic fibrosis epidemiologic studies and clinical trials: a statistical perspective for the clinical researcher. *J Cyst Fibros* 2017; 16: 318–326.

141 Stanojevic S, Quanjer P, Miller MR, *et al.* The Global Lung Function Initiative: dispelling some myths of lung function test interpretation. *Breathe* 2013; 9: 462–474.

142 Wainwright CE, Elborn JS, Ramsey BW, *et al.* Lumacaftor–ivacaftor in patients with cystic fibrosis homozygous for Phe508del CFTR. *N Engl J Med* 2015; 373: 220–231.

143 Elborn JS, Ramsey BW, Boyle MP, *et al.* Efficacy and safety of lumacaftor/ivacaftor combination therapy in patients with cystic fibrosis homozygous for Phe508del CFTR by pulmonary function subgroup. *Lancet Respir Med* 2016; 4: 617–626.

144 Konstan MW, Wagener JS, VanDevanter DR, *et al.* Risk factors for rate of decline in FEV1 in adults with cystic fibrosis. *J Cyst Fibros* 2012; 11: 405–411.
145 Fretzayas A, Douros K, Moustaki M, *et al.* Applications of lung clearance index in monitoring children with cystic fibrosis. *World J Clin Pediatr* 2019; 8: 15–22.
146 Hardaker KM, Panda H, Hulme K, *et al.* Abnormal preschool Lung Clearance Index (LCI) reflects clinical status and predicts lower spirometry later in childhood in cystic fibrosis. *J Cyst Fibros* 2019; 18: 721–727.
147 Bouma N JH, Andrinopoulou ER, Tiddens H. Airway disease on chest computed tomography of preschool children with cystic fibrosis is associated with school aged bronchiectasis. *Pediatr Pulmonol* 2020; 55: 141–148.