ORIGINAL ARTICLE

Generic PROMIS item banks in adults with hemophilia for patient-reported outcome assessment: Feasibility, measurement properties, and relevance

Isolde A. R. Kuijlaars PhD[1] | Lorynn Teela MSc[2] |
Lize F. D. van Vulpen MD, PhD[1] | Merel A. Timmer PhD[1] |
Michiel Coppens MD, PhD[3] | Samantha C. Gouw MD, PhD[4,5] |
Marjolein Peters MD, PhD[4] | Marieke J. H. A. Kruip MD, PhD[6] |
Marjon H. Cnossen MD, PhD[7] | Jelmer J. Muis MSc[2,4] | Evelien S. van Hoorn MSc[8] |
Lotte Haverman PhD[2] | Kathelijn Fischer MD, PhD[1]

[1]Van Creveldkliniek, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

[2]Amsterdam University Medical Centers, Emma Children's Hospital, Child and Adolescent Psychiatry & Psychosocial Care, Amsterdam Reproduction and Development, Amsterdam Public Health, University of Amsterdam, Amsterdam, The Netherlands

[3]Department of Vascular Medicine, Amsterdam Cardiovascular Sciences, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, The Netherlands

[4]Emma Children's Hospital, Amsterdam University Medical Centers, Pediatric Hematology, University of Amsterdam, Amsterdam, The Netherlands

[5]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

[6]Erasmus MC, Department of Hematology, Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands

[7]Erasmus MC, Department of Pediatric Hematology, Sophia Children's Hospital, Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands

[8]Erasmus MC, Department of Public Health, Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands

Correspondence
Isolde A. R. Kuijlaars, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands.
Email: i.a.r.kuijlaars-2@umcutrecht.nl

Handling Editor: Pantep Angchaisuksiri

Abstract

Background: Legacy hemophilia-specific questionnaires are considered too long, show floor-/ceiling effects, and/or include irrelevant questions. Patient Reported Outcomes Measurement Information System (PROMIS) item banks, including Computer Adaptive Tests (CATs) and short forms, were designed for more efficient outcome assessment.

Objectives: Evaluate the feasibility, measurement properties, and relevance of seven PROMIS CATs and two short forms in patients with hemophilia.

Patients/Methods: In this cross-sectional study, Dutch adults with hemophilia completed nine PROMIS item banks electronically. Feasibility was assessed by number of items and floor/ceiling effects. Reliability was determined as the proportion of reliable scores (standard error ≤3.2). Construct validity was assessed by comparison with legacy instruments and expected differences between subgroups. Relevance of item banks was determined by proportions of limited scores.

Results: Overall, 142 of 373 invited patients (mean age, 47 [range, 18-79], 49% severe hemophilia, 46% receiving prophylaxis) responded. Per CAT item bank, mean number of items answered varied from 5 (range, 3-12) to 9 (range, 5-12), with floor effects in pain interference (26% lowest scores) and depression (18% lowest scores). Construct validity and reliability were good for physical function, pain interference, satisfaction with social roles and activities, and fatigue. The CAT physical function showed the most limited scores (38%). The self-efficacy short forms showed ceiling effects (22%-28%) and no relation with the legacy instruments.

Conclusions: The PROMIS CATs physical function, pain interference, satisfaction with social roles and activities, and fatigue are feasible, reliable, and valid alternatives to legacy instruments for patients with hemophilia, with few items and low floor-/ceiling effects.

KEYWORDS

hemophilia, patient-reported outcome, questionnaire, reliability, validity

Essentials

- This study investigates Patient Reported Outcomes Measurement Information System (PROMIS) Computer Adaptive Tests (CATs) and short forms in hemophilia.
- This is a multicenter study in 142 Dutch adults with hemophilia.
- Four PROMIS CATs are feasible and valid alternatives to legacy instruments.
- The PROMIS CAT domain physical function is most relevant to patients with hemophilia.

1 | INTRODUCTION

Clinical management and therapeutic options for hemophilia have greatly improved in the past decades in resource-rich countries, from prophylactic clotting factor replacement therapy to prevent bleeding,[1] introduced in the Netherlands in 1968, to current ongoing hemophilia gene therapy trials[2] and upcoming nonreplacement therapy development and implementation since 2017.[3] Logically, outcome has also improved, from reduced life expectancy and development of painful crippling arthropathy at an early age to a near-normal life expectancy and participation in contact sports.[4,5] Currently, comprehensive care with a focus on physical and psychosocial health is standard care.[1] Appropriate patient-reported outcomes are essential to evaluate these and novel interventions in individual patients and should cover the wide range of consequences of hemophilia.[6]

Patient Reported Outcomes Measurement Information System (PROMIS) is a set of universal, person-centered item banks that evaluates and monitors physical, mental, and social health in adults and children, currently available in many languages (eg, English, Dutch, German, Spanish, French, Hebrew, Chinese, and Korean).[7] Other than Classical Test Theory (CTT)-based legacy instruments like the Haemophilia Activities List (HAL) and RAND-36, with sum and component scores that are a sum of all individual ordinal items of the questionnaire, PROMIS item banks were developed according to the Item Response Theory (IRT). According to IRT, the score is determined on the basis of the specific item and scoring option levels, taking rating of individual answers into account. In addition, an important advantage of IRT is the application of Computer Adaptive Tests (CATs), where the next item presented to the patient depends on the response to earlier items. As a consequence, it is not necessary to answer all items of the patient-reported outcome measure. This system lowers the burden of outcome assessment by administering a limited number of more relevant questions with a higher reliability.[8] For example, the HAL contains 42 items versus a mean of 4 to 6 PROMIS physical functioning CAT items.[9] When CATs are not available or information technology (IT) facilities and budget are limited, static PROMIS short forms with a selection of items are a reliable alternative.[10] An additional advantage is that PROMIS item banks are generic and patients do not need to complete different questionnaires for every comorbidity, resulting in a lower burden for the patient. This aspect is increasingly relevant to patients with hemophilia who experience an increasing life expectancy and as a result acquire more comorbidities associated with aging.[5] Furthermore, the occurrence of floor and ceiling effects, a frequent limitation of the HAL and Short Form-36 (SF-36), are minimized in PROMIS item banks based on item selection over the whole score range.[9,11-14] A large set of PROMIS item banks has been translated and validated in the Dutch general population and several patient populations.[15-18] However, PROMIS item banks have not yet been validated in adults with hemophilia and were to date seldom applied in hemophilia research projects.[19-21] Therefore, this study aimed to evaluate (i) feasibility, (ii) measurement properties, and (iii) relevance of nine PROMIS CATs and short forms for Dutch adult patients with hemophilia. We hypothesized that PROMIS CATs and short forms

are feasible alternatives to legacy instruments for patients with hemophilia.

2 | MATERIALS AND METHODS

2.1 | Study design and study population

This study was a cross-sectional multicenter study in three Dutch Hemophilia Treatment Centers: Van Creveldkliniek in Utrecht, Amsterdam University Medical Center in Amsterdam, and Erasmus University Medical Center in Rotterdam. Data collection occurred from December 2020 to February 2021 in adult patients who participated in the Hemophilia in the Netherlands-6 (HiN-6) nationwide survey study, for which patients were invited from June 2018 until July 2019, and gave permission to be contacted for follow-up studies.[22] Inclusion criteria were mild to severe hemophilia A or B and ≥18 years at HiN-6 assessment. Exclusion criteria were relevant self-reported changes in health between HiN-6 and PROMIS assessment. Relevant changes were defined as started or stopped with prophylactic treatment, started with emicizumab or gene therapy, changes in health status like stroke or major bleeds with remaining complaints, and joint surgery and other major surgeries between HiN-6 and PROMIS assessment. Patients were invited to participate by email and received a personal link to the research website to sign online informed consent and to complete PROMIS questionnaires. After 2 weeks, patients received one reminder email. Included patients gave written informed consent for the HiN-6 survey and electronic informed consent for the collection of PROMIS data. Patients were informed about the project by the Netherlands Hemophilia Patient Society. A sample size of ≥100 has been recommended for a validation study.[23]

Patient characteristics and data from five legacy questionnaires (HAL, RAND-36, Haemophilia & Exercise Project-Test-Questionnaire [HEP-test-Q], Validated Hemophilia Regimen Treatment Adherence Scale—Prophylaxis [VERITAS-Pro], Patient Activation Measure-13 [PAM-13]) were extracted from the HiN-6 study. Legacy questionnaires were completed online (84%) or on paper.

The Medical Research Ethical Committee of the University Medical Center Utrecht reviewed the study (protocol number 20-691/C).

2.2 | Measurements

2.2.1 | PROMIS item banks

Nine Dutch PROMIS item banks were selected by nine members of the PROMIS in Hemophilia Care' workgroup (IK, MT, LV, MP, SG, MHC, LH, KF, MH), including physicians, physical therapists, and psychologists. Seven item banks were assessed as CAT: V1.2, physical function; V1.1, pain interference; V1.0, depression; V1.0, anxiety; V2.0, ability to participate in social roles and activities (participation);

V2.0, satisfaction with social roles and activities (satisfaction with participation); and V1.0, fatigue.[15] For two items banks no CAT was available; these were assessed as short form with 8 questions: V1.0, self-efficacy for managing medications and treatment (self-efficacy medications); and V1.0, self-efficacy for managing symptoms (self-efficacy symptoms). All item banks use a 5-point Likert scale. The CATs automatically stopped when the standard error (SE) was ≤2.2 (95% reliability) and/or a maximum of 12 items was administered. PROMIS total scores are calculated by transforming the item scores into T scores, based on US population data, with a mean of 50 and a standard deviation (SD) of 10. For all item banks, higher scores represent more presence of the construct (eg, more pain interference or better physical function). The scores of the short forms were calculated in the PROMIS Assessment Center Scoring Service. All item banks cover a 7-day recall period, except from the physical function and participation, which do not use a recall period, and the self-efficacy item banks, which ask the current level of confidence.

2.2.2 | Legacy instruments from HiN-6

The HAL is a validated instrument for assessment of self-reported limitations in activities and participation in patients with hemophilia. It contains 42 items, distributed over seven domains (lying down/sitting/kneeling/standing, functions of the legs, functions of the arms, use of transportation, self-care, household tasks, and leisure activities and sports). Patients score the items on a 6-point Likert scale (impossible, always, mostly, sometimes, rarely, never), with a not applicable (N/A) scoring option for some items. Domain scores, component scores, and sum scores are converted to a normalized domain score ranging from 0 (worst possible functional abilities) to 100 (best possible functional abilities). Domain and component scores were calculated only if a minimum of 50% of items of a domain or component were scored on the 6-point Likert scale. The HAL uses a recall period of 1 month.[24,25] The internal consistency of the HAL was high (Cronbach's α, 0.97-0.98).[24,26]

The RAND-36 measures health-related quality of life across 8 domains ('physical functioning, role limitations due to physical health problems, bodily pain, general health, energy/fatigue, social functioning, role limitations due to emotional health problems, and emotional well-being), and construct validity has been studied in patients with hemophilia.[27] In six of eight domains, patients score the items on a 3- to 6-point Likert scale, and in two of eight domains, patients score yes or no. Scores range from 0 to 100, with higher scores indicating better health status. The recall period varies from at this moment to the last four weeks.[28,29] The internal consistency of the RAND-36 was high (Cronbach's α, 0.78-0.95).[30]

The HEP-test-Q is a validated questionnaire for the assessment of subjective physical performance in patients with hemophilia. The HEP-test-Q consists of 25 items pertaining to four domains (mobility, strength & coordination, endurance, and body perception). The response options are a 5-point Likert scale (never to always). Subscales and the total score were transformed to a scale ranging

from 0 to 100, with high scores indicating better physical performance. The HEP-test-Q used a recall period of 4 weeks, except for two items assessing physical activity: at this moment and compared to the last year.[31] The internal consistency of the HEP-test-Q was high (Cronbach's α, 0.96).[31]

The VERITAS-Pro is a validated questionnaire for the assessment of prophylactic treatment adherence in patients with hemophilia. The 24-item questionnaire consists of six subscales (time, dose, plan, remember, skip, communicate). The response options were a 5-point Likert scale (always to never). The score ranges from 100 to 0, with an optimum score of 0. The VERITAS-Pro has a recall period of 3 months.[32] The VERITAS-Pro was available only for patients on prophylactic treatment. The internal consistency of the VERITAS-Pro was high (Cronbach's α, 0.92).[32]

The PAM-13 measures patient knowledge, skills, and confidence for self-management. All 13 items have five possible responses, with scores ranging from 1 (disagree strongly) to 4 (agree strongly) or 0 (not applicable). The PAM-13 has a calibrated scale range from 38.6 to 53.0 (on a theoretical 0-100 point scale, with 100 as the optimum score). The PAM-13 does not specify a recall period.[33,34] In the general population, internal consistency of the PAM-13 was high (Cronbach's α, 0.88).[34]

2.2.3 | Patient characteristics

Patient characteristics analyzed included age at HiN-6 participation, type of hemophilia (A or B), severity of the disease (mild [factor VIII/IX activity, 0.06-0.40 IU/mL], moderate [factor VIII/IX activity, 0.01-0.05 IU/mL] or severe [factor VIII/IX activity, <0.01 IU/mL]), clotting factor regimens (prophylaxis yes/no), inhibitor status (current/former/never) and comorbidities (HIV yes/no, hepatitis C current/past/unknown, and other comorbidities).

2.3 | Statistical analyses

SPSS, version 25 (IBM, Armonk, NY, USA) was used for data analyses. Complete case analyses were performed in the case of missing data.

Patient characteristics were presented as proportions or means (SD).

2.3.1 | Feasibility

To determine the feasibility of PROMIS CATs and short forms, floor and ceiling effects were evaluated. Floor effects were defined as >15% of the patients reported the lowest possible score, and ceiling effects were defined as >15% reported the highest possible score.[35] In addition, for the CATs the number of items (mean [SD], range) completed by patients with hemophilia were evaluated. For the legacy instruments, floor and ceiling effects were evaluated and

number of items described. Both the floor and ceiling effects and number of items were compared between the PROMIS item banks and the legacy instruments. Data on time to administer the legacy instruments and PROMIS CATs and short forms were not available.

2.3.2 | Measurement properties—construct validity and reliability

Construct validity was studied by testing hypotheses regarding the relationship of PROMIS items banks with the legacy instruments (convergent validity) as well as regarding expected differences between subgroups (known-group validity). Hypotheses were defined a priori based on the literature[13,14,31,36-39] and expert opinion (KF, MT, MP, MC, SG, MHC, MK) and are presented in Supporting Information.

To test hypotheses regarding convergent validity, correlations between PROMIS item banks and the legacy instruments were calculated. Spearman's correlations were calculated because some data showed skewed distributions. Correlation coefficients of ≥0.9 were considered as a very strong correlation, 0.7 to 0.89 as strong, 0.4 to 0.69 as moderate, 0.10 to 0.39 as weak, and <0.10 as negligible.[40]

To test the hypotheses regarding known-group validity, differences in PROMIS T scores between a priori defined groups (severe vs nonsevere [mild and moderate] hemophilia and young adults [18-29 years] vs adults [≥30 years]) were tested with unpaired t tests.

The reliability of the CATs was evaluated by calculating the proportion of T scores with an SE ≤ 3.2. In IRT, the reliability varies across levels of the measured construct and is shown as the SE. An SE of ≤3.2 signifies a reliability of 90%, which has been considered a minimum requirement for use of patient-reported outcome measures (PROMs) in individual patients.[41] This SE cutoff point deviates from the stopping rule of ≤2.2 as described for the PROMIS CATs. To assess reliability of the legacy instruments of the patients with hemophilia, internal consistency estimates (Cronbach's α) were calculated.

2.3.3 | Relevance

To determine which item banks were relevant to adults with hemophilia, descriptive analyses (mean T scores and SD, range) were performed for the PROMIS item banks. For the PROMIS CATs pain interference, physical function, depression, anxiety, participation, and fatigue T scores were categorized in the following categories: within normal limits, mildly- (0.5 SD), moderately (1 SD), or severely (2 SD) deviant. For the PROMIS CAT satisfaction with participation T scores were categorized as very high (+2 SD), high (+1 SD), average, low (−1 SD), and very low (−2 SD) for the construct being measured. Reference data from the general Dutch male population were used to determine the score cutoff points for these seven PROMIS item banks, according to data from the Dutch-Flemish PROMIS national center (personal communication, CB Terwee, April 29, 2021, and

August 26, 2021). For the PROMIS short forms self-efficacy medications and self-efficacy symptoms T scores were categorized as very high (+2 SD), high (+1 SD), average, low (–1 SD), and very low (–2 SD) for the construct being measured. In the absence of Dutch reference data, score cutoff values from the general US population were used to categorize these two PROMIS scores. In the general population, 84.1% scored within normal limits or the mildly deviant categories or the average, high, or very high categories.[42]

Finally, a synthesis of the results on feasibility, measurement properties, and relevance was generated.

3 | RESULTS

3.1 | Patient characteristics

In total, 373 adult patients with hemophilia were invited to participate in the study, 162 patients (43%) signed informed consent, but 6 did not proceed to answer the questionnaires, and 14 reported relevant changes in their health status since participating in the HiN-6 study and were therefore excluded. Eventually, 142 adult patients with hemophilia were included and started to complete the PROMIS item banks resulting in a response rate of 38%. Of these 142 patients, 133 (94%) completed all nine PROMIS item banks. Patient characteristics are shown in Table 1. The mean age was 47.4 (range, 18-79) years, and 49% had severe hemophilia. One-third (34%) of the patients reported no comorbidities. The most common reported comorbidities were hepatitis C (51%), hypertension (20%), HIV (8%), hypercholesterolemia (8%), and (a history of) cancer (7%). The mean (SD) time between the data collection for the HiN-6 study (legacy instruments) and the current study (PROMIS item banks) was 2.2 (±0.3) years and varied from 1.0 to 2.6 years.

3.2 | PROMIS item banks and legacy instruments

3.2.1 | Feasibility

Table 2 presents data on the number of items and floor and ceiling effects of the nine Dutch PROMIS item banks. The mean number of questions answered per CAT item bank varied from 5.2 (range, 3-12) for satisfaction with participation to 8.7 (range, 5-12) for anxiety. In total, the legacy instruments contained 141 items, and the mean total number of PROMIS items completed was 57 (±13). Details on the number of items for the legacy instruments are shown in the Supporting Information.

Floor effects were observed in two PROMIS item banks: the CATs pain interference (26% minimum scores) and depression (18% minimum scores). Patients had to administer the maximum of 12 CAT items when reporting minimum scores. Ceiling effects were observed in two PROMIS item banks: the short forms self-efficacy medications (28% maximum scores) and self-efficacy symptoms (22% maximum scores). Details on the proportions of lowest and highest scores for

TABLE 1 Patient characteristics

Patient characteristics (n = 142)	Mean (SD) or %
Age (years)	47.4 (17.1)
Hemophilia A	86.5
Hemophilia severity	
Mild	33.1
Moderate	18.3
Severe	48.6
Prophylaxis	45.8
Inhibitor	
Current	2.1
Former	8.5
Comorbidities	
No comorbidities	34.7
HIV positive	7.7
Hepatitis C	
Current	1.4
Former	49.3
Unknown	0.7
Other comorbidities[a]	
1	22.5
>1	20.4

Abbreviation: SD, standard deviation.

[a]Other comorbidities included hypertension (20%), hypercholesterolemia (8%), cancer (ever) (7%), heart diseases other than angina pectoris (6%), diabetes type 2 (6%), lung diseases (5%), psychological complaints (4%), liver failure (2%), ischemic stroke (1%), hemorrhagic stroke (1%), osteoporosis (1%), and kidney disease (1%).

the legacy instruments are shown in the Supporting Information. Ceiling effects were observed for the RAND-36 domains physical functioning (26%), social functioning (49%), pain (28%), and role limitations due to physical health problems (64%), and for the HAL sum score (22%). Floor effects were observed for the VERITAS-Pro domains time (20%) and remember (22%).

3.2.2 | Measurement properties—construct validity and reliability

Results of construct validity and hypotheses testing of the PROMIS item banks compared to legacy instruments are shown in Table 3. For PROMIS CATs physical function and pain interference domains, correlations with the legacy instruments were strong and met the predefined criteria for convergent validity. For the PROMIS CAT satisfaction with participation domain, correlations were moderate and met the predefined criteria for convergent validity. The correlation between the PROMIS CAT fatigue and the RAND-36 energy/fatigue domains was −0.59, which was almost consistent with the hypothesis (r > −0.6) and considered as confirmed by the authors. The correlations between PROMIS CATs depression and anxiety and

TABLE 2 Feasibility and reliability of the PROMIS CATs and short forms: floor and ceiling effects and numbers of items completed

PROMIS item bank	Floor %	Ceiling %	Number of items Mean (SD)		Min	Max	Reliability (SE ≤ 3.2) %	n
Computer Adapted Tests (CATs)								
Physical function	-	0.7	6.0	3.5	3	12	95.1	142
Pain interference	26.1	-	6.1	4.2	3	12	73.9	142
Depression	17.6	-	8.6	3.2	5	12	82.4	142
Anxiety	11.4	-	8.7	2.7	5	12	87.1	140
Participation	-	9.4	6.4	2.9	3	12	90.6	139
Satisfaction with participation	-	2.2	5.2	2.4	3	12	97.8	139
Fatigue	2.2	-	6.4	2.5	4	12	100.0	138
Short forms								
Self-efficacy medications	1.5	27.8					43.6	133
Self-efficacy symptoms	0.7	22.4					70.1	134

Note: Participation: ability to participate in social roles and activities; satisfaction with participation: satisfaction with social roles and activities; self-efficacy medications, self-efficacy for managing medications and treatment; self-efficacy symptoms, self-efficacy for managing symptoms.

Abbreviations: CATs, Computer Adaptive Tests; PROMIS, Patient Reported Outcomes Measurement Information System; SD, standard deviation; SE, standard error.

the RAND-36 emotional well-being domain were moderate and did not meet the predefined criteria (r > −0.6). The correlations between the PROMIS CAT participation domain and the legacy instruments RAND-36 social functioning domain, RAND-36 domain role limitations due to physical health problems and the HAL complex lower-extremity component were weak to moderate and did not meet the predefined criteria (r > 0.6).

The correlation between the PROMIS short form domain self-efficacy symptoms and PAM-13 was weak, and the correlations between the PROMIS short form self-efficacy medications and VERITAS-Pro time and remember scales were negligible and did not meet the predefined criteria (r > 0.4).

The hypotheses regarding expected differences between subgroups were confirmed for the PROMIS CATs domains physical function and participation (Figure 1 and Table 3). Compared to patients with severe hemophilia, patients with nonsevere hemophilia had better physical function (53.0 vs 45.0; P < .001) and better ability to participate in social roles and activities (54.4 vs 50.5; P < .001). Compared to patients aged ≥30 years, patients aged 18 to 29 years had better physical function (57.9 vs 46.4; P < .001) and better ability to participate in social roles and activities (56.3 vs 51.3; P < .001). For the PROMIS physical function domain, the minimal important change is 2 to 8, and the differences were considered to be clinically relevant.[43] For the PROMIS participation domain, data on the minimal important change were not available.

The reliability varied between the different PROMIS item banks (Table 2). For all PROMIS CATs and short forms, >70% of the T scores were reliable (SE ≤ 3.2, 90% reliable), except for the PROMIS

short form domain self-efficacy medications (44%). The internal consistency of the legacy instruments was good, with Cronbach's α between 0.76 and 0.97. Details on the internal consistency for the legacy instruments are shown in the Supporting Information.

3.2.3 | Relevance

Table 4 presents the T scores for the PROMIS item banks. In addition, Figures 2 and 3 show the distribution of the scores according to the score cutoff values. The PROMIS CAT domain physical function (38%) was most frequently scored as limited. Adult patients with hemophilia reported lower scores than the general Dutch male population for physical function and satisfaction with participation. For all other PROMIS item banks, adult patients with hemophilia scored similar or better, compared to the general population.

3.3 | Synthesis of results on feasibility, measurement properties and relevance

Table 5 presents a synthesis of the results on feasibility, measurement properties, and relevance for the item banks. The number of items for the PROMIS CATs was lower than the entire legacy instruments, but on domain level the number of items was similar or higher, except for the PROMIS CAT domain physical function. Minimum and maximum scores occurred equally or less frequently in the PROMIS CATs than in the legacy instruments, except for

TABLE 3 Predefined hypotheses and results of validity testing according to PROMIS item banks show that convergent validity was confirmed for the PROMIS CATs domains physical function, pain interference, and satisfaction with participation and known-group validity was confirmed for the PROMIS CATs domains physical function and participation

Convergent validity				
PROMIS item bank	Legacy instrument	Predefined correlation	Spearman's correlation	Confirmed (Yes/No)
Physical function	RAND-36 Physical functioning	>0.6	0.85	Yes
	HAL	>0.4	0.84	Yes
	HEP-test-Q	>0.6	0.81	Yes
Pain interference	RAND-36 Pain	>−0.6	−0.72	Yes
Depression	RAND-36 Emotional well-being	>−0.6	−0.52	No
Anxiety	RAND-36 Emotional well-being	>−0.6	−0.46	No
Participation	RAND-36 Social functioning	>0.6	0.39	No
	RAND-36 Role limitations due to physical health problems	>0.6	0.44	No
	HAL complex lower extremity	>0.6	0.44	No
Satisfaction with participation	RAND-36 Social functioning	>0.4	0.46	Yes
Fatigue	RAND-36 Energy/fatigue	>−0.6	−0.59	No
Self-efficacy medications	VERITAS-Pro Time	>−0.4	−0.08	No
	VERITAS-Pro Remember	>−0.4	0.01	No
Self-efficacy symptoms	PAM-13	>0.4	0.37	No
Known-group validity				
PROMIS item bank	Differences between			
Physical function	• Severe and nonsevere hemophilia • Young adults (18-29 years) and adults (≥30 years)			Yes
Participation	• Severe and nonsevere hemophilia • Young adults (18–29 years) and adults (≥30 years)			Yes

Note: In the nonsevere categories, all persons with mild and moderate hemophilia were included.

Participation: ability to participate in social roles and activities; satisfaction with participation: satisfaction with social roles and activities; self-efficacy medications: self-efficacy for managing medications and treatment; self-efficacy symptoms: self-efficacy for managing symptoms.

Abbreviations: HAL, Haemophilia Activities List; HEP-test-Q, Haemophilia & Exercise Project-Test-Questionnaire; PAM-13, Patient Activation Measure-13; PROMIS, Patient Reported Outcomes Measurement Information System; VERITAS-Pro, Validated Hemophilia Regimen Treatment Adherence Scale – Prophylaxis.

the PROMIS CAT domain depression. Convergent validity of the PROMIS CATs domains physical function, pain interference, satisfaction with participation, and fatigue was confirmed by hypothesis testing. Convergent validity of the PROMIS CATs domains depression, anxiety, and participation was, in this study, not confirmed. For the PROMIS CAT domains participation and physical function, known-group validity was confirmed, as both were able to discriminate between different age and severity categories. The reliability of the CATs was good. The PROMIS CAT domain physical function was considered to be most relevant for adult patients with hemophilia, as most limitations were reported in this domain. The PROMIS short form domain self-efficacy symptoms was reliable and shorter than the PAM-13, but showed a considerable ceiling effect and convergent validity was not confirmed. The PROMIS short form domain self-efficacy medications was not a feasible and reliable alternative to the VERITAS-Pro and measured a different construct.

4 | DISCUSSION

This study aimed to determine the feasibility, validity, and relevance of nine PROMIS item banks in 142 adult Dutch adults with hemophilia. The PROMIS CATs were considered to be feasible, with a low number of items and limited floor effects. The number of CAT items (mean number of CAT items, 5-9) was substantially lower than in the legacy instruments, which varies from 13 items for the entire PAM-13 to 42 for the entire HAL. The PROMIS CAT domain physical function was more feasible than the legacy instruments and was most relevant for adult patients with hemophilia. In addition, the PROMIS CATs domains pain interference, satisfaction with participation, and fatigue were feasible alternatives to the legacy instruments. The PROMIS CAT domain participation was a feasible tool to discriminate between different age and severity categories. The PROMIS CATS on mental health did not meet the predefined correlation criteria with the legacy instruments. The current results
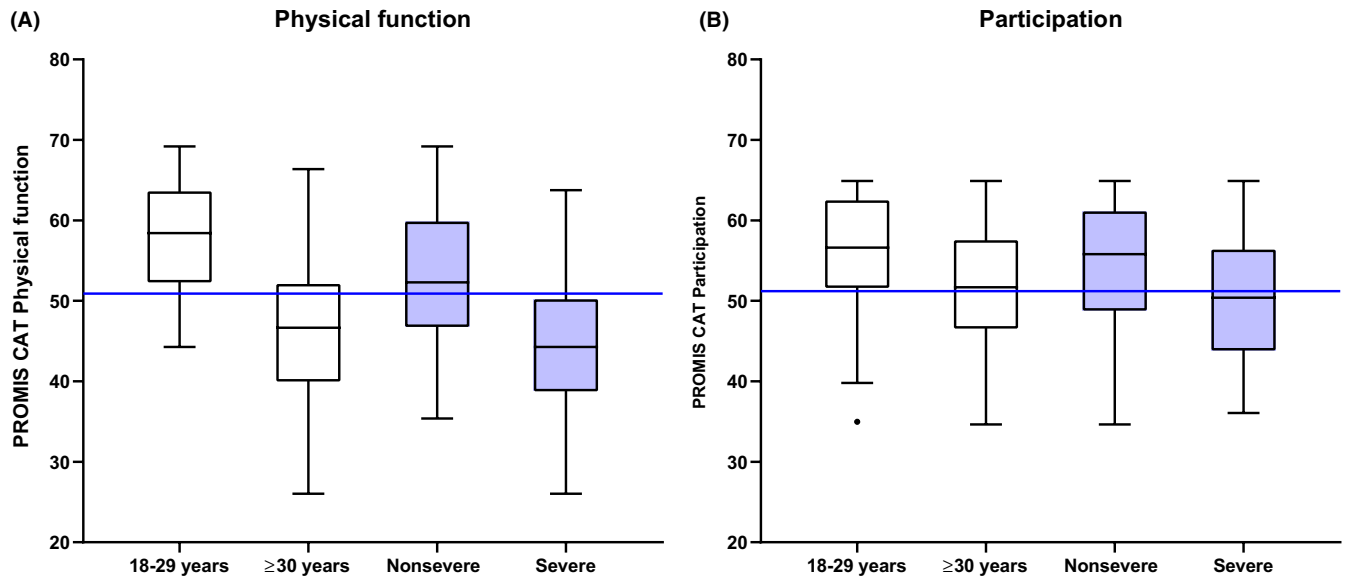
FIGURE 1 T-scores on the PROMIS CAT (A) physical function and (B) participation according to age and hemophilia severity. The blue lines show the mean score of the general adult Dutch male population on the PROMIS CAT domains physical function (50.9) and participation (51.2). Participation: ability to participate in social roles and activities. CAT, Computer Adaptive Test; PROMIS, Patient Reported Outcomes Measurement Information System

TABLE 4 T scores on the PROMIS CATs and short forms

PROMIS item bank	PROMIS T score			
	Mean (SD)	Min	Max	n
Computer Adapted Tests (CATs)				
Physical function	49.1 9.5	26.0	69.2	142
Pain interference	51.0 7.7	41.0	70.2	142
Depression	47.3 7.5	37.1	68.9	142
Anxiety	47.8 7.7	35.9	79.7	140
Participation	52.5 8.2	34.7	64.9	139
Satisfaction with participation	50.0 7.2	29.3	65.7	139
Fatigue	46.9 9.2	28.8	74.2	138
Short forms				
Self-efficacy medications	49.9 9.5	19.0	60.6	133
Self-efficacy symptoms	51.8 8.7	23.2	63.5	134

Note: Interpretation: PROMIS total scores are calculated by transforming the item-scores into T scores with 50 (based on the US population mean) with a SD of 10. For all item banks, higher scores represent more of the construct (eg, more pain interference or better physical function).

Participation: ability to participate in social roles and activities; satisfaction with participation: satisfaction with social roles and activities; self-efficacy medications: self-efficacy for managing medications and treatment; self-efficacy symptoms: self-efficacy for managing symptoms.

Abbreviations: CATs, computer adaptive tests; PROMIS, Patient Reported Outcomes Measurement Information System; self-efficacy symptoms, self-efficacy for managing symptoms.

do not support the use of the PROMIS short forms on self-efficacy in adults with hemophilia.

4.1 | Internal and external validity

The generalizability of the study to other populations with comparable treatment regimens was promoted by inclusion of a heterogeneous group of adult Dutch patients with hemophilia aged 18 to 79 years of all severities. However, a higher proportion of patients with severe hemophilia (49%) were included in the current study in comparison to all Dutch patients with hemophilia (33%).[22] In addition, the effect of data collection in an online survey on the generalizability was unclear.

The choice of legacy instruments is an important factor in testing convergent validity. However, the legacy instruments were already collected for the HiN-6 study and were the best available legacy data. For the PROMIS short form domain self-efficacy medications, higher correlations with the VERITAS-Pro were expected, although the focus of the PROMIS short form domain self-efficacy medications is more on confidence in managing medication schedules and the VERITAS-Pro on adherence to prophylactic treatment in hemophilia.[32] Besides a narrow data range and ceiling effects, which always lower correlations, the lack of correlation may have been affected by the differences between the management of medications for patients with hemophilia compared to other diseases as well as by the multifactorial character of adherence to prophylaxis.[39,44] The correlation between the PROMIS short form domain self-efficacy symptoms and PAM-13 was also lower than expected. This may be explained by a difference in focus of these instruments: where the PROMIS short form domain self-efficacy symptoms focuses on
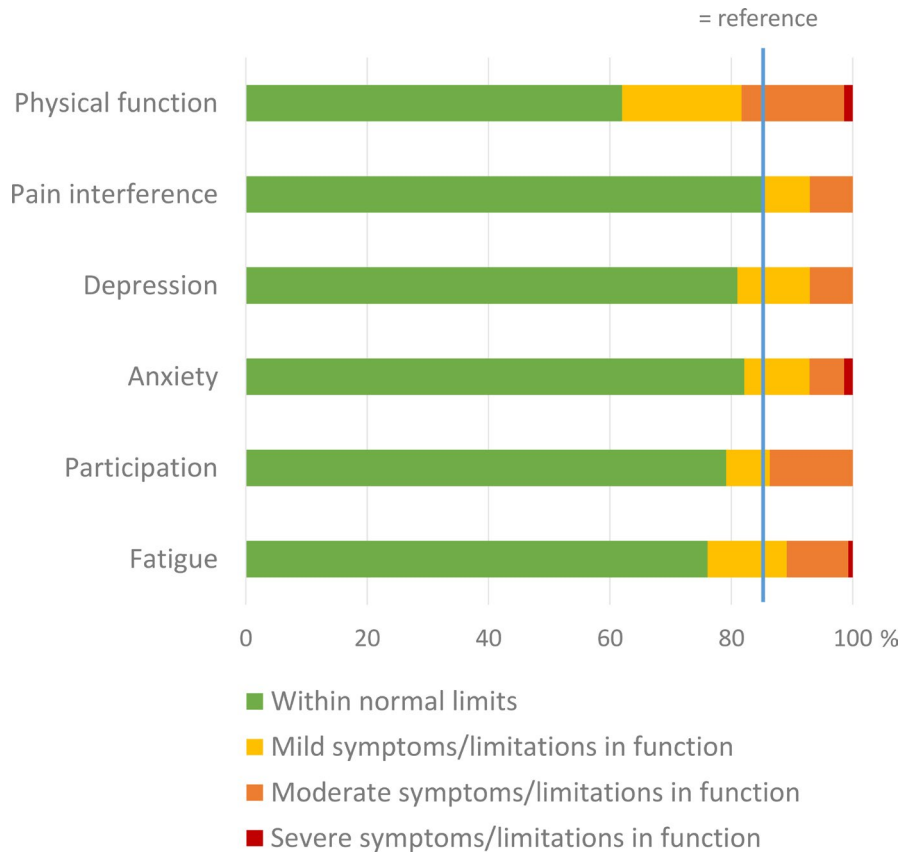
FIGURE 2 Scores on six PROMIS CATs in patients with hemophilia. PROMIS T scores were presented in four categories according to score cutoff points: within normal limits (green), mild (0.5 SD) (yellow), moderate (1 SD) (orange), and severe (2 SD) (red) symptoms/limitations in function. As depicted, 84% (blue line) of the general adult Dutch male population scores within normal limits or mild symptoms. Participation: ability to participate in social roles and activities. Interpretation: For example, for the PROMIS CAT domain physical function, 62% of the patients with hemophilia scored within normal limits (green), 19% reported mild limitations (yellow), 18% reported moderate limitations (orange), and 1% reported severe limitations (red). For the general Dutch male population, 84% scored within the normal limits or mild limitations (blue reference line). In conclusion, patients with hemophilia reported more moderate/severe limitations (19%) on physical functioning than the general Dutch male population (16%). CAT, Computer Adaptive Test; PROMIS, Patient Reported Outcomes Measurement Information System
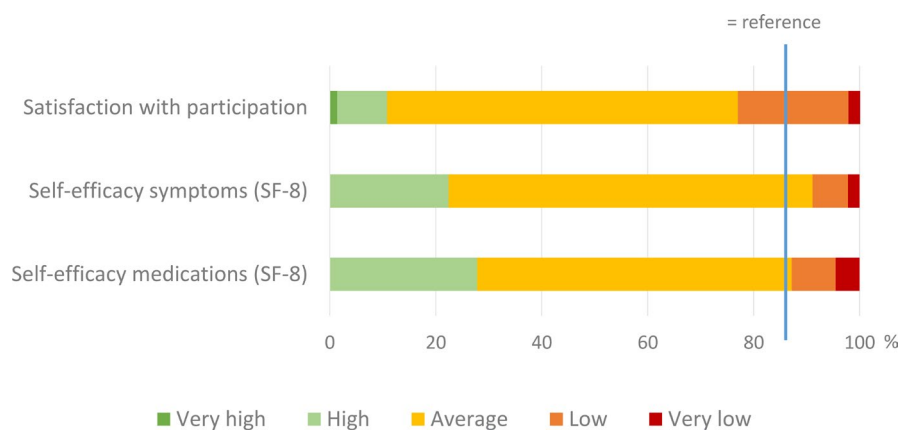

FIGURE 3 Scores on three PROMIS CATs and short forms in patients with hemophilia. PROMIS T scores were presented in five categories according to score cutoff points: very high (+2 SD) (green), high (+1 SD) (light green), average, low (−1 SD) (orange), and very low (−2 SD) (red). As depicted, 84% (blue line) of the adult Dutch male (satisfaction with participation) or general US population (self-efficacy short forms) scores within the very high, high, or average categories. Satisfaction with participation: satisfaction with social roles and activities. CAT, Computer Adaptive Test; PROMIS, Patient Reported Outcomes Measurement Information System

rpth
research & practice
in thrombosis & haemostasis

TABLE 5 Synthesis of the results on feasibility, measurement properties and relevance of the PROMIS CATs and short forms shows that the PROMIS CATs domains physical function, pain interference, satisfaction with participation, and fatigue were feasible alternatives to the legacy instruments

PROMIS item bank	Legacy instrument	Feasibility			Measurement properties				Relevance
		N items PROMIS	N items legacy instruments	Floor/ ceiling	Convergent validity	Known-group validity	Reliability		
Physical function	RAND-36 Physical functioning	6.0	10	+	+	+	++		+
	HAL		42						
	HEP-test-Q		25						
Pain interference	RAND-36 Pain	6.1	2	±	+	N/A	+		–
Depression	RAND-36 Emotional well-being	8.6	5	–	–	N/A	+		–
Anxiety	RAND-36 Emotional well-being	8.7	5	+	–	N/A	+		–
Participation	RAND-36 Social functioning	6.4	2	+	–	+	++		–
	RAND-36 RP		4						
	HAL complex lower extremity		9						
Satisfaction with participation	RAND-36 Social functioning	5.2	2	+	+	N/A	++		+
Fatigue	RAND-36 Energy/fatigue	6.4	4	+	+	N/A	++		–
Self-efficacy medications	VERITAS-Pro Time	8	4	–	–	N/A	–		–
	VERITAS-Pro Remember		4						
Self-efficacy symptoms	PAM-13	8	13	–	–	N/A	+		–

Participation: ability to participate in social roles and activities; satisfaction with participation: satisfaction with social roles and activities; self-efficacy medications: self-efficacy for managing medications and treatment; self-efficacy symptoms: self-efficacy for managing symptoms.

Abbreviations: CATs, computer adaptive tests; HAL, Haemophilia Activities List; HEP-test-Q, Haemophilia & Exercise Project-Test-Questionnaire; PAM-13, Patient Activation Measure-13; PROMIS, Patient Reported Outcomes Measurement Information System; RP, role limitations due to physical health problems; VERITAS-Pro, Validated Hemophilia Regimen Treatment Adherence Scale – Prophylaxis.

Interpretation: N items: for the CATs the mean number of CAT items was shown. Floor/ceiling: + no floor- and/or ceiling effects, ± similar floor- and/or ceiling effects as legacy instrument, - more floor- and/or ceiling effects than legacy instrument. Convergent validity: + predefined hypothesis was confirmed, - predefined hypothesis was not confirmed. Known-group validity: + predefined hypothesis was confirmed, - predefined hypothesis was not confirmed. Reliability: ++ >90% of the scores was reliable (SE ≤ 3.2), + >70% of the scores was reliable (SE ≤ 3.2), - <70% of the scores was reliable (SE ≤ 3.2). Relevance: + patients with hemophilia had more limited scores than the general population, - patients with hemophilia had fewer or similar limited scores than the general population.

confidence in managing symptoms, the PAM-13 has a broader view on self-management.[33]

In addition, the comparison of PROMIS item banks with legacy instruments may have been negatively affected by the extended interval (1.0-2.6 years) between the assessments collected for the HiN-6 and PROMIS studies. However, we did observe a high correlation for the domain physical function after exclusion of patients with major health changes, which were identified by an anchor question focused on physical health. The lack of correlation for the PROMIS CATs domains anxiety and depression could be attributable to changes in mental health during COVID-19 or the high scores, as the symptoms evaluated within these domains are generally less prevalent in hemophilia.[40,45,46] The lack of correlation between the PROMIS CAT domain participation and the legacy instruments was also affected by high scores and a narrow data range, which was also observed for the participation item bank in the Dutch general population.[16,47]

A limitation of the study is that reliability of the PROMIS item banks (SE) could not be compared with the legacy instruments (Cronbach's α), which is a result of different measurement theories for the legacy instruments and PROMIS (CTT vs IRT). However, it is expected that the PROMIS item banks measure more precisely at the lower and upper ends of the score ranges.[8] For example, the RAND-36 domain role limitations due to physical health problems consists of only two items and had a large ceiling effect, which will result in less measurement precision. In contrast, the total PROMIS participation item bank consists of 35 items and a selection of relevant items will be used in the CAT.

Finally, the use of reference data from the general population influenced the distribution of the categories of the PROMIS T scores. Proportions of abnormal scores were similar or lower than in Dutch men or the general US population, except for physical function and satisfaction with participation. This may be explained by a tendency that patients with lifelong conditions like hemophilia report higher health states than the general population, known as the disability paradox, suggesting the impact of hemophilia may be underestimated if general population references are used.[48] In the absence of Dutch male reference data, reference data of the general US population were used for the PROMIS item banks self-efficacy medications and self-efficacy symptoms, which may have affected the results.[49]

4.2 | Comparison with other studies

The reported floor and ceiling effects for the legacy instruments were comparable to earlier reports of the HAL and SF-36 in Dutch and Swedish patients with hemophilia.[11-13] The PROMIS T scores (physical function, pain interference, depression, anxiety, participation, and fatigue) in the current study were also comparable to T scores in North American patients with hemophilia, although in North American patients higher correlations were reported between mental health domains and the EQ-5D-5L anxiety/depression.[21] In addition, strong correlations were reported between the PROMIS-29 domains depression and anxiety and the RAND-36 emotional well-being domain in the HiN-6 study, with data collected at the same time point.[50]

In the current study, the PROMIS CAT domain pain interference was limited in only 15% of the patients, which may be a result of a high reference value in the general Dutch male population (mean, 54.7) in contrast to 50 for the general US population. These findings are in contrast with reports of increased pain in a European study in 903 patients with hemophilia (age, 36; 35% receiving prophylaxis) and a recent study in 46 young Canadians (weighted mean age, 21; all receiving prophylaxis), measured with the SF-36.[46,51] Using US population references would have resulted in a score for increased pain interference in 33% of adults with hemophilia.

The current results partly support the recommendations of the recent HaemoValue initiative. Based on expert opinion only, the core outcome set for hemophilia care includes five of the currently investigated PROMIS item banks—physical function, pain interference, depression, anxiety, and participation—but excluded fatigue and satisfaction with participation.[52]

Finally, the recently developed Patient Reported Outcomes, Burdens and Experiences (PROBE) questionnaire was not included in the current study as this questionnaire was not part of the HiN-6 study. The PROBE provides the opportunity to measure patient-important reported outcomes (demographic data, general health problems, hemophilia-related health problems, and health-related quality of life) in patients with hemophilia and people without a bleeding disorders.[53] In comparison to PROBE, PROMIS has the advantage of application of CATs as well as item banks specifically developed for children. A disadvantage of PROMIS is the lack of hemophilia-specific items. Both arguments must be considered when choosing between patient-reported outcome instruments in studies and clinical practice.

4.3 | Clinical implications and future research

Why should we use PROMIS above the legacy instruments in patients with hemophilia? As 66% of patient with hemophilia reported ≥1 comorbidity, the use of generic PROMIS item banks will be an efficient tool for outcome assessment while including the ability to consider effects of and/or comparison according to comorbidities. For research purposes especially, the PROMIS CAT domain physical function is more feasible than the legacy instruments and is relevant to patients when assessing disabilities at group level. In addition, the PROMIS CAT domain participation may be considered useful for research purposes for comparison of groups of patients with hemophilia. However, in day-to-day care for individual patients, all health domains may be of interest in a comprehensive care setting. The PROMIS CATs domains pain interference, satisfaction with participation, and fatigue are expected to result in more precise measurement in the lower and upper ends of the score range with more relevant items for each individual patient, in comparison with the RAND-36 with only a few items on each domain. In addition, based on the current data the PROMIS short forms on self-efficacy for managing chronic conditions are not recommended for use in hemophilia care. Possibly due to the study design (2.2 years between questionnaires and COVID-19 pandemic) the current results do not support the use of the PROMIS CATs depression and anxiety domains as an alternative to the RAND-36 emotional well-being domain (5 items, Cronbach's $\alpha = 0.87$). This is in contrast with results from other studies that do support the use of PROMIS depression and anxiety domains in patients with hemophilia.[21,50]

What work should be done before implementation of PROMIS CATs in day-to-day care and research? Several issues need to be addressed. First, further testing of smallest detectable changes and minimal important changes of PROMIS item banks is needed to improve the interpretability of scores in a setting of routine follow-up assessment. Second, the stopping rule of PROMIS CATs should be evaluated to improve feasibility, as people had to administer the maximum of 12 CAT items when they had no pain or depression symptoms. Finally, good facilities for digital administration of CATs like a PROMs mobile app or routine data collection from the electronic medical records are essential. Especially if IT facilities and budget for using CATs are limited, PROMIS short forms are an alternative for the CATs.

5 | CONCLUSION

PROMIS CATs are feasible and may lower the burden of outcome assessment by reducing the number of questions needed to assess various aspects of health compared to legacy instruments. The PROMIS CATs domains physical function, pain interference, satisfaction with participation, and fatigue are feasible, reliable, and valid alternatives

to legacy instruments for adult patients with hemophilia, with a low number of items and low floor and ceiling effects. For the implementation of PROMIS CATs in hemophilia care with lifelong routine assessment, data on the smallest detectable changes and minimal important changes and validation in children and young adults are essential.

AUTHOR CONTRIBUTIONS

IARK, LT, LFDvV, MAT, MC, SCG, MP, MJHAK, MHC, LH, and KF contributed to the design of the study. IARK performed the statistical analyses. IARK, LT, LH, and KF wrote the first draft of the paper. All authors contributed to interpretation of the data, modification of statistical analyses, and the writing of the manuscript.

ORCID

Isolde A. R. Kuijlaars https://orcid.org/0000-0003-2920-2258
Lorynn Teela https://orcid.org/0000-0002-8067-2792
Lize F. D. van Vulpen https://orcid.org/0000-0003-3242-5524
Merel A. Timmer https://orcid.org/0000-0003-1910-999X
Michiel Coppens https://orcid.org/0000-0001-6891-9062
Samantha C. Gouw https://orcid.org/0000-0002-1957-4122
Marieke J. H. A. Kruip https://orcid.org/0000-0002-0265-4871
Marjon H. Cnossen https://orcid.org/0000-0003-1557-2995
Evelien S. van Hoorn https://orcid.org/0000-0001-9661-2746
Lotte Haverman https://orcid.org/0000-0001-7849-0562
Kathelijn Fischer https://orcid.org/0000-0001-7126-6613

TWITTER

Isolde A. R. Kuijlaars @IsoldeKuijlaars
Lize F. D. Vulpen @lizevanvulpen
Marjon H. Cnossen @CnossenM
Lotte Haverman @LotteHaverman

REFERENCES

1. Srivastava A, Santagostino E, Dougall A, et al. WFH guidelines for the management of hemophilia, 3rd edition. *Haemophilia.* 2020;26(S6):1-158. https://doi.org/10.1111/hae.14046
2. Miesbach W, O'Mahony B, Key NS, Makris M. How to discuss gene therapy for haemophilia? A patient and physician perspective. *Haemophilia.* 2019;25(4):545-557.
3. Oldenburg J, Mahlangu JN, Kim B, et al. Emicizumab prophylaxis in hemophilia A with inhibitors. *N Engl J Med.* 2017;377(9):809-818.
4. Versloot O, Timmer MA, de Kleijn P, et al. Sports participation and sports injuries in Dutch boys with haemophilia. *Scand J Med Sci Sport.* 2020;30(7):1256-1264.
5. Shapiro S, Makris M. Haemophilia and ageing. *Br J Haematol.* 2019;184(5):712-720.
6. Manco-Johnson MJ, Warren BB, Buckner TW, Funk SM, Wang M. Outcome measures in haemophilia: beyond ABR (annualized bleeding rate). *Haemophilia.* 2021;27(S3):87-95.
7. Cella D, Riley W, Stone A, et al. Initial adult health item banks and first wave testing of the Patient-Reported Outcomes Measurement Information System (PROMIS™) network: 2005–2008. *J Clin Epidemiol.* 2010;63(11):1179-1194.
8. Fries JF, Witter J, Rose M, Cella D, Khanna D, Morgan-DeWitt E. Item response theory, computerized adaptive testing, and PROMIS: assessment of physical function. *J Rheumatol.* 2014;41(1):153-158.
9. Kollmorgen RC, Hutyra CA, Green C, Lewis B, Olson SA, Mather RC. Relationship between PROMIS computer adaptive tests and legacy hip measures among patients presenting to a tertiary care hip preservation center. *Am J Sports Med.* 2019;47(4):876-884.
10. Cella D, Choi SW, Condon DM, et al. PROMIS® adult health profiles: efficient short-form measures of seven health domains. *Value Heal.* 2019;22(5):537-544. doi:10.1016/j.jval.2019.02.004
11. Kuijlaars IAR, van Emst M, van der Net J, Timmer MA, Fischer K. Assessing the test–retest reliability and smallest detectable change of the Haemophilia Activities List. *Haemophilia.* 2021;27(November):108-112.
12. Fischer K, Poonnoose P, Dunn AL, et al. Choosing outcome assessment tools in haemophilia care and research: a multidisciplinary perspective. *Haemophilia.* 2017;23:11-24.
13. Fischer K, Nijdam A, Holmström M, et al. Evaluating outcome of prophylaxis in haemophilia: objective and self-reported instruments should be combined. *Haemophilia.* 2016;22(2):e80-e86. https://doi.org/10.1111/hae.12901
14. Driban JB, Morgan N, Price LL, Cook KF, Wang C. Patient-Reported Outcomes Measurement Information System (PROMIS) instruments among individuals with symptomatic knee osteoarthritis: a cross-sectional study of floor/ceiling effects and construct validity. *BMC Musculoskelet Disord.* 2015;16(253):1-9. https://doi.org/10.1186/s12891-015-0715-y

15. Terwee CB, Roorda LD, De Vet HCW, et al. Dutch-Flemish translation of 17 item banks from the Patient-Reported Outcomes Measurement Information System (PROMIS). *Qual Life Res.* 2014;23(6):1733-1741.

16. Terwee CB, Crins MHP, Boers M, de Vet HCW, Roorda LD. Validation of two PROMIS item banks for measuring social participation in the Dutch general population. *Qual Life Res.* 2019;28(1):211-220. https://doi.org/10.1007/s11136-018-1995-0

17. Crins MHP, Roorda LD, Smits N, et al. Calibration and validation of the Dutch-Flemish PROMIS pain interference item bank in patients with chronic pain. *PLoS One.* 2015;10(7):1-18.

18. Crins MHP, van der Wees PJ, Klausch T, van Dulmen SA, Roorda LD, Terwee CB. Psychometric properties of the PROMIS Physical Function item bank in patients receiving physical therapy. *PLoS One.* 2018;13(2):1-14.

19. Kempton CL, Michaels Stout M, Barry V, et al. Validation of a new instrument to measure disease-related distress among patients with haemophilia. *Haemophilia.* 2021;27(1):60-68.

20. Pinto P, Paredes A, Pedras S, et al. Sociodemographic, clinical, and psychosocial characteristics of people with hemophilia in Portugal: findings from the first national survey. *TH Open.* 2018;02(01):e54-67.

21. Barry V, Buckner TW, Lynch ME, et al. An evaluation of PROMIS health domains in adults with haemophilia: a cross-sectional study. *Haemophilia.* 2021;27(3):375-382.

22. Hassan S, van Balen EC, Smit C, et al. Fifty years of hemophilia treatment in the Netherlands (accepted for publication). *J Thromb Haemost.* 2021;19(10):2394–2406.

23. Mokkink LB, Patrick DL, Alonso J, Bouter LM, Terwee CB. COSMIN Study Design checklist for Patient-reported outcome measurement instruments. 2019;(July).

24. van Genderen FR, Westers P, Heijnen L, et al. Measuring patients' perceptions on their functional abilities: validation of the Haemophilia Activities List. *Haemophilia.* 2006;12(1):36-46. https://doi.org/10.1111/j.1365-2516.2006.01186.x

25. van Genderen FR, van Meeteren NLU, van der Bom JG, et al. Functional consequences of haemophilia in adults: the development of the Haemophilia Activities List. *Haemophilia.* 2004;10(5):565-571.

26. Brodin E, Baghaei F, Elfvinger P, Lindvall K, Sunnerhagen KS. The Swedish version of the haemophilia activity list. *Haemophilia.* 2011;17(4):662-668.

27. Limperg PF, Terwee CB, Young NL, et al. Health-related quality of life questionnaires in individuals with haemophilia: a systematic review of their measurement properties. *Haemophilia.* 2017;23(4):497-510.

28. VanderZee KI, Sanderman R, Heyink JW, De Haes H. Psychometric qualities of the RAND 36-item health survey 1.0: a multidimensional measure of general health status. *Int J Behav Med.* 1996;3(2):104-122.

29. van der Zee KI, Sanderman R. Het meten van de algemene gezondheidstoestand met de Rand-36; 2012.

30. Solovieva S, Santavirta N, Santavirta S, Konttinen YT. Assessing quality of life in individuals with hereditary blood coagulation disorders. *Qual Life Res.* 2004;13(5):987-1000.

31. Von Mackensen S, Czepa D, Herbsleb M, Hilberg T. Development and validation of a new questionnaire for the assessment of subjective physical performance in adult patients with haemophilia - the HEP-Test-Q. *Haemophilia.* 2010;16(1):170-178.

32. Duncan NA, Kronenberger W, Roberson C, Shapiro A. VERITAS-Pro: a new measure of adherence to prophylactic regimens in haemophilia. *Haemophilia.* 2010;16(2):247-255.

33. Hibbard JH, Mahoney ER, Stockard J, Tusler M. Development and testing of a short form of the patient activation measure. *Health Serv Res.* 2005;40(6p1):1918-1930.

34. Rademakers J, Maindal HT, Steinsbekk A, Gensichen J, Brenk-Franz K, Hendriks M. Patient activation in Europe: an international comparison of psychometric properties and patients' scores on the short form Patient Activation Measure (PAM-13). *BMC Health Serv Res.* 2016;16(1):1-7. https://doi.org/10.1186/s12913-016-1828-1

35. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res.* 1995;4(4):293-307.

36. Oude Voshaar MAH, ten Klooster PM, Glas CAW, et al. 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. *Rheumatol (United Kingdom).* 2015;54(12):2221-2229.

37. Kroenke K, Baye F, Lourens SG. Comparative validity and responsiveness of PHQ-ADS and other composite anxiety-depression measures. *J Affect Disord.* 2018;2019(246):437-443. https://doi.org/10.1016/j.jad.2018.12.098

38. Van Der Meij E, Anema JR, Huirne JAF, Terwee CB. Using PROMIS for measuring recovery after abdominal surgery: a pilot study. *BMC Health Serv Res.* 2018;18(1):1-9.

39. Bingham CO, Gutierrez AK, Butanis A, et al. PROMIS Fatigue short forms are reliable and valid in adults with rheumatoid arthritis. *J Patient-Reported Outcomes.* 2019;3(1):14.

40. Schober P, Schwarte LA. Correlation coefficients: appropriate use and interpretation. *Anesth Analg.* 2018;126(5):1763-1768.

41. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales. A Practical Guide to Their Development and Use*, 5th ed. Oxford University Press; 2014.

42. Health Measures. *PROMIS® Score Cut Points* [Internet]. [cited 2021 Apr 13]. Available from: https://www.healthmeasures.net/score-and-interpret/interpret-scores/promis/promis-score-cut-points

43. Hung M, Bounsanga J, Voss MW, Saltzman CL. Establishing minimum clinically important difference values for the Patient-Reported Outcomes Measurement Information System Physical Function, hip disability and osteoarthritis outcome score for joint reconstruction, and knee injury and osteoarthritis outcome score for joint reconstruction in orthopaedics. *World J Orthop.* 2018;9(3):41-49.

44. Schrijvers LH, Kars MC, Beijlevelt-van der Zande M, Peters M, Schuurmans MJ, Fischer K. Unravelling adherence to prophylaxis in haemophilia: a patients' perspective. *Haemophilia.* 2015;21(5):612-621.

45. Vindegaard N, Benros ME. COVID-19 pandemic and mental health consequences: systematic review of the current evidence. *Brain Behav Immun.* 2020;89(May):531-542. https://doi.org/10.1016/j.bbi.2020.05.048

46. Royal S, Schramm W, Berntorp E, et al. Quality-of-life differences between prophylactic and on-demand factor replacement therapy in European haemophilia patients. *Haemophilia.* 2002;8(1):44-50.

47. van Leeuwen LM, Tamminga SJ, Ravinskaya M, et al. Proposal to extend the PROMIS® item bank v2.0 "Ability to Participate in Social Roles and Activities": item generation and content validity. *Qual Life Res.* 2020;29(10):2851-2861. https://doi.org/10.1007/s11136-020-02540-3

48. O'Hara J, Martin AP, Nugent D, et al. Evidence of a disability paradox in patient-reported outcomes in haemophilia. *Haemophilia.* 2021;(February):245-252.

49. Terwee CB, Crins MHP, Roorda LD, et al. International application of PROMIS computerized adaptive tests: US versus country-specific item parameters can be consequential for individual patient scores. *J Clin Epidemiol.* 2021;134:1-13.

50. Balen EC, Haverman L, Hassan S, et al. Validation of PROMIS Profile-29 in adults with hemophilia in the Netherlands. *J Thromb Haemost.* 2021;19(11):2687-2701.

51. St-Louis J, Urajnik DJ, Ménard F, et al. Generic and disease-specific quality of life among youth and young men with hemophilia in Canada. *BMC Hematol.* 2016;16(1):13. https://doi.org/10.1186/s12878-016-0052-x

52. van Balen EC, O'Mahony B, Cnossen MH, et al. Patient-relevant health outcomes for hemophilia care: development of an international standard outcomes set. *Res Pract Thromb Haemost.* 2020;2021:1-13.

53. Skinner MW, Chai-Adisaksopha C, Curtis R, et al. The Patient Reported Outcomes, Burdens and Experiences (PROBE) Project: development and evaluation of a questionnaire assessing patient reported outcomes in people with haemophilia. *Pilot Feasibility Stud*. 2018;4(1):1-10.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.