On outcomes for hemophilia

Balen, E.C. van

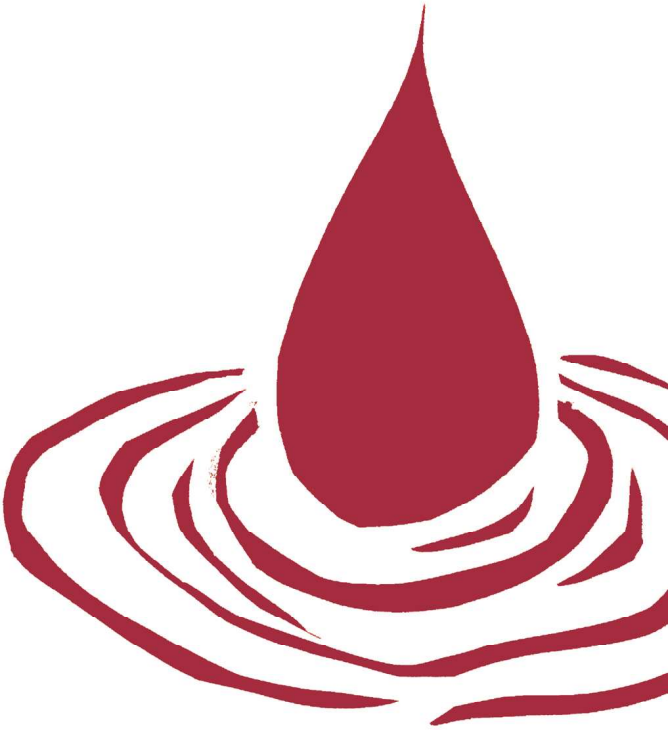
CHAPTER 6

Validation of PROMIS Profile-29 in adults with hemophilia in the Netherlands

Erna C. van Balen, Lotte Haverman, Shermarke Hassan, Elisabeth M. Taal, Cees Smit, Mariëtte H. Driessens, Erik A.M. Beckers, Michiel Coppens, Jeroen Eikenboom, Louise H. Hooimeijer, Frank W.G. Leebeek, Lize F.D. van Vulpen, Saskia E.M. Schols, Caroline B. Terwee, Frits R. Rosendaal, Johanna G. van der Bom, Samantha C. Gouw

J Thromb Haemost. 2021;19:2687–2701.

Lees dit artikel online:



Abstract

Background

The PROMIS Profile-29 questionnaire is widely used worldwide, but it has not yet been validated in the Netherlands, nor in persons with hemophilia. The aim of this study was to validate the Dutch-Flemish version of the PROMIS-29 Profile v2.01 in adults with hemophilia.

Methods

Dutch males with hemophilia (all severities) completed questionnaires that contained socio-demographic and clinical characteristics, the PROMIS-29, RAND-36, and the Hemophilia Activities List (HAL). Structural validity of each subscale was assessed with Confirmatory Factor Analysis (CFA). Internal consistency was calculated for each subscale with sufficient model fit in CFA. Construct validity was assessed by testing hypotheses about 1) correlations of each PROMIS-29 subscale with corresponding scales of RAND-36 and domains of HAL, and 2) mean differences in T-scores between subgroups with different hemophilia severities, self-reported joint impairment, and hiv infection status. We considered ≥75 percent of data in accordance with the hypotheses evidence for construct validity.

Results

In total, 770 persons with hemophilia participated in this cross-sectional study. CFA revealed sufficient structural validity for five subscales: Physical Function, Depression, Sleep Disturbance, Ability to Participate in Social Roles and Activities and Pain Interference. Internal consistency was high and Cronbach's alpha ranged from 0.79 for Sleep Disturbance to 0.96 for Pain Interference. Differences between clinical subgroups were in the expected direction. Construct validity was confirmed for Physical Function, Anxiety, Depression, Fatigue, Sleep Disturbance, and Pain Intensity.

Conclusion

This study revealed sufficient evidence for structural validity, internal consistency, and construct validity for most PROMIS Profile-29 subscales among people with hemophilia in the Netherlands.

Introduction

The congenital bleeding disorder hemophilia causes recurrent bleeds into joints and muscles due to a deficiency in coagulation factor VIII (hemophilia A) or factor IX (hemophilia B). The condition predominantly affects males and is classified into mild (0.05-0.40 IU/mL), moderate (0.01-0.05 IU/mL) and severe (<0.01 IU/mL) hemophilia, depending on the activity of factor VIII or IX. Individuals with severe hemophilia often suffer from spontaneous bleeds into joints and muscles, while those with mild hemophilia typically bleed when triggered by trauma or surgery.[1] Treatment consists of coagulation factor replacement by intravenous injection to treat bleeds (episodic treatment) or to prevent bleeds (prophylaxis, defined as regular administration of an hemostatic agent, usually administered intravenously or subcutaneously). Recently, non-factor replacement products have been marketed and gene therapy is currently under study.[1]

Early forms of treatment had devastating effects on the hemophilia community: through contaminated plasma-derived blood products, many patients were infected with hiv in the 1980s and / or hepatitis C (HCV) before the 1990s.[2] The availability of treatment has resulted in a near-normal life expectancy and improved outcomes,[3] but a potential side-effect of factor replacement therapy is the development of neutralizing antibodies ('inhibitors') against the infused coagulation factor. Regular prophylaxis with factor replacement products is not effective in patients with inhibitors, and since recently, prophylaxis with non-factor replacement products helps reduce the burden of bleeding.[1] In addition, joint damage (hemophilic arthropathy), pain and disability are still relatively common, especially among older males affected by severe hemophilia, due to recurrent joint bleeding. Large differences in joint status and pain exist between individuals. It is important to measure and monitor these outcomes in persons with hemophilia in order to personalize health care.

Patient-reported outcomes (PROs) are any aspect of a patient's health that come directly from the patient without interpretation of the patient's responses by a physician or anyone else.[4] In hemophilia, PROs have been measured with hemophilia-specific instruments such as the Hemophilia Activities List (HAL),[5, 6] Haemo-QoL-A [7] and Hemofilia-QoL [8] as well as with generic instruments such as the RAND-36 [9] or EQ-5D. Two systematic reviews reported that the measurement properties of hemophilia-specific instruments have not been studied sufficiently, in particular structural validity, responsiveness and hypothesis-testing.[10, 11] Whether to use disease-specific or generic tools for hemophilia PROs depends on the goal of measuring such outcomes.

An alternative approach to measuring patient-reported outcomes is to use generic instruments based on Item Response Theory (IRT), which has several advantages over other generic instruments. First, instruments using IRT-based scoring take the difficulty of items into account, thereby providing more valid and reliable scores.[12] Second, IRT-based item banks, consisting of large sets of questions, can be used as short forms of

any length (consisting of the best performing items from an item bank) or as computerized adaptive tests (CAT). In a CAT, the computer selects relevant questions based on the answer to the previous question, resulting in even more efficient and precise, but comprehensive assessment of a construct of interest. The use of patient-reported outcome measures (PROMs) in clinical practice is increasing. Using different PROMS for different patients and implementing many different PROMs in electronic health records may pose a burden on researchers and clinicians. Therefore, the availability of valid and precise generic PROMs for domains that are relevant across medical conditions (such as pain, fatigue, physical function) would be highly beneficial.

The Patient-Reported Outcomes Measurement Information System (PROMIS®), developed in the United States, is the most extensively validated measurement system of item banks in the world.[13-15] PROMIS profiles have been developed that consist of a collection of short forms derived from IRT-based item banks, covering seven patient-relevant domains. Profiles offer quick assessment of several domains of health-related quality of life (HRQoL).[16] Available profiles are the Profile-29, Profile-43 and Profile-57, which measure seven domains with 4, 6 or 8 items, respectively.[16] As a generic tool, PROMIS-29 has the advantage of making results comparable across diseases and the general population.[12]

Before using an instrument in a new population or language, it should be validated [4] by assessing its measurement properties. The measurement properties can be divided into three domains: validity (content validity, construct validity, hypotheses-testing), reliability (internal consistency, measurement error and test-retest reliability) and responsiveness.[17] A hierarchy of measurement properties can be defined.[18] Content validity is considered the most important measurement property, defined as the degree to which the content of an instrument is an adequate reflection of the construct to be measured.[18] It can be assessed in a qualitative study in which the relevance, comprehensiveness and comprehensibility of the items of a PROM are assessed, for example by cognitive debriefing in the target population.[19] The next measurement properties that should be evaluated are structural validity and internal consistency.[17] Structural validity is the degree to which the scores of an instrument are an adequate reflection of the dimensionality of the construct to be measured [18] and is assessed with confirmatory factor analysis.[4] Internal consistency is the degree of interrelatedness of items [18] as assessed with Cronbach's alpha.[4] Finally, other measurement properties are to be evaluated, such as test-retest reliability (the extent to which scores are stable over time in stable participants), construct validity (the degree to which the scores of an instrument are consistent with formulated hypotheses about relationships to scores of other instruments, or differences between relevant groups, based on the assumption that the instrument validly measures the construct to be measured), and responsiveness (the ability of an instrument to detect a change of the construct over time).

Item banks that underlie the PROMIS Profiles were translated into Dutch and showed sufficient linguistic, content and conceptual equivalence.[20] A next step is to evaluate the measurement properties of the item banks and their derivative short forms. PROMIS Profiles have been validated in several countries and in a number of conditions,[21-23] but not yet in hemophilia.

Therefore, this study aimed to validate the Dutch-Flemish version of the PROMIS-29 Profile v2.01 ('PROMIS-29') in Dutch adults with hemophilia by assessing its structural validity, internal consistency, and construct (convergent and discriminative) validity.

Methods

Data were collected as part of the Dutch nation-wide 'Hemophilia in the Netherlands 6' study (HiN-6). HiN-6 is the latest in a series of six cross-sectional studies that have been conducted since 1972.[3, 24, 25] Approval was obtained from the Medical Ethical Committee at Leiden University Medical Center, the Netherlands (registration number NL59114.058.17).

Participants and procedures

All adult males with mild, moderate or severe congenital hemophilia A or B with levels of Factor VIII of IX <0.40 IU/mL registered at one of the six Dutch hemophilia treatment centers were invited by letter to participate between June 2018 and July 2019.

Participants received a questionnaire through a secure e-mail link or in hard copy, depending on their preference. Answers were stored in the Castor Electronic Data Capture system.[26] Clinical characteristics were collected from electronic medical records. Participants signed written informed consent for extraction of data from electronic medical records, but this was not required for participation in the questionnaire.

Measures

Self-reported sociodemographic and clinical data collected through the questionnaire were: age, education level (categorized in ISCED levels [27]), and perceived impairment in joint function. Joint impairment was assessed with a single question that was used in previous HiN surveys. Joint impairment was defined as 'do you have any chronic joint problems due to hemophilia' (yes / no). Clinical characteristics collected from electronic medical records were type and severity of hemophilia, treatment type (prophylaxis, episodic), inhibitor status, and hiv and HCV status. Clinical characteristics were taken from medical records if the participant had signed written informed consent for use of these data. If medical record data were not available, self-reported data from the questionnaire were used. Hemophilia severity was known for all responders and non-responders.

Dutch-Flemish PROMIS-29 Profile v2.01

PROMIS Profiles are derived from full PROMIS item banks that were developed in the U.S. general population and patient groups.[13] PROMIS Profiles were shown to be reliable and correlate highly with full item banks.[16] The PROMIS-29 Profile v2.01 (PROMIS-29) measures seven domains of health-related quality of life (HRQoL) that are often considered important by patients:[16] Physical Function; Anxiety; Depression; Fatigue; Sleep Disturbance; Ability to Participate in Social Roles and Activities; and Pain Interference. Each domain is measured with four items. The PROMIS-29 also contains a single item on Pain Intensity, resulting in a total of 29 items. Each item is scored from 1 to 5; a higher score indicates a higher degree of the construct being measured. For the subscales Physical Function and Ability to Participate in Social Roles and Activities this means that a higher score indicates better HRQoL, while for the other subscales a higher score indicates worse HRQoL.[16] Domain scores were calculated as T-scores using the Health Measures Scoring Service,[28] resulting in a normalized score with a mean of 50 and a standard deviation of 10 in the reference population (the US general population). T-scores were only calculated for a domain if at least one item of that domain was completed; T-scores were considered missing if none of the items was completed.

RAND-36

RAND-36 version 1 is a generic measure that assesses health status using 36 items. It consists of eight health concepts with multi-item scales: Physical functioning (10 items); Social functioning (2 items); Role limitations caused by physical health problems (4 items); Role limitations caused by emotional problems (3 items); Emotional well-being (5 items); Pain (2 items); General health perceptions (5 items); Energy / Fatigue (4 items); and an additional single item measuring Change in perceived health during the past 12 months.[29] Items were scored on a three to six point Likert scale. As per the standard scoring instructions, subscale scores were calculated if a participant had completed at least half of the items of that subscale.[30] If fewer than half of the items were completed, subscale scores were considered missing. Subscale scores were converted to a 0-100 point scale.[9] A higher score indicates a better health status. The RAND-36 was reported to have good internal consistency and discriminative validity in the Dutch general population [31] and in several hemophilia populations.[32, 33]

Hemophilia Activities List (HAL)

The HAL version 2.0 is a hemophilia-specific instrument, developed in the Netherlands, that measures self-perceived functional abilities in adults due to hemophilia, in the previous month. It consists of 42 items in seven subdomains: Lying / sitting / kneeling / standing (8 items), Functions of the legs (9 items), Functions of the arms (4 items), Use of transportation (3 items), Self-care (5 items), Household tasks (6 items), Leisure activities and sports (7 items). Items are scored on a 6-point Likert scale.[5, 6] Scores

were calculated according to the standard instructions (i.e. a domain score was calculated if less than half of the items were missing) and converted to a 0-100 point scale, with a higher score indicating better functional status. The HAL has sufficient content validity and construct validity but its structural validity is not known.[11]

Statistical analyses

Descriptive statistics (means, standard deviations (SD), N) were used to describe participant characteristics. Mean scores, SDs, the proportion of best and worst scores and percentage of missing scores for each domain or subscale were described for all measures. If proportions of best and worst scores were >30 percent, these were considered substantial ceiling or floor effects, respectively.[21]

Structural validity, internal consistency and construct validity were investigated as defined by the COSMIN taxonomy[18] and reported according to the COSMIN reporting guideline for studies on measurement properties.[34] A sample size of at least 100 participants is considered adequate for these analyses.[35]

Structural validity was assessed with confirmatory factor analysis (CFA) for each PROMIS domain separately. Model parameters were estimated with the Weighted Least Square Mean and Variance Adjusted Estimators (WLSMV) for ordinal data.[36] Model fit was assessed using the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI) and the Root Mean Squared Error of Approximation (RMSEA). Model fit was considered sufficient if CFI or TLI were >0.95, or RMSEA<0.06.[37] Internal consistency was calculated for each domain with sufficient model fit and considered sufficient if Cronbach's alpha was ≥0.70.[37]

Hypotheses were formulated a priori for construct validity (convergent and discriminative) for each domain. We considered ≥ 75 percent of results in accordance with the hypotheses evidence for construct validity.[37] Convergent validity was assessed with Pearson's correlations. We expected strong correlations ($r \geq 0.70$ or $r \geq -0.70$) between similar subscales of PROMIS-29 with RAND-36 subscales and HAL domains, based on published literature [38-40] and expert judgment (authors EvB and SG), as shown in Table 1. All other correlations were expected to be ≤ 0.60.

Discriminative validity was assessed by comparing mean T-scores between relevant clinical groups. Clinical subgroups were defined based on: hemophilia severity (mild compared to severe hemophilia); self-reported joint impairment in one or more of the six main joints (left and right ankles, knees, elbows; no / yes) and hiv infection (no / yes). Mean differences between mild and severe hemophilia were adjusted for age, mean differences between absent and present joint impairment were adjusted for age and severity using UNIANOVA. The comparison of mean T-scores for individuals with and without hiv were restricted to those born in 1985 or earlier, because the risk of hiv infection was considered negligible for younger patients.

Table 1: Hypotheses for construct validity (convergent and discriminative)

	Convergent validity		Discriminative validity		
PROMIS-29 subscale	Pearson's r ≥0.70	Pearson's r ≤0.60	MID	△ T-score ≥ MID	△ T-score < MID
Physical function	RAND-36 Physical Functioning	All other RAND-36 subscales (n=8)	2.0	Mild - severe hemophilia	No - yes hiv infection
	HAL LSKS	All other HAL domains (n=2)		No - yes joint impairment	
	HAL Legs				
	HAL Arms				
	HAL Transportation				
	HAL Household				
Anxiety	RAND-36 Emotional well-being	All other RAND-36 subscales (n=8)	-2.3	n.a.	Mild - severe hemophilia
		All HAL domains			No - yes joint impairment
					No - yes hiv infection
Depression	RAND-36 Emotional well-being	All other RAND-36 subscales (n=8)	-3.0	n.a	Mild - severe hemophilia
		All HAL domains			No - yes joint impairment
					No - yes hiv infection
Fatigue	RAND-36 Energy / Fatigue	All other RAND-36 subscales (n=8)	-2.0	No - yes hiv infection	Mild - severe hemophilia
		All HAL domains			No - yes joint impairment
Sleep Disturbance	n.a.	All RAND-36 subscales	-1.0	n.a	Mild - severe hemophilia
		All HAL domains			No - yes joint impairment
					No - yes hiv infection
Ability to Participate in Social Roles and Activities	RAND-36 Social Functioning	All other RAND-36 subscales (n=6)	1.0	No - yes joint impairment	Mild - severe hemophilia
	RAND-36 Role limitations – physical	All other HAL domains (n=5)			No - yes hiv infection
	RAND-36 Role limitations - emotional				
	HAL Household				
	HAL Leisure and Sports				

Table 1 (Continued)

	Convergent validity		Discriminative validity		
Pain Interference	RAND-36 Physical Functioning	All other RAND-36 subscales (n=7)	-2.0	Mild - severe hemophilia	No - yes hiv infection
	RAND-36 Pain	All other HAL domains (n=3)		No - yes joint impairment	
	HAL LSKS				
	HAL Legs				
	HAL Transportation				
	HAL Household				
Pain Intensity	RAND-36 Pain	All other RAND-36 subscales (n=8)	-1.0	Mild - severe hemophilia	No - yes hiv infection
		All HAL domains		No - yes joint impairment	

MID: Minimal important difference; HAL: Hemophilia Activities List; LSKS: Lying / sitting / kneeling / standing; n.a: not applicable

The following differences in mean T-scores were considered relevant differences between groups, based on published minimally important differences or changes for other patient groups (MID): ≥ 2 for Physical function,[41] ≥ -2.3 for Anxiety,[42] ≥ -3.0 for Depression,[42] ≥ -2 for Fatigue,[43] ≥ -1 for Sleep Disturbance,[43] ≥ 1 for Ability to Participate in Social Roles and Activities,[43] ≥ -2.0 for Pain Interference,[44] and ≥ -1 for Pain Intensity.[45] Because the MID is specific for each domain, a difference of, for example, 2 points may be a relevant difference in one domain, but not in another. Based on literature [46, 47] and clinical experience (authors SG, MD), we expected to find the following relevant differences: between mild and severe hemophilia and between absent and present joint impairment for Physical Function; between not hiv-infected and hiv-infected for Fatigue; between absent and present joint impairment for Ability to Participate in Social Roles and Activities; between mild and severe and between absent and present joint impairment for Pain Interference and for Pain Intensity (Table 1).
All analyses were performed with IBM SPSS version 25, except for Confirmatory Factor Analysis, which was performed in R, version 3.6.1 (package 'lavaan').

Results

Participants

Of 1746 Dutch adults with hemophilia who were invited to participate, 808 completed the questionnaires partially or in full (response 46.3 percent). The final sample for analysis consisted of 770 participants for whom one or more PROMIS-29 T-scores were calculated. For 598 of 770 participants (77.7 percent) clinical data from electronic medical records were available. Mean age was 48.9 (SD 17.2) years. Half of the participants (49.9 percent) had mild hemophilia, 15.6 percent had moderate and 34.5 percent had severe hemophilia, which is representative of the total Dutch hemophilia population (55.8, 13.2 and 30.1 percent, respectively). Clinical and socio-demographic characteristics are shown in Table 2.

Table 2: Participant characteristics (n = 770)

Clinical characteristics		
Hemophilia severity*	**N**	**%**
Mild	384	49.9
Moderate	120	15.6
Severe	266	34.5
Type of hemophilia	**N**	**%**
Hemophilia A	669	86.9
Hemophilia B	92	11.9
No hemophilia*	3	0.4
Unknown†	6	0.7
Prophylaxis (severe hemophilia)	**N**	**%**
Yes	233	87.6

Table 2: Participant characteristics (n = 770) (Continued)

Clinical characteristics		
No	30	11.3
Missing	3	1.1
Hiv infection	**N**	**%**
Yes	22	2.9
No	721	93.6
Unknown	27	3.5
HCV infection	**N**	**%**
Never infected	418	54.3
Past infection	231	30.0
Current infection	8	1.0
Past or current infection‡	2	0.6
Unknown	111	14.4
Inhibitor	**N**	**%**
Never	637	82.7
Past	68	8.8
Current	12	1.6
Unknown§	53	6.9
Joint impairment¶	**N**	**%**
Yes	338	43.9
No	379	49.2
Unknown	53	6.9
Demographic characteristics	**Mean**	**SD**
Age in years††	48.9	17.2
Education‡‡	**N**	**%**
Primary education	44	5.7
Secondary education	397	51.6
Tertiary education	298	38.7
Missing / prefer not to say	31	4.0

Clinical characteristics were taken from electronic medical records if participant had provided informed consent for extraction of data. If electronic medical record data were not available and participants did not complete the questions, status is unknown. Hemophilia severity was available from electronic medical records for all eligible persons (responders and non-responders)

* Three participants indicated on the questionnaire that they no longer had hemophilia, which might be because of a liver transplant (n=1) or participation in a gene therapy trial, but the exact reason is unknown.

† Five participants did not know their type of hemophilia (A or B), and one person skipped this question. Medical record data was missing for these individuals.

‡ Five individuals had a past or current HCV infection, but current infection status could not be established.

§ Inhibitor data from the medical record were not available for 53 participants because they did not provide informed consent for extraction of data.

¶ Joint impairment was self-reported chronic joint impairment in any joint (yes / no).

†† For three participants, age was missing and no electronic medical record was available.

‡‡ Education level was categorized according to ISCED levels: Primary education (ISCED level 1), Secondary education ISCED levels 2 and 3), Tertiary education (ISCED levels 6 and 7).

Description of measures

Table 3 shows mean, minimum and maximum scores, standard deviations, floor and ceiling effects and percentage of missing scores of all measures from the questionnaires. Mean T-scores for PROMIS-29 were better than the U.S. general population average for all subscales except Physical function, which was worse (48.9). Distributions of all PROMIS-29 domain scores were skewed toward better scores, i.e. scores >50 for the subscales Physical Function and Ability to Participate in Social Roles and Activities, and <50 for all other subscales (Figure 1). Five of seven PROMIS-29 subscales and Pain Intensity showed substantial ceiling effects of >30 percent patients with the best scores, while this was the case for five of eight RAND subscales and for all HAL-domains. PROMIS-29 had fewer missing answers than RAND-36 and HAL.

Structural validity

PROMIS-29 showed sufficient CFA model fit (CFI or TLI >0.95, or RMSEA<0.06) for Physical Function (CFI 0.95, TLI 0.85, RMSEA 0.13), Depression (CFI 1.00, TLI 0.99, RMSEA 0.02), Sleep Disturbance CFI 0.94, TLI 0.82, RMSEA 0.05), Ability to Participate in Social Roles and Activities (CFI 1.00, TLI 1.00, RMSEA 0.00) and Pain Interference (CFI 0.99, TLI 0.98, RMSEA 0.05). The subscales Anxiety and Fatigue did not show sufficient model fit (Table 4).

Internal consistency

Internal consistency was sufficient (Cronbach's alphas ≥0.70) for all five PROMIS-29 subscales with sufficient model fit in CFA. For four of them, Cronbach's alphas were ≥0.90: Physical function, Depression, Ability to Participate in Social Roles and Activities, and Pain Interference (Table 4). No Cronbach's alphas were calculated for Anxiety and Fatigue, because model fit was not sufficient.

Table 3: Characteristics of PROMIS-29, RAND-36 and HAL for adult men with hemophilia

	N*	Mean (SD)†	Range (min-max)	Worst score (%)‡	Best score (%)‡	Missing (%)§
PROMIS-29						
Physical Function	765	48.9 (9.6)	22.9-56.9	1.3	**51.9**	0.6
Anxiety	744	48.0 (8.2)	40.3-81.4	0.1	**43.2**	3.4
Depression	744	46.4 (7.8)	41.0-79.3	0.3	**59.1**	3.4
Fatigue	738	46.6 (9.6)	33.7-75.8	0.5	21.0	4.2
Sleep Disturbance	738	46.5 (7.9)	32.0-73.3	0.3	5.6	4.2
Ability to Participate in Social Roles and Activities	729	54.2 (8.9)	27.5-64.2	0.6	**30.6**	5.3
Pain Interference	726	49.6 (9.0)	41.6-75.6	0.6	**47.4**	5.7
Pain Intensity	724	2.4 (2.5)	0-10	0.1	**31.6**	6.0
RAND-36						
Physical functioning	734	77.9 (27.4)	0-100	0.8	**31.9**	2.3
Social functioning	705	83.5 (20.7)	0-100	0.5	**43.0**	8.4
Role limitations - physical	710	76.5 (37.5)	0-100	13.1	**61.7**	7.7
Role limitations - emotional	702	84.9 (31.6)	0-100	8.1	**71.8**	8.7
Emotional well-being	698	77.2 (15.6)	0-100	0.1	3.6	9.2
Energy / Fatigue	698	64.7 (17.8)	0-100	0.3	1.2	9.1
Pain	698	77.4 (22.5)	0-100	0.5	**31.6**	9.0
General health perceptions	694	64.5 (22.3)	0-100	0.6	4.3	0.0
Change in health	763	50.4 (19.8)	0-100	2.7	4.8	0.9
HAL						
Lying / sitting / kneeling / standing	709	77.6 (26.5)	7.5-100	0.0	**37.3**	7.1
Functions of the legs	694	74.0 (31.3)	0-100	1.6	**38.8**	9.1
Functions of the arms	688	83.9 (24.5)	0-100	0.6	**50.9**	10.3
Use of transportation	680	85.8 (24.7)	0-100	0.4	**55.6**	11.6
Self-care	681	90.8 (18.3)	5-100	0.0	**59.0**	11.4
Household tasks	647	87.4 (21.8)	0-100	0.4	**51.7**	12.5
Leisure activities and sports	614	82.0 (24.9)	0-100	0.5	**39.1**	13.1

* The number of participants for whom a score could be computed as described in the methods section.
† Higher scores on RAND-36 and HAL indicate better health status and better physical functioning, higher scores on PROMIS-29 indicate more of the construct being measured (e.g. more Physical Function and Ability to Participate in Social Roles and Activities, or more Anxiety, Depression, Fatigue, Sleep Disturbance and Pain)
‡ Worst and best possible scores were calculated if at least one item had been completed. Floor and ceiling effects are defined as the percentage of participants with the worst and the best scores possible. Floor and ceiling effects are considered present if >30 percent (in **bold**).
§ Percentage of participants for whom all items on a domain are missing.

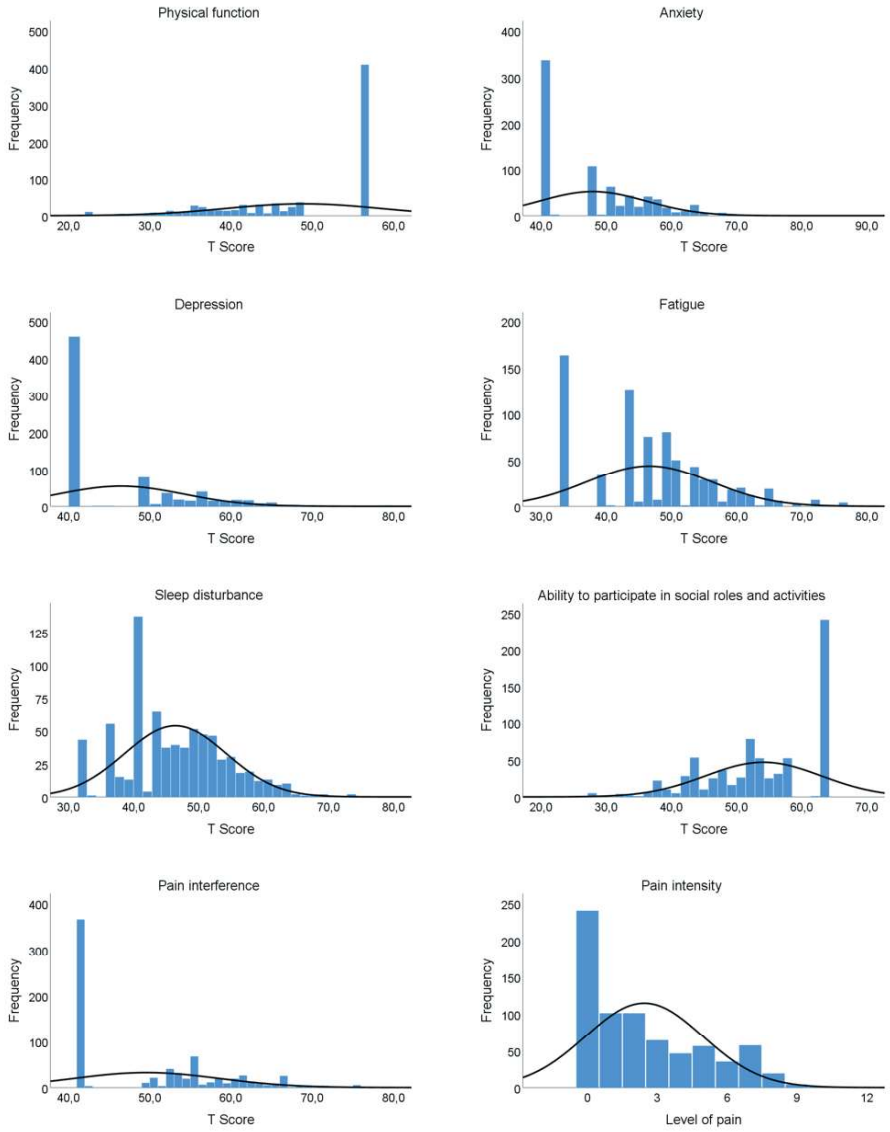


Figure 1: Distribution of T-scores on PROMIS-29 subscales

Frequencies of T-scores for each PROMIS-domain, and level of pain for Pain Intensity. The black curve indicates the normal distribution based on the frequencies. A higher score indicates more of the construct being measured.

Table 4: Structural validity and internal consistency of PROMIS-29

	N	CFI	TLI	RMSEA	Cronbach's alpha
PROMIS-29					
Physical function	752	**0.95**	0.85	0.13	0.94
Anxiety	735	0.88	0.63	0.15	-
Depression	727	**1.00**	**0.99**	**0.02**	0.93
Fatigue	728	0.85	0.56	0.24	-
Sleep Disturbance	713	0.94	0.82	**0.05**	0.79
Ability to Participate in Social Roles and Activities	717	**1.00**	**1.00**	**0.00**	0.93
Pain Interference	715	**0.99**	**0.98**	**0.05**	0.96

CFI: comparative fit index, TLI: Tucker-Lewis Index, RMSEA: root mean square error of approximation, Sufficient fit, indicated in **bold**: CFI or TLI >0.95, or RMSEA<0.06. Good internal consistency is defined as Cronbach's alpha ≥0.70. Fit parameters were rounded to two decimal places.

Construct validity

Results for convergent validity are shown in Table 5. For the subscales Anxiety, Depression, Fatigue, Sleep Disturbance, and Pain Intensity all correlations were in accordance with the hypotheses for convergent validity. For the subscales Physical Function 12 out of 16 correlations were as hypothesized, while for Ability to Participate in Social Roles and Activities this was the case for 11 out of 16 correlations. Nine out of 16 correlations were in accordance with the hypotheses for Pain Interference.

Table 5: Pearson's r for correlations between RAND-36 and PROMIS-29 subscales (convergent validity)

		PROMIS-29							
		Physical function	Anxiety	Depression	Fatigue	Sleep Disturbance	Ability to participate	Pain Interference	Pain Intensity
RAND-36	Physical functioning	**0.91**	-0.31	-0.36	-0.37	-0.28	0.59	**-0.70**	-0.59
	Social functioning	0.52	-0.57	-0.60	-0.58	-0.42	**0.70**	-0.59	-0.53
	Role limitations – physical	0.63	-0.36	-0.40	-0.48	-0.32	**0.62**	-0.66	-0.57
	Role limitations - emotional	0.33	-0.48	-0.56	-0.44	-0.37	**0.50**	-0.37	-0.33
	Emotional well-being	0.28	**-0.74**	**-0.75**	-0.55	-0.50	0.56	-0.35	-0.33
	Energy / Fatigue	0.40	-0.55	-0.59	**-0.72**	-0.50	0.60	-0.46	-0.42
	Pain	0.63	-0.33	-0.38	-0.46	-0.31	0.55	**-0.82**	**-0.80**
	General health perceptions	0.59	-0.46	-0.47	-0.54	-0.43	0.62	-0.63	-0.55
	Change in health	0.29	-0.14	-0.16	-0.17	-0.11	0.27	-0.28	-0.28
HAL	Lying / sitting / kneeling / standing	**0.79**	-0.24	-0.27	-0.29	-0.28	0.53	**-0.63**	-0.53
	Functions of the legs	**0.85**	-0.23	-0.28	-0.28	-0.25	0.53	**-0.65**	-0.56
	Functions of the arms	**0.73**	-0.30	-0.31	-0.35	-0.28	0.56	-0.64	-0.54
	Use of transportation	**0.77**	-0.25	-0.30	-0.30	-0.25	0.54	**-0.58**	-0.48
	Self-care	0.66	-0.30	-0.32	-0.33	-0.27	0.54	-0.60	-0.53
	Household tasks	**0.80**	-0.31	-0.35	-0.36	-0.29	**0.60**	**-0.70**	-0.56
	Leisure activities and sports	0.80	-0.29	-0.34	-0.36	-0.29	**0.60**	-0.70	-0.57
Hypotheses confirmed (%)		75	100	100	100	100	69	56	100

Correlations in **bold** were expected to be ≥ 0.70 or ≥ -0.70. All other correlations were expected to be ≤0.60.

Table 6: Differences in mean PROMIS-29 T-scores for clinical subgroups (discriminative validity)

		Severe – mild hemophilia		Yes – no joint impairment		Yes – no hiv infection		Hypotheses confirmed (%)§
	MID	unadjusted (95% CI)	adjusted (95% CI)*	unadjusted (95% CI)	adjusted (95% CI)†	unadjusted (95% CI)	adjusted (95% CI)‡	
Physical function	2.0	**8.6 (7.0; 10.0)**	**10.2 (9.1; 11.4)**	**10.8 (9.6; 11.9)**	**6.1 (4.8; 7.4)**	10.5 (6.4; 14.6)	1.5 (-1.9; 4.8)	100
Anxiety	-2.3	-0.4 (-1.7; 0.9)	-0.8 (-2.1; 0.5)	-2.5 (-3.7; -1.3)	-2.7 (-4.2; -1.2)	-2.6 (-6.2; 1.0)	-1.7 (-5.8; 2.3)	67
Depression	-3.0	-1.7 (-2.9; -0.4)	-2.1 (-3.3; -0.8)	-2.7 (-3.9; -1.6)	-1.8 (-3.3; -0.4)	-3.4 (-6.7; -0.1)	-1.7 (-5.6; 2.2)	100
Fatigue	-2.0	-1.9 (-3.4; -0.3)	-2.2 (-3.7; -0.6)	-3.6 (-5.0; -2.1)	-3.5 (-5.3; -1.7)	**-4.7 (-8.8; -0.7)**	**-2.5 (-6.9; 1.9)**	33
Sleep Disturbance	-1.0	-1.2 (-2.5; 0.0)	-1.4 (-2.7; -0.2)	-2.2 (-3.3; -1.0)	-1.8 (-3.2: -0.3)	-3.7 (-7.0; -0.3)	-2.8 (-6.5; 1.0)	0
Ability to Participate in Social Roles and Activities	1.0	4.4 (3.0; 5.8)	5.4 (4.0; 6.7)	**5.7 (4.4; 7.0)**	**3.1 (1.6; 4.6)**	8.7 (5.0; 12.5)	3.5 (-0.2; 7.3)	33
Pain Interference	-2.0	**-6.0 (-7.4; -4.7)**	**-7.2 (-8.5; -5.8)**	**-8.2 (-9.4; -7.0)**	**-5.7 (-7.2; -4.3)**	-7.3 (-11.1; -3.5)	-0.9 (-4.7; 3.0)	100
Pain Intensity	-1.0	**-1.5 (-1.9; -1.2)**	**-1.8 (-2.1; -1.4)**	**-1.9 (-2.3; -1.6)**	**-1.4 (-1.8; -0.9)**	-1.9 (-2.9; -0.8)	-0.6 (-1.7; 0.5)	100

Differences in **bold** were hypothesized to be > MID. 95% CI: 95% confidence interval. For the subscales Physical Function and Ability to Participate in Social Roles and Activities a positive difference means that persons with mild hemophilia, no joint impairment or no hiv infection have more of these constructs than persons with severe hemophilia, joint impairment or hiv infection. For the other subscales negative differences indicate less of these constructs for persons with mild hemophilia, no joint impairment or no hiv infection.

* Adjusted for age.

† Adjusted for age and severity.

‡ Mean difference between individuals with severe hemophilia born in 1985 or earlier, with or without hiv.

§ Hypotheses confirmed for discriminative validity.

Unadjusted and adjusted differences in mean T-scores between clinical groups (discriminative validity) are shown in Table 6. All differences between groups were in the expected direction, i.e. participants with mild hemophilia, no joint damage and no hiv infection had better scores for all subscales. Adjusting for age resulted in a larger difference between mild and severe hemophilia, and adjusting for age and disease severity resulted in smaller differences between individuals with and without joint impairment. Finally, differences became smaller when hiv-infected participants were compared with non-infected participants with severe hemophilia born in or before 1985.

The evidence for discriminative validity was strongest for Physical Function, Depression, Pain Interference and Pain Intensity: all differences between subgroups were as hypothesized. For Anxiety, two of three differences between groups were as hypothesized, and for Fatigue and Ability to Participate in Social Roles and Activities one difference was as hypothesized. None of the differences between groups were in accordance with the hypotheses for Sleep Disturbance.

In total, six subscales showed evidence for construct validity (≥75 percent hypotheses confirmed): Physical Function, (79 percent), Anxiety (95 percent), Depression (100 percent), Fatigue (89 percent), Sleep Disturbance (84 percent) and Pain Intensity (100 percent). Two subscales did not meet the criterium for ≥75 percent of hypotheses confirmed: for Ability to Participate in Social Roles and Activities and Pain Interference 63 percent of hypotheses were confirmed.
Table 7 summarizes the evidence for structural validity, internal consistency and construct validity.

Table 7: Summary of the evidence for structural validity, internal consistency and construct validity (convergent and discriminative)

PROMIS-29 subscale	Structural validity	Internal consistency	Construct validity
Physical function	+	+	+
Anxiety	-	0	+
Depression	+	+	+
Fatigue	-	0	+
Sleep Disturbance	+	+	+
Ability to Participate in Social Roles and Activities	+	+	-
Pain Interference	+	+	-
Pain Intensity	n/a	n/a	+

+ indicates evidence for the measurement property according to pre-specified criteria; - indicates that the evidence for the measurement property did not meet pre-specified criteria; 0: not assessed because of limited structural validity; n/a: measurement property not applicable (1 item)

Discussion

This study is the first validation of the Dutch-Flemish version of the PROMIS Profile-29, as well as the first validation of this Profile among persons with hemophilia. Using consensus-based standards for evaluating validity, we aimed to assess structural validity, internal consistency and construct validity of the PROMIS-29 Profile v2.01 in Dutch adults with hemophilia. In a representative sample of the Dutch hemophilia population, our analyses showed sufficient evidence for structural validity and internal consistency for five of seven subscales and sufficient evidence for construct validity for five subscales and for Pain Intensity.

In the confirmatory factor analysis, model fit was not sufficient for Anxiety and Fatigue, potentially indicating a lack of unidimensionality,[48] i.e. that these subscales may measure more than one construct for people with hemophilia. An explanation may be that CFA modelling assumes a normal distribution of the data. Our results, however, showed skewed distributions for all subscales. This may have influenced fit statistics.[48] In contrast to our findings, a previous validation of PROMIS-29 among kidney transplant recipients found excellent structural validity for all subscales,[23] even with similarly skewed distributions.

We found evidence for sufficient internal consistency for five subscales, with Cronbach's alphas >0.90 for the subscales Physical Function, Depression, Ability to Participate in Social Roles and Activities, and Pain Interference. Consistent with our findings, two previous studies in kidney transplant recipients and populations with rheumatoid arthritis, osteoarthritis, systemic lupus erythematosus reported similarly high Cronbach's alphas for all subscales.[21, 23]

Overall, the subscales Physical Function, Anxiety, Depression, Fatigue, Sleep Disturbance and Pain Intensity showed evidence for construct validity (i.e. >75 percent of results in accordance with the hypotheses). Fewer hypotheses were confirmed for the subscales Ability to Participate in Social Roles and Activities and Pain Interference (63 percent).

Correlations lower than the expected 0.70 were found for Ability to Participate in Social Roles and Activities with the RAND-36 Role limitations caused by physical or emotional health problems (0.62 and 0.50, respectively). The hypothesis for the former correlation was based on a Dutch study among 30 abdominal surgery patients that reported a correlation of 0.72 between the SF-36 subscale Role limitations caused by physical health problems and the 8-item PROMIS Ability to Participate in Social Roles and Activities short form.[40] Though the correlation we report is below the 0.70 threshold, it is of the same order of magnitude and the difference may be due to random variation or to differences in the underlying constructs being measured.

Lower correlations were also found between PROMIS-29 Ability to Participate in Social Roles and Activities with HAL Household tasks (0.60) and Leisure and sports

(0.60). This may mean that these constructs differ more than anticipated, resulting in fewer hypotheses for convergent validity confirmed. Indeed, HAL subscales measure several aspects of self-perceived functional ability, while PROMIS Ability to Participate in Social Roles and Activities measures participation.

Some subscales that were not expected to correlate highly with RAND-36 and HAL (i.e. expected to be ≤ 0.60) showed correlations above the threshold of 0.60. This was the case for the correlation between Physical Function with RAND-36 Pain (0.63), RAND-36 Role limitations caused by physical health problems (0.63) and HAL Self-care (0.66), and for Ability to Participate in Social Roles and Activities with RAND-36 General health perceptions (0.62), and for Pain Interference with RAND-36 Role limitations caused by physical health problems (-0.66), with RAND-36 General health perceptions (-0.63) and with HAL Functions of the arms (-0.64). We used a relatively low expected correlation of ≤ 0.60 between subscales that do not measure the same construct to distinguish them from the correlations ≥ 0.70 expected between subscales that measure the same construct, but this resulted in fewer hypotheses confirmed (especially for Ability to Participate in Social Roles and Activities and Pain Interference), and thus lower evidence of construct validity. This strict criterium may have led to quite conservative conclusions.

Also interesting is that most correlations between Pain Interference and HAL subscales were of similar strength, between -0.58 and -0.66. Though below the 0.70 threshold, the subscales perceived functional ability (HAL) [5, 6] and Pain Interference with functional ability (PROMIS) [16] may measure similar constructs after all.

We found unexpected differences larger than the MID for some subscales. For example, differences between all clinical groups were larger than expected for Sleep Disturbance. Sleep Disturbance is not routinely studied in hemophilia, but a qualitative study reported that pain may affect sleep disturbance.[49] Persons with severe hemophilia, joint impairment and hiv are more likely to experience pain due to recurrent bleeding, which may explain part of the observed differences. However, confidence intervals of the observed differences were wide. Also, the correlation between PROMIS-29 Sleep Disturbance and RAND-36 pain was low (-0.31), making a substantial influence of pain on sleep disturbance less likely. Differences between mild and severe hemophilia and for different hiv infection status were also larger than expected for Ability to Participate in Social Roles and Activities, while we only expected to find differences for joint impairment. Since effective treatment is available, persons with severe hemophilia should be able to lead near-normal lives, and for this reason were expected to have similar levels of social participation as individuals with mild hemophilia. Our results indicate that this may not be the case. Indeed, hemophilia is reported to have a negative impact on employment and education,[50] and may also have affected the Ability to Participate in Social Roles and Activities. Individuals with hiv infection may have a more severe bleeding phenotype than those without hiv: persons with a more severe bleeding phenotype may have received more plasma-derived treatment products in

the past, and contracted hiv as a result, compared to persons with severe hemophilia with a milder bleeding phenotype. A more severe bleeding phenotype may also have resulted in more joint impairment and lower participation. Unfortunately, we did not have reliable information on bleeding phenotype and were therefore unable to correct for this confounder. It should be noted that the number of individuals with hiv was small (n=22), resulting in less reliable estimates of T-scores in this subgroup.

A potential limitation of this study is that the response rate of the HiN-6 study was limited (46.3 percent). This may have led to some bias. First, fewer people had only primary education (5.7 percent) and more had secondary education (51.6 percent) compared to the general Dutch population (21 and 40 percent, respectively).[51] If people with a higher education were better able to manage their hemophilia, this could have resulted in higher scores on PROMIS subscales. This may, in part, explain our finding that mean scores on many PROMIS-29 subscales were higher than the general population average of 50. Second, persons with more health-related problems due to hemophilia may have been more likely to participate because they were more motivated to complete a questionnaire about their health. This would have resulted in low scores. However, our results showed large proportions of participants with the highest scores on several subscales, indicating few health problems. Therefore, we believe selection bias due to health problems was unlikely to have impacted the findings of this study.

Content validity of PROMIS-29 was reported to be good in several other populations.[13, 14] Our results also provide some evidence for content validity of PROMIS-29 among persons with hemophilia: the number of missing answers was low, which may indicate that items were relevant to participants.[52] On the other hand, PROMIS-29 showed large proportions of best scores for most subscales, which may indicate a lack of content validity: best scores may indicate that items were not relevant to measure the domain for this population and that more 'difficult' items may be missing.[52] The large proportion of best scores on most subscales (except Fatigue and Sleep Disturbance) leads to a loss in measurement precision in well-functioning individuals. The 4-item short forms that comprise PROMIS-29 may therefore not be optimal for persons with hemophilia. Because PROMIS item banks are IRT-based, they are flexible and another selection of items can be considered. For example, a longer or a custom short form with more 'difficult' items from the item bank or a Computerized Adaptive Test may solve these ceiling effects and still yield comparable results.[12] Unfortunately, Dutch CATs were not available yet at the time of our study, but have become available recently.[53, 54]

In our study, five subscales met all criteria for structural validity and internal consistency and five and Pain Intensity met all the criteria for hypotheses-testing for construct validity. Small changes in the methods regarding the cut-offs of correlations and the percentage of hypotheses confirmed may have had profound effects on the conclusions.

Other studies that validated PROMIS-29 in different populations did not formulate hypotheses for construct validity, which may lead to less transparent and less consistent

interpretation of the results.[21-23] Yet, hypothesis-testing for construct validity depends on sufficient knowledge about the constructs being measured with all subscales. However, limited literature was available that quantified correlations with other instruments or differences between groups. Despite the lack of explicit hypotheses in other studies, the magnitude of differences between relevant subgroups is similar.[21-23] This indicates generalizability across diseases.

Ideally, PROMs are used that measure the most relevant outcomes for a specific population. A consensus-based standard set of relevant outcomes for persons with hemophilia was published recently,[55] along with recommendations for instruments to measure these outcomes. The set included the five patient-reported outcomes Ability to engage in normal daily activities, Chronic pain, Sustainability of physical functioning, Social functioning, and Mental health. The latter four can be measured with the PROMIS Profile-29 subscales that were validated in the current study: Pain Interference and Pain Intensity; Physical function; Ability to Participate in Social Roles and Activities; and Anxiety and Depression, respectively. For an even more comprehensive assessment, Social functioning may be measured with the PROMIS domain Self-efficacy for managing social interactions, and Mental health with the subscales General Life Satisfaction and Positive Affect. Ability to engage in normal daily activities may be measured with PROMIS Self-efficacy for Managing Chronic Conditions - Managing Daily Activities. PROMIS item banks or short forms for these subscales may be validated for comprehensive assessment of the standard set of outcomes for hemophilia. The standard set of outcomes did not include the domains fatigue and sleep disturbance, which may not need to be prioritized for measurement, though they may still be important in some patients or certain situations.

Which tools to use (disease-specific or generic) depends on the goal of measuring outcomes and the type of outcomes. Some outcomes, such as degree of hemophilic arthropathy, are disease-specific and need to be assessed with disease-specific instruments. Functional outcomes such as those measured with PROMIS item banks (e.g. physical function, fatigue) are of a more generic nature. For clinical care aimed at improving outcomes, generic tools may be the most suitable, while in other cases disease-specific tools may be necessary. Still, in many cases, a combination of generic tools where possible, supplemented with disease-specific tools where needed, may be the most suitable for comprehensive measurement of all outcomes that are relevant for hemophilia.

Conclusion

This study found sufficient evidence for structural validity, internal consistency and construct validity of the PROMIS-29 subscales Physical Function, Depression and Sleep Disturbance in adult persons with hemophilia in the Netherlands. Construct validity

was also sufficient for Anxiety, Fatigue and Pain Intensity. These results indicate that PROMIS short forms that measure these domains may be used in hemophilia populations. Future studies should explore whether the use of custom short forms or CAT can solve observed ceiling effects.

References

1 Srivastava A, Santagostino E, Dougall A, Kitchen S, Sutherland M, Pipe SW, et al. WFH Guidelines for the Management of Hemophilia, 3rd edition. *Haemophilia*. 2020; **26**: 1-2. 10.1111/hae.14046.

2 Mauser-Bunschoten EP, Bresters D, van Drimmelen AA, Roosendaal G, Cuypers HT, Reesink HW, et al. Hepatitis C infection and viremia in Dutch hemophilia patients. *J Med Virol*. 1995; **45**: 241-6. 10.1002/jmv.1890450302.

3 Plug I, van der Bom JG, Peters M, Mauser-Bunschoten EP, de Goede-Bolder A, Heijnen L, et al. Thirty years of hemophilia treatment in the Netherlands, 1972-2001. *Blood*. 2004; **104**: 3494-500. 10.1182/blood-2004-05-2008.

4 De Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in medicine*. Cambridge University Press, 2011.

5 van Genderen FR, van Meeteren NL, van der Bom JG, Heijnen L, de Kleijn P, van den Berg HM, et al. Functional consequences of haemophilia in adults: the development of the Haemophilia Activities List. *Haemophilia*. 2004; **10**: 565-71. 10.1111/j.1365-2516.2004.01016.x.

6 van Genderen FR, Westers P, Heijnen L, de Kleijn P, van den Berg HM, Helders PJ, et al. Measuring patients' perceptions on their functional abilities: validation of the Haemophilia Activities List. *Haemophilia*. 2006; **12**: 36-46. 10.1111/j.1365-2516.2006.01186.x.

7 Rentz A, Flood E, Altisent C, Bullinger M, Klamroth R, Garrido RP, et al. Cross-cultural development and psychometric evaluation of a patient-reported health-related quality of life questionnaire for adults with haemophilia. *Haemophilia*. 2008; **14**: 1023-34. 10.1111/j.1365-2516.2008.01812.x.

8 Arranz P, Remor E, Quintana M, Villar A, Díaz JL, Moreno M, et al. Development of a new disease-specific quality-of-life questionnaire to adults living with haemophilia. *Haemophilia*. 2004; **10**: 376-82. 10.1111/j.1365-2516.2004.00918.x.

9 Hays RD, Sherbourne CD, Mazel RM. The RAND 36-Item Health Survey 1.0. *Health Econ*. 1993; **2**: 217-27.

10 Limperg PF, Terwee CB, Young NL, Price VE, Gouw SC, Peters M, et al. Health-related quality of life questionnaires in individuals with haemophilia: a systematic review of their measurement properties. *Haemophilia*. 2017; **23**: 497-510. 10.1111/hae.13197.

11 Timmer MA, Gouw SC, Feldman BM, Zwagemaker A, de Kleijn P, Pisters MF, et al. Measuring activities and participation in persons with haemophilia: A systematic review of commonly used instruments. *Haemophilia*. 2018; **24**: E33-E49. 10.1111/hae.13367.

12 Cella D, Gershon R, Lai JS, Choi S. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. *Quality of Life Research*. 2007; **16**: 133-41. 10.1007/s11136-007-9204-6.

13 Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, 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. 10.1016/j.jclinepi.2010.04.011.

14 Ader DN. Developing the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care*. 2007; **45**: S1-S2. DOI 10.1097/01.mlr.0000260537.45076.74.

15 Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, 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**: S3-s11. 10.1097/01.mlr.0000258615.42478.55.

16 Cella D, Choi SW, Condon DM, Schalet B, Hays RD, Rothrock NE, et al. PROMIS (R) Adult Health Profiles: Efficient Short-Form Measures of Seven Health Domains. *Value in Health*. 2019; **22**: 537-44. 10.1016/j.jval.2019.02.004.

17 Prinsen CAC, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" – a practical guideline. *Trials*. 2016; **17**: 449. 10.1186/s13063-016-1555-2.

18 Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol 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. 10.1016/j.jclinepi.2010.02.006.

19 Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res*. 2018; **27**: 1159-70. 10.1007/s11136-018-1829-0.

20 Terwee CB, Roorda LD, de Vet HCW, Dekker J, Westhovens R, van Leeuwen J, et al. Dutch–Flemish translation of 17 item banks from the Patient-Reported Outcomes Measurement Information System (PROMIS). *Quality of Life Research*. 2014; **23**: 1733-41. 10.1007/s11136-013-0611-6.

21 Katz P, Pedro S, Michaud K. Performance of the Patient-Reported Outcomes Measurement Information System 29-Item Profile in Rheumatoid Arthritis, Osteoarthritis, Fibromyalgia, and Systemic Lupus Erythematosus. *Arthritis Care Res (Hoboken)*. 2017; **69**: 1312-21. 10.1002/acr.23183.

22 Rose AJ, Bayliss E, Huang W, Baseman L, Butcher E, Garcia RE, et al. Evaluating the PROMIS-29 v2.0 for use among older adults with multiple chronic conditions. *Qual Life Res*. 2018; **27**: 2935-44. 10.1007/s11136-018-1958-5.

23 Tang E, Ekundayo O, Peipert JD, Edwards N, Bansal A, Richardson C, et al. Validation of the Patient-Reported Outcomes Measurement Information System (PROMIS)-57 and -29 item short forms among kidney transplant recipients. *Qual Life Res*. 2019; **28**: 815-27. 10.1007/s11136-018-2058-2.

24 Smit C, Rosendaal FR, Varekamp I, Brocker-Vriends A, Van Dijck H, Suurmeijer TP, et al. Physical condition, longevity, and social performance of Dutch haemophiliacs, 1972-85. *Bmj*. 1989; **298**: 235-8. 10.1136/bmj.298.6668.235.

25 Triemstra AH, Smit C, HM VDP, Briet E, Rosendaal FR. Two decades of haemophilia treatment in the Netherlands, 1972-92. *Haemophilia*. 1995; **1**: 165-71. 10.1111/j.1365-2516.1995.tb00061.x.

26 Castor EDC. Castor Electronic Data Capture. 2019.

27 UNESCO Institute for Statistics. International Standard Classification of Education: ISCED 2011. Montreal, 2012.

28 Cella D, Gershon R, Bass M, Rothrock N. HealthMeasures Scoring Service. Chicago: Northwestern University, 2007.

29 Hays RD, Morales LS. The RAND-36 measure of health-related quality of life. *Ann Med*. 2001; **33**: 350-7. 10.3109/07853890109002089.

30 Van der Zee KI, Sanderman R. Het meten van de algemene gezondheidstoestand met de RAND-36, een handleiding., 2e druk edn. Groningen: Rijksuniversiteit Groningen, Noordelijk Centrum voor Gezondheidsvraagstukken, 2012.

31 Aaronson NK, Muller M, Cohen PD, Essink-Bot ML, Fekkes M, Sanderman R, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. *J Clin Epidemiol*. 1998; **51**: 1055-68.

32 Solovieva S, Santavirta N, Santavirta S, Konttinen YT. Assessing Quality of Life in Individuals with Hereditary Blood Coagulation Disorders. *Quality of Life Research*. 2004; **13**: 987-1000.

33 Buckner TW, Wang M, Cooper DL, Iyer NN, Kempton CL. Known-group validity of patient-reported outcome instruments and hemophilia joint health score v2.1 in US adults with hemophilia: results from the Pain, Functional Impairment, and Quality of life (P-FiQ) study. *Patient Prefer Adherence*. 2017; **11**: 1745-53. 10.2147/ppa.s141392.

34 Gagnier JJ, Lai J, Mokkink LB, Terwee CB. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. *Qual Life Res*. 2021. 10.1007/s11136-021-02822-4.

35 Mokkink LB, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, De Vet HC, et al. COSMIN Study Design checklist for Patient-reported outcome measurement instruments. Amsterdam, 2019.

36 Li CH. Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behav Res Methods*. 2016; **48**: 936-49. 10.3758/s13428-015-0619-7.

37 Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018; **27**: 1147-57. 10.1007/s11136-018-1798-3.

38 Batt K, Recht M, Cooper DL, Iyer NN, Kempton CL. Construct validity of patient-reported outcome instruments in US adults with hemophilia: results from the Pain, Functional Impairment, and Quality of life (P-FiQ) study. *Patient Prefer Adherence*. 2017; **11**: 1369-80. 10.2147/ppa.s141390.

39 Choi SW, Podrabsky T, McKinney N, Schalet BD, Cook KF, Cella D. Prosetta Stone Analysis Report. A Rosetta Stone for Patient-reported Outcomes Chicago: Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University 2015.

40 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**: 128. 10.1186/s12913-018-2929-9.

41 Hays RD, Spritzer KL, Fries JF, Krishnan E. Responsiveness and minimally important difference for the patient-reported outcomes measurement information system (PROMIS) 20-item physical functioning short form in a prospective observational study of rheumatoid arthritis. *Ann Rheum Dis*. 2015; **74**: 104-7. 10.1136/annrheumdis-2013-204053.

42 Lee AC, Driban JB, Price LL, Harvey WF, Rodday AM, Wang C. Responsiveness and Minimally Important Differences for 4 Patient-Reported Outcomes Measurement Information System Short Forms: Physical Function, Pain Interference, Depression, and Anxiety in Knee Osteoarthritis. *J Pain*. 2017; **18**: 1096-110. 10.1016/j.jpain.2017.05.001.

43 Katz P, Pedro S, Alemao E, Yazdany J, Dall'Era M, Trupin L, et al. Estimates of Responsiveness, Minimally Important Differences, and Patient Acceptable Symptom State in Five Patient-Reported Outcomes Measurement Information System Short Forms in Systemic Lupus Erythematosus. *ACR Open Rheumatol*. 2020; **2**: 53-60. 10.1002/acr2.11100.

44 Katz P, Kannowski CL, Sun L, Michaud K. Estimation of Minimally Important Differences and Patient Acceptable Symptom State Scores for the Patient-Reported Outcomes Measurement Information System Pain Interference Short Form in Rheumatoid Arthritis. *ACR Open Rheumatol*. 2020; **2**: 320-9. 10.1002/acr2.11141.

45 Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. *Eur J Pain*. 2004; **8**: 283-91. 10.1016/j.ejpain.2003.09.004.

46 Plug I, Peters M, Mauser-Bunschoten EP, de Goede-Bolder A, Heijnen L, Smit C, et al. Social participation of patients with hemophilia in the Netherlands. *Blood*. 2008; **111**: 1811-5. 10.1182/blood-2007-07-102202.

47 Lindvall K, Von Mackensen S, Berntorp E. Quality of life in adult patients with haemophilia--a single centre experience from Sweden. *Haemophilia*. 2012; **18**: 527-31. 10.1111/j.1365-2516.2012.02765.x.

48 Cook KF, Kallen MA, Amtmann D. Having a fit: impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. *Qual Life Res*. 2009; **18**: 447-60. 10.1007/s11136-009-9464-4.

49 Rambod MP, Sharif FP, Molazem ZP, Khair KP. Pain Experience in Hemophilia Patients: A Hermeneutic Phenomenological Study. *Int J Community Based Nurs Midwifery*. 2016; **4**: 309-19.

50 Forsyth AL, Gregory M, Nugent D, Garrido C, Pilgaard T, Cooper DL, et al. Haemophilia Experiences, Results and Opportunities (HERO) Study: survey methodology and population demographics. *Haemophilia*. 2014; **20**: 44-51. 10.1111/hae.12239.

51 Maslowski R. Onderwijs. In: Den Ridder J, Josten E, Boelhouwer J, Van Campen C, eds. *De Sociale Staat van Nederland*. Den Haag: Sociaal en Cultureel Planbureau, 2020.

52 Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007; **60**: 34-42. 10.1016/j.jclinepi.2006.03.012.

53 Flens G, Smits N, Terwee CB, Dekker J, Huijbrechts I, de Beurs E. Development of a Computer Adaptive Test for Depression Based on the Dutch-Flemish Version of the PROMIS Item Bank. *Eval Health Prof*. 2017; **40**: 79-105. 10.1177/0163278716684168.

54 Flens G, Smits N, Terwee CB, Dekker J, Huijbrechts I, Spinhoven P, et al. Development of a Computerized Adaptive Test for Anxiety Based on the Dutch-Flemish Version of the PROMIS Item Bank. *Assessment*. 2019; **26**: 1362-74. 10.1177/1073191117746742.

55 Balen EC, O'Mahony B, Cnossen MH, Dolan G, Blanchette VS, Fischer K, et al. Patient-relevant health outcomes for hemophilia care: Development of an international standard outcomes set. *Res Pract Thromb Haemost*. 2021; **5**: e12488. 10.1002/rth2.12488.