



Assessing exertional dyspnea in patients with idiopathic pulmonary fibrosis



Jeffrey J. Swigris^{a,*}, David L. Streiner^b, Kevin K. Brown^a,
Amanda Belkin^a, Kathy E. Green^c, Frederick S. Wamboldt^d on
behalf of the IPFnet Investigators

^a Autoimmune Lung Center and Interstitial Lung Disease Program, National Jewish Health, Denver, CO, USA

^b Department of Psychiatry and Behavioural Neurosciences, McMaster University, Ontario, Canada

^c Morgridge College of Education, University of Denver, Denver, CO, USA

^d Division of Pulmonary and Critical Care Medicine, Sleep and Behavioral Health Sciences Section, National Jewish Health, Denver, CO, USA

Received 24 September 2013; accepted 16 December 2013
Available online 24 December 2013

KEYWORDS

Dyspnea;
Idiopathic pulmonary
fibrosis;
Quality of life;
Functional capacity

Summary

Background: Dyspnea is a hallmark symptom of idiopathic pulmonary fibrosis (IPF), and dyspnea induced physical activity limitation is a prominent driver of quality of life impairment among IPF patients.

Methods: We examined response data for the 21 physical activity items (the first 21 of 24) from the University of California San Diego Shortness of Breath Questionnaire (UCSD) collected at baseline in a recently conducted IPF trial. We used Rasch analysis and hypothesis testing with conventional statistical methodology to achieve three objectives: 1) to examine the items to identify the one characteristic that distinguishes one from another; 2) to assess these items for their ability to measure dyspnea severity in IPF; 3) to use the items to develop a dyspnea ruler.

Results: The sample comprised 178 subjects. The 21 items fit the Rasch model. There was very strong correlation between Rasch item severity and their metabolic equivalents (METs) values ($r = -0.86$, $p < 0.0001$). With the sample stratified on scores from the 21 items, there were significant between group differences in FVC%, DLCO% and distance walked during the six-minute walk test. The dyspnea ruler can be used to put dyspnea levels in a more easily understood clinical context.

Conclusions: The first 21 items from the UCSD compose a unidimensional dyspnea-with-activity scale and are both sensibly ordered and distinguished from each other by their

* Corresponding author. Autoimmune Lung Center and Interstitial Lung Disease Program, National Jewish Health, 1400 Jackson Street, Denver, CO 80206, USA. Tel.: +1 (303) 398 1621; fax: +1 (303) 398 1040.
E-mail address: swigrisj@njc.org (J.J. Swigris).

METS values. These 21 items can be used confidently to formulate clinically-relevant inferences about IPF patients and should be considered for use as a meaningful endpoint in IPF research.

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Idiopathic pulmonary fibrosis (IPF) is a progressive, fibrosing interstitial lung disease of unknown etiology. IPF insinuates itself into patients' lives, ultimately leaving them short of breath when performing the most basic daily activities [1]. In IPF, dyspnea is the most common, and for the majority of patients, the most debilitating symptom and the primary driver of quality of life (QOL) impairment [2]. Thus, dyspnea—and by extension the ability to meet the energy demands of day-to-day activities—is an outcome worth considering in trials of therapy for IPF.

The University of California San Diego Shortness of Breath Questionnaire (UCSD) is a patient-reported outcome (PRO), 21 of whose 24 items ask respondents to rate the dyspnea they perceived while performing various physical activities during the previous week [3]. The last three items focus on manifestations of dyspnea unrelated to physical activity (e.g., effects on emotional health). The UCSD has been used as a secondary endpoint in IPF trials, and there are data from a single study to support its validity as an instrument capable of tracking dyspnea in IPF patients [4]. Compared with other dyspnea indexes that have also been used in IPF studies (e.g., the Borg scale, Medical Research Council Breathlessness scale, the Baseline/Transition Dyspnea Index), the UCSD includes more items and response options, and thus may assess a person's dyspnea severity with greater precision.

When investigators study the validity of PROs, analyses predominantly focus on the relationship between PRO scores and concurrently collected tests of disease severity, or these analyses look for expected differences in PRO scores between subgroups of the study sample defined by measures of disease severity [5]. What is rarely studied are the items themselves; specifically, what characteristics make one item more difficult for a patient to endorse than another item—without this information, “the understanding of what is being measured [by a PRO] is unsatisfyingly primitive” [6].

In this study, we examined what the first 21 items of the UCSD measure. We hypothesized that what differentiates one item from another is the metabolic equivalents (METs) linked to the physical activity each item inquires about. We analyzed response data collected at baseline in the Sildenafil Trial of Exercise Performance in Idiopathic Pulmonary Fibrosis (STEP-IPF) to achieve three goals: 1) to test this hypothesis; 2) to examine the ability of scores from these 21 items to distinguish subgroups with different levels of IPF severity and 3) to generate a “dyspnea ruler” that places scores from these items in a clinically relevant context.

Methods

STEP-IPF was a placebo-controlled trial designed to examine the effects of sildenafil in patients with severe IPF

[7]. Baseline data, including percent predicted forced vital capacity and diffusing capacity of the lung for carbon monoxide (FVC% and DLCO% respectively) and distance walked during a six-minute walk test (6MWD), from 178 of the 180 STEP-IPF participants were suitable for analysis.

The UCSD

For the UCSD, respondents rate themselves from 0 (“Not at all”) to 5 (“Maximally or unable to do because of breathlessness”) in two areas: 1) how short of breath they are while performing various activities (21 items); and 2) how much shortness of breath, fear of hurting themselves by overexerting, and fear of shortness of breath limit them in their daily lives (3 items). See [Supplement](#) for a copy of the UCSD. Scores for the entire instrument range from 0 to 120; thus scores for the first 21 items range from 0 to 105, with higher scores indicating greater dyspnea [3].

Analyses

Rasch analysis

Rasch analysis is a statistical method used with increasing frequency to evaluate the performance characteristics of individual PRO items and entire PROs [8–11]. In Rasch analysis, PRO items are first calibrated on a linear difficulty scale, from most likely (easiest) to least likely (most difficult) to be endorsed. Although other terms are sometimes used, here, we refer to these item calibrations as item difficulties. Once items are calibrated, the underlying mathematics of the Rasch model incorporate a patient's responses to the aggregate of items to locate him on the same scale at a position corresponding to his level of the “thing” being measured [11]. Here, we refer to that position as patient severity. Both item difficulty and patient severity are measured in log odds or logits.

Rasch analysis is based on the principle of Guttman scaling; that is, agreement with an item implies agreement with any less difficult items (e.g., if a patient reports shortness of breath after climbing one flight of stairs, she should report being short of breath after climbing two flights of stairs). Rasch uses only the difference between two parameters—item difficulty and patient severity—to model the probability of responses to each item. The analysis generates several statistics that are used to assess the fit of items to model expectations. In the specific case of the UCSD, two things are expected: 1) patients with more severe dyspnea will be located higher on the scale than patients with less severe dyspnea; and 2) as a patient's dyspnea severity increases, for any given item, the probability of him choosing a response option that suggests more severe dyspnea also increases. For the UCSD (and its six response options: 0, 1, 2, 3, 4, 5), an item's difficulty is where a response of “0” or a response of “5” is equally

likely. Rasch generates a plethora of output, including an informative figure, called an item map. This shows item difficulty and patient severity along opposite sides of a vertical line. [12] The item map is likened to a ruler: just as lower-value numbers on a ruler connote less length, patients at the lower end of the map have less severe dyspnea, and items at the lower end of the map are less difficult to endorse and thus connote less severe dyspnea than items at the higher end.

We subjected response data for the 21 items to Rasch analysis. We examined the ordering of responses for each item and rescored items with improperly ordered response thresholds by collapsing response categories as needed. Once all items had properly ordered response categories, we assessed the fit of the aggregate of items and of individual items to the Rasch model. There are no absolute criteria, but perhaps the most commonly used measure of item fit to the Rasch model—and the one we employed—is the infit mean square statistic (an infit mean square statistic from 0.5 to 1.5 is considered useful for measurement, while those greater than 2.0 degrade measurement) [13]. Finally, for each of the 21 items, we assessed whether responses from subgroups within the cohort deviated from model-derived, expected values. This is accomplished with an analysis called DIF or differential item functioning.

By providing a mathematical formulation of (certain) fundamental measurement properties, Rasch analysis can be used to determine whether a dataset conforms to the requirements of fundamental measurement [14]. One property—unidimensionality, holds that items function in unison to measure a single construct (here, it would be dyspnea).

Exploratory factor analysis (EFA)

To further confirm that the 21 UCSD conform to this fundamental property, we subjected them to exploratory factor analysis. The EFA allowed us to assess the relationship among the first 21 items of the UCSD and to determine if they all target the same underlying construct, dyspnea-with-activity.

Examining METS as a determinant of item severity

We hypothesized the intensity of physical demands tied to an item distinguishes it from other items. We tested whether item difficulties (from the Rasch analysis) were related to physical activity demand intensities measured in units of energy cost, or METS. As such, items highest on the scale (connoting the most severe dyspnea) would be those requiring the lowest intensity of physical demands. For example, the item “How short of breath do you get while eating?” is expected to be more difficult than the item “How short of breath do you get while climbing a hill?” This is because most IPF patients, except perhaps those with the least severe disease, will have some shortness of breath when climbing a hill, but only the most severe IPF patients will experience shortness of breath with a low-intensity activity, like eating.

We identified METS for each item [15]. To examine the relationship between an item’s METS value and its difficulty, we used the Pearson correlation coefficient and linear regression (regressed item difficulty on item METS) [5]. Precise METS values for any physical activity depend on

the intensity with which it is completed; thus, for many activities, the Compendium of Physical Activity [15] provides a range of METS values. For our analyses, for items with ranges of METS values, we used the median.

Validity analyses for 21 items

Using the extreme groups approach, we stratified the cohort into three subgroups based on their 21-item score (≤ 25 th percentile, 25–75th percentile or ≥ 75 th percentile) and compared values for disease severity measures (e.g., FVC%, DLCO% and 6MWD) across subgroups using analysis of variance (ANOVA) and *p*-value-corrected pairwise comparisons between subgroups. Next, we stratified the cohort into three subgroups with differing functional capacities, as defined by 6MWD (≤ 25 th percentile, 25–75th percentile or ≥ 75 th percentile) and compared mean 21-item UCSD scores across the three 6MWD subgroups via the same method.

Dyspnea ruler

This ruler shows the relationships between patient dyspnea severity (as measured by raw score for the 21 UCSD items), item difficulty (from the Rasch analysis) and METS values. See [Supplement](#) for details.

The Rasch analysis was run with Winsteps, Version 3.69.1.14 (www.winsteps.com). All other statistical analyses were run using SAS version 9.2 (SAS Inc., Cary, NC). We considered *p* less than 0.05 to represent statistical significance.

Results

There were 180 subjects enrolled in STEP-IPF, and 178 had complete, analyzable UCSD response data at baseline. Their baseline characteristics are found in [Table 1](#).

EFA

The results from the factor analysis confirmed unidimensionality; i.e., the presence of a single dominant factor (first factor eigenvalue = 11.39, second factor eigenvalue = 2.38). The scree plot is in the [Supplement](#).

Rasch analysis

To correct disorder, response categories were collapsed for five items (Hill, Eating, Dishes, MowLawn and SexualAct). [Table 2](#) displays Rasch Rating Scale Model fit statistics and METS values for each item. All items had mean square infit

Table 1 Baseline characteristics of study cohort.

Characteristic	Total subjects = 178
Age, yrs	69.0 (9.0)
Male, %	84
Time since diagnosis, yrs	2.0 (1.9)
FVC%	56.9 (14.2)
DLCO%	26.4 (6.1)
6MWD, meters	265.4 (117.1)
UCSD SOB questionnaire	47.1 (21.3)

Table 2 Rasch fit statistics and METS values for the 21 UCSD physical activity items.

Item	Logit	MNSQ Infit	MNSQ Outfit	METS
While eating	1.72 ± 0.11	1.03	0.98	1.5
At rest	1.61 ± 0.12	1.04	0.99	1.0
Standing up from chair	1.51 ± 0.10	1.28	1.53	1.3–1.8
Dressing	1.46 ± 0.10	0.96	1.79	2.0–2.5
Doing dishes	1.28 ± 0.09	0.76	0.86	1.8–2.5
Brushing teeth	1.23 ± 0.11	0.85	0.73	2.0
Shaving and/or brushing hair	1.19 ± 0.11	0.88	0.85	2.0
Showering/bathing	0.69 ± 0.09	0.96	0.96	2.0
Doing laundry	0.26 ± 0.08	0.71	0.71	2.0–2.3
Shopping	−0.01 ± 0.08	0.82	0.83	2.3
Picking up and straightening	−0.06 ± 0.09	0.79	0.79	2.5
Making bed	−0.33 ± 0.09	0.65	0.64	3.3
Watering lawn	−0.34 ± 0.08	1.13	1.07	1.5–3.0
Sweeping/vacuuming	−0.42 ± 0.08	0.74	0.73	2.3–3.8
Walk on the level at your own pace	−0.65 ± 0.10	1.05	1.03	2.0–3.5
Walk on the level with others your own age	−0.89 ± 0.08	1.12	1.17	2.5–4.3
Sexual activities	−0.93 ± 0.08	1.50	1.46	1.8–2.8
Washing car	−1.32 ± 0.08	0.78	0.75	3.5
Mowing lawn	−1.59 ± 0.08	1.65	1.93	5.5
Walking up stairs	−2.19 ± 0.10	1.24	1.36	4.0–8.0
Walking up a hill	−2.21 ± 0.10	1.23	1.37	6.3–7.3

MNSQ = mean square; METS = metabolic equivalents.

statistics less than 2.0 and all but “Mowing the Lawn” were 0.5–1.5. The person separation index (similar to Cronbach’s alpha) for the 21 items was 0.95, suggesting excellent ability to discriminate between subjects with differing levels of dyspnea severity. The item map is displayed in Fig. 1. Another item map that includes thresholds between response options for each item is in the Supplement. No items demonstrated significant DIF.

Relationship between item severity and METS

Fig. 2 shows the association between Rasch item difficulty and item METS values ($r = -0.86$, $p < 0.0001$). From the linear regression, METS values accounted for nearly 75% of the variance in item severity (METS $\beta = -0.72$, $p < 0.0001$, $R^2 = 0.74$).

Validity analyses

With the sample stratified on raw score for the 21 items, subjects with the least severe dyspnea had higher FVC%, DLCO% and 6MWD than subjects in either of the other two strata (Table 3). Conversely, with the sample stratified on 6MWD, we observed that subjects who walked the greatest distance had the lowest UCSD scores (Table 4).

Dyspnea ruler

The UCSD Dyspnea Ruler, displayed in Fig. 3, shows the relationship between 21-item UCSD score, locations for the thresholds between response options “2” and “3” for each item, and loosely-defined METS ranges. In STEP-IPF, 12 subjects died prior to study end (24 weeks), and a total of 15 deaths occurred by week 28. All but one of the 15 had a 21-item UCSD score less than 28 at baseline. On the Dyspnea ruler, two items (MowLawn and Stairs) have “2–3” threshold positions near 28 and METS values around 5.5–6.0. The Rasch model predicts that all the subjects who died, except the one whose score was less than 28 (it was 20), would have rated their dyspnea as at least moderate when performing such 5.5–6.0 MET-activities, like mowing the lawn or climbing a flight of stairs. However, the majority of subjects who did not die rated their dyspnea the same, yielding a positive predictive value of only 8%. In contrast, among the 13 subjects who rated their dyspnea as less than moderate for such items, 12 lived through the end of follow-up, yielding a negative predictive value of 92% for this criterion. Six subjects who died had scores greater than 55; the Rasch model predicts all six would have rated their dyspnea as at least moderate when bathing (Bathe “2–3” threshold at 55)—a roughly 2 MET-activity.

In exploratory analyses, we generated dyspnea rulers for the UCSD using 12- and 24-week data, and the results were reassuringly similar (data not shown) to the ruler derived from the baseline data.

Discussion

We examined the first 21 (dyspnea-with-activity) items from the UCSD in a sample with physiologically severe IPF and found that scores from these 21 items can distinguish subgroups of patients with differing levels of IPF severity.

In contrast to conventional “validation studies,” in which investigators focus only on the scores a PRO yields, we examined the characteristics of the items themselves. The UCSD is usually referred to as a simple dyspnea questionnaire, but we showed that its first 21 items compose a hierarchical dyspnea-with-activity scale, with the position of items on the scale largely dependent on their METS values.

When we stratified the sample into three subgroups based on 21-item UCSD score, there were significant differences between subgroups for FVC%, DLCO% and 6MWD—the aggregate of the 21 items discriminated between subjects with varying IPF severity. The six-minute walk test, specifically 6MWD, is increasingly used as a clinical disease-severity metric and a research endpoint measure of functional capacity in patients with IPF. When we stratified the sample into three subgroups based on 6MWD, we found significant differences between subgroups in 21-item UCSD scores.

The current study builds on prior work examining the UCSD in IPF [4]. Here, we conducted the first-ever exploration of what its first 21 items measure. We found that, in IPF, the UCSD (at least its first 21 items) “behaves” as we would expect: the more severe the IPF, the more severe the dyspnea while performing a given physical activity, even

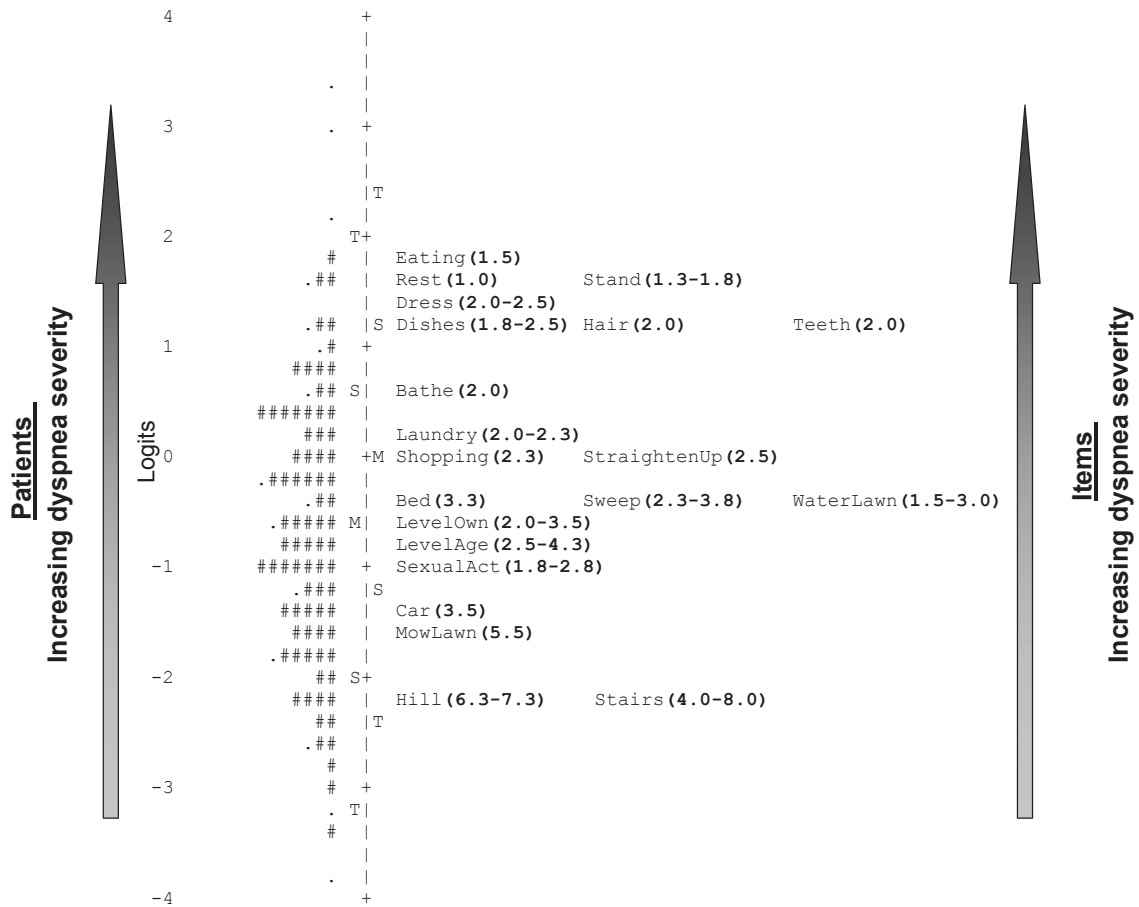


Figure 1 Item map for the 21 physical activity items from the UCSD. Footnote. The vertical line separates subjects on the left from items on the right. The units are logits. Each subject is positioned along the vertical line at his dyspnea severity value. “#” = two subjects. “.” = one subject. The “M” on the left of the line = sample’s mean dyspnea level. The “S”s on the left = one standard deviation. The “T”s = two standard deviations from the sample mean. Each item is positioned along the vertical line at its average item difficulty measure. The “M” on the right of the line = mean difficulty of the 21 items. The “S”s on the right = one standard deviation. The “T”s = two standard deviations from the mean. UCSD = University of California San Diego Shortness of Breath Questionnaire.

those activities which, for many people without respiratory disease, are not all that dyspnea-inducing. Confirming a PRO “behaves” as expected in the target population and being able to confidently make meaningful inferences

about a respondent based on his score are the basic tenets of “validation.”

Although some disagree, most experts hold that items conforming to the Rasch model are sample

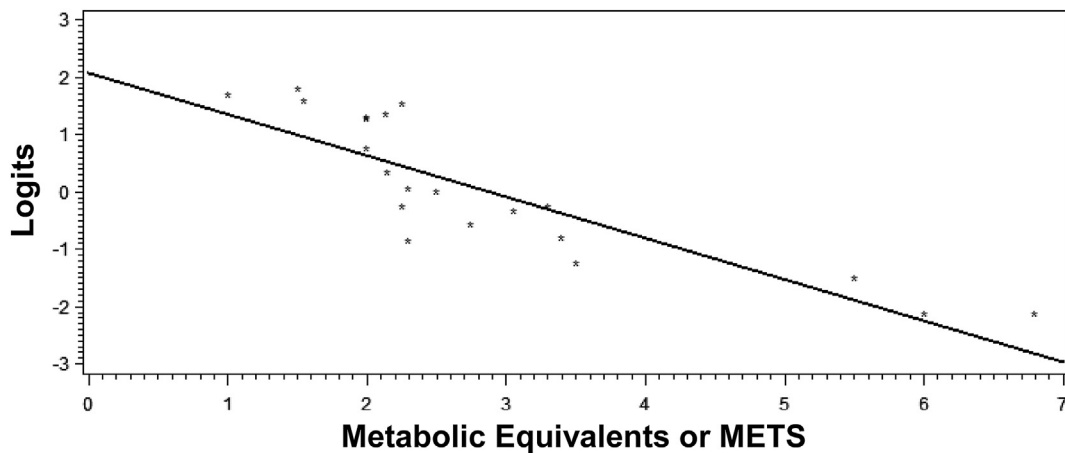


Figure 2 Relationship between item severity (in logits) and METS values. Footnote. The line is the regression line.

Table 3 Mean \pm standard error values of FVC%, DLCO% and 6MWD for subgroups defined by UCSD scores.

	Group 1 Score < 26 N = 44	Group 2 26 \leq Score \leq 57 N = 94	Group 3 Score > 57 N = 40	p Values
FVC%	60.3 \pm 2.2	57.3 \pm 1.5	53.0 \pm 2.1	Overall effect < 0.0001 1 vs. 2 = 0.09 1 vs. 3 = 0.01 2 vs. 3 = 0.26
DLCO%	27.8 \pm 1.0	26.5 \pm 0.6	24.7 \pm 0.9	Overall effect < 0.0001 1 vs. 2 = 0.11 1 vs. 3 = 0.01 2 vs. 3 = 0.23
6MWD (meters)	335.2 \pm 17.3	260.6 \pm 11.3	212.2 \pm 16.5	Overall effect < 0.0001 1 vs. 2 = 0.01 1 vs. 3 < 0.0001 2 vs. 3 = 0.0004

FVC% = percent predicted forced vital capacity; DLCO% = percent predicted diffusing capacity of the lung for carbon monoxide; 6MWD = distance walked during six-minute walk test; UCSD = University of California San Diego Shortness of Breath Questionnaire.

independent—their item severities will not vary from sample to sample [16]. Because they fit the Rasch model, like METS—for a given activity, every person receives the same MET value [17]—the first 21 items of the UCSD are independent of the severity a sample's dyspnea (or by extension, the severity of IPF). In this way, item difficulties are analogous to bar height in a high-jump competition [12]: in the high-jump, the probability of successfully clearing the bar depends only on the jumper's ability relative to the bar height; our analysis shows us that the probability of a patient responding to a UCSD item a certain way depends only on the severity of his dyspnea relative to the item's difficulty (or by extension, the METS associated with it).

Items that fit the Rasch model function like a ruler: IPF patients with more severe dyspnea will be located higher up the ruler (or item map), but like a ruler, the values on the line (i.e., the item difficulties) are static, regardless of application. So, based on a patient's position on the Dyspnea ruler (i.e., her 21-item UCSD score), because of the relationship between item and METS as depicted, one can make inferences about her ability to complete any number of physical activities, so long as you know the METS value associated with it. For example, a patient with a 21-item UCSD score of 30 would likely rate an activity like sweeping

the sidewalk (4.0 METS) as causing a moderate degree of dyspnea. The converse is true: the patient who says they get moderately short of breath when performing an activity like sacking grass or leaves (4.0 METS) would likely score around 30 on the 21-item UCSD.

This study has limitations. Subjects in the STEP-IPF trial—as required for inclusion—had DLCOs less than 35% predicted. However, because item difficulty values are independent of the sample to which they are applied, the Dyspnea Ruler—and inferences made based on 21-item UCSD scores—should be applicable to IPF patients of any severity. We caution against over-extending the results of this single study, but they may have clinical implications: the 21 dyspnea-with-activity items of the UCSD could be administered to patients and their results plotted on the Dyspnea Ruler. Not much can be said about the prognosis of patients who are at least moderately dyspneic performing 5.5–6.0 MET activities, but the data suggest patients less than moderately dyspneic performing such activities have very good short-term prognosis. Additional research could shed more light on other applications of these data.

Our goal was not to decrease UCSD items to the lowest number possible; 24, 21 or fewer items of this type are equally trivial in terms of respondent burden. But even the 21-item instrument could potentially be improved by

Table 4 Mean \pm standard error score for 21 UCSD items for subgroups defined by 6MWD.

	Group 1 6MWD > 357 m N = 40	Group 2 182 m \leq 6MWD \leq 357 m N = 90	Group 3 6MWD < 182 m N = 44	p Values
21 UCSD items	67.2 \pm 1.3	39.9 \pm 0.9	17.3 \pm 1.3	Overall effect < 0.0001 1 vs. 2 < 0.0001 1 vs. 3 < 0.0001 2 vs. 3 < 0.0001

6MWD = distance walked during six-minute walk test; UCSD = University of California San Diego Shortness of Breath Questionnaire.

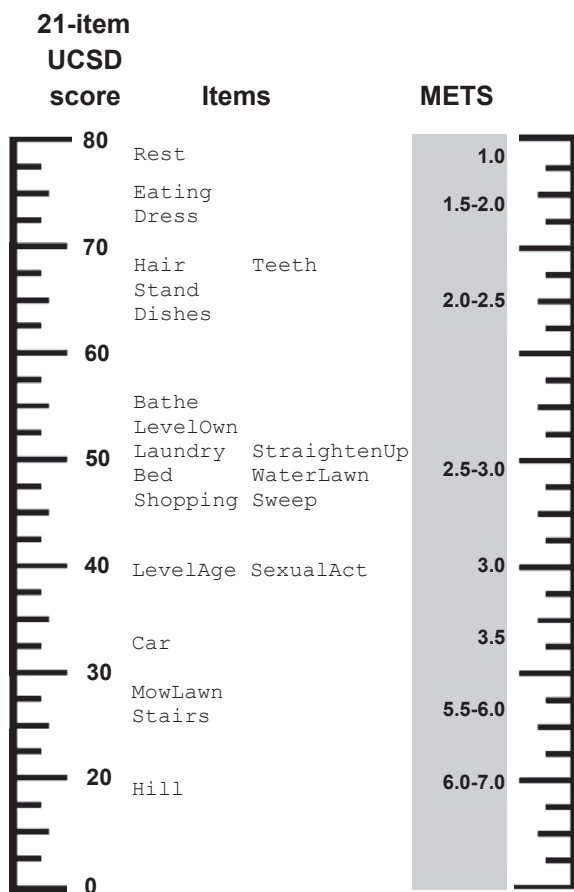


Figure 3 Dyspnea ruler. Footnote. The units are raw scores from the 21-item UCSD (left) and metabolic equivalents (METS on the right). Items are positioned at their respective “2–3” thresholds; that is, the point at which a patient is equally likely to respond “2” or “3” to an item. To use the ruler, first calculate the patient’s 21-item UCSD score, and find it on the ruler. For items located at that level on the ruler, the patient is predicted to report a moderate degree of dyspnea (coinciding with a response of “2” or “3”). By extension, he would report moderate dyspnea for other activities with similar METS values. For items (and other higher-METS activities) located lower on the ruler, he would report at least moderate dyspnea; for items (and other lower-METS activities) located higher on the ruler, he would report no greater than moderate dyspnea.

eliminating certain items and/or perhaps adding others. Although doing so is beyond the scope of this study, because we now understand more clearly what determines item difficulty, this could be accomplished relatively simply.

Conclusion

The 21 physical activity items from the UCSD formulate a dyspnea-with-activity scale whose components are sensibly ordered according to their METS values, capable of discriminating subgroups of IPF patients with differing severities of IPF (and dyspnea). The Dyspnea ruler can be used for at least two purposes: 1) to place dyspnea in an understandable, real-world context; and 2) to promote formulation of inferences about IPF patients that extend

beyond available data. These novel findings add data supporting the usefulness of the UCSD in IPF.

Funding

The project described was supported by grants from the NHLBI: U10HL080413 (data coordinating center), U10HL080274, U10HL080370, U10HL080371, U10HL080383, U10HL080411, U10HL080509, U10HL080510, U10HL080513, U10HL080543, U10HL080571, U10HL080685 (clinical centers) and NCATS: CTSA Grant UL1 TR000154. Dr. Swigris is supported in part by a Career Development Award from the NIH (K23 HL092227).

Author contributions

Dr. Swigris: contributed to study conceptualization and planning, analyzing data, generating intellectual content for the manuscript and critiquing and approving the final content.

Dr. Streiner contributed to analyzing data, generating intellectual content for the manuscript and critiquing and approving the final content.

Dr. Brown: contributed to generating intellectual content for the manuscript and critiquing and approving the final content.

Miss Belkin: contributed to generating intellectual content for the manuscript and critiquing and approving the final content.

Dr. Green: contributed to interpreting analyzed data, generating intellectual content for the manuscript and critiquing and approving the final content.

Dr. Wamboldt: contributed to study conceptualization and planning, generating intellectual content for the manuscript and critiquing and approving the final content.

Conflict of interest

Dr. Brown was an investigator in the STEP-IPF trial. None of the other authors have any real or potential conflicts of interest to disclose.

Acknowledgments

The following IPFnet members participated in the STEP-IPF study:

Protocol Chairs—*National Jewish Health*: M. Schwarz; *Sansum Clinic, Santa Barbara*: D.A. Zisman. **IPFnet Steering Committee Chair**—*University of Iowa*: G. Hunninghake. **Clinical Centers**—*Cleveland Clinic*: J. Chapman, M. Olman, S.Lubell; *Duke University Medical Center*: L.D. Morrison, M.P. Steele, T. Haram; *Emory University*: J. Roman, R. Perez, T. Perez; *Mayo Clinic, Rochester*: J.H. Ryu, J.P. Utz, A.H. Limper, C.E. Daniels, K. Meiras, S. Walsh. *National Jewish Health*: K.K. Brown, M. Schwarz, MD, C. Bair, D. Kervitsky; *Tulane University*: J.A. Lasky, S. Ditta. *University of Alabama, Birmingham*: J. deAndrade, V.J. Thanickal, M. Stewart; *University of California, Los Angeles*: D.A. Zisman, J. Lynch, E. Calahan, P. Lopez; *University of California, San Francisco*: T.E. King, Jr., H.R. Collard, J.A.

Golden, P.J. Wolters, R. Jeffrey; *University of Chicago*: I. Noth, D.K. Hogarth, N. Sandbo, M.E. Streck, S.R. White, C. Brown, I. Garic, S. Maleckar; *University of Michigan*: F.J. Martinez, K.R. Flaherty, M. Han, B. Moore, G.B. Toews, D. Dahlgren; *University of Washington*: G. Raghu, J. Hayes, M. Snyder; *Vanderbilt University*: J.E. Loyd, L. Lancaster, W. Lawson, R. Greer, W. Mason; *Weill Medical College of Cornell University*: R.J. Kaner, V. Monroy, M. Wang. **Core Lab Chairs—Radiology**: *National Jewish Health*: D.A. Lynch; *Pathology*: *Mayo Clinic*: T. Colby. **Data Coordinating Center—Duke Clinical Research Institute**: K.J. Anstrom, R.C. Becker, E.L. Eisenstein, N.R. MacIntyre, L.D. Morrison, J. Rochon, M.P. Steele, J.S. Sundry, L. Davidson-Ray, P. Dignacco, R. Edwards, R. Anderson, R. Beci, S. Calvert, K. Cain, T. Gentry-Bumpass, D. Hill, M. Ingham, E. Kagan, J. Kaur, C. Matti, J. McClelland, A. Meredith, T. Nguyen, J. Pesarchick, R.S. Roberts, W. Tate, T. Thomas, J. Walker, D. Whelan, J. Winsor, Q. Yang, E. Yow. **NHLBI**—H.Y. Reynolds, X. Tian, J. Kiley.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.rmed.2013.12.009>.

References

- [1] Raghu G. Idiopathic pulmonary fibrosis: guidelines for diagnosis and clinical management have advanced from consensus-based in 2000 to evidence-based in 2011. *Eur Respir J* 2011;37:743–6.
- [2] Swigris JJ, Gould MK, Wilson SR. Health-related quality of life among patients with idiopathic pulmonary fibrosis. *Chest* 2005;127:284–94.
- [3] Eakin EG, Resnikoff PM, Prewitt LM, et al. Validation of a new dyspnea measure: the UCSD Shortness of Breath Questionnaire. University of California, San Diego. *Chest* 1998; 113:619–24.
- [4] 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 Oct;106(10):1447–55.
- [5] Doward LC, McKenna SP. Defining patient-reported outcomes. *Value Health* 2004;7(Suppl. 1):S4–8.
- [6] Stenner A, Smith III M. Testing construct theory. *Percept Motor Skills* 1982;1982:415–26.
- [7] Zisman DA, Schwarz M, Anstrom KJ, et al. A controlled trial of sildenafil in advanced idiopathic pulmonary fibrosis. *N Engl J Med* 2010;363:620–8.
- [8] Yorke J, Moosavi SH, Shulldham C, et al. Quantification of dyspnoea using descriptors: development and initial testing of the Dyspnoea-12. *Thorax* 2010;65:21–6.
- [9] Yorke J, Jones PW, Swigris JJ. Development and validity testing of an IPF-specific version of the St George's Respiratory Questionnaire. *Thorax* 2010;65:921–6.
- [10] Meguro M, Barley EA, Spencer S, et al. Development and validation of an improved, COPD-specific version of the St. George Respiratory Questionnaire. *Chest* 2007;132:456–63.
- [11] Jones PW, Chen WH, Wilcox TK, et al. Characterizing and quantifying the symptomatic features of COPD exacerbations. *Chest* 2011;139:1388–94.
- [12] Bond T, Fox C. Applying the Rasch model: fundamental measurement in the human sciences. Mahway, New Jersey: Lawrence Erlbaum Associates; 2007.
- [13] Linacre J. What do infit and outfit, mean-square and standardized mean? *Rasch Meas Trans* 2002;16:878.
- [14] Sick J. Rasch measurement in language education, Part 5: assumptions and requirements of Rasch measurement. *SHIKEN – JALT Test Eval SIG Newsl* 2010;23–9.
- [15] Ainsworth BE, Haskell WL, Herrmann SD, et al. 2011 Compendium of physical activities: a second update of codes and MET values. *Med Sci Sports Exerc* 2011;43:1575–81.
- [16] Linacre J. The Rasch model cannot be “disproved”! *Rasch Meas Trans* 1996;10:512–4.
- [17] Hu F. Assessment of physical activity in nutritional epidemiology. In: Willett W, editor. *Nutritional epidemiology*. New York: Oxford University Press; 2013. pp. 241–59.