RHEUMATOLOGY

Rheumatology 2015;54:2221–2229
doi:10.1093/rheumatology/kev265
Advance Access publication 29 July 2015

Original article

# Validity and measurement precision of the PROMIS physical function item bank and a content validity–driven 20-item short form in rheumatoid arthritis compared with traditional measures

**Martijn A. H. Oude Voshaar[1], Peter M. ten Klooster[1], Cees A. W. Glas[2], Harald E. Vonkeman[1,3], Erik Taal[1], Eswar Krishnan[4], Hein J. Bernelot Moens[5], Maarten Boers[6], Caroline B. Terwee[6], Piet L. C. M. van Riel[7] and Mart A. F. J. van de Laar[1,3]**

## Abstract

**Objective.** To evaluate the content validity and measurement properties of the Patient-Reported Outcome Measurement Information System (PROMIS) physical function item bank and a 20-item short form in patients with RA in comparison with the HAQ disability index (HAQ-DI) and 36-item Short Form Health Survey (SF-36) physical functioning scale (PF-10).

**Methods.** The content validity of the instruments was evaluated by linking their items to the International Classification of Functioning, Disability and Health (ICF) core set for RA. The measures were administered to 690 RA patients enrolled in the Dutch Rheumatoid Arthritis Monitoring registry. Measurement precision was evaluated using item response theory methods and construct validity was evaluated by correlating physical function scores with other clinical and patient-reported outcome measures.

**Results.** All 207 health concepts identified in the physical function measures referred to activities that are featured in the ICF. Twenty-three of 26 ICF RA core set domains are featured in the full PROMIS physical function item bank compared with 13 and 8 for the HAQ-DI and PF-10, respectively. As hypothesized, all three physical function instruments were highly intercorrelated (*r* 0.74–0.84), moderately correlated with disease activity measures (*r* 0.44–0.63) and weakly correlated with age (*r*s 0.07–0.14). Item response theory–based analysis revealed that a 20-item PROMIS physical function short form covered a wider range of physical function levels than the HAQ-DI or PF-10.

**Conclusion.** The PROMIS physical function item bank demonstrated excellent measurement properties in RA. A content-driven 20-item short form may be a useful tool for assessing physical function in RA.

**Key words:** PROMIS, physical function, HAQ, item bank.

[1]Arthritis Center Twente, Department of Psychology, Health & Technology, Enschede, [2]Department of Research Methodology, Measurement and Data Analysis, University of Twente, [3]Arthritis Center Twente, Department of Rheumatology and Clinical Immunology, Medisch Spectrum Twente, Enschede, The Netherlands, [4]Stanford ARAMIS Program, Stanford University School of Medicine, Palo Alto, CA, USA, [5]Department of Rheumatology, Ziekenhuisgroep Twente, [6]Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam and [7]Department of Rheumatic Diseases, Radboud University Medical Centre, Nijmegen, The Netherlands

Submitted 9 April 2014; revised version accepted 23 June 2015

Correspondence to: Martijn A. H. Oude Voshaar, Department of Psychology, Health & Technology, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands.
E-mail: a.h.oudevoshaar@utwente.nl

**Rheumatology key messages**

- The patient-reported outcome measurement information system physical function item bank was shown to have content and construct validity in RA.
- A physical function short form was proposed to comprehensively assess physical function in RA.
- A 20-item physical function short form yielded more reliable scores than previously validated questionnaires in RA.

## Introduction

Physical function is a core outcome domain in RA [1]. Physical function is usually assessed with fixed-length questionnaires such as the HAQ disability index (HAQ-DI) and the 36-item Short Form Health Survey (SF-36) physical functioning scale (PF-10) [2]. Although these measures possess generally favourable measurement properties, key shortcomings associated with their use in RA are floor and ceiling effects and patient and administrator burden [2, 3]. Accurate assessment of all patients, irrespective of their level of physical impairment is therefore a major topic of ongoing research.

Item response theory (IRT)–based item banking enables computerized adaptive tests (CATs) or tailored short forms that facilitate tailoring of the difficulty of questions posed to the specific population or individual patient, allowing reliable assessment of patients with all levels of functioning. The Patient-Reported Outcome Measurement Information System (PROMIS) project has developed item banks for a great number of patient-reported outcome domains, including physical function [4, 5]. In a previous study, the Dutch–Flemish PROMIS physical function item bank was calibrated in a sample of Dutch patients with patients with RA [6]. In the current study we examined the content and construct validity of the full PROMIS physical function item bank in RA. We also developed a 20-item short form of the physical function item bank that optimally reflects those aspects of physical function that are relevant for patients with RA and compared its measurement precision with the HAQ-DI and PF-10.

## Methods

### Patients and sampling design

The current study is a secondary analysis of baseline data from a larger, longitudinal study aimed at calibrating the Dutch–Flemish PROMIS physical function item bank in Dutch RA patients [6]. Data were collected in the Dutch Rheumatoid Arthritis Monitoring (DREAM) registry. The DREAM registry is an observational multicentre cohort study that monitors the course of RA patients undergoing different treatment regimens in 12 hospitals in the Netherlands. Clinical and patient-reported outcomes of patients with RA are monitored over time using a web-based data collection and storage system. Patient-reported outcomes are administered preceding every visit to the clinic. RA patients of three participating DREAM hospitals were informed about the study and invited to participate after logging in to their personal DREAM web portal. Informed consent was provided according to the Declaration of Helsinki and digitally obtained from all participating patients. According to the Dutch law on medical research in humans, approval by an ethics committee was not required for this study.

The calibration study followed a longitudinal design where data were collected preceding three scheduled clinic visits. Patients completed randomly allocated questionnaires containing approximately 40 partially overlapping items from the 121-item PROMIS physical function item bank at each time point [7] along with the HAQ-DI and PF-10. Questionnaires and the allocation of patients to questionnaires was organized in such a way that patients answered all items exactly once after three clinic visits. Analyses in the present study were performed on the expected *a posteriori* estimates obtained at baseline for each patient (i.e. ~40 items per patient). In this study, expected *a posteriori* estimates are expressed on a scale where 0 reflects the average level of physical function of the participating RA patients at time point 3. The baseline scores reported in this study were higher than at time point 3 [T1 mean 0.55 (S.D. 1.51)]. More elaborate information about the data collection design, item calibration and model fit analyses are described elsewhere [6].

### Measures

Patients completed the HAQ-DI, the SF-36 version 2 health survey and single-item measures of general health, disease activity, pain, stiffness and fatigue alongside the PROMIS physical function items. For a subgroup of patients visiting the outpatient clinic around the time of completing the questionnaire, clinical measures were available to calculate the 28-joint DAS (DAS28).

#### *PROMIS physical function item bank*

The PROMIS physical function item bank measures self-reported current capability of physical activities. The item bank contains 121 questions assessing the functioning of the upper extremities (dexterity), lower extremities (walking or mobility) and central regions (neck, back), as well as instrumental activities of daily living, such as running errands. Each item is scored on a 5-point rating scale, with higher scores indicating better functioning.

#### *HAQ disability index*

The HAQ-DI contains 20 items measuring physical disabilities over the past week in eight categories of daily living: dressing and grooming, rising, eating, walking, hygiene, reach, grip and activities [8]. Each item is scored on a 4-point rating scale from 0 (without any difficulty) to 3 (unable to do). Disability scores were calculated according to the alternative scoring rule, which does not take the use of aids and help from others into account [9]. The highest

scores from each category score are averaged to produce a total score between 0 and 3, with higher values indicating more disability.

*SF-36 health survey PF-10*

The PF-10 is one of the eight scales of the SF-36 and consists of 10 items measuring perceived current limitations in a variety of physical activities on a 3-point response scale from 1 (yes, limited a lot) to 3 (no, not limited at all). In the current study a norm-based scoring procedure was employed where 50 (S.D. 10) represents the average level of physical function in the 1998 United States general population [10].

*DAS for 28 joints*

The DAS28 score is a pooled clinical index that includes a tender and swollen joint count of 28 joints, the ESR and the patient's global assessment of general health [11].

*Numerical rating scales*

Patient-reported general health, disease activity, pain, stiffness and fatigue were each assessed using 10-point numerical ratings scales, with higher values reflecting worse health status.

## Analysis

*Content validity*

A patient-reported outcome measure has content validity if its items all refer to aspects of the construct to be measured that are relevant to the construct and to the intended study population. Together, the items must comprehensively reflect the construct to be measured and must be suitable for the intended purpose of an instrument [12]. The content validity of the PROMIS physical function item bank as well as the HAQ-DI and PF-10 for measuring physical function in patients with RA was evaluated by linking their items to the International Classification of Functioning, Disability and Health (ICF) codes using proposed linking rules [13, 14]. The ICF is an exhaustive taxonomy of functioning that categorizes health concepts into dimensions, chapters and codes of functioning, with each successive step representing increasingly specific health concepts. For example, walking more than a mile is a health concept that belongs to the ICF dimension activities and participation, the chapter on mobility and the second-level code walking and moving. Health concepts embodied in the items of the various instruments were considered to be relevant indicators of function in patients with RA if they belonged to the ICF activities domain and were included in the ICF core set for RA or were specifically mentioned by patients as important omissions from the core set in a subsequent qualitative study [15, 16]. The ICF core set for RA represents the minimum set of ICF codes needed to comprehensively represent the experience of patients with RA. It was developed using systematic literature review and expert consensus. Because of their relevance to physical function in RA and inclusion in the core set, ICF chapters writing (d170) and sports (d920) were considered relevant as well. The comprehensiveness of the instruments was determined by evaluating the representation of ICF second-level codes in the instruments and by evaluating the proportional distribution of health concepts across the mobility, activities of daily living and domestic life domains.

*Measurement precision*

Measurement precision across the different levels of physical function measured by the various instruments was evaluated by obtaining test information functions for various evaluated instruments. The standard error of estimation of a scale at a given level of physical function, $\theta$, is inversely related to the amount of statistical information provided by that scale at that specific level of physical function [$SE\theta = 1/\sqrt{I(\theta)}$]. Global reliability of the HAQ-DI and PF-10 was evaluated with an IRT-based index of reliability. The index is the ratio of the variance of the expected *a posteriori* $\theta$ estimates and the total variance of $\theta$ {$\rho = \mathrm{Var}[E(\theta|r)]/\mathrm{Var}(\theta)$}, where $r$ is a person's response pattern [17]. Values $\geqslant 0.70$ and $\geqslant 0.90$ were considered to indicate adequate reliability for group-level and individual-level applications, respectively.

*Construct validity*

Construct validity was evaluated by calculating Pearson correlation coefficients between the physical function instruments and other patient-reported outcome measures and by comparing scores between known groups. Since all evaluated physical function instruments assess similar constructs, the pattern of correlations with clinical and patient-reported outcomes was hypothesized to be similar for the HAQ-DI, PF-10 and PROMIS physical function item bank. Moreover, the physical function measures were hypothesized to be highly intercorrelated (i.e. $>0.60$). We further hypothesized that the strongest correlations would be observed among the three physical function instruments, followed by the correlation between physical function scores and other patient-reported outcomes (well-being, fatigue, stiffness, patient-reported disease activity and pain). Weak correlations (i.e. $r \leqslant 0.30$) were expected with age. The relative validity of the three physical function instruments in discriminating between patients in clinical remission and those with active disease according to the DAS28 was analysed by comparing ratios of *F*-statistics from univariate one-way analysis of variance tests [18, 19]. DAS28 scores $<2.6$ were considered to reflect clinical remission.

*Short-form development*

A 20-item PROMIS short form was developed by selecting item content to optimally reflect the ICF chapters featured in the core set for RA. An attempt was made to select exactly one item for each ICF code. As described in more detail below, the PROMIS physical function item bank has items reflecting 23 of 26 relevant ICF codes. ICF codes d630 (preparing meals) and d550 (eating) and codes d4660 (moving around in different locations) and d455 (moving around) overlapped considerably and therefore all item content related to preparing meals and eating and to moving around and moving around in different

locations was pooled. ICF code d570 (looking after one's health) was judged to be least strongly related to physical function and no item was included from this ICF code in the short form. To optimize the short form in terms of measurement precision, one item within each pool of possible items was selected based on its contribution to the measurement precision of the short form according to the item information functions of the full item bank. The items selected for the short form are presented in Table 1.

## Results

Patient characteristics are summarized in Table 2. A total of 690 unique patients completed at least one questionnaire from the item bank and the other patient-reported outcomes. There were no missing data. On average, patients had low levels of disease activity and high levels of physical function. Substantial ceiling effects were observed on the HAQ-DI (53%), but not on any of the other physical function measures.

### Content validity

In total, 207 health concepts were identified in the 151 items of the three physical function instruments (Table 3). Only 15 (7%) of these could not be linked to the ICF. In all cases these health concepts were too broad to be linked (e.g. usual physical activities or vigorous activities). All health concepts that were linked to the ICF either featured in the ICF core set for RA or were identified as important by patients with RA. Therefore all the evaluated items of the HAQ-DI, PF-10 and PROMIS physical function item bank refer to health concepts that are relevant indicators of physical function in RA. The generic nature of the PF-10 is reflected in its items being predominantly linked to ICF codes within the mobility domain and sports chapter, whereas the HAQ-DI has more items reflecting basic activities of daily living, such as eating and washing, and mobility items related to use of the hands and arms (Table 2). The PROMIS physical function item bank most comprehensively reflects all areas of physical function that are relevant to RA patients according to the ICF core set. Except for assisting others (d660), driving (d475) and moving around using equipment (d465), all the relevant ICF activity domains that are featured in the ICF core set for RA are represented in the full PROMIS physical function item bank.

### Measurement precision

The global reliability of the HAQ-DI and PF-10 was high, with reliability coefficients of 0.89 and 0.90, respectively. The precision of the full PROMIS physical function item bank surpassed that of the HAQ-DI and PF-10 for all levels of physical function by virtue of having more items than either of the other instruments (data not shown). However, the precision of the PROMIS physical function item bank was high across all physical function levels, while physical function scores $>0$, which were observed in 47% of patients, could not be precisely measured with the HAQ-DI or PF-10.

The information functions belonging to the 20-item short form and individual physical function instruments are plotted in Fig. 1 against the distribution of physical function scores observed at baseline. The short form outperformed the HAQ-DI and PF-10 across most levels of physical function. However, for patients with physical function scores $>3$ S.D. below the mean of the scale, the HAQ-DI outperformed the short-form. The PF-10 provides comparatively little statistical information since it has only half the number of items of the other two instruments.

### Construct validity

Table 4 summarizes the correlations between the PROMIS physical function item bank, the HAQ-DI and the PF-10 and other patient-related variables. The three physical function instruments were highly intercorrelated and the pattern of correlations with other clinical and patient-reported measures was also highly similar. In addition, the association between the three physical function instruments and the other measures was of the expected magnitude for all comparisons, except for the correlation between age and the HAQ-DI. Taken together, these results were in accordance with the pre-specified hypotheses, indicating adequate convergent and construct validity for the PROMIS physical function item bank. In Table 5, the relative validity of the three instruments in discriminating between active disease and remission is presented. The HAQ-DI and PROMIS items were about equally efficient, while the PF-10 was less efficient in distinguishing between levels of disease activity.

## Discussion

The data of this study demonstrated that the PROMIS physical function item bank assesses a similar construct to the HAQ-DI and PF-10. In addition, the content-driven 20-item PROMIS physical function short form assesses a wider range of physical functioning than do the two traditional instruments, while its item content more comprehensively reflects the physical function domain relevant for RA.

The proposed 20-item short form comprehensively assesses aspects of physical function that are represented in the ICF core set for RA. Specific items included in the short form were chosen to cover a wide range of physical function levels. The measurement precision of the PROMIS physical function short form surpassed that of HAQ-DI and PF-10 and better captured the various levels of function that were observed in the current study. Its use is therefore recommended in studies with patients with RA, particularly for patients with well-controlled disease. Those interested in using the proposed short form in their studies can report a mean score, provided there are few missing data. In the case of missing responses, IRT-based pattern scoring can also be performed on the observed data using appropriate resources at the PROMIS assessment centre. This latter scoring system has the advantage that results may be expressed on the standard PROMIS metric, so that results can be compared with other studies utilizing (other applications

**Table 1** Items chosen for 20 item short form

| ICF core set for RA second-level codes | PROMIS item included in short form |
|---|---|
| D170, writing | Are you able to write with a pen or pencil? |
| D410, changing basic body position | Are you able to stand up from an armless straight chair? |
| D420, maintaining a body position | Are you able to sit on the edge of a bed? |
| D420, transferring oneself | Are you able to transfer from a bed to a chair and back? |
| D430, lifting and carrying objects | Are you able to carry a laundry basket up a flight of stairs? |
| D440, fine hand use | Are you able to open and close a zipper? |
| D445, hand and arm use | Are you able to push open a heavy door? |
| D450, walking | Does your health now limit you in hiking a couple of miles (3 km) on uneven surfaces, including hills? |
| D455, moving around | Are you able to carry a laundry basket up a flight of stairs? |
| D450, moving around in different locations | Does your health now limit you in doing strenuous activities such as back-packing, skiing, playing tennis, bicycling or jogging? |
| D465, moving around using equipment | — |
| D470, using transportation | Are you able to run a short distance, such as to catch a bus? |
| D475, driving | — |
| D510, washing oneself | Are you able to wash and dry your body? |
| D520, caring for body parts | Are you able to brush your teeth? |
| D530, toileting | Are you able to sit on and get up from the toilet? |
| D540, dressing | Are you able to pull on trousers? |
| D550, eating | Are you able to hold a plate of food? |
| D560, drinking | Are you able to open a new milk carton? |
| D570, looking after one's health | — |
| D620, acquisition of goods and services | Are you able to run errands and shop? |
| D630, preparing meals | See eating. |
| D640, doing housework | Are you able to wash dishes, pots, and utensils by hand while standing at a sink? |
| D650, caring for household objects | Are you able to do yard work like raking leaves, weeding, or pushing a lawn mower? |
| D660, assisting others | — |
| D9201, sports | Are you able to exercise for an hour? |

ICF: International Classification of Functioning, Disability and Health; PROMIS: Patient-Reported Outcome Information System.

**Table 2** Baseline characteristics

| Characteristic | *n* (%) |
|---|---|
| Female | 439 (63.6) |
| Age, mean (S.D.) years | 56.8 (11.9) |
| DAS28 | 2.1 (1.2) |
| General health | 25.1 (23.7) |
| Disease activity | 28.2 (23.5) |
| Fatigue | 33.2 (29.8) |
| Stiffness | 18.8 (41.4) |
| Pain | 27.7 (22.8) |
| HAQ-DI | 0.6 (0.5) |
| PF-10 | 42.9 (10.5) |

DAS28: 28-joint DAS; HAQ-DI: Health Assessment Questionnaire Disability Index; PF-10: 36-item Short Form Health Survey physical function scale.

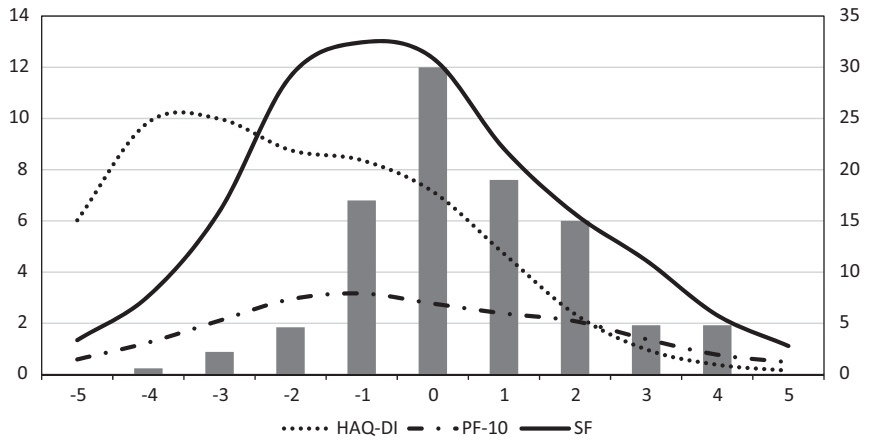
of) the PROMIS physical function item bank. Global reliability indices were high for both the HAQ-DI and PF-10, which corresponds to previous research where reliability coefficients sufficient for individual comparisons were observed as well [20, 21]. The instrument-specific information functions showed that the HAQ-DI outperformed PF-10 and the 20-item short form in discriminating between patients with the poorest levels of function. This corresponds to the results from the analysis of content validity that showed that many of its items refer to basic activities of daily living such as eating or dressing unassisted and mobility issues related to use of the hands. Consistent with previous work in RA, the analysis of local measurement precision showed that the items of the HAQ-DI less reliably measure higher levels of function [22], which were commonly experienced by patients with RA in our sample. In fact, 53% of patients scored the best possible HAQ-DI score, which is well over the commonly accepted criterion for ceiling effects of 15% [23]. Although ceiling effects of the HAQ-DI are a well-known and often reported shortcoming in various settings, the proportion of patients scoring the best possible score was even higher than reported in previous studies in RA [2, 3]. This likely reflects improved general status of current patients as a result of aggressive and early treatment. These findings underscore the need for personalized assessment of physical function and suggest that practical applications of the PROMIS physical function item bank, such as the proposed short form, allow for more precise measurement of patients with various levels of function and reduce floor and ceiling effects. The precision of such applications will

**Table 3** Content validity of the HAQ-DI, PF-10 and PROMIS physical function item bank according to the ICF RA core set

| ICF core set for RA second level codes | PROMIS PF, *n* (%) | HAQ-DI, *n* (%) | PF-10, *n* (%) |
|---|---|---|---|
| D170, writing | 1 (0.6) | — | — |
| D410, changing basic body position | 13 (8.1) | 5 (17.9) | 3 (15.7) |
| D420, maintaining a body position | 10 (6.3) | — | — |
| D420, transferring oneself | 2 (1.3) | — | — |
| D430, lifting and carrying objects | 16 (10.0) | 2 (7.7) | 3 (15.7) |
| D440, fine hand use | 17 (10.6) | 4 (15.4) | |
| D445, hand and arm use | 14 (8.8) | 1 (3.8) | 2 (10.5) |
| D450, walking | 10 (6.3) | 1 (3.8) | 3 (15.7) |
| D455, moving around | 19 (11.9) | 1 (3.8) | 3 (15.7) |
| D450, moving around in different locations | 2 (1.3) | — | — |
| D465, moving around using equipment | — | — | — |
| D470, using transportation | 1 (0.6) | — | — |
| D475, driving | — | — | — |
| Mobility total | 105 (65) | 14 (53) | 14 (74) |
| D510, washing oneself | 8 (5.0) | 3 (11.5) | 1 (5.2) |
| D520, caring for body parts | 4 (2.5) | — | — |
| D530, toileting | 1 (0.6) | — | — |
| D540, dressing | 13 (8.1) | 3 (11.5) | 1 (5.2) |
| D550, eating | 4 (2.5) | 2 (7.7) | — |
| D560, drinking | 1 (0.6) | 1 (3.8) | — |
| D570, looking after one's health | 1 (0.6) | — | — |
| Self-care total | 32 (19) | 7 (26) | 2 (10.0) |
| D620, acquisition of goods and services | 3 (1.9) | 2 (7.7) | — |
| D630, preparing meals | 1 (0.6) | — | — |
| D640, doing housework | 8 (5.0) | 1 (3.8) | — |
| D650, caring for household objects | 4 (2.5) | 1 (3.8) | — |
| D660, assisting others | — | — | — |
| Domestic life total | 16 (7.5) | 3 (11.5) | 0 (0.0) |
| D9201, sports | 7 (4.4) | — | 3 (15.7) |
| Total | 160 | 28 | 19 |

ICF: International Classification of Functioning, Disability and Health; HAQ-DI: Health Assessment Questionnaire Disability Index; PF-10: 36-item Short Form Health Survey physical functioning scale; PROMIS: Patient-Reported Outcome Information System.

**Fig. 1** Measurement precision of the various forms in relation to observed physical function levels




Test information functions of the 20-item PROMIS physical function short form, HAQ-DI and PF-10 in relation to the percentage of patients at each of the physical function score levels at baseline. Physical function scores on the *x*-axes are centred around the average level of physical function of patients, with higher values indicating better functioning. Information represents measurement precision, with higher values representing more reliable measurement [$SE\theta = 1/\sqrt{I}/(\theta)$]. HAQ-DI: Health Assessment Questionnaire Disability Index; PF-10: 36-item Short Form Health Survey physical functioning scale; PROMIS: Patient-Reported Outcome Information System; SF: short form.

TABLE 4 Pearson correlations between the psychical function measures and other variables

| Instrument | HAQ-DI ($n = 677$) | PF-10 ($n = 669$) | Pain ($n = 383$) | General health ($n = 223$) | Disease activity ($n = 147$) | Fatigue ($n = 74$) | Stiffness ($n = 74$) | Age ($n = 690$) |
|---|---|---|---|---|---|---|---|---|
| PROMIS physical function (~40 items) | 0.76 | 0.84 | −0.52 | −0.53 | −0.46 | −0.47 | −0.63 | 0.14 |
| HAQ-DI | 1 | −0.74 | 0.52 | 0.48 | 0.50 | 0.46 | 0.62 | −0.07[a] |
| PF-10 | −0.74 | 1 | −0.52 | −0.44 | −0.46 | −0.52 | −0.46 | 0.11 |
| Hypothesized correlation | >0.60 | >0.60 | 0.30–0.60 | 0.30–0.60 | 0.30–0.60 | 0.30–0.60 | 0.30–0.60 | 0.10–0.30 |

[a]Not significant at $P < 0.05$. HAQ-DI: Health Assessment Questionnaire Disability Index; PF-10: 36-item Short Form Health Survey physical functioning scale; PROMIS: Patient-Reported Outcome Information System.

TABLE 5 Relative validity of the instruments in differentiating between patients in remission and those with active disease according to the DAS28

| Instrument | Remission ($n = 37$) | Active disease ($n = 79$) | $F$ | RV |
|---|---|---|---|---|
| PROMIS physical function (~40 items), mean (S.D.) | 2.06 (1.89) | −0.06 (1.86) | 31.87 | 1.00 |
| HAQ-DI, mean (S.D.) | 0.35 (0.51) | 0.87 (0.51) | 31.57 | 0.99 |
| PF-10, mean (S.D.) | 75.90 (20.48) | 54.86 (23.73) | 20.99 | 0.65 |

*F*: *F*-statistic from univariate one-way analysis of variance; HAQ-DI: Health Assessment Questionnaire Disability Index; PF-10: 36-item Short Form Health Survey physical functioning scale; PROMIS: Patient-Reported Outcome Information System; RV: relative validity (ratio of *F*-statistic compared with observed PROMIS physical function).

depend on the number of administered items and the success of tailoring these applications to the samples' or individual patients' level of functioning. Previous research has shown that a preliminary CAT version of the PROMIS physical function item bank approached the reliability of the full bank and outperforms an equal number of HAQ-DI or PF-10 items [24]. Based on the current results, in particular, gains in the precision of measuring higher levels of functioning are to be expected in such applications, which are most relevant in light of the increased physical health status of contemporary RA patients.

All 121 items of the PROMIS physical function item bank refer to ICF codes that are relevant for measuring function in RA according to the ICF core set for RA. Moreover, the item bank covers almost all ICF chapters that feature in the core set. These results support the content validity of the item bank as a whole. The item bank currently lacks items reflecting the ICF codes assisting others, driving and moving around using equipment. Moreover, many items in the item bank measure relatively simple activities of daily living, which is reflected in very high measurement precision at lower levels of function. The item bank might benefit from more items reflecting more difficult physical activities, such as assisting others or walking around using equipment, which is a research topic that is currently being addressed [25].

The pattern of observed correlations with other measures was highly similar for all three evaluated physical function instruments, which supports the construct validity of the PROMIS physical function item bank in RA. Previous studies of the HAQ-DI and PF-10 [2, 26] also found moderate correlations with other patient-reported outcomes and disease severity measures and low, but significant, correlations with age.

CATs are often viewed as the ultimate goal of item banking. However, an advantage of tailored short forms is that researchers may precisely specify the content they wish to assess. In this study we showed that a short form could be developed that more comprehensively assesses the relevant aspects of physical function for patients with RA and additionally surpasses the measurement precision of the established HAQ-DI and PF-10. It would be interesting for future studies to evaluate whether CATs provide additional measurement precision beyond what is achieved with the fixed short form. A limitation of the current study is that some measurement properties of the short form, such as its global reliability and construct validity, could not be evaluated since none of the patients in our sample filled out all 20 items included in the short form.

The sensitivity of individual items to changes in functional status is also an important aspect of content validity that could not be tested using the data of this study. Future studies should evaluate the responsiveness of individual PROMIS physical function items.

Taken together, the results of the current study suggest that the PROMIS physical function item bank will result in

improved measurement precision and less patient burden in the assessment of self-reported physical function in patients with RA. The proposed 20-item short form better targets the most frequently observed levels of physical function in this study among typical Dutch outpatients with RA. Therefore researchers and clinicians interested in assessing physical function in RA patients with relatively high levels of physical function might consider using it in their studies or clinical practice.

## Acknowledgements

All authors made substantial contributions to the conception or design of the study or the acquisition, analysis or interpretation of data for the study. All authors contributed to drafting the manuscript and approved the final version.

*Funding*: This study was funded by the Dutch Arthritis Association (project number 11-1-203).

*Disclosure statement*: C.B.T. is coordinator of the Dutch–Flemish PROMIS group (http://www.dutchflemishpromis.nl), which aims to validate and implement PROMIS instruments in the Netherlands and Flanders. All other authors have declared no conflicts of interest.

## References

1 Felson DT, Anderson JJ, Boers M *et al*. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. Arthritis Rheum 1993;36:729–40.

2 Oude Voshaar MA, Ten Klooster PM, Taal E, van de Laar MA. Measurement properties of physical function scales validated for use in patients with rheumatoid arthritis: a systematic review of the literature. Health Qual Life Outcomes 2011;9:99.

3 Stucki G, Stucki S, Bruhlmann P, Michel BA. Ceiling effects of the Health Assessment Questionnaire and its modified version in some ambulatory rheumatoid arthritis patients. Ann Rheum Dis 1995;54:461–5.

4 Cella D, Riley W, Stone A *et al*. 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.

5 Cella D, Yount S, Rothrock N *et al*. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap Cooperative Group during its first two years. Med Care 2007;45(Suppl 1):S3–S11.

6 Oude Voshaar AH, Ten Klooster PM, Glas CAW *et al*. Calibration of the PROMIS physical function item bank in Dutch patients with rheumatoid arthritis. PLoS One 2014;9:e92367.

7 Oude Voshaar MA, Ten Klooster PM, Taal E, Krishnan E, van de Laar MA. Dutch translation and cross-cultural adaptation of the PROMIS® physical function item bank and cognitive pre-test in Dutch arthritis patients. Arthritis Res Ther 2012;14:R47.

8 Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137–45.

9 Fries JF. The Health Assessment Questionnaire (HAQ) and the Improved HAQ. Stanford, CA, USA: Stanford University School of Medicine, Division of Immunology and Rheumatology, 2009.

10 Ware JE Jr, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473–83.

11 Prevoo ML, van 't Hof MA, Kuper HH *et al*. Modified disease activity scores that include twenty-eight-joint counts: development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum 1995;38:44–8.

12 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;63:737–45.

13 Cieza A, Geyh S, Chatterji S *et al*. ICF linking rules: an update based on lessons learned. J Rehabil Med 2005;37:212–8.

14 Cieza A, Brockow T, Ewert T *et al*. Linking health-status measurements to the international classification of functioning, disability and health. J Rehabil Med 2002;34:205–10.

15 Stucki G, Cieza A, Geyh S *et al*. ICF Core Sets for rheumatoid arthritis. J Rehabil Med 2004;44(Suppl):87–93.

16 Coenen M, Cieza A, Stamm TA *et al*. Validation of the International Classification of Functioning, Disability and Health (ICF) Core Set for rheumatoid arthritis from the patient perspective using focus groups. Arthritis Res Ther 2006;8:R84.

17 Scheerens J, Glas CAW, Thomas SM. Educational Evaluation, Assessment, and Monitoring: a Systemic Approach. Lisse, The Netherlands: Swets and Zeitlinger, 2003.

18 McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care 1993;31:247–63.

19 Liang MH, Larson MG, Cullen KE, Schwartz JA. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. Arthritis Rheum 1985;28:542–7.

20 ten Klooster PM, Taal E, van de Laar MA. Rasch analysis of the Dutch Health Assessment Questionnaire disability index and the Health Assessment Questionnaire II in patients with rheumatoid arthritis. Arthritis Rheum 2008;59:1721–8.

21 Kosinski M, Zhao SZ, Dedhiya S, Osterhaus JT, Ware JE Jr. Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. Arthritis Rheum 2000;43:1478–87.

22 Fries JF, Krishnan E, Rose M, Lingala B, Bruce B. Improved responsiveness and reduced sample size requirements of PROMIS physical function scales with item response theory. Arthritis Res Ther 2011;13:R147.

23 Terwee CB, Bot SD, de Boer MR *et al*. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34–42.

24 Rose M, Bjorner JB, Becker J, Fries JF, Ware JE. Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). J Clin Epidemiol 2008;61:17–33.

25 Fries JF, Bharathi L, Siemons L *et al*. Extending the floor and the ceiling for assessment of physical function (disability). Arthritis Rheumatol 2014;66:1378–87.

26 Wolfe F, Michaud K, Pincus T. Development and validation of the Health Assessment Questionnaire II: a revised version of the health assessment questionnaire. Arthritis Rheum 2004;50:3296–305.

# Clinical vignette

Rheumatology 2015;54:2229
doi:10.1093/rheumatology/kev317
Advance Access publication 28 August 2015

## Atypical skin involvement in Erdheim–Chester disease

A 49-year-old woman was referred to our hospital because of oedema of the lower extremities and exertional dyspnoea. On physical examination, multiple subcutaneous nodules from the neck to the trunk were observed (Fig. 1A). Serum biochemistry showed hypoalbuminaemia, and CRP was within normal limits. CT demonstrated numerous subcutaneous mediastinal and intra-abdominal small granular nodules, with periaortic thickening and pulmonary effusion (Fig. 1B and C). PET-CT revealed uptake of fludeoxyglucose in the bilateral femoral and lower leg bones and multiple subcutaneous nodules (Fig. 1D). A biopsy specimen from the skin nodules demonstrated foamy monocyte hyperplasia, and immunohistochemistry revealed these cells as positive for CD68 and CD163 and negative for CD1a and S100. She was negative for $BRAF^{V600E}$ mutation. Currently she has been in stable with IFN-$\alpha$ therapy.

Erdheim–Chester disease (ECD) is a rare (~500 known cases worldwide) systemic non-Langerhans cell histiocytosis with multi-organ involvement, and the common clinical features include skeletal, CNS, cardiac and retroperitoneal involvement. Xanthelasma, a common cutaneous involvement in ECD, is seen in 25% of patients, but papulonodular lesions are less frequent [1, 2]. A somatic mutation in $BRAF^{V600E}$ was identified recently, suggesting a pivotal role of this pathway in the pathogenesis of ECD [2]. Although she had no typical symptoms such as bone pain, PET-CT was helpful for detection of the bone lesions in ECD.

*Funding:* No specific funding was received from any funding bodies in the public, commercial or not-for-profit sectors to carry out the work described in this article.

*Disclosure statement:* The authors have declared no conflicts of interest.

**Chinatsu Takai[1], Yoko Wada[1] and Ichiei Narita[1]**

[1]Department of Clinical Nephrology and Rheumatology, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan

Correspondence to: Yoko Wada, Department of Clinical Nephrology and Rheumatology, Niigata University Graduate School of Medical and Dental Sciences, 1-757, Asahimachi-dori, Chuo-ku, Niigata, 951-8510, Japan.
E-mail; yoko.wada@gmail.com

## References

1 Cavalli G, Guglielmi B, Campochiaro C *et al*. The multifaceted clinical presentations and manifestations of Erdheim-Chester disease: comprehensive review of the literature and 10 new cases. Ann Rheum Dis 2013;72:1691–5.

2 Haroche J, Arnaud L, Cohen-Aubart F *et al*. Erdheim-Chester disease. Curr Rheumatol Rep 2014;16:412.

FIG. 1 Skin nodules, CT and PET-CT findings of the case