Journal of Clinical Epidemiology 165 (2024) 111203

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Equipercentile equating of scores from common patient-reported outcome measures of physical function in patients with cancer

Maria Rothmund[a,b], Micha J. Pilz[a], Lisa Schlosser[c], Juan I. Arraras[d], Mogens Groenvold[e,f], Bernhard Holzner[g], Marieke van Leeuwen[h], Morten Aa. Petersen[e], Heike Schmidt[i], Teresa Young[j], Matthias Rose[k], David Cella[l], Johannes M. Giesinger[a,*], on behalf of the EORTC Quality of Life Group

[a]*Department of Psychiatry, Psychotherapy, Psychosomatics, and Medical Psychology, University Clinic of Psychiatry II, Innsbruck Medical University, Anichstraße 35, A-6020 Innsbruck, Austria*

[b]*Institute of Psychology, University of Innsbruck, Innrain 52, A-6020 Innsbruck, Austria*

[c]*Data Lab Hell GmbH, Europastraße 2a, A-6170 Zirl, Austria*

[d]*Oncology Department, Hospital Universitario de Navarra, C/Irunlarrea 3, ES-31008 Pamplona, Spain*

[e]*Palliative Care Research Unit, Department of Geriatrics and Palliative Medicine GP, Bispebjerg & Frederiksberg Hospital, University of Copenhagen, Bispebjerg bakke 23B, Copenhagen, Denmark*

[f]*Department of Public Health, University of Copenhagen, Oester Farimagsgade 5, DK-1353 Copenhagen, Denmark*

[g]*Department of Psychiatry, Psychotherapy, Psychosomatics, and Medical Psychology, University Clinic of Psychiatry I, Innsbruck Medical University, Anichstraße 35, A-6020 Innsbruck, Austria*

[h]*Division of Psychosocial Research & Epidemiology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands*

[i]*University Clinic and Outpatient Clinic for Radiotherapy and Institute of Health and Nursing Science, Medical Faculty of Martin Luther University Halle-Wittenberg, Halle (Saale), Germany*

[j]*Supportive Oncology Research Team, East & North Hertfordshire NHS Trust incorporating Mount Vernon Cancer Centre, Northwood, Middlesex HA6 2RN, United Kingdom*

[k]*Center for Patient-Centered Outcomes Research (CPCOR), Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany*

[l]*Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, USA*

Accepted 24 October 2023; Published online 31 October 2023

Abstract

Objective: To provide equipercentile equating of physical function (PF) scores from frequently used patient-reported outcome measures (PROMs) in cancer patients to facilitate data pooling and comparisons.

Study Design and Setting: Adult cancer patients from five European countries completed the European Organization for Research and Treatment of Cancer (EORTC) computer adaptive test (CAT) Core, EORTC Quality of Life Questionnaire Version 3.0 (QLQ-C30), Functional Assessment of Cancer Therapy - General (FACT-G), 36-item Short Form Health Survey (SF-36), and the Patient-Reported Outcomes Measurement Information System (PROMIS) Physical Function 20a short form. The R package ‘‘equate’’ was used to establish conversion tables of PF scores on those measures with a bivariate rank correlation of at least 0.75.

Results: In total, 953 patients with cancer (mean age 58.9 years, 54.7% men) participated. Bivariate rank correlations between PF scores from the EORTC CAT Core, EORTC QLQ-C30, SF-36, and PROMIS were all above 0.85, but below 0.69 for the FACT-G. Conversion tables were established for all measures but the FACT-G. These tables indicate which score from one PROM best matches the score from another PROM and provide standard errors of converted scores.

Conclusion: Our analysis indicates that linking of PF scores from both EORTC measures (CAT and QLQ-C30) with PROMIS and SF-36 is possible, whereas the physical domain of the FACT-G seems to be different. The established conversion tables may be used for

Funding: This research was funded by the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Group (QLG), grant number 2016 011. The EORTC QLG business model involves license fees for commercial use of their instruments. Academic use of EORTC instruments is free of charge. If you are interested in using any of the EORTC quality of life instruments, please see: https://qol.eortc.org/questionnaires/

* Corresponding author. Department of Psychiatry, Psychotherapy, Psychosomatics, and Medical Psychology, University Hospital of Psychiatry II, Medical University of Innsbruck, Anichstrasse 35, A-6020 Innsbruck, Austria. Tel.: +43-512-504-23691.

E-mail address: johannes.giesinger@i-med.ac.at (J.M. Giesinger).

https://doi.org/10.1016/j.jclinepi.2023.10.019

0895-4356/© 2023 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

comparing results or pooling data from clinical studies using different PROMs. © 2023 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Keywords: Patient-reported outcome measures; Linking; Equipercentile equation; Physical functioning; Cancer; EORTC CAT; QLQ-C30; PROMIS; SF-36; FACT-G

1. Introduction

Patient-reported outcome measures (PROMs) have become a cornerstone of outcome assessment in cancer clinical trials [1,2]. PROMs provide a standard methodology for assessing health-related quality of life; a multidimensional construct including physical function (PF), role/social functioning, emotional functioning, and specific cancer symptoms and treatment side effects [3]. PF reflects the ability to pursue activities required for maintaining functional independence and independent living [3,4]. Guidance documents from regulatory authorities highlight patient-reported PF as a key outcome in cancer clinical trials [3] that has been shown to be more sensitive to treatment effects than clinician-reported performance status and to provide prognostic information on patient survival that goes beyond clinical or laboratory data [5].

The most frequently used PROM in cancer clinical trials and clinical practice [6−8] is the European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire Version 3.0 (QLQ-C30), a multidimensional questionnaire assessing PF alongside other functional domains and symptoms [9]. The recently released EORTC computer adaptive test (CAT) Core allows for the assessment of the same outcomes with more precision and an extended measurement range [10,11]. Other widely used PROMs, the 36-item Short Form Health Survey (SF-36) [12] and the Functional Assessment of Cancer Therapy - General (FACT-G) [13], contain subscales for PF, while the Patient-Reported Outcomes Measurement Information System (PROMIS) provides an item bank for PF [14]. These PF PROMs have been shown to be strongly correlated [15,16], despite some conceptual differences [17]. For example, the FACT-G physical well-being (PWB) subscale focuses on symptom burden rather than physical capacity, mobility, and activities of daily living [18].

The availability of various PROMs allows to select the most suitable measure for specific purposes. However, the variety in terms of content and metrics compromises comparisons of study results and data aggregation in meta-analyses. For example, a PF score of 50 points on the metric of one PROM does not equal to the same value from another PROM, as these values may indicate different levels of the measured construct. Thus, such scores cannot be compared directly or included in a pooled data analysis.

To overcome this limitation, linking methods have been introduced to convert scores from one PROM to the metric of another PROM [19] or to create a common metric [20]. Linking studies have been published for a number of commonly used PROMs in cancer medicine [20,21] but have not included the EORTC QLQ-C30 or EORTC CAT Core, which hinders comparing or pooling PRO data collected with these measures.

In an ongoing project by the EORTC Quality of Life Group, we evaluate the possibility of linking scores from EORTC measures with other common PROMs in cancer research. The project comprises a qualitative content comparison to assess conceptual overlap and quantitative analyses establishing and evaluating the actual linking algorithm. The content analysis of the included PF measures has been published previously [17]. In this article, we present linking algorithms for the PF domain of the following scales and item banks that were included in this analysis: the EORTC QLQ-C30 PF scale, the EORTC CAT PF item bank, the SF-36 PF scale, PROMIS PF v2.0 20a short form, the FACT-G PWB scale, and a combined scale FACT/PROMIS-PF4.

2. Methods

2.1. Sample

In this prospective study, we recruited a heterogeneous sample of patients having mixed cancer diagnoses from five European centers, the Medical University of Innsbruck (Austria), the Medical Faculty of Martin Luther University Halle-Wittenberg (Germany), the University Hospital of Navarra (Spain), the Netherlands Cancer Institute (Netherlands), and the Mount Vernon Cancer Centre (United Kingdom). Additional inclusion criteria were any treatment status (on-treatment and off-treatment), age $\geq$18 years, written informed consent, sufficient command of the primary language of the patient's country of residence, and no cognitive impairment interfering with questionnaire completion. Patients were approached during hospital visits or contacted via e-mail. PROMs were administered electronically using the software CHES [22]. In a subgroup of patients who completed the initial assessment at the start of chemotherapy, a second assessment was conducted at the third chemotherapy cycle (or earlier, in case of treatment discontinuation). All other patients completed the PROMs only at a single time point.

The study was approved by the ethics committee of the Medical University of Innsbruck (AN20140368) and conducted in accordance with Good Clinical Practice and the ethical principles of the Declaration of Helsinki.

What is new?

Key findings

- The present study investigates the possibilities to link patient-reported outcome measures that are commonly used to evaluate physical functioning in cancer patients. Equipercentile equation is used to establish linking tables for score conversion.
- Our analysis suggests that PF scores from the EORTC CAT Core / EORTC QLQ-C30, SF-36, and PROMIS can be linked with adequate precision for use in group-based comparisons.
- Quantitative analyses indicate that linking between these measures and the FACT-G seems not feasible. This supports the findings of a previous qualitative analysis that the FACT-G is conceptually distinct.

What this adds to what was known?

- Previous linking analysis on patient-reported outcome measures in oncology did not include the EORTC QLQ-C30 or EORTC CAT Core.

What is the implication and what should change now?

- Linking methods allow for conversion of scores between two measures. The equating tables may help to pool data from different studies when conducting meta-analyses.

2.2. Physical function measures

The PROM battery completed by the patients included the EORTC QLQ-C30 PF scale, the EORTC CAT PF item bank, the SF-36 PF scale, the PROMIS PF v2.0 20a short form, the FACT-G PWB scale, and a 4-item combination FACT/PROMIS-PF4. Full titles and short descriptions of all included PROMs are provided in Table 1.

2.3. Statistical analysis

Spearman rank correlation coefficients were calculated to investigate the strength of the bivariate correlation between the measures, and scatter plots are presented. Correlations exceeding 0.75 [35] were considered sufficient to link scores on group level. Furthermore, we calculated group standardized mean difference (SMD) for each measure for observed values based on sex, age (above/below 70 years), and treatment status (on/off treatment). The SMD was calculated by dividing the mean group difference by the pooled standard deviation (SD) of the two groups. In line with Dorans [36], a difference in SMD between measures (with equal reliability) that is larger than 0.11 indicates that group-specific equating procedures may be needed. Given the assumption of equal reliability, we expect comparisons of SMDs only to be meaningful for measures with comparable numbers of items. Unidimensionality of the measures has been evaluated using a confirmatory factor analysis.

To establish linking relationships between the included PROMs, we used the equipercentile equating approach which aligns scores from two measures with the same percentile ranks [37]. We do not provide linking tables to convert scores between the EORTC CAT Core and EORTC QLQ-C30, since conversion for these two scales can be done directly via the item response theory (IRT) model underlying the EORTC CAT Core [24]. For our analysis, we used the R package "equate" [38] which permits the definition of nonlinear relationships between scales with discrete scores using a cumulative distribution function approximated by percentile ranks. Polynomial log-linear presmoothing [39] was used to reduce the impact of sampling and measurement error. Standard errors were estimated via parametric bootstrapping with resampling based on the smoothed score distribution. Furthermore, the mean absolute error (MAE) was calculated as a measure of precision of the linking procedure [40]. To investigate variation in linking precision across patient groups, we also calculated the MAE separately for patient groups defined by sex, age (above/below 70 years), and treatment status (on/off treatment). Standardized MAE (SMAE) was calculated by dividing the MAE by the SD of the target measure in the total sample. Smaller MAEs and SMAEs indicate more precise score prediction, while large SMAEs reflect standard errors. SMAEs below 0.5 [41−43] were considered to indicate sufficiently invariant linking precision.

3. Results

3.1. Patient characteristics

The final sample consisted of 953 patients with different cancer diagnoses. The mean age was 58.9 years (SD = 13.4), and 54.7% were men. The most frequent diagnoses were lymphoma (11.9%), breast cancer (10.9%), and thyroid cancer (9.9%). At assessment, 73.2% of patients were receiving anticancer treatment. For further sample characteristics, see Table 2.

3.2. Descriptive statistics and bivariate correlations for PROMs

The EORTC CAT Core PF T-score ranged from 8.0 to 67.0, and the mean score was 44.2 (SD = 10.7). The scores on the EORTC QLQ-C30 PF scale ranged between 0 and 100, with a mean of 76.9 (SD = 22.0). The T-scores obtained from the PROMIS PF 20a short form ranged between 17.0 and 63.0, with a mean score of 45.9 (SD = 9.2). The SF-36 covered the whole 0−100 metric

Table 1. Short titles, full titles, and descriptions of all included patient-reported outcome measures

Short title	Full title	Description
EORTC CAT Core PF	Physical functioning item bank of the European Organization for Research and Treatment of Cancer Computer Adaptive Tests for Core domains — Physical Functioning item bank	The EORTC CAT Core [10,23] consists of 14 item banks that have been developed to measure the same functional health domains and symptoms as the EORTC QLQ-C30 v3.0 [9] with higher precision and an extended measurement range. The item bank for PF consists of 31 items, including the five PF items from the EORTC QLQ-C30 [24]. The EORTC measures use a four-point rating scale as the response format, with categories ranging from 'Not at all' to 'Very much,' without referring to a specific recall period.
EORTC QLQ-C30 PF	Physical functioning scale within the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire — Core module	Scores for the EORTC QLQ-C30 are calculated as sum scores from all items of a subscale, which are then transformed to a metric from 0 to 100 through linear regression [25]. Results from the EORTC CAT Core are based on a T-score metric with normative mean of 50 points and a standard deviation (SD) of 10 points. The underlying normative data were collected from the European general population [26]. The item banks can be used for computer-adaptive assessments [24] or to create static short forms that target specific patient populations [27], with both providing scores that are fully compatible with those from the original EORTC QLQ-C30. In this study, we asked patients all 31 items in the PF item bank.
PROMIS PF 20a (v2.0)	Patient-Reported Outcomes Measurement Information System, Short form for Physical Functioning with 20 items (version 2.0)	The PROMIS Physical Function 20a [28] is a generic short form based on the PROMIS Physical Function Item Bank v2.0 [29]. Six items from the PROMIS Cancer Item Bank v1.1 are included in the Physical Function 20a short form. No recall period is specified in the short form. The questions are answered on a five-point Likert scale ranging from 'Not at all' to 'Cannot do' or 'Without any difficulty' to 'Unable to do.' Scores are given on a T-score metric based on the general population of the United States [30].
SF-36 PF	Short Form 36, Physical Function subscale	The SF-36 is a 36-item HRQOL measure [12,31] that comprises eight individual domains including PF. The PF scale is conceptualized as the 'performance of or capacity to perform a variety of activities that are normal for an individual in good health' [32], including self-care, mobility, and physical activities. The questionnaire items assess limitations of physical activities and functioning owing to health conditions. The PF domain comprises 10 items rated on three-point rating scale (response categories: 'Yes, limited a lot,' 'Yes, limited a little,' and 'No, not limited at all') and no specific recall period. Please note that in this analysis, we refer to the PF scale and not to the physical component summary score. Scores for the SF-36 PF scale are given on a 0–100 metric and calculated through linear transformation of the sum of the item scores.
FACT-G PWB	Functional Assessment of Cancer Therapy — General, Physical Well-being subscale	The FACT-G [33] is a cancer-specific questionnaire comprising four domains, including a physical well-being scale (FACT-G PWB). The 27 items on the FACT-G version 4.0 [13] are rated on a five-point Likert scale ranging from 'Not at all' to 'Very much,' referring to a 7-day recall period. The PWB scale consists of seven items assessing fatigue, pain, and nausea, two items on treatment side effects and feeling ill in general, an item on being bedbound, and one item on the interference with patients' family life. Scores on the FACT-G are calculated as sum scores. The FACT-G PWB score thus ranges between 0 and 28.
FACT/PROMIS-PF4	Ad-hoc scale with two items from the FACT-G PWB and PROMIS PF 20a (v2.0)	To improve linking with PROMIS, Kaat et al. [34] suggested a combined FACT/PROMIS-PF5 scale, which consists of two FACT-G items ('trouble meeting the needs of my family,' 'spend time in bed') and three PROMIS items (PFA11 'chores,' PFA53 'errands,' PFA1 'vigorous activities'). The PROMIS item PFA53 was not part of our assessment, but our data allowed to approximate the scale suggested by Kaat et al. [34] by combining the remaining four items, hereafter FACT/PROMIS-PF4, to estimate linking options for this combined scale. Sum scores for this scale range from 0 to 16.

Table 2. Characteristics of the study sample (N = 953)

Variable	N	%
Sex		
Female	432	45.3
Male	521	54.7
Age		
Below 70 yr	752	78.9
70 yr or above	201	21.1
Diagnoses		
Lymphoma	113	11.9
Breast cancer	104	10.9
Thyroid cancer	94	9.9
Naso-/oropharyngeal	89	9.3
Colorectal cancer	79	8.3
Lung cancer	72	7.6
Prostate cancer	70	7.3
Leukemia	50	5.2
Gynecological cancers	46	4.8
Pancreas cancer	36	3.8
Neuroendocrine tumor	34	3.6
Gastric cancer	20	2.1
Bladder cancer	15	1.6
Skin cancer	15	1.6
Anal cancer	13	1.4
Other solid tumors	77	8.1
Other hematological malignancies	26	2.7
Treatment status		
On treatment	698	73.2
Off treatment	255	26.8
Marital status		
Single	140	14.8
Married/partnership	666	70.4
Separated/divorced/widowed	140	14.8
Missing	7	
Highest level of education		
Less than compulsory	19	2.0
Compulsory	273	29.0
Post compulsory but below university	436	46.3
University level or above	213	22.6
Missing	12	
Employment status		
Retired	427	45.4
Employed full time	245	26.0
Employed part time	85	9.0
Self-employed	68	7.2
Unemployed	32	3.4
Homemaker/parent/carer	32	3.4
Student/in training	14	1.5
Other	38	4.0
Missing	12	
Country		
Amsterdam (NL)	44	4.6
Innsbruck (AT)	448	47.0
London (UK)	182	19.1
Pamplona (ES)	91	9.5
Halle (DE)	188	19.7

with a mean score of 71.0 (SD = 25.9). Our sample reported the whole range on the FACT-G PWB, from 0 to 28, while the mean score was 20.9 (SD = 6.2). Scores on the FACT/PROMIS-PF4 ranged between 0 and 16, with a mean of 11.3 (SD = 3.9). The results from the confirmatory factor analysis evaluating the unidimensionality of the measures are shown in Supplementary File 3.

The highest Spearman rank correlation was found between the two EORTC measures: 0.94 (95% confidence interval [95% CI] 0.93−0.94). Other correlations of the EORTC CAT Core instrument were 0.90 (95% CI 0.88−0.91) with the PROMIS PF 20a short form, 0.89 (95% CI 0.87−0.90) with the SF-36 PF scale, and 0.65 (95% CI 0.60−0.69) with the FACT-G PWB scale. Comparable correlations were found between the EORTC QLQ-C30 PF scale and PROMIS (0.86; 95% CI 0.84−0.88), SF-36 (0.85; 95% CI 0.03−0.87), and the FACT-G PWB (0.67; 95% CI 0.63−0.71). Further bivariate correlations were 0.90 (95% CI 0.89−0.91) between the SF-36 and the PROMIS measure, 0.69 (95% CI 0.65−0.73) for the FACT-G PWB with the SF-36, and 0.69 (95%CI 0.66−0.73) for the FACT-G with the PROMIS measure. Scatter plots detailing the bivariate associations are shown in Fig. 1.

In comparison to the FACT-G PWB scale, the FACT/PROMIS-PF4 showed higher correlations with the comparator measures, that is, 0.89 (95% CI 0.87−0.90) with PROMIS, 0.81 (95% CI 0.79−0.84) with the SF-36, and 0.80 (95% CI 0.77−0.82) with the EORTC CAT Core as well as the QLQ-C30. The correlation between FACT-G PWB and FACT/PROMIS-PF4 was 0.82 (95% CI 0.80−0.85).

3.3. Standardized mean differences for age, sex, and treatment status

SMDs in the comparison between men and women ranged from 0.07 (QLQ-C30) to 0.14 (EORTC CAT Core and PROMIS) and 0.15 (FACT-G, SF-36, and FACT/PROMIS-PF4). Comparing patients above and below the age of 70 years showed the following SMDs: FACT-G −0.03, FACT/PROMIS-PF4 0.10, QLQ-C30 0.12, SF-36 0.17, PROMIS 0.18, and EORTC CAT Core 0.29. Differences between patients on and off treatment showed SMDs of 0.48 for the FACT-G, 0.52 for the QLQ-C30 and the SF-36, 0.54 for the EORTC CAT Core, 0.55 for PROMIS, and 0.60 for the FACT/PROMIS-PF4. Further details are reported in Supplementary File 3.

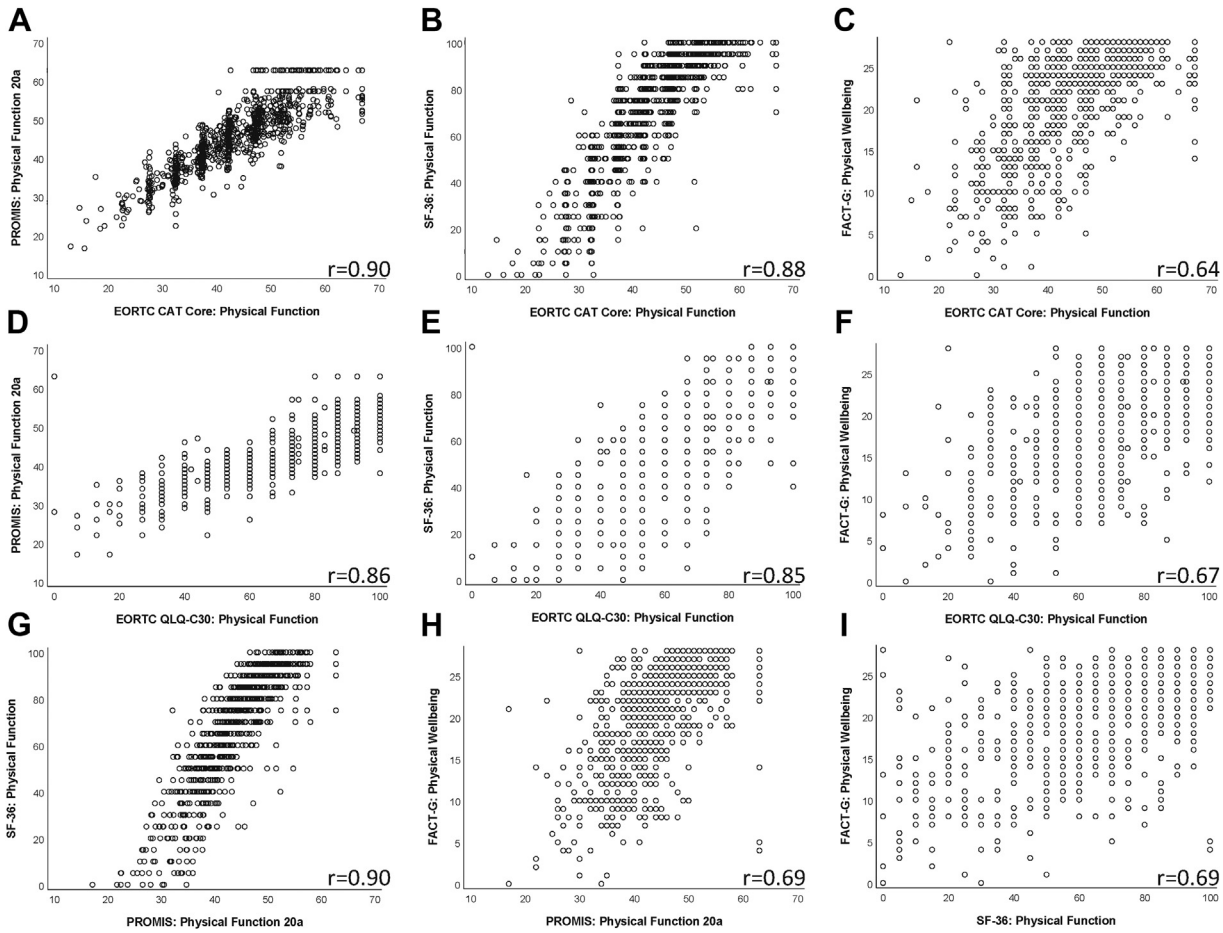

Fig. 1. Scatterplot and rank correlation for the bivariate associations of the EORTC CAT Core Physical Functioning item bank (A–C) and the EORTC QLQ-C30 Physical Functioning scale (D–F) with the three comparator measures, as well as the bivariate correlations between the SF-36 Physical Function subscale, the PROMIS Physical Function short form 20a (v2.0), and the FACT-G Physical Well-being subscale (G–I). EORTC, European Organization for Research and Treatment of Cancer; CAT, computer adaptive test; QLQ, Quality of Life Questionnaire; PROMIS, Patient-Reported Outcomes Measurement Information System; SF-36, 36-item Short Form Health Survey; FACT-G, Functional Assessment of Cancer Therapy - General.

All differences in SMDs across measures and groups were below or close to the predefined threshold of 0.11 [36], with the exception of SMDs in the age group comparisons that showed differences in SMDs of up to 0.32 (FACT-G −0.03 vs. EORTC CAT Core 0.29). In this comparison, those scales with more items and, thus, presumably higher reliability showed higher SMDs than scales with fewer items. Since variation in reliability compromises the interpretation of differences in SMDs, we do not consider these results as sufficient evidence for the need for group-specific linking.

3.4. Linking using equipercentile equating

The difference between the observed mean scores and the mean scores predicted from the comparator measures was always less than one point on the metric of the measures. SDs were also very similar for observed and predicted scores. Details are shown in Table 3.

MAEs and SMAEs for the overall sample, as well as age groups, sex, and treatment status, are presented in Table 4. All SMAEs for conversions between the EORTC measures, PROMIS, and the SF-36 were below 0.5, except for the prediction of PROMIS scores based on the QLQ-C30 PF scale (SMAE = 0.54) in the subgroup of patients off treatment. In contrast, predictions of EORTC, PROMIS, or SF-36 scores based on the FACT-G mostly showed SMAE values above 0.5; SMAEs for conversions based on or predicting FACT/PROMIS-PF4 scores were larger than 1.0.

Details on the variation in standard errors across the measurement continuum of the PROMs are shown in Figures 2–4. Score conversion tables are shown in Supplementary File 1 for the EORTC CAT Core, QLQ-C30, PROMIS, and SF-36. The FACT-G was not included owing to the low bivariate correlations. Conversion tables for the FACT/PROMIS-PF4 are shown in Supplementary File 2.

To use the conversion tables, the observed scores from the measure to be converted are rounded to the nearest integer. If this integer is, for example, 50 points on the T-score metric of the PROMIS measure, the predicted score on the T-score metric of the EORTC CAT Core would be

Table 3. Descriptive statistics for observed and predicted scores for the PRO measures

PRO measure	Mean	SD
EORTC CAT Core PF		
Observed	44.23	10.68
Predicted by PROMIS PF 20a SF	44.23	10.67
Predicted by SF-36 Physical Function	44.73	11.76
Predicted by FACT-G Physical Well-being	44.01	10.15
Predicted by FACT/PROMIS-PF4	44.36	10.25
EORTC QLQ-C30 PF		
Observed	76.94	21.99
Predicted by PROMIS PF 20a SF	76.83	21.68
Predicted by SF-36 Physical Function	76.85	22.15
Predicted by FACT-G Physical Well-being	77.04	22.06
Predicted by FACT/PROMIS-PF4	77.27	21.07
PROMIS PF 20a SF		
Observed	45.89	9.20
Predicted by EORTC CAT Core PF	45.89	9.20
Predicted by EORTC QLQ-C30 PF	46.87	10.76
Predicted by SF-36 Physical Function	46.29	10.09
Predicted by FACT-G Physical Well-being	45.72	8.77
Predicted by FACT/PROMIS-PF4	46.00	8.81
SF-36 Physical Function		
Observed	71.00	25.86
Predicted by EORTC CAT Core PF	70.90	25.32
Predicted by EORTC QLQ-C30 PF	71.13	25.85
Predicted by PROMIS PF 20a SF	70.88	25.48
Predicted by FACT-G Physical Well-being	71.06	25.96
Predicted by FACT/PROMIS-PF4	71.27	25.07
FACT-G Physical Well-being		
Observed	20.90	6.23
Predicted by EORTC CAT Core PF	20.90	6.09
Predicted by EORTC QLQ-C30 PF	20.98	6.34
Predicted by PROMIS PF 20a SF	20.89	6.15
Predicted by SF-36 Physical Function	20.91	6.33
Predicted by FACT/PROMIS-PF4	21.02	5.90
FACT/PROMIS-PF4		
Observed	11.30	3.89
Predicted by EORTC CAT Core PF	11.30	3.99
Predicted by EORTC QLQ-C30 PF	10.73	4.87
Predicted by PROMIS PF 20a SF	11.29	4.00
Predicted by SF-36 Physical Function	11.06	4.40
Predicted by FACT-F Physical Well-being	11.37	3.71

Abbreviations: PRO, patient-reported outcome; SD, standard deviation; EORTC, European Organization for Research and Treatment of Cancer; CAT, computer adaptive test; QLQ, Quality of Life Questionnaire; PF, physical function; PROMIS, Patient-Reported Outcomes Measurement Information System; SF-36, 36-item Short Form Health Survey; FACT-G, Functional Assessment of Cancer Therapy - General.

48.86 points (standard error 0.54 points). In another example, the maximum value of 100 points on the metric of the SF-36 PF scale is converted to 67.09 points (standard error 0.77 points) on the EORTC CAT Core metric.

4. Discussion

Our study showed high correlations between the PF domains of the EORTC CAT Core, EORTC QLQ-C30, PROMIS, and SF-36 suggesting that scores from these measures can be linked to each other. The correlations of the FACT-G PWB scale with these three scales were substantially lower. SMAEs were lower for conversions between EORTC CAT Core, EORTC QLQ-C30, PROMIS, and SF-36 (all but one below 0.5 across all subgroups analyzed), compared to the FACT-G or the FACT/PROMIS-PF4 scale.

In a previous qualitative content analysis [17], we investigated the item content of these measures using the international classification of functioning (ICF) framework [44] with an established analytical approach [45]. We found that most of the items from the EORTC CAT Core, SF-36, PROMIS PF 20a short form, and FACT-G could be assigned to ICF categories within the ICF component 'Activities and Participation.' For all PROMs but the FACT-G, the first-level category 'Mobility' covered more than half of the item content, followed by 'Self-care' for the EORTC, SF-36, and PROMIS 20a short form. For the cancer-specific PROMIS item bank, 'Domestic life' was the second most frequently used first-level category. The FACT-G PWB was the most heterogeneous measure with items related to the components 'Activities and Participation' and 'Body Functions' and additional content that could not be categorized within the ICF framework. These results suggested that there is conceptually sufficient overlap in content to allow for meaningful linking of scores from the EORTC CAT Core, SF-36, and PROMIS, whereas the FACT-G is conceptually distinct [17,18].

The quantitative analysis in this article supports these findings on (dis)similarity from the qualitative content analysis. Correlations between the EORTC QLQ-C30, the EORTC CAT Core, SF-36, and PROMIS ranged between 0.85 and 0.90 (0.94 between the two EORTC measures), clearly exceeding the threshold of 0.75 that has been applied previously for identifying scales where group-level linking is possible [35,46]. In fact, the observed correlations of the EORTC CAT Core with the SF-36 and PROMIS were even above the threshold for individual-level linking of 0.866 suggested by Dorans [36]. Also, the correlations between the EORTC QLQ-C30 PF scale with these two comparator measures were close to this value. In contrast, correlations for the FACT-G PWB with the other measures ranged from 0.64 to 0.69.

Regarding the need for age-, sex-, and treatment status-specific linking procedures, the SMDs indicated that these

Table 4. Mean absolute error and standardized mean absolute error of the linking procedures

	Total sample		Age <70		Age ≥70		Men		Women		On treatment		Off treatment	
Direction of score conversion	**MAE**	**SMAE**	**MAE**	**SMAE**	**MAE**	**SMAE**	**MAE**	**SMAE**	**MAE**	**SMAE**	**MAE**	**SMAE**	**MAE**	**SMAE**
EORTC CAT Core PF														
Predicted by PROMIS PF 20a SF	3.82	0.36	3.88	0.37	3.48	0.33	3.89	0.37	3.67	0.35	3.83	0.36	3.84	0.36
Predicted by SF-36 Physical Function	4.53	0.43	4.71	0.44	3.79	0.36	4.80	0.45	4.20	0.40	4.33	0.41	5.03	0.47
Predicted by FACT-G Physical Well-being	7.00	0.66	7.24	0.68	5.82	0.55	7.22	0.68	6.68	0.63	6.98	0.66	6.99	0.66
Predicted by FACT/PROMIS-PF4	16.04	1.51	16.39	1.54	14.37	1.35	16.94	1.60	14.84	1.40	15.30	1.44	15.52	1.46
EORTC QLQ-C30 PF														
Predicted by PROMIS PF 20a SF	7.85	0.36	7.83	0.36	7.78	0.35	7.72	0.35	7.97	0.36	8.36	0.38	6.30	0.29
Predicted by SF-36 Physical Function	8.38	0.38	8.38	0.38	8.38	0.38	8.49	0.39	8.05	0.37	8.96	0.41	6.78	0.31
Predicted by FACT-G Physical Well-being	12.46	0.57	12.38	0.56	12.54	0.57	12.82	0.58	11.94	0.54	13.43	0.61	9.69	0.44
Predicted by FACT/PROMIS-PF4	32.96	1.50	32.94	1.50	32.96	1.50	33.09	1.51	32.62	1.48	34.13	1.55	23.67	1.08
PROMIS PF 20a S														
Predicted by EORTC CAT Core PF	3.28	0.36	3.34	0.36	3.01	0.33	3.15	0.34	3.40	0.37	3.23	0.35	3.42	0.37
Predicted by EORTC QLQ-C30 PF	4.26	0.47	4.40	0.48	3.68	0.40	4.25	0.46	4.25	0.46	3.98	0.43	4.95	0.54
Predicted by SF-36 Physical Function	3.52	0.38	3.66	0.40	3.01	0.33	3.43	0.37	3.61	0.39	3.29	0.36	4.15	0.45
Predicted by FACT-G Physical Well-being	5.53	0.60	5.63	0.61	4.79	0.52	5.44	0.59	5.66	0.62	5.44	0.59	5.64	0.62
Predicted by FACT/PROMIS-PF4	14.00	1.53	14.44	1.58	12.28	1.34	13.76	1.50	14.28	1.56	13.20	1.44	14.33	1.56
SF-36 Physical Function														
Predicted by EORTC CAT Core PF	8.69	0.34	8.73	0.34	8.48	0.33	8.47	0.33	8.83	0.34	9.20	0.36	7.25	0.28
Predicted by EORTC QLQ-C30 PF	9.96	0.38	9.95	0.38	9.98	0.39	9.76	0.38	10.02	0.39	10.50	0.41	8.56	0.33
Predicted by PROMIS PF 20a SF	8.10	0.31	7.95	0.31	8.89	0.34	7.96	0.31	8.23	0.32	8.66	0.33	6.54	0.25
Predicted by FACT-G Physical Well-being	14.87	0.57	14.86	0.57	14.45	0.56	14.18	0.55	15.70	0.61	16.26	0.63	11.02	0.43
Predicted by FACT/PROMIS-PF4	39.31	1.52	39.25	1.52	39.23	1.52	37.66	1.46	41.17	1.59	40.68	1.57	28.80	1.11
FACT-G Physical Well-being														
Predicted by EORTC CAT Core PF	3.79	0.61	3.84	0.61	3.57	0.57	3.60	0.58	3.98	0.64	4.10	0.66	2.97	0.48
Predicted by EORTC QLQ-C30 PF	3.70	0.59	3.70	0.59	3.63	0.58	3.65	0.58	3.70	0.59	3.91	0.63	3.15	0.50
Predicted by PROMIS PF 20a SF	3.54	0.57	3.52	0.56	3.54	0.57	3.44	0.55	3.67	0.59	3.82	0.61	2.76	0.44
Predicted by SF-36 Physical Function	3.66	0.59	3.66	0.59	3.54	0.57	3.51	0.56	3.84	0.61	3.99	0.64	2.80	0.45
Predicted by FACT/PROMIS-PF4	9.28	1.48	9.20	1.47	9.59	1.53	9.06	1.45	9.53	1.52	9.75	1.56	6.48	1.04
FACT/PROMIS-PF4														
Predicted by EORTC CAT Core PF	6.01	1.54	5.96	1.52	6.26	1.60	5.88	1.50	6.13	1.57	5.93	1.52	4.78	1.22
Predicted by EORTC QLQ-C30 PF	6.61	1.69	6.60	1.69	6.63	1.70	6.46	1.65	6.72	1.72	6.39	1.63	5.76	1.47
Predicted by PROMIS PF 20a SF	6.13	1.57	6.12	1.57	6.20	1.59	5.93	1.52	6.36	1.63	6.06	1.55	5.04	1.29
Predicted by SF-36 Physical Function	6.27	1.60	6.26	1.60	6.33	1.62	6.06	1.55	6.46	1.65	6.15	1.57	5.40	1.38
Predicted by FACT-G Physical Well-being	5.93	1.52	5.88	1.51	6.10	1.57	5.82	1.50	6.04	1.55	6.00	1.54	4.36	1.12

Abbreviations: EORTC, European Organization for Research and Treatment of Cancer; CAT, computer adaptive test; QLQ, Quality of Life Questionnaire; PF, physical function; PROMIS, Patient-Reported Outcomes Measurement Information System; SF-36, 36-item Short Form Health Survey; FACT-G, Functional Assessment of Cancer Therapy - General; MAE, mean absolute error; SMAE, standardized mean absolute error (i.e., mean absolute error divided by the standard deviation of the observed scores in the total sample).

may not be needed for the EORTC CAT Core, EORTC QLQ-C30, SF-36, and PROMIS, whereas for other scales the results were inconclusive as larger differences in SMDs may be a result of varying reliability.

Recently, a study by Lee et al. [46] comparing the five-item PF scale of the EORTC QLQ-C30 and the four-item PF scale of the PROMIS-29 highlighted the similar conceptualization of the domains in these measures. Content differences were found to only reflect variations in difficulties of the tasks assessed. The two PF scales showed a correlation of 0.78 which is somewhat lower than the correlation of 0.86 observed in our study. This difference may reflect a difference in reliability of the assessments given the lower number of PF items in their study using

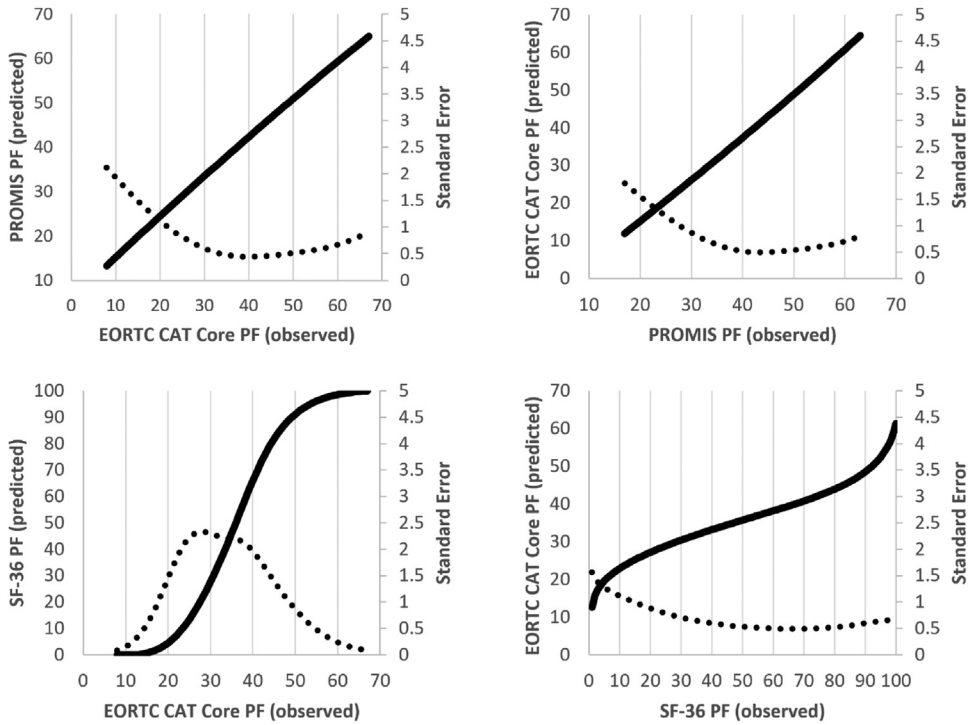

Fig. 2. Linking functions (bold lines) and standard errors (dotted lines) to predict the PROMIS and SF-36 physical functioning scores from the EORTC CAT Core Physical Functioning item bank and vice versa. EORTC, European Organization for Research and Treatment of Cancer; CAT, computer adaptive test; PROMIS, Patient-Reported Outcomes Measurement Information System; SF-36, 36-item Short Form Health Survey.

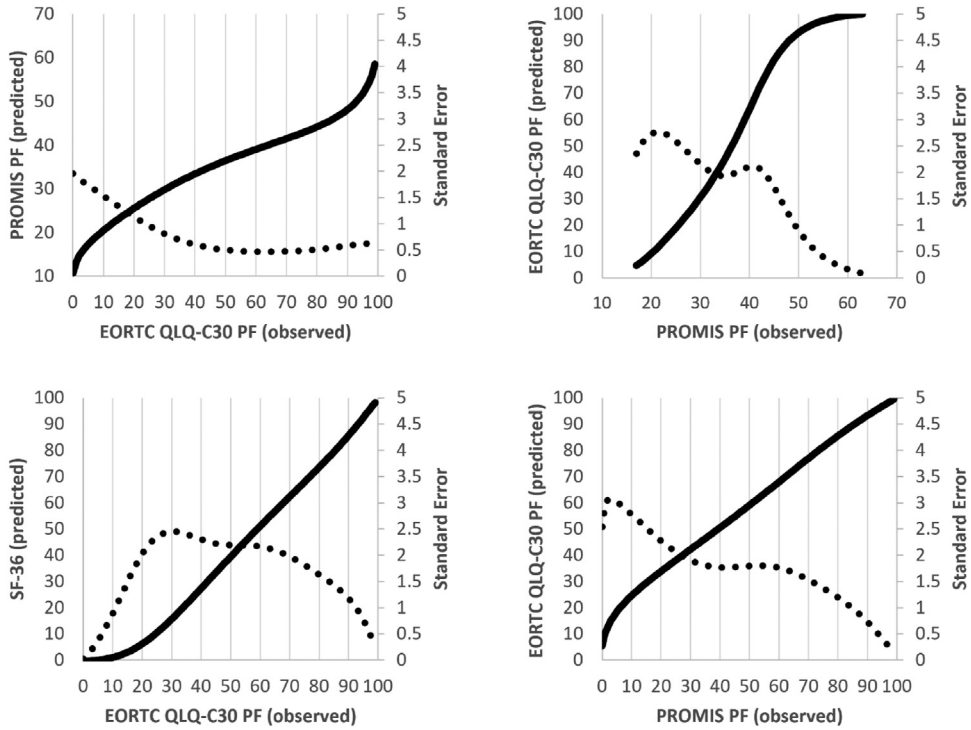

Fig. 3. Linking functions (bold lines) and standard errors (dotted lines) to predict the PROMIS and SF-36 physical functioning scores from the EORTC QLQ-C30 Physical Functioning scale and vice versa. EORTC, European Organization for Research and Treatment of Cancer; QLQ, Quality of Life Questionnaire; PROMIS, Patient-Reported Outcomes Measurement Information System; SF-36, 36-item Short Form Health Survey.

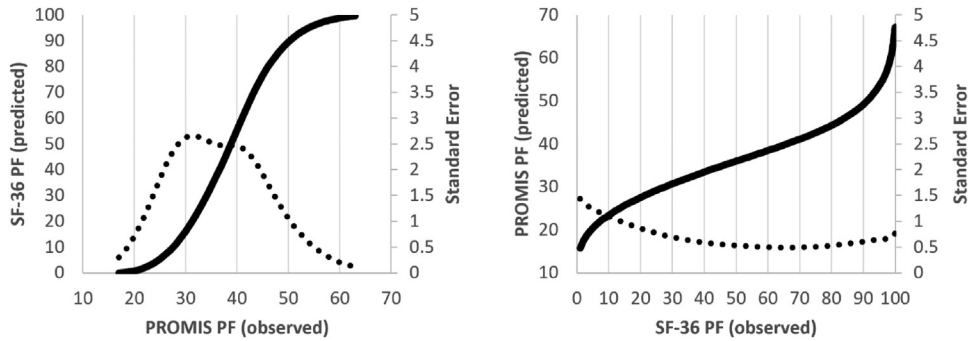

Fig. 4. Linking functions (bold lines) and standard errors (dotted lines) for equipercentile equation between PROMIS and SF-36 Physical Functioning scores. PROMIS, Patient-Reported Outcomes Measurement Information System; SF-36, 36-item Short Form Health Survey.

the 4-item short form by PROMIS compared to ours using the 20-item short form.

Whereas our analysis covers several widely used PF measures among patients with cancer, this selection is still limited, particularly regarding PROMIS. Here, we only assessed a single short form, albeit the one with the most items. Because all PROMIS short form and CATs based on the same item bank provide scores on the same metric, and the same is true for the EORTC short form and CATs, our conversion tables are applicable for score conversion between any measures based on these item banks. The selection of the short form was informed by considerations on response burden which only allowed the inclusion of a certain number of items. At the same time, the use of a single-group design in which all patients completed all measures at the same time point is considered optimal for linking analyses [36].

Another limitation is the imbalanced sample which only allowed for certain analysis of possible variation in linking precision across patient groups. This imbalance extends to the distribution of patients across centers and countries. Nonetheless, the coverage of different countries and languages goes beyond most previous linking studies, and score distributions indicated good coverage of the measurement range of the investigated PROMs.

Whereas standard errors of predicted scores are reasonably small in the mid-range of the score distribution, extreme scores can only be linked with very limited precision. Therefore, linking scores in patient populations with high performance levels, such as young cancer survivors, and those with very poor levels, such as palliative care patients, is not recommended with the current conversion tables.

Whereas there are other analytical techniques available for this purpose, such as IRT or regression models, we think that equipercentile equating provides some important advantages over more complex methods. Most importantly, this method relies on scale-level data rather than item-level data which makes it more applicable in situations where data with individual response sets are not available. This is mostly the case for meta-analyses or other analysis comparing or merging results from different PROMs. Next, all PROMs have their own scoring algorithms; these can be simple, such as for the EORTC QLQ-C30 PF, FACT-G PWB, and SF-36 PF scales, or more complex, relying on IRT models and maximum likelihood estimation like the EORTC CAT Core and PROMIS measures. IRT-based linking methods require a joint model that comprises items from different PROMs; this may not fit the same unidimensional model or may only do so if the item characteristics (e.g., item difficulty and slope) of the original item bank are altered. Such alterations of item characteristics, however, will result in incompatibility of the official and commonly used scoring algorithms and the scores obtained from the linking model. Such incompatibilities are avoided when equipercentile equating is applied. Furthermore, previous studies found that differences in precision obtained by different linking methods are mostly small [47–49] whereas the strength of the correlation can have a substantial impact [19].

In conclusion, our analysis suggests that PF scores from the EORTC CAT Core/EORTC QLQ-C30, SF-36, and PROMIS can be linked with adequate precision for use in group-based comparisons. Conceptually, linking is supported by previous qualitative findings [17] suggesting conceptual overlap among these measures. The FACT-G PWB scale is not only conceptually distinct, but it is also statistically less associated with these other scales, suggesting that linking to the other measures may not be meaningful. The equating tables from our analysis may be used for linking PF scores from the EORTC CAT Core/EORTC QLQ-C30, SF-36, and PROMIS. This may be used, for example, to pool data from different studies or conduct meta-analyses. Furthermore, our results inform future studies on linking procedures by highlighting possible variations in linking precision across patient groups; this warrants further investigation.

CRediT authorship contribution statement

Conceptualization: Giesinger, Petersen, Groenvold. Data curation: Rothmund, Giesinger. Formal analysis: Schlosser, Rothmund, Giesinger. Funding acquisition: Giesinger. Data collection: Arraras, van Leeuwen, Schmidt, Young, Pilz, Holzner. Project administration: Pilz, Giesinger. Supervision: Giesinger, Cella, Rose. Visualization: Rothmund. Writing—original draft: Rothmund, Giesinger. Writing—review and editing: Rothmund, Pilz, Schlosser,

Arraras, Groenvold, Holzner, van Leeuwen, Petersen, Schmidt, Young, Rose, Cella, Giesinger.

Data availability

The authors do not have permission to share data.

Declaration of competing interest

David Cella reports to hold a compensated position as President of FACIT.org and an uncompensated position as President of the PROMIS Health Organization and member of its Board of Directors. Bernhard Holzner reports holding intellectual property rights on the software tool CHES. All other coauthors declared no conflicts of interest.

Acknowledgments

We would like to thank Anna Thurner and Imad Al-Naesan for their help with data collection, and we are grateful to Tobias Hell, PhD (Data Lab Hell GmbH; tobias.hell@datalabhell.at) for his support in the statistical analysis. We thank Analisa Avila, MPH, ELS, of Edanz (www.edanz.com/ac) for editing a draft of this manuscript. Furthermore, we would like to express our gratitude to all patients who completed the comprehensive questionnaire assessments required for this study.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jclinepi.2023.10.019.

References

[1] European Medicines Agency. Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man: the use of patient-reported outcome (PRO) measures in oncology studies. London: European Medicines Agency; 2016.

[2] Food and Drug Administration. Principles for selecting, developing, modifying, and adapting patient reported outcome instruments for use in medical device evaluation: guidance for industry and food and drug administration staff, and other stakeholders. Silver Spring: Food and Drug Administration; 2022.

[3] Kluetz PG, Slagle A, Papadopoulos EJ, Johnson LL, Donoghue M, Kwitkowski VE, et al. Focusing on core patient-reported outcomes in cancer clinical trials: symptomatic adverse events, physical function, and disease-related symptoms. Clin Cancer Res 2016;22:1553−8.

[4] Garber CE, Greaney ML, Riebe D, Nigg CR, Burbank PA, Clark PG. Physical and mental health-related correlates of physical function in community dwelling older adults: a cross sectional study. BMC Geriatr 2010;10:6.

[5] Brown JC, Harhay MO, Harhay MN. Physical function as a prognostic biomarker among cancer survivors. Physical function as a prognostic biomarker among cancer survivors. Br J Cancer 2015;112:194−8.

[6] Giesinger JM, Efficace F, Aaronson N, Calvert M, Kyte D, Cottone F, et al. Past and current practice of patient-reported outcome measurement in randomized cancer clinical trials: a systematic review. Value Health 2021;24:585−91.

[7] Smith AB, Cocks K, Parry D, Taylor M. Reporting of health-related quality of life (HRQOL) data in oncology trials: a comparison of the European organization for research and treatment of cancer quality of life (EORTC QLQ-C30) and the functional assessment of cancer therapy-general (FACT-G). Qual Life Res 2014;23:971−6.

[8] Blazeby JM, Avery K, Sprangers M, Pikhart H, Fayers P, Donovan J. Health-related quality of life measurement in randomized clinical trials in surgical oncology. J Clin Oncol 2006;24:3178−86.

[9] Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European organization for research and treatment of cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993;85:365−76.

[10] Petersen MA, Aaronson NK, Arraras JI, Chie WC, Conroy T, Costantini A, et al. European organisation for research and treatment of cancer (EORTC) quality of life group. The EORTC CAT core-the computer adaptive version of the EORTC QLQ-C30 questionnaire. Eur J Cancer 2018;100:8−16.

[11] Vachon H, Petersen MA, Pe M, Reijneveld JC, Groenvold M. Another look at floor and ceiling effects in the EORTC QLQ-C30 physical functioning subscale and possible solutions. Cancer 2022;128:2384−5.

[12] Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Medical Care 1992;60:473−83.

[13] Webster K, Cella D, Yost K. The functional assessment of chronic illness therapy (FACIT) measurement system: properties, applications, and interpretation. Health Qual Life Outcomes 2003;1:79.

[14] Jensen RE, Potosky AL, Reeve BB, Hahn E, Cella D, Fries J, et al. Validation of the PROMIS physical function measures in a diverse US population-based cohort of cancer patients. Qual Life Res 2015;24:2333−44.

[15] Kuenstner S, Langelotz C, Budach V, Possinger K, Krause B, Sezer O. The comparability of quality of life scores. a multitrait multimethod analysis of the EORTC QLQ-C30, SF-36 and FLIC questionnaires. Eur J Cancer 2002;38:339−48.

[16] Hartmann C, Fischer F, Klapproth CP, Röhle R, Rose M, Karsten MM. PROMIS-29 and EORTC QLQ-C30: an empirical investigation towards a common conception of health. Qual Life Res 2023;32:749−58.

[17] Schurr T, Loth F, Lidington E, Piccinin C, Arraras JI, Groenvold M, et al. European organisation for research, treatment of cancer quality of life group (EORTC QLG). Patient-Reported outcome measures for physical function in cancer patients: content comparison of the EORTC CAT core, EORTC QLQ-C30, SF-36, FACT-G, and PROMIS measures using the international classification of functioning, disability and health. BMC Med Res Methodol 2023;23:21.

[18] Luckett T, King MT, Butow PN, Oguchi M, Rankin N, Price MA, et al. Choosing between the EORTC QLQ-C30 and FACT-G for measuring health-related quality of life in cancer clinical research: issues, evidence and recommendations. Ann Oncol 2011;22:2179−90.

[19] Schalet BD, Lim S, Cella D, Choi SW. Linking scores with patient-reported health outcome instruments: a VALIDATION study and comparison of three linking methods. Psychometrika 2021;86:717−46.

[20] Schalet BD, Revicki DA, Cook KF, Krishnan E, Fries JF, Cella D. Establishing a common metric for physical function: linking the HAQ-DI and SF-36 PF subscale to PROMIS(®) physical function. J Gen Intern Med 2015;30:1517−23.

[21] Holzner B, Bode RK, Hahn EA, Cella D, Kopp M, Sperner-Unterweger B, et al. Equating EORTC QLQ-C30 and FACT-G scores and its use in oncological research. Eur J Cancer 2006;42:3169−77.

[22] Evaluation Software Development GmbH. Computer based health evaluation system - CHES. Innsbruck. Available at www.ches.pro. Accessed November 28, 2023.

[23] Petersen MA, Aaronson NK, Conroy T, Costantini A, Giesinger JM, Hammerlid E, et al. European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Group. International validation of the EORTC CAT Core: a new adaptive instrument for measuring core quality of life domains in cancer. Qual Life Res 2020;29:1405−17.
[24] Petersen MA, Groenvold M, Aaronson NK, Chie WC, Conroy T, Costantini A, et al. Development of computerized adaptive testing (CAT) for the EORTC QLQ-C30 physical functioning dimension. Qual Life Res 2011;20:479−90.
[25] Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D. Bottomley A, on behalf of the EORTC Quality of Life Group. The EORTC QLQ-C30 Scoring Manual. 3rd Edition. Brussels: EORTC; 2001.
[26] Liegl G, Petersen MA, Groenvold M, Aaronson NK, Costantini A, Fayers PM, et al, EORTC Quality of Life Group. Establishing the European Norm for the health-related quality of life domains of the computer-adaptive test EORTC CAT Core. Eur J Cancer 2019;107:133−41.
[27] Petersen MA, Vachon H, Groenvold M, European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Group. Development of a diverse set of standard short forms based on the EORTC CAT Core item banks. Qual Life Res 2023;32:2037−45.
[28] Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al, PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol 2010;63:1179−94.
[29] Northwestern University. HealthMeasures: transforming how health is measured. Available at http://www.healthmeasures.net/. Accessed February 10, 2023.
[30] Rose M, Bjorner JB, Gandek B, Bruce B, Fries JF, Ware JE Jr. The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. J Clin Epidemiol 2014;67:516−26.
[31] Ware JE Jr. SF-36 health survey update. Spine 2000;25:3130−9.
[32] Ware JE, Brook R, Davies AR, Williams KN, Stewart A, Rogers WH, et al. In: Conceptualization and measurement of health for adults in the health insurance study, Vol. I. Santa Monica: Rand Corporation; 1980. Model of Health and Methodology.
[33] Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The functional assessment of cancer therapy scale: development and validation of the general measure. J Clin Oncol 1993;11:570−9.
[34] Kaat AJ, Schalet BD, Rutsohn J, Jensen RE, Cella D. Physical function metric over measure: an illustration with the patient-reported outcomes measurement information system (PROMIS) and the Functional Assessment of Cancer Therapy (FACT). Cancer 2018;124:153−60.
[35] Choi SW, Schalet B, Cook KF, Cella D. Establishing a common metric for depressive symptoms: linking the BDI-II, CES-D, and PHQ-9 to PROMIS depression. Psychol Assess 2014;26:513−27.
[36] Dorans NJ. Equating, concordance, and expectation. Appl Psychol Meas 2004;28:227−46.
[37] Kolen MJ, Brennan RL. Test equating, scaling, and linking methods and practices. J Educ Meas 2004;43:169−72.
[38] Albano AD. Equate: an R package for observed-score linking and equating. J Stat Software 2016;74:1−36.
[39] Holland PW, Thayer DT. Univariate and bivariate loglinear models for discrete test score distributions. J Educ Behav Stat 2000;25:133−83.
[40] Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res 2005;30:79−82.
[41] Thompson NR, Lapin BR, Steinmetz MP, Benzel EC, Katzan IL. Mapping PROMIS physical function and pain interference to the modified low back pain disability questionnaire. Qual Life Res 2022;31:3467−82.
[42] Edelen MO, Rodriguez A, Herman P, Hays RD. Crosswalking the patient-reported outcomes measurement information system physical function, pain interference, and pain intensity scores to the roland-morris disability questionnaire and the oswestry disability index. Arch Phys Med Rehabil 2021;102:1317−23.
[43] Hays RD, Qureshi N, Edelen M, Rodriguez A, Slaughter M, Herman PM. Crosswalking the national institutes of health impact stratification score to the PEG. Arch Phys Med Rehabil 2023;104:425−9.
[44] World Health Organisation. International classification of functioning, disability and health. Geneva: ICF; 2001.
[45] Cieza A, Brockow T, Ewert T, Amman E, Kollerits B, Chatterji S, et al. Linking health-status measurements to the international classification of functioning, disability and health. J Rehabil Med 2002;34:205−10.
[46] Lee MK, Schalet BD, Cella D, Yost KJ, Dueck AC, Novotny PJ, et al. Establishing a common metric for patient-reported outcomes in cancer patients: linking patient reported outcomes measurement information system (PROMIS), numerical rating scale, and patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). J Patient Rep Outcomes 2020;4:1−11.
[47] Blackwell CK, Tang X, Elliott AJ, Thomes T, Louwagie H, Gershon R, et al. Developing a common metric for depression across adulthood: linking PROMIS depression with the edinburgh postnatal depression scale. Psychol Assess 2021;33:610−8.
[48] Klaufus L, Tang X, Verlinden E, van der Wal M, Haverman L, Luijten M, et al. Linking the RCADS-25 to the PROMIS® pediatric item banks anxiety and depressive symptoms in a general dutch population sample. Qual Life Res 2022;31:1587−95.
[49] Cook KF, Schalet BD, Kallen MA, Rutsohn JP, Cella D. Establishing a common metric for self-reported pain: linking BPI pain interference and SF-36 bodily pain subscale scores to the PROMIS pain interference metric. Qual Life Res 2015;24:2305−18.