Improving outcome assessment covering functions, activities and participation in patients with haemophilia

Isolde Aldegonda Rudolf Kuijlaars

Improving outcome assessment covering functions, activities and participation in patients with haemophilia

Isolde Aldegonda Rudolf Kuijlaars

Improving outcome assessment covering functions, activities and participation in patients with haemophilia

ISBN: 978-94-6419-280-3
Layout and design: Marilou Maes, persoonlijkproefschrift.nl
Printing: Gildeprint Enschede, gildeprint.nl

Improving outcome assessment covering functions, activities and participation in patients with haemophilia

Het verbeteren van meten van

functies, activiteiten en participatie bij patiënten met hemofilie

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op donderdag 14 oktober 2021 des middags te 4.15 uur

door

Isolde Aldegonda Rudolf Kuijlaars

geboren op 29 juni 1991

te Veldhoven

Promotor:

Prof. dr. R.E.G. Schutgens

Copromotoren:

Dr. K. Fischer

Dr. J. van der Net

TABLE OF CONTENTS

CHAPTER 1

General introduction

HAEMOPHILIA

Haemophilia is an X-linked recessive bleeding disorder, caused by a deficiency of coagulation factor VIII (FVIII, haemophilia A) or factor IX (FIX, haemophilia B). The prevalence of haemophilia is 1 in 10,000 births. The severity of the disease is determined by the level of clotting factor VIII or FIX, categorized into severe (FVIII/FIX activity <0.01 IU/ml), moderate (FVIII/FIX activity 0.01 - 0.05 IU/ml) or mild (FVIII/FIX activity 0.06 - 0.40 IU/ml) haemophilia. Patients with mild haemophilia experience bleeds only after major trauma or surgery, moderate haemophilia may cause spontaneous bleeding or bleeds after minor trauma or surgery, whereas severe haemophilia leads to spontaneous bleeding. Most of these bleeds occur in joints and muscles [1]. The ankles, knees and elbows are the most frequently affected joints [2]. Joint bleeds affect the cartilage directly, as well as indirectly through synovial inflammation, and eventually result in haemophilic arthropathy [3]. Haemophilic arthropathy leads to pain and loss of range of motion which has an impact on activities, participation and health-related quality of life [1,4].

Haemophilia treatment improved dramatically over the last decades. In 1964, the first haemophilia comprehensive care center was established in the Netherlands by dr. Simon van Creveld. Prophylactic clotting factor replacement therapy is an effective treatment to reduce bleeding frequency and was introduced soon afterwards in the Netherlands in 1968 [1]. Before the introduction of prophylaxis, persons with haemophilia (PWH) suffering from a bleed were imposed to (long-term) bed rest and immobilization of the affected joint. Nowadays, boys participate in sports similar to their peers [5]. On top of that, extended half-life concentrates and non-replacement therapy were implemented in day-to-day care and even gene therapy is introduced in haemophilia trials as a next step to improve outcomes and reduce patient burden of PWH [6–8]. These improvements in treatment strategies have a great impact for PWH through bleed prevention and consequently maintaining joint health.

OUTCOME ASSESSMENT

Assessing outcomes is essential to evaluate any treatment from an individual patient level in day-to-day care or at group level to compare treatment strategies. As PWH experience less bleeds and life expectancy increased to almost normal [9], a shift is needed from assessing only bleeding frequency and joint status to a broader health assessment. This can be achieved through outcome assessment on the whole International Classification of Functioning (ICF) spectrum. According to the World Health Organization (WHO) ICF framework, health is an interaction between 'body functions and structure', 'activities and participation' and 'contextual factors' [10]. Figure 1 gives an overview of the ICF and recommended outcome measures for each domain. Outcome measures studied in this thesis are outlined in Box 1. The Hemophilia Joint Health Score (HJHS) is a physical examination of the joints and the (paediatric) Haemophilia Activities List (pedHAL and HAL) are patient reported outcome measures (PROMs) about limitations in activities and participation. All outcome measures were recommended for both clinical use and research goals [11].

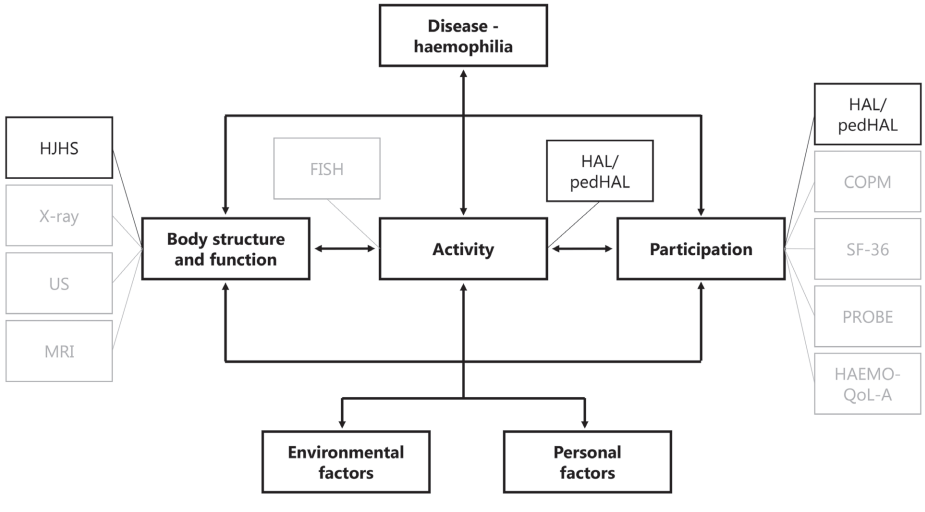

Figure 1. International Classification of Functioning and Health (ICF) model, with domain-related outcome measures (in black: outcome measures within the scope of this thesis; in grey: outcome measures beyond the scope of this thesis). COPM, Canadian Occupational Performance Measure; FISH, Functional Independence Score in Hemophilia; HAEMO-QoL-A, hemophilia-specific quality-of-life questionnaire for adults; HAL, Haemophilia Activities List; HJHS, Hemophilia Joint Health Score; MRI, magnetic resonance imaging; PedHAL, peadiatric Haemophilia Activities List; PROBE, Patient-Reported Outcomes, Burdens and Experiences; SF-36, 36-Item Short Form Survey Instrument; US, ultrasound. Adapted from: Srivastava A, Santagostino E, Dougall A et al. WFH Guidelines for the Management of Hemophilia, 3rd edition. Haemophilia 2020 Aug;26 Suppl 6:1-158.

Box 1. Outline of the Hemophilia Joint Health Score (HJHS), peadiatric Haemophila Activities List (pedHAL) and Haemophilia Activities List (HAL)

HJHS [11,12]	• Physical examination of both elbows, knees and ankles • 9 items: swelling, duration swelling, muscle atrophy, crepitus on motion, flexion loss, extension loss, joint pain, strength, global gait • HJHS total score (0 – 124) • 30-60 minutes
pedHAL [13]	• 53 items • 7 domains: sitting/kneeling/standing, functions of the legs, functions of the arms, use of transportation, self-care, household tasks, leisure activities and sports • Sum and domain scores (0 – 100) • 10 minutes
HAL [14,15]	• 42 items • 7 domains: lying down/sitting/kneeling/standing, functions of the legs, functions of the arms, use of transportation, self-care, household tasks, leisure activities and sports • Scores (0 – 100): sum, upper extremity component, basic lower extremity component, complex lower extremity component • 10 minutes

PSYCHOMETRICS

Appropriate outcome assessment is essential in medicine and measurement properties of existing measurement instruments should be known by caregivers and researchers for adequate interpretation of outcomes. Because many synonyms are used in psychometrics, measurement properties in this thesis were defined according the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) taxonomy (Box 2) [16]. Some important measurement properties of traditionally used legacy instruments in haemophilia care were not established, such as the reliability and responsiveness of the HJHS and HAL [17,18].

Box 2. Definitions of measurement properties according the COSMIN taxonomy

Reliability	The degree to which the measurement is free from measurement error.
Validity	The degree to which an outcome measure measures the construct it purports to measure.
Responsiveness	The ability of an outcome measure to detect change over time in the construct to be measured.
Interpretability[a]	The degree to which one can assign qualitative meaning—that is, clinical or commonly understood connotations—to an instrument's quantitative scores or change in scores.

[a] Interpretability is not considered a measurement property, but an important characteristic of a measurement instrument

Most legacy instruments like the HJHS, pedHAL and HAL were developed according the Classical Test Theory (CTT). CTT has the disadvantage that all items need to be answered by every patient, resulting in lengthy questionnaires including patient-irrelevant items. In contrast, in Item Response Theory (IRT) it is not necessary to answer all items of the outcome measure. This is applied in Computer Adaptive Testing (CAT), where selection of the next item depends on the response on the earlier items. This lowers the burden of outcome assessment by a smaller number but more relevant questions, while increasing measurement precision [19]. A promising way of outcome assessment based on IRT is Patient Reported Outcomes Measurement Information System (PROMIS). PROMIS is a set of universal, person-centered measures about physical, mental, and social health in adults and children [20]. The feasibility and validity of PROMIS is not yet studied for PWH.

OUTLINE OF THESIS

The general aim of this thesis is to optimize outcome assessment of functions, activities and participation in PWH in two ways:

- Improving the interpretation of legacy instruments;
- Reducing the time-investment of completing outcome assessment for PWH, caregivers and researchers.

HJHS

In order to recommend the optimal frequency of HJHS assessment, in **Chapter 2** we evaluate changes in joint health in adult PWH over a 5- to 10- year period as measured by the HJHS. **Chapter 3** is our first step to develop a shorter and/or more convenient version of the HJHS for the measurement of joint function in children and young adults (aged 4-30 years) with haemophilia, by combining real-life data (n=499) and expert opinion.

pedHAL

In **Chapter 4** we explore the optimal frequency of administering the pedHAL, assess child-parent agreement and identify which pedHAL domains yielded most information in Dutch PWH. In **Chapter 5** we evaluate which items of the pedHAL are redundant to construct a shorter version of the pedHAL for the measurement of activities and participation in children and youth with haemophilia, by pooling international data (n=315).

HAL

In **Chapter 6** we evaluate the test-retest reliability and interpretability of the HAL. **Chapter 7** evaluates which items of the HAL are redundant to construct a shorter version of the HAL for the measurement of activities and participation in adults with haemophilia, by pooling international data.

PROMIS

As a next step to lower the burden of outcome assessment for PWH, caregivers and researchers, in **Chapter 8** we evaluate the feasibility, measurement properties and relevance of PROMIS item banks.

REFERENCES

1 Srivastava A, Santagostino E, Dougall A, et al. WFH Guidelines for the Management of Hemophilia, 3rd edition. Haemophilia *2020*; 26(Suppl 6):1–158.

2 Stephensen D, Tait R, Brodie N, et al. Changing patterns of bleeding in patients with severe haemophilia A. Haemophilia *2009*; 15: 1210–1214.

3 van Vulpen LFD, Mastbergen SC, Lafeber FPJG, Schutgens REG. Differential effects of bleeds on the development of arthropathy – basic and applied issues. Haemophilia *2017*; 23: 521–527.

4 Fischer K, van der Bom JG, Mauser-Bunschoten EP, Roosendaal G, van den Berg HM. Effects of haemophilic arthropathy on health-related quality of life and socio-economic parameters. Haemophilia *2005*; 11: 43–48.

5 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: 1256–1264.

6 Miesbach W, O'Mahony B, Key NS, Makris M. How to discuss gene therapy for haemophilia? A patient and physician perspective. Haemophilia *2019*; 25: 545–557.

7 Oldenburg J, Mahlangu JN, Kim B, et al. Emicizumab Prophylaxis in Hemophilia A with Inhibitors. N Engl J Med *2017*; 377: 809–818.

8 Mancuso M, Santagostino E. Outcome of Clinical Trials with New Extended Half-Life FVIII/IX Concentrates. J Clin Med *2017*; 6: 39.

9 Shapiro S, Makris M. Haemophilia and ageing. Br J Haematol *2019*; 184: 712–720.

10 World Health Organization. International Classification of Functioning, Disability and Health: Children & Youth Version: ICF-CY *2007*.

11 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.

12 Feldman BM, Funk SM, Bergstrom B-MM, et al. Validation of a new pediatric joint scoring system from the international hemophilia prophylaxis study group: Validity of the hemophilia joint health score. Arthritis Care Res *2011*; 63: 223–230.

13 Groen WG, van der net J, Helders PJM, Fischer K. Development and preliminary testing of a Paediatric Version of the Haemophilia Activities List (pedhal). Haemophilia *2010*; 16: 281–289.

14 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: 565–571.

15 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: 36–46.

16 Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol *2010*; 63: 737–745.

17 Timmer MA, Gouw SC, Feldman BM, et al. Measuring activities and participation in persons with haemophilia: A systematic review of commonly used instruments. Haemophilia *2018*; 24: e33–49.

18 Gouw SC, Timmer MA, Srivastava A, et al. Measurement of joint health in persons with haemophilia: A systematic review of the measurement properties of haemophilia-specific instruments. Haemophilia *2019*; 25: e1–10.

19 Fries JF, Witter J, Rose M, et al. Item response theory, computerized adaptive testing, and promis: Assessment of physical function. J Rheumatol *2014*; 41: 153–158.

20 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: 1179–1194.

HJHS

CHAPTER 2

Monitoring joint health in haemophilia: factors associated with deterioration

Isolde A.R. Kuijlaars;[1,2]
Merel A. Timmer;[1,2,3]
Piet de Kleijn; [1,3]
Martijn F. Pisters;[2,3,4]
Kathelijn Fischer[1]

[1] Van Creveldkliniek, University Medical Center Utrecht, Utrecht, The Netherlands
[2] Physical Therapy Sciences, program in Clinical Health Sciences, University Medical Center Utrecht, Utrecht, The Netherlands
[3] Physical Therapy Research, Department of Rehabilitation, Physiotherapy Science & Sport, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands
[4] Center for Physical Therapy Research and Innovation in Primary Care, Leidsche Rijn Julius Health Care Centers, Utrecht, The Netherlands

Haemophilia. 2017;23: 934-940

ABSTRACT

Introduction

Joint bleeds in patients with haemophilia may result in haemophilic arthropathy. Monitoring joint health is essential for identifying early signs of deterioration and allow timely adjustment of treatment.

Aim

The aim was to describe changes in joint health over 5-10 years follow-up and identify factors associated with joint health deterioration in patients with haemophilia.

Methods

A post-hoc analysis was performed from previous cohort studies in patients with moderate/severe haemophilia, ≥16 years. Joint health of ankles, knees and elbows was measured with the Haemophilia Joint Health Score (HJHS) from 2006-2008 (T0) to 2011-2016 (T1). Analyses were performed on patient level (ΔHJHS-total) and joint level (ΔHJHS-joint). Deterioration was defined as ΔHJHS-total ≥4 and ΔHJHS-joint ≥2.

Results

Sixty-two patients (median age 25, 73% severe haemophilia, median [interquartile range] 0.0 [0.0;2.0] joint bleeds between T0-T1) were included. After median 8 years, HJHS-total deteriorated in 37% and HJHS-joint in 17%. Ankle joints (31%) showed deterioration more often than elbows (19%) and knees (3%). Deterioration of HJHS-total was only associated with severe haemophilia. Deterioration of HJHS-joint was weakly associated with a lower HJHS at baseline and more self-reported limitations in activities, and strongly with more joint bleeds between T0-T1 and presence of synovitis.

Conclusion

In 37% of patients with moderate/severe haemophilia and low joint bleeding rates, joint health deteriorated over 5-10 years. Ankle and elbow joints showed deterioration most frequently. Factors found in the current study help to identify which joints need frequent monitoring in patients with haemophilia with access to early prophylaxis.

INTRODUCTION

Persons with haemophilia (PWH) experience intra-articular and intramuscular bleeding; eventually joint bleeds may result in haemophilic arthropathy (HA) [1]. The mechanism of HA is multifactorial; joint bleeds affect cartilage directly, as well as indirectly through synovial inflammation [2]. HA leads to pain, loss of range of motion and muscle atrophy resulting in loss of activities and restrictions in participation [1,3]. In the Netherlands prophylactic clotting factor replacement therapy was introduced in 1968. This medical treatment is proven to be effective: it prevents bleeds and subsequent arthropathy [4,5].

Monitoring joint health is essential for identifying early signs of deterioration as it enables adjustments in clotting factor replacement therapy, physical therapy, use of walking aids or prescription of braces to limit further decline. Previous studies detected no or minimal changes in joint health over the years measured with the radiologic Pettersson score and World Federation of Hemophilia (WFH) physical examination score in patients treated with prophylaxis [6,7]. However, early joint alterations remain undetected on the Pettersson score as X-ray only shows osteochondral changes [8]. Furthermore, in a paediatric population the WFH physical examination score is less sensitive than the Haemophilia Joint Health Score (HJHS), which is developed more recently [9] and is the recommended tool for routine patient follow-up assessments of joint health [10]. The WFH guidelines recommend annual use of the HJHS during regular evaluations [1], although assessment must be performed by a trained physiotherapist and is time consuming. Data on the occurrence and rate of deterioration in HJHS scores in adults with low bleeding rates are lacking. Identifying patients and/or joints at risk for deterioration may help individualize monitoring schedules and promote efficiency without jeopardizing the quality of care.

Known factors related to joint health in haemophilia are severity of disease, use of prophylactic clotting factor replacement, number of joint bleeds, radiological status, synovitis and Body Mass Index (BMI) [1,2,4,11]. In addition, limitations in activities could predict joint health deterioration, as demonstrated in patients with osteoarthritis [12].

The aim of this study was to describe changes in joint health over a five to ten years follow-up and identify factors associated with joint health deterioration in adult patients with moderate or severe haemophilia.

MATERIALS AND METHODS

Study design and study population

This study was a post-hoc analysis using HJHS data collected for previous cohort studies and data from medical files. Studies used for our post-hoc analysis reported the HJHS in patients with moderate (1-5 IU/dL factor VIII/IX activity) or severe (<1 IU/dL factor VIII/IX activity) haemophilia treated at the Van Creveldkliniek in Utrecht. In this study we included data of subjects with two HJHS measurements with an interval of at least five years. For each subject the last available HJHS score was selected to get the follow-up period as long as possible. Patients aged <16 years at first measurement were excluded.

Regardless of study participation, all patients visited the clinic at least annually, including evaluation and documentation of treatment and bleeding. The HJHS at T0 was derived from the studies by Den Uijl *et al.* (2013, 2014) and Fischer *et al.* (2013) [13-15]. For the follow-up measurement (T1) the HJHS was derived from the studies by Nijdam *et al.* (2016) [16] and routine measurements documented in medical files. This resulted in a data collection period from January 2006 – August 2008 (T0) up to November 2011 – May 2016 (T1). These previous studies were approved by the Medical research ethics committee (MREC) of the University Medical Centre Utrecht (06-248, 06-002, 11-442) and informed consent included permission for subsequent analyses of joint outcome data.

The potential factors self-reported limitations in activities (Haemophilia Activity List [HAL]) and radiological status were obtained from Den Uijl *et al.* (2013), Den Uijl *et al.* (2014) and Fischer *et al.* (2013) [13-15]. In addition, patient characteristics, severity of disease, number of joint bleeds, use of prophylaxis, presence of synovitis and BMI were extracted from patient logs and medical files.

Measurements

Outcome

The primary outcome was joint health of elbows, knees and ankles measured with the HJHS 2.1, which consists of eight item scores on joint level and a global gait score. Scores range from 0 to 20 per joint and the global gait score ranges from 0 to 4, resulting in a HJHS-total score (0 to 124). A higher score indicates worse joint health [17]. In this study the HJHS-total score and the HJHS-joint scores of the HJHS version 2.1 were reported.

Scores of T0 were measured with HJHS 1.0 and were converted to HJHS 2.1 by recoding of original range of motion data according to the manual.

Since this tool was developed for detection of early joint changes the manual of the HJHS does not prescribe how items have to be scored in case of joint replacement or arthrodesis [17]. It was decided to score joints after joint replacement or arthrodesis similar to joints without joint replacement or arthrodesis, and to correct for a history of surgery in the statistical analyses.

Factors associated with joint health deterioration

Disease severity and medication use. Severity of disease was reported as moderate or severe. The use of prophylactic clotting factor between T0 and T1 was reported in four categories: (1) no prophylaxis, (2) continuous prophylaxis, (3) non-compliant use of prophylaxis (according to the notes in the medical file) and (4) change from prophylactic clotting factors to on demand use or vice versa.

Joint bleeds. The number of joint bleeds between T0 and T1 was reported per joint for elbows, knees and ankles. Joint bleeds were defined as any complaint in elbows, knees or ankles requiring treatment with clotting factor concentrate.

Joint status. Joint health at baseline (T0) was measured with the HJHS 2.1 [17]. The radiological status of the joints at baseline (T0) was scored by means of the Pettersson score [18]. Knees, elbows and ankles were evaluated with a maximum score of 13 points per joint. Higher scores reflect more severe arthropathy [18]. Pettersson scores available within 2.5 years of T0 measurement of the HJHS were included. For consistency, all Pettersson scores were performed by two radiologists. The presence of synovitis between T0 to T1 was reported per joint. Synovitis was considered present when documented in the patient file and treated according to the local synovitis protocol in which synovitis is defined as a painless swelling and warmth of the joint on clinical exam.

Age, BMI and limitations in activities. Age in years was reported at baseline. BMI (kg/m^2) was calculated with the height and weight. Self-reported limitations in activities at baseline (T0) were measured with the HAL [19,20]. The HAL is a validated 42-item haemophilia-specific self-administered questionnaire assessing self-reported limitations in activities in eight domains. Normalized scores range from 0 to 100, where 100 represents no limitations in activities [20].

Patient characteristics
Type of disease (haemophilia A or B), regimen of prophylaxis, presence of Hepatitis C Virus and/or Human immunodeficiency virus and history of surgery (joint replacement or arthrodesis) were reported as patient characteristics.

Statistical analyses

Descriptive results were presented as proportions or medians (interquartile ranges [IQR]). Analyses were conducted on patient level (HJHS-total) and joint level (HJHS-joint). To account for correlation of joint scores within patients, all analyses on joint level were performed using multilevel models [21]. Change (Δ) scores between T0 and T1 were calculated for the HJHS-total score, HJHS-joint score and for the elbow, knee and ankle joints separately *(ΔHJHS = HJHS T1 – HJHS T0)*. Cut-off points for clinical relevant changes were ≥|4| for the HJHS-total score and ≥|2| on joint level. Cut-off points were based on expert opinion (KF, MT) and a published range of 0-3 points on the HJHS-total score in young adults without haemophilia [22]. Differences in HJHS-total scores between T0 and T1 were tested by the non-parametric Wilcoxon signed-rank test. At joint level differences in HJHS scores were tested with a univariate three-level regression including the level measurement point, patient and joint.

Individual factors associated with ΔHJHS-total score were determined with univariate linear regression analyses. Multicollinearity between the determinants was checked. Subsequently, to determine factors associated with the ΔHJHS-total score a multivariate linear regression analysis was performed. Determinants were selected stepwise backward. Variables were removed if $p>0.10$.

Factors associated with ΔHJHS-joint score were determined with univariate and multivariate two-level regression analyses, including adjustment for joint type (elbow, knee or ankle). The best fitting model was chosen based on the lowest Akaike Information Criterium (AIC) value [21]. All analyses for determining factors associated with ΔHJHS were adjusted for time between HJHS measurement at T0 and T1 and history of joint surgery. Unstandardized β with 95% confidence intervals (95%-CI) were presented.

Sensitivity analyses were done with other cut-off scores (ΔHJHS-joint ≥|3|, ΔHJHS-total ≥|6|) for HJHS changes. In addition, the multivariate two-level regression was performed excluding the joints with a history of surgery.

Multiple imputations were used to impute missing data in this study [23]. Ten imputed data sets were created, which were analyzed separately. The results of the ten analyses were combined with the Rubin's rules [23].

SPSS version 22 (IBM Corp., Armonk, New York, USA) was used for the statistical analyses.

RESULTS

Patients and joint characteristics

Sixty-two patients were included in this post-hoc analysis. Table 1 and 2 show the patient and joint characteristics. Median age at baseline was 25.1 (mean age 28.4), ranging from 16 to 58 years. Forty-five patients had severe haemophilia. The follow-up period varied from 5.1 to 10.1 years, with a median of 8.0 years. A total of 372 joints were measured, including nine joints after total joint replacement or arthrodesis. About half of the joints (47.8%) had ≥1 joint bleed between T0 and T1. The percentage of joints with ≥1 joint bleed and the median number of joint bleeds was highest for the ankle joints. Pettersson scores and HAL scores were missing for respectively 45.2% (n=28) and 22.6% (n=14) of the patients. Joint bleeds of the elbows and knees were missing in 3.2% (n=2) and of the ankles in 4.8% (n=3) of the patients. Missing data were Missing at Random (MAR).

Table 1. Patient characteristics

Patient characteristics (n=62)	Median (IQR), n (%)
Age (years)	25.1 (20.8 ; 33.4)
Body Mass Index (kg/m^2)	24.0 (22.3 ; 27.4)
Haemophilia A	56 (90.3)
Severe haemophilia	45 (72.6)
Clotting factor	
No prophylaxis	17 (27.4)
Continuous prophylaxis	23 (37.1)
Non-compliant use of prophylaxis	14 (22.6)
Change prophylaxis to on demand or vice versa	8 (12.9)
Frequency of prophylaxis per week	3.0 (2.3 ; 3.0)
Dose of prophylaxis, IU	1000 (1000 ; 1000)
HCV-positive	12 (19.4)
HIV-positive	4 (6.5)
History of joint surgery	6 (9.7)

HCV = Hepatitis C Virus; HIV = Human immunodeficiency virus; HJHS = Haemophilia Joint Health Score.

Table 2. Joint characteristics at baseline and during follow-up

	Median (IQR), %			
	Elbow (n=124)	**Knee (n=124)**	**Ankle (n=124)**	**Total (n=372)**
Baseline characteristics				
HJHS	0.0 (0.0 ; 2.0)	0.0 (0.0 ; 1.0)	1.0 (0.0 ; 3.0)	0.0 (0.0 ; 2.0)
HJHS-joint level score ≥ 2	27.4	15.3	38.7	17.5
During follow-up (T0 – T1)				
Joint bleeds	0.0 (0.0 ; 2.0)	0.0 (0.0 ; 1.0)	1.0 (0.0 ; 3.0)	0.0 (0.0 ; 2.0)
≥ 1 joint bleed	45.2	41.9	56.5	47.8
Synovitis	4.0	2.4	3.2	3.2
Before and during follow-up period				
History of joint surgery	0.8	1.6	4.8	2.4

HJHS = Haemophilia Joint Health Score.

Change in HJHS

Changes in HJHS-total scores (ΔHJHS ≥|4|) and HJHS-joint scores (ΔHJHS ≥|2|) are shown in Figure 1. HJHS-total score increased significantly ($p<0.001$) from a median of 8.5 (IQR 3.8;14.8) at T0 to 11.0 (IQR 4.0;19.0) at T1. In 37.1% (n=23) of the patients the HJHS-total score increased by a minimum 4 points over time. HJHS-joint score remained stable with median scores of 0.0 (IQR 0.0;2.0) at T0 to 0.0 (IQR 0.0;3.0) at T1. In 17.5% (n=65) of the joints the HJHS-joint score deteriorated by a minimum of 2 points. Ankle joints (30.6%) showed deterioration more often than the elbows (18.5%) and knees (3.2%). The HJHS scores for knee joints did not change significantly from T0 to T1 (3.2% deterioration, $p=0.060$). Improvement of joint health was found in a small proportion of the patients and joints (9,7% and 8,3%, respectively). Sensitivity analyses with higher cut-off scores (ΔHJHS-total ≥|6| and ΔHJHS-joint ≥|3|) showed higher rates of joints which stayed constant during follow-up (HJHS-total 66.1%, elbow 86.3%, knee 92.7% and ankle 71.8%). Ankle and elbow joints deteriorated more often.

In addition, a flow chart (Figure 2) was made to show the follow-up of joints without joint impairment at baseline. Of the joints without impairment (HJHS-joint ≤1) at baseline, with ≤1 joint bleed and no synovitis during follow-up, 91.9% of the joints maintained HJHS-joint scores ≤1 at T1.

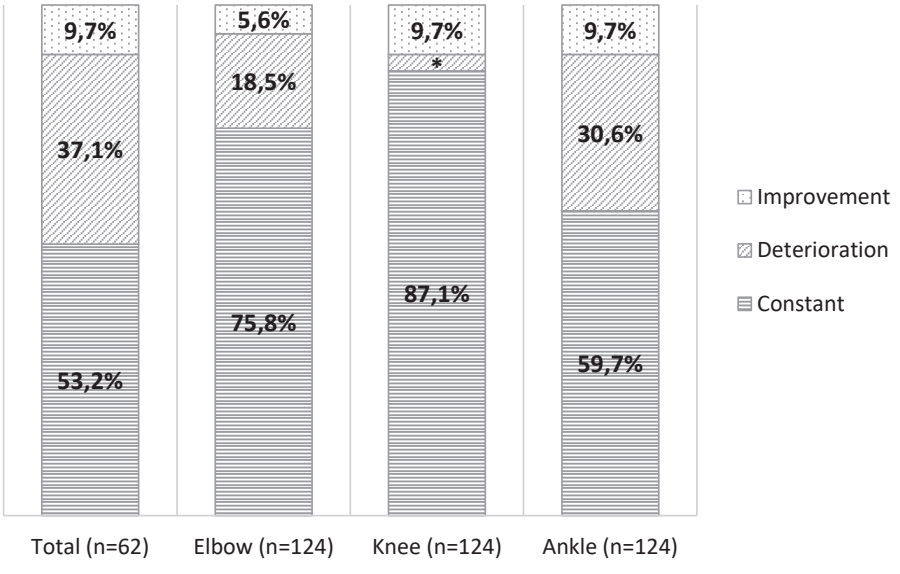

Figure 1. Change of HJHS-total score and joint scores. ΔHJHS-total score deterioration ≥ 4; improvement ≥ -4; constant from -3 to 3. ΔHJHS-joint score deterioration ≥ 2; improvement ≥ -2; constant from -1 to 1. * = 3.2%.

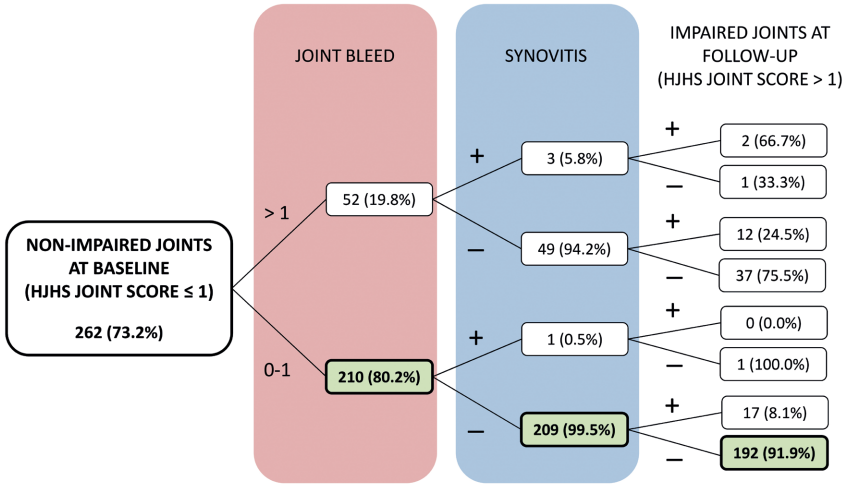

Figure 2. Observed development of HJHS-joint scores in non-impaired joints, stratified by reported joint bleeds (> 1) and presence of synovitis during a five to ten year follow-up. + present; - absent. Complete case analysis for joints.

Factors associated with joint health deterioration

Multicollinearity was found between radiological status and joint health at baseline. Since joint health is more often available in daily care, joint health at baseline was included in the multivariate analyses. Use of prophylaxis correlated with severity of disease. Because use of prophylaxis correlated most with the other factors, this factor was not included in the multivariate analyses.

Factors associated with overall change in joint health over time

Univariate linear regression analyses resulted in two factors significantly associated with ΔHJHS; severity of disease and total number of joint bleeds between T0-T1. In the multivariate linear regression model severe haemophilia was the only factor associated with joint health deterioration (β [95%-CI]: 4.60 [1.07;8.13], p=0.011). The univariate and multivariate linear regression models studying the potential factors associated with ΔHJHS-total score are presented in the Supplementary material.

Factors associated with change in joint health on joint level over time

The multivariate two-level regression models of factors associated with ΔHJHS-joint score are presented in Table 3, data on the univariate two-level regression models are shown in Supplementary material. Univariate two-level regression analyses resulted in six factors significantly associated with ΔHJHS; severity of disease, joint health at baseline, limitations in activities, joint type , number of joint bleeds and presence of synovitis. In the multivariate two-level regression analyses six factors were independently associated with deterioration of HJHS-joint; better joint health at baseline, lower BMI, more limitations in activities, joint type, higher number of joint bleeds and presence of synovitis. The association between ΔHJHS and time between T0 and T1 was not significant (p=0.179), signifying that follow-up times of five or ten years did not influence the HJHS scores. Random slopes for the variables number of joint bleeds and joint health at baseline were added, which improved the model fit.

The sensitivity analysis of the multivariate two-level regression model excluding the joints with a history of joint surgery yielded similar results.

Table 3. Multivariate two-level regression model for Δ Haemophilia Joint Health Score (HJHS) at joint level

Potential factors	β (95%-CI)	p-value
Baseline characteristics		
Severe haemophilia (compared to moderate)	—	—
Joint health (per point)	-0.31 (-0.45 ; -0.18)	<0.001
Age (per year)	—	—
Body Mass Index (per kg/m^2)	-0.05 (-0.11 ; +0.01)	0.093
Limitations in activities (HAL, per point)	-0.04 (-0.07 ; -0.02)	<0.001
Joint type (knee = reference)		
Elbow	0.54 (+0.12 ; +0.97)	0.012
Ankle	1.23 (+0.79 ; +1.66)	<0.001
During follow-up (T0 – T1)		
Joint bleeds (per bleed)	0.21 (+0.10 ; +0.33)	<0.001
Presence of synovitis	1.78 (+0.54 ; +3.01)	0.005
Parameters used for adjustment of the model		
Time between HJHS measurement (per year)	-0.13 (-0.28 ; +0.03)	0.112
History of joint surgery	1.19 (-0.54 ; +2.91)	0.179

DISCUSSION

This study describes changes in joint health in PWH over a five to ten years period as measured by the HJHS. After a median of 8.0 years HJHS scores decreased in 37.1% of patients (≥4 points) and in 17.5% of joints (≥2 points). Deterioration was most prevalent in ankle joints (30.6%). The majority (91.9%) of joints without impairment at baseline, with ≤1 joint bleed and no synovitis during follow-up showed no deterioration during five to ten years follow-up.

Both on HJHS-total and HJHS-joint level factors associated with change in joint health were identified. Patients with severe haemophilia were more likely to show deterioration on the HJHS-total score than patients with moderate haemophilia. No other factors were associated with deterioration of the HJHS-total score. At joint level, the presence of synovitis, joint type and increased joint bleeds were the most important factors associated with deterioration. This information may be used to determine which joints need more frequent monitoring.

Internal and external validity

Results of the present study depend on both the population included and psychometric properties of the HJHS, which have not been widely investigated for adult PWH. The majority of the study population had very limited joint changes and low bleeding rates (median number of joint bleeds 0.0/joint [IQR 0.0;2.0] during follow up of median 8 years) due to access to early prophylaxis. This is the population that the HJHS was designed for, but limits representativeness of these findings in settings with more prevalent arthropathy and/or higher bleeding rates.

In this study several patients had undergone joint surgery. Currently the HJHS manual does not give directions on how to score joints after surgery. Since the HJHS is recommended for adult patients nowadays, agreement among health professionals and researchers about scoring of joints in patients after joint surgery is needed. Given the uncertainty how to scores these joint, we performed a sensitivity analysis, which showed similar results.

For this longitudinal study with limited change rates adequate responsiveness of the HJHS is essential. Currently, information on responsiveness is still insufficient but the evidence regarding responsiveness is emerging. The HJHS was able to measure improvement in joint status three months after radiosynovectomy [24] and was able to distinguish between severe and non-severe haemophilia and different treatment groups [9,13,16,25]. In the current study, sensitivity analysis of the cut-off scores for changes of the HJHS showed that higher cut-off scores resulted in more joints which are indicated as stable joints.

Finally, in the present study HJHS assessments were performed by two physical therapists who were experienced with the HJHS and trained together to calibrate HJHS assessment.

While most studies analyze joint health at patient level, we focused on both patient and joint level. Since most joints were unaffected, HJHS sum scores were low. Analyses at joint level gave more specific information about joint conditions. The more direct association of joint specific factors on joint health may explain why at total and joint level different factors were associated with HJHS changes.

Comparison with other studies

In the current study most patients showed minimal HJHS changes over time. This is in line with the minimal changes in joint health measured with the WFH physical examination score and radiologic Pettersson score in previous reports on young adults in Sweden and the Netherlands, who also have access to early prophylaxis [6,7].

The observation that the ankle was the most affected joint in the current study is in accordance with earlier observations [26]. It is hypothesized that physical abilities and activity levels of PWH increased after the institution of early prophylactic replacement therapy. The increased participation in sports and activities could have resulted in higher impact on ankle joints and thus a higher bleeding frequency in ankles compared to knees and elbows [26]. The number of bleeds at joint level during 8 years of follow-up was very low and only ankles, knees and elbows were considered. Overall joint bleed rates were not calculated and cannot be compared with other studies.

Clinical implications and future research

Results suggest that not all patients need high frequent monitoring of all six joints by means of a complete HJHS. Time was not associated with joint health deterioration during a follow-up of five to ten years. This implicates that monitoring all six joints every five years in PWH on long-term prophylaxis with low bleeding rates seems to be a safe interval when there are no reported bleeds or synovitis. However, less frequent monitoring of joints will only be safe in the presence of reliable and timely bleeding reporting. Particularly joints suffering from repeated bleeding and/or synovitis are at risk for deterioration and should be closely and frequently monitored. The 3 factors 'better joint health at baseline', 'lower BMI' and 'more self- reported limitations in activities' were of minor interest for clinical practice because of small coefficients. In the absence of data on the optimum interval for monitoring joint health following synovitis and/or frequent bleeding, we suggest to follow the WFH guidelines and recommend monitoring these joints at least annually. Less frequent monitoring of joints without bleeding/synovitis saves time that can be used for efficient care in acute situations in PWH, especially by physical therapists. For more exact determination of the optimum interval, we suggest a study including more frequent measurements in this patient group. Although systematic and repeated joint health assessment by a trained physiotherapist does not directly improve joint health, it allows for early detection of changes and therefore adaptation of prophylactic treatment initiation of physiotherapy

treatment. Moreover, confronting PWH with joint changes may promote adherence to prophylaxis [27].

CONCLUSION

Joint health deteriorated over 5-10 years in 37.1% of the patients with moderate or severe haemophilia and low joint bleeding rates. Ankle and elbow joints showed deterioration most frequently. Deterioration in joint health was associated with joint type, increased joint bleeding and presence of synovitis. Joints without impairment that suffered ≤1 joint bleed and no synovitis during follow-up remained healthy during five to ten years follow-up. Monitoring all six joints every five years seems to be a safe interval when there are no reported bleeds or synovitis in PWH with access to early prophylaxis.

REFERENCES

1 Srivastava A, Brewer AK, Mauser-Bunschoten EP, et al. Guidelines for the management of hemophilia. *Haemophilia.* 2013; 19: e1-47.

2 Jansen NW, Roosendaal G, Lafeber FP. Understanding haemophilic arthropathy: an exploration of current open issues. *Br J Haematol.* 2008; 143: 632-40.

3 Heijnen L, de Kleijn P, Roosendaal G, van Rinsum AC. Orthopedische behandeling en revalidatie bij problemen van het houdings- en bewegingsapparaat bij hemofiliepatiënten. In: Leebeek FWG & Mauser- Bunschorten EM, ed. Nederlandse Vereniging van Hemofiliebehandelaars (NVHB). Richtlijn Diagnostiek en behandeling van hemofilie en aanverwante hemostasestoornissen. Alphen aan de Rijn (NL): Van Zuiden Communications B.V.; 2009. p. 91-100 (in Dutch).

4 Iorio A, Marchesini E, Marcucci M, Stobart K, Chan AK. Clotting factor concentrates given to prevent bleeding and bleeding-related complications in people with hemophilia A or B. *Cochrane Database Syst Rev.* 2011; (9):CD003429. doi: CD003429.

5 van Creveld S. Prophylaxis of joint hemorrhages in hemophilia. *Acta Haematol.* 1971; 45: 120-7.

6 Lofqvist T, Nilsson IM, Berntorp E, Pettersson H. Haemophilia prophylaxis in young patients--a long-term follow-up. *J Intern Med.* 1997; 241: 395-400.

7 Fischer K, van der Bom JG, Mauser-Bunschoten EP, et al. Changes in treatment strategies for severe haemophilia over the last 3 decades: effects on clotting factor consumption and arthropathy. *Haemophilia.* 2001; 7: 446-52.

8 Funk MB, Schmidt H, Becker S, et al. Modified magnetic resonance imaging score compared with orthopaedic and radiological scores for the evaluation of haemophilic arthropathy. *Haemophilia.* 2002; 8: 98-103.

9 Feldman BM, Funk SM, Bergstrom BM, et al. Validation of a new pediatric joint scoring system from the International Hemophilia Prophylaxis Study Group: validity of the hemophilia joint health score. *Arthritis Care Res (Hoboken).* 2011; 63: 223-30.

10 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.

11 Carpenter SL, Chrisco M, Johnson E. The Effect of Overweight and Obesity on Joint Damage in Patients with Moderate or Severe Hemophilia. *Blood.* 2015; 108: 4064-.

12 de Rooij M, van der Leeden M, Heymans MW, et al. Prognosis of Pain and Physical Functioning in Patients With Knee Osteoarthritis: A Systematic Review and Meta-Analysis. *Arthritis Care Res (Hoboken).* 2016; 68: 481-92.

13 Fischer K, Steen Carlsson K, Petrini P, et al. Intermediate-dose versus high-dose prophylaxis for severe hemophilia: comparing outcome and costs since the 1970s. *Blood.* 2013; 122: 1129-36.

14 den Uijl I, Biesma D, Grobbee D, Fischer K. Turning severe into moderate haemophilia by prophylaxis: are we reaching our goal? *Blood Transfus.* 2013; 11: 364-9.

15 den Uijl I, Biesma D, Grobbee D, Fischer K. Outcome in moderate haemophilia. *Blood Transfus.* 2014; 12 Suppl 1: s330-6.

16 Nijdam A, Foppen W, de Kleijn P, et al. Discontinuing early prophylaxis in severe haemophilia leads to deterioration of joint status despite low bleeding rates. *Thromb Haemost.* 2016; 115: 931-8.

17 Hemophilia Joint Health Score 2.1 Instruction Manual [Internet]. Available from: http://www.ipsg.ca/working-groups/Physical-Health-and-Joint-Function-(Formerly-Physical-Therapy).

18 Pettersson H, Ahlberg A, Nilsson IM. A radiologic classification of hemophilic arthropathy. *Clin Orthop Relat Res.* 1980; (149): 153-9.

19 van Genderen FR, van Meeteren NL, van der Bom JG, et al. Functional consequences of haemophilia in adults: the development of the Haemophilia Activities List. *Haemophilia.* 2004; 10: 565-71.

20 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: 36-46.

21 Hox JJ, Moerbeek M. Multilevel Analysis Techniques and Applications. 2nd revised ed. Abingdon: Taylor & Francis Group; 2010. p. 11-39.

22 Sluiter D, Foppen W, de Kleijn P, Fischer K. Haemophilia Joint Health Score in healthy adults playing sports. *Haemophilia.* 2014; 20: 282-6.

23 Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol.* 2006; 59: 1087-91.

24 Teyssler P, Taborska K, Kolostova K, Bobek V. Radiosynoviorthesis in hemophilic joints with yttrium-90 citrate and rhenium-186 sulfide and long term results. *Hell J Nucl Med.* 2013; 16: 44-9.

25 Khawaji M, Astermark J, Berntorp E. Lifelong prophylaxis in a large cohort of adult patients with severe haemophilia: a beneficial effect on orthopaedic outcome and quality of life. *Eur J Haematol.* 2012; 88: 329-35.

26 Stephensen D, Tait RC, Brodie N, et al. Changing patterns of bleeding in patients with severe haemophilia A. *Haemophilia.* 2009; 15: 1210-4.

27 Schrijvers LH, Uitslager N, Schuurmans MJ, Fischer K. Barriers and motivators of adherence to prophylactic treatment in haemophilia: a systematic review. *Haemophilia.* 2013; 19: 355-61.

SUPPLEMENTARY MATERIAL

Supplemental table 1. Univariate and multivariate linear regression models for Δ Haemophilia Joint Health Score (HJHS) total score

Potential factors	Univariate linear regression model		Multivariate linear regression model	
	β (95%-CI)	p-value	β (95%-CI)	p-value
Baseline characteristics				
Severe haemophilia (compared to moderate)	4.60 (+1.00 ; +8.20)	0.013	4.60 (+1.07 ; +8.13)	0.011
Joint health (per point)	0.02 (-0.20 ; +0.23)	0.892	—	—
Age (per year)	-0.06 (-0.24 ; +0.12)	0.481	—	—
Body Mass Index (per kg/m^2)	-0.18 (-0.63 ; +0.27)	0.422	—	—
Limitations in activities (HAL, per point)	-0.10 (-0.28 ; +0.09)	0.292	—	—
During follow-up (T0 – T1)				
Joint bleeds (per bleed)	0.19 (+0.00 ; +0.37)	0.046	—	—
Presence of synovitis	2.02 (-2.61 ; +6.66)	0.386	—	—
Parameters used for adjustment of the model				
Time between HJHS measurement (per year)	-0.70 (-1.86 ; +0.46)	0.233	-0.89 (-1.97 ; +0.21)	0.112
History of joint surgery	5.42 (-0.21 ; +11.04)	0.059	5.69 (+0.42 ; +10.97)	0.034

Interpretation: After median of 8.0 years, the HJHS total score increased by 4.60 points in patients with severe haemophilia compared to patients with moderate haemophilia.

Supplemental table 2. Univariate and multivariate linear regression models for Δ Haemophilia Joint Health Score (HJHS) joint score

Factors	Univariate two-level regression model		Multivariate two-level regression model[a]	
	β (95%-CI)	p-value	β (95%-CI)	p-value
Baseline characteristics				
Severe haemophilia (compared to moderate)	0.56 (+0.00 ; +1.12)	0.049	—	—
Joint health (per point)	-0.16 (-0.31 ; -0.01)	0.042	-0.31 (-0.45 ; -0.18)	<0.001
Age (per year)	0.01 (-0.02 ; +0.03)	0.695	—	—
Body Mass Index (per kg/m^2)	-0.03 (-0.09 ; +0.04)	0.451	-0.05 (-0.11 ; +0.01)	0.093
Limitations in activities (HAL, per point)	-0.04 (-0.06 ; -0.01)	0.004	-0.04 (-0.07 ; -0.02)	<0.001
Joint type (knee = reference)				
Elbow	0.69 (+0.18 ; +1.20)	0.008	0.54 (+0.12 ; +0.97)	0.012
Ankle	1.43 (+0.91 ; +1.94)	<0.001	1.23 (+0.79 ; +1.66)	<0.001
During follow-up (T0 – T1)				
Joint bleeds (per bleed)	0.18 (+0.06 ; +0.31)	0.003	0.21 (+0.10 ; +0.33)	<0.001
Presence of synovitis	1.89 (+0.63 ; +3.15)	0.003	1.78 (+0.54 ; +3.01)	0.005
Parameters used for adjustment of the model				
Time between HJHS measurement (per year)	-0.12 (-0.29 ; +0.06)	0.192	-0.13 (-0.28 ; +0.03)	0.112
History of joint surgery	-0.80 (-2.28 ; +0.67)	0.285	1.19 (-0.54 ; +2.91)	0.179

[a] Variances of the model for the random effects were: intercept 0.24; slope number of joint bleeds 0.10; slope joint health at T0 0.07. The residual variance was 2.57. Interpretation: After median 8.0 years, the HJHS joint score increases by 1.05 points after 5 joint bleeds; 1.78 points in case of presence of synovitis; 0.54 and 1.23 points in elbow and ankle joints respectively compared to knee joints. The HJHS joint score deteriorated less when patients had more joint impairment at baseline: i.e. 0.93 points less when the HJHS at T0 was 3 points higher. In addition, the HJHS deteriorated only 1 point when BMI was 20 kg/m^2 higher and 1 point when the HAL was 25 points higher.

CHAPTER 3

Evaluating international Haemophilia Joint Health Score (HJHS) results combined with expert opinion: options for a shorter HJHS

Isolde A.R. Kuijlaars;[1] Janjaap van der Net;[2] Brian M. Feldman;[3,4,5] Magnus Aspdahl;[6] Melanie Bladen;[7] Wypke de Boer;[8] Rubén Cuesta-Barriuso;[9,10,11] Ruth E.D. Matlary;[12] Sharon M. Funk;[13] Pamela Hilliard;[3] Judy A. John;[14] Christine L. Kempton;[15] Piet de Kleijn;[1] Marilyn Manco-Johnson;[13] Pia Petrini;[6] Pradeep Poonnoose;[16] Jean St-Louis;[17] Sylvia Thomas;[18] Merel A. Timmer;[1] Sonata Saulyte Trakymiene;[19] Leo van Vlimmeren;[20] Kathelijn Fischer[1]

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

[2] Center for Child Development, Exercise and Physical Literacy, Children's Hospital of the University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

[3] Child Health Evaluative Sciences Program, Research Institute, The Hospital for Sick Children, Toronto, Ontario, Canada

[4] Department of Pediatrics, Faculty of Medicine and the Institute of Health Policy Management & Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

[5] Division of Rheumatology, The Hospital for Sick Children, Toronto, Ontario, Canada

[6] Department of Pediatrics, Clinic of Coagulation Disorders, Karolinska University Hospital, Stockholm, Sweden

[7] Haemophilia Center, Great Ormond Street Hospital for Children NHS Foundation Trust, London, United Kingdom

[8] Department of Rehabilitation, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands

[9] Department of Physiotherapy, European University of Madrid, Madrid, Spain

[10] Royal Victoria Eugenia Foundation, Madrid, Spain

[11] Fishemo CEE, Spanish Federation of Hemophilia, Madrid, Spain

[12] Department of Haematology, Oslo University Hospital, Oslo, Norway

[13] Hemophilia and Thrombosis Center, University of Colorado Anschutz Medical Campus, Aurora, Colorado, United States of America

[14] Department of PMR, Christian Medical College, Vellore, India

[15] Hemophilia of Georgia Center for Bleeding & Clotting Disorders of Emory, Emory University School of Medicine, Atlanta, Georgia, United States

[16] Department of Orthopaedics, Christian Medical College, Vellore, India

[17] CHU Sainte-Justine, Montreal, Canada

[18] Department of Radiology, Clementino Fraga Filho University Hospital, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

[19] Clinic of Children's diseases, Faculty of Medicine, Vilnius University, Vilnius, Lithuania

[20] Department of Rehabilitation, Paediatric Physical Therapy, Radboud university medical center, Nijmegen, The Netherlands

Haemophilia. 2020;26: 1072-1080.

ABSTRACT

Introduction

The Hemophilia Joint Health Score (HJHS) was developed to detect early changes in joint health in children and adolescents with haemophilia. The HJHS is considered by some to be too time consuming for clinical use and this may limit broad adoption.

Aim

This study was a first step to develop a shorter and/or more convenient version of the HJHS for the measurement of joint function in children and young adults with haemophilia, by combining real-life data and expert opinion.

Methods

A cross-sectional multicenter secondary analysis on pooled data of published studies using the HJHS (0-124, optimum score 0) in persons with haemophilia A/B aged 4-30 was performed. Least informative items, scoring options and/or joints were identified. An expert group of 19 international multidisciplinary experts evaluated the results and voted on suggestions for adaptations in a structured meeting (consensus set at ≥80%).

Results

Original data on 499 persons with haemophilia from 7 studies were evaluated. Median age was 15.0 years [range 4.0-29.9], 83.2% had severe haemophilia and 61.5% received prophylaxis. Median (IQR) HJHS total was 6.0 (1.0-17.0). The items 'duration swelling' and 'crepitus' were identified as clinically less informative and appointed as candidates for reduction.

Conclusion

Analysis of 499 children and young adults with haemophilia showed that the HJHS is able to discriminate between children and adults and different treatment regimens. Reduction of the items 'duration swelling' and 'crepitus' resulted in the $HJHS_{short}$, which had the same discriminative ability. Additional steps are needed to achieve a substantially shorter HJHS assessment.

INTRODUCTION

Musculoskeletal assessment is an important component of the comprehensive care program for persons with haemophilia (PWH) [1]. The Hemophilia Joint Health Score (HJHS) is recommended to evaluate joint health in clinical care and research [2]. The HJHS 1.0 was developed in 2003 and further developed in versions 2.0 and 2.1 by the International Prophylaxis Study Group (IPSG) Musculoskeletal Health Expert Working Group (EWG) for detection of early joint changes in children and youth with haemophilia [3,4]. Also in adolescents the HJHS is a feasible and reliable tool [5].

HJHS assessment including scoring takes approximately 45-60 minutes per patient, which has been felt, by some, to be impractical and infeasible, especially in a busy clinical setting. Therefore, there is a demand for a shorter version of the HJHS for clinical practice and time limited settings. With more studies using the HJHS being conducted internationally over the last 15 years, there is an opportunity to determine which items, scoring options and joints are universally important for different patient populations.

This study was a first step to develop a shorter and/or more convenient version of the HJHS for the measurement of joint function in children and young adults (aged 4-30 years) with haemophilia, by combining real-life data and expert opinion.

MATERIALS AND METHODS

This study was a multicenter secondary analysis of pooled data of published studies using the HJHS. After statistical analyses of the pooled data, results of this study were discussed and suggestions for adaptations were formally voted on, in an international expert meeting on October 3rd, 2019 in Utrecht (The Netherlands). This blended methodology was chosen since no criterion standard for the construct 'joint health' is available. In addition, consensus between HJHS developers, users and investigators is needed for adaptations of the HJHS, as well as implementation of recommended adaptations.

The Medical Research Ethical Committee (MREC) of the University Medical Center Utrecht reviewed and approved the study (17-499/C).

Pooling of published HJHS data

A literature search identified 48 studies published between 2006-2019, which used either the HJHS 1.0, HJHS 2.0 or HJHS 2.1 in PWH. Forty-three of these studies had unique data. Inclusion criteria were PWH A (FVIII) or B (FIX) of all severities, aged 4-30 years. PWH were excluded if there were fewer than 5 complete items or fewer than 4 complete joints on the HJHS assessment. First, studies with <20 eligible PWH (n=9), without full text papers (n=1) or without HJHS data for all joints (n=5) were excluded. The first validation study of the HJHS was also excluded [4]. Second, of the remaining 27 studies, the authors (IK, JN, BF, KF) selected studies to create a heterogeneous mix of PWH from different countries and treatment regimens, taking into account existing collaborations with authors of the studies like the IPSG Musculoskeletal Health EWG members. A sample size of 500 has been recommended for this sort of work [6].

Authors of 16 papers were invited to share the original HJHS data (with scores on the item level) and patient characteristics. Two authors declined the invitation because HJHS item scores were unavailable. Seven authors did not reply to our request to share the data. Eventually, data of all children (4-17 years) and young adults (18-30 years) from seven remaining studies were included in the analysis [7–13]. One study used HJHS version 1.0 [8], one study used HJHS version 2.0 [7] and the other studies used HJHS version 2.1 [9–13].

Measurements

Patient characteristics collected for the included datasets were age at HJHS assessment, type of haemophilia (A or B), severity of the disease (mild [factor 0.06 IU/ml-0.40 IU/ml], moderate [factor 0.01 IU/ml- 0.05 IU/ml] or severe [factor <0.01 IU/ml]), clotting factor regimens (prophylaxis yes/no and start prophylaxis before age of 3 years yes/no) and current inhibitor status for each individual patient.

HJHS

The HJHS 2.1 is the most recent version and consists of assessments of swelling (0-3), duration of swelling (0-1), muscle atrophy (0-2), crepitus on motion (0-2), flexion loss (0-3), extension loss (0-3), joint pain (0-2) and strength (0-4) for elbows, knees and ankles and a global gait score (0-4). Scores range from 0 to 20 per joint and the global gait score ranges from 0 to 4, resulting in a total HJHS score from 0 to 124 points [14]. A higher score indicates worse joint health. Scores of version HJHS 1.0 were converted to HJHS 2.1 by recoding of the original data for the items flexion loss, extension loss and

gait (per joint) and deleting of the items axial alignment and instability. For datasets of version HJHS 2.0 and HJHS 2.1 scores on 'flexion loss' and 'extension loss' were copied.

Statistical analyses

Patient characteristics were presented as proportions or medians (interquartile ranges [IQR:P25-P75]). Descriptive analyses (median [IQR], proportions of score categories) were performed for the HJHS total scores and joint scores. HJHS scores were compared for two age groups (children [4-17 years] vs. adults [18-30 years]) and two different treatment regimens, defined as less intensive treatment (Romania, Pakistan, Lithuania, Brazil, USA) vs. intensive treatment (the Netherlands, UK, Canada) according to access to (early) prophylaxis (see Table 1).

To identify redundant items the following aspects were evaluated.

1. *Inter-item* correlations were evaluated. Inter-item correlations calculated with Spearman's rho <0.2 indicated items which do not correlate with any of the others and >0.9 indicated item redundancy [15].
2. Component loadings on exploratory factor analyses were evaluated. Factor loadings <0.5 were considered indicators of item redundancy [15]. Model fit was evaluated with the Root Mean Square Error of Approximation [RMSEA].
3. Internal consistency calculated with Cronbach's α and internal consistency after item deletion were evaluated on joint level. Cronbach's α should be between 0.7- 0.9; a higher Cronbach's α after item deletion was considered a reason to eliminate an item [15]. Global gait was included for the knees and ankles.
4. *Item-total* correlations for total joint scores were evaluated. Item-total correlations calculated with Spearman's rho <0.3 were indicators for an item that did not contribute to measurement of the construct [15].
5. Proportions of zero and maximum scores on HJHS items were analyzed for each joint (elbow, knee, ankle) to detect floor- and ceiling effects (≥85% zero or maximum scores on items) [4,15], in two age groups (children vs. adults) and two different treatment regimens, defined as less intensive treatment (Romania, Pakistan, Lithuania, Brazil, USA) vs. intensive treatment (the Netherlands, UK, Canada).

After item deletion a shortened HJHS total score ($HJHS_{short}$) was calculated. To evaluate the ability to discriminate between various patient groups, median HJHS total scores ($HJHS_{full}$ and $HJHS_{short}$) were calculated for children vs. adults and PWH receiving less

intensive treatment vs. intensive treatment. In addition, proportions of PWH with affected joints were calculated with a cut-off point of ≥4 score for $HJHS_{full}$ and ≥3 score for $HJHS_{short}$, as HJHS scores up to 3 were shown in healthy subjects based on the items crepitus and flexion loss [16].

For the comparison of the $HJHS_{full}$ and $HJHS_{short}$, scores were normalized from 0 to 100. Spearman's correlations, two-way mixed consistency Intraclass Correlation Coefficient (ICC) and Limits of Agreement (LoA) were calculated with the normalized scores.

To identify which scoring options were scored, endorsement for all scoring options for each item (% of options) was evaluated. To identify which joints were affected, descriptive analyses on joint level were performed.

Expert meeting

Nineteen international experts participated in the 1-day expert meeting. A purposive sample of members of the IPSG Musculoskeletal Health EWG, experienced users of the HJHS and investigators of the included studies was selected. The expert group included eight physicians and eleven physical therapists. The expert meeting started with a presentation of the analysis of the pooled data and published literature on the HJHS scoring system [17], crepitus in healthy subjects [16,18] and HJHS use for monitoring joint changes [19]. This was followed by three structured discussion sessions about the topics: item reduction, scoring options and number of joints assessed in the HJHS. Each discussion included five steps: presentation of results of the pooled data and statements for voting; questions for clarification of the results; first voting; discussion; final voting. Experts voted anonymously with the online tool Mentimeter.com. Results of each vote were shown to the experts when all experts completed the vote on a statement. If at least 80% of the experts agreed, consensus was reached about a statement. BF moderated the discussion sessions.

RESULTS

Patient characteristics

In the seven studies, 499 PWH A or B (children [n=325]; young adults [n=174]) were included. Four PWH were excluded because they had <5 completed items or <4 completed joints assessed on the HJHS. The data are from Romania, the Netherlands, United Kingdom, Pakistan, Lithuania, Brazil, Canada and the United States. Patient

characteristics according to age and treatment intensity are shown in Table 1. Median age at the time of HJHS assessment was 15.0 years (IQR 10.4-21.3, range 4.0-29.9). Most PWH had severe haemophilia (n=415, 83.2%). More than half of the PWH (n=307, 61.5%) used prophylaxis and 38.1% of these PWH had received early prophylaxis. Seventeen PWH (3%) had an inhibitor at HJHS assessment.

For the data from Pakistan about treatment regimen and from the United States about start of prophylaxis assumptions were made, after contact with the authors and published patient characteristics.

Table 1. Patient characteristics

		Intensive treatment (n=220)	Less intensive treatment (n=279)
		n=183	**n=142**
Children (n=275)	Age (years), median (IQR)	11.6 (8.9-14.7)	11.8 (9.0-15.0)
	Haemophilia severity, %		
	Mild	0	2.8
	Moderate	8.2	19.0
	Severe	91.8	78.2
	Prophylaxis, %	92.3	33.1
	Early prophylaxis (<3 years) / prophylaxis, %	67.8[a]	26.1[b]
		n=37	**n=137**
Adults (n=174)	Age (years), median (IQR)	24.6 (20.9-27.2)	23.8 (20.9-26.8)
	Haemophilia severity, %		
	Mild	0	7.3
	Moderate	0	20.4
	Severe	100	72.3
	Prophylaxis, %	81.1	44.5
	Early prophylaxis (<3 years) / prophylaxis, %	14.3	0

[a] Missing n=20; [b] missing n=1.

HJHS total and joint scores

The median (IQR) HJHS total score was 6.0 (1.0-17.0), with a range of 0-63. Twenty-one percent of the PWH had a total score of 0 (children 26%; young adults 10%). Young adults

had higher HJHS scores (11.5 [4.0-23.0]) than children (5.0 [0.0-12.0]). PWH receiving less intensive treatment showed higher HJHS scores (12.0 [5.0-26.0]) than PWH receiving intensive treatment (2.0 [0.0-7.0]). The ankles were the most affected joints, followed by the knees and elbows (see Figure 1).

Discussion session 1: HJHS items

Figure 2 shows the process of discussion session 1 from the statistical analyses of the pooled data up to the validation of the $HJHS_{short}$.

Selection of items eligible for item reduction

Reduction in item number was the first technique explored to reduce the time needed for HJHS assessment.

Inter-item correlations suggested no items were eligible for item reduction, since items did not show correlations >0.9 or <0.2. Swelling' and 'duration swelling' showed the strongest correlation (r = 0.78-0.80) for elbows, knees and ankles.

The exploratory factor analyses suggested no items were eligible for item reduction. A 3-factor model was selected which included all HJHS items of the elbows, knees and ankles. Three factors were identified, namely elbows, knees and ankles. The model fit of the 3-factor model was good (RMSEA = 0.05). The highest factor loading for each item was >0.5. In addition, each joint was analyzed separately using 1-factor models. The model fits of the 1-factor models were moderate (RMSEA = 0.07-0.08). For all items the factor loadings were >0.5.

The internal consistency analyses suggested no items were eligible for item reduction. HJHS items were strongly related (Cronbach's α = 0.78-0.87) without a distinct increase in Cronbach's α after item deletion of separate items, except from item deletion of global gait in the knee joint.

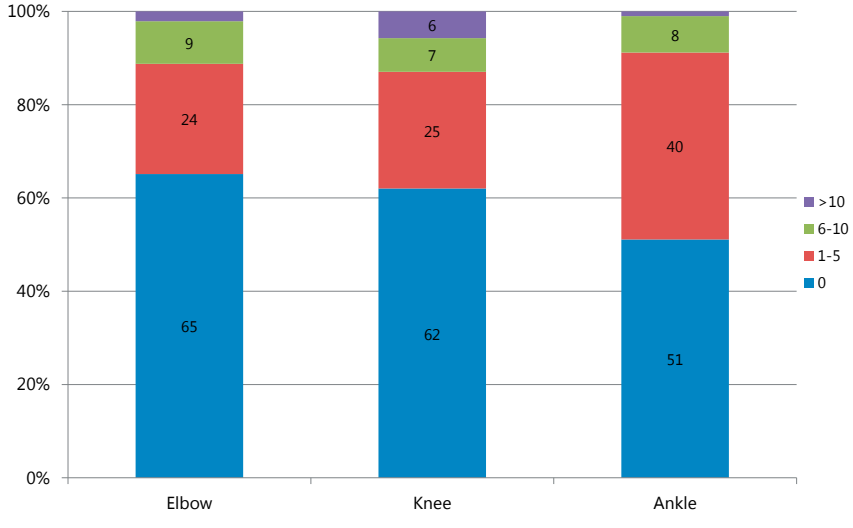

Figure 1. HJHS joint scores for all PWH (n=499). HJHS joint scores >10: 2% for the elbow, 1% for the ankle

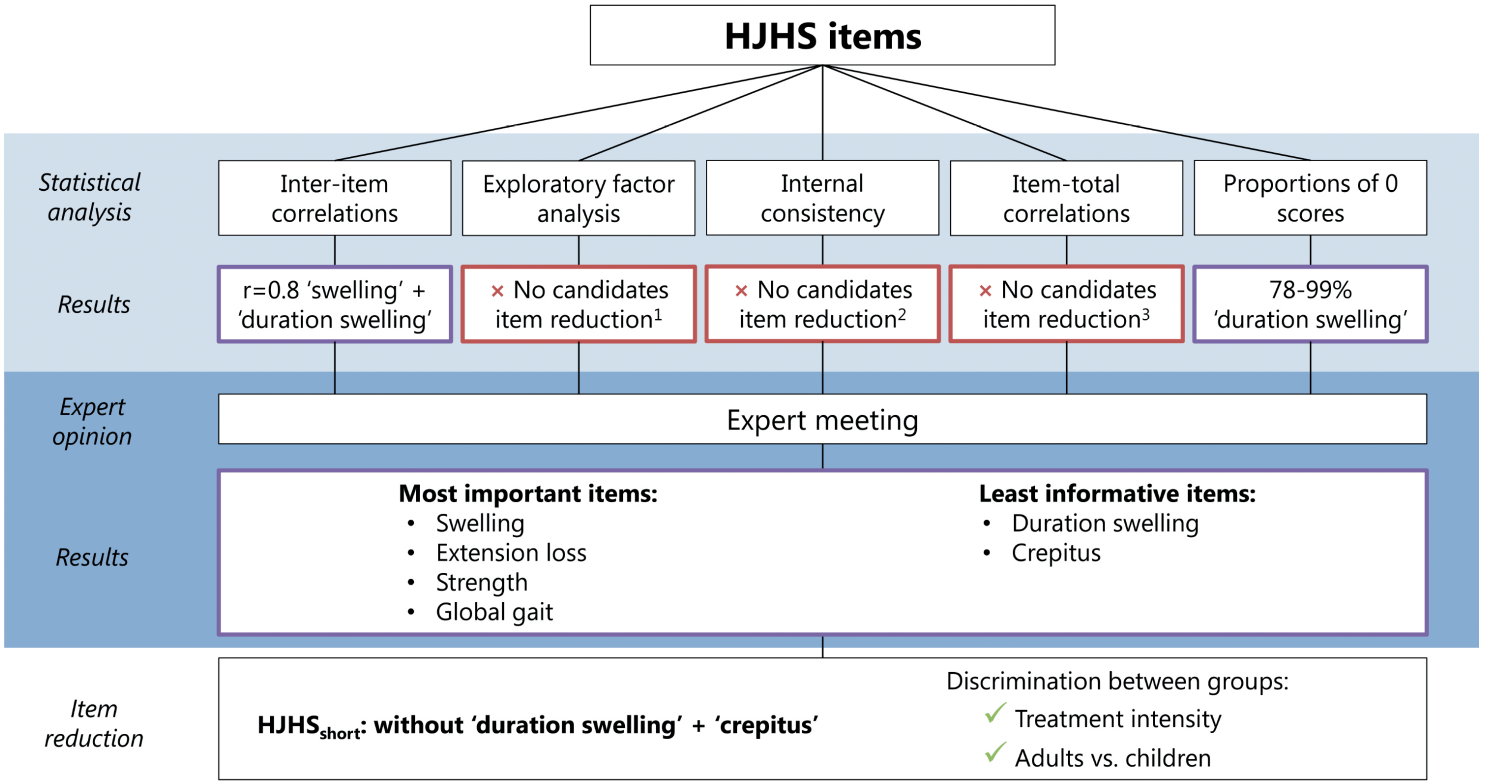

Figure 2. Flow chart of discussion session 1. Criteria for item reduction were: 1 Factor loadings <0.5; 2 Cronbach's alpha <0.7 or >0.9; 3 correlations <0.3

In addition, *item-total* correlations showed high correlations (r = 0.37-0.69), thus identifying no candidates for item reduction.

Proportions of zero scores on HJHS items were analyzed in four groups stratified by age and treatment intensity. The proportions of zero scores were highest for 'duration swelling' (varying from 78 to 98% for the different joints) and lowest for 'global gait' (35-64%). The other items had proportions of zero scores of 63-99%. Proportions of zero scores were higher in children and more intensively treated PWH.

Proportions of zero scores are shown in Table 2. Detailed data of the 3-factor exploratory factor analysis and 1-factor exploratory analyses are shown in the Supplementary material.

Expert voting 1

The results of the voting are shown in Table 3. The experts reached consensus that 'duration swelling' (95%) and 'crepitus' (95%) are redundant items. Experts discussed the reliability of the item 'duration swelling' and the potential impact of recall bias on reliability, considering this is part of the clinical history rather than physical examination of the joint. For 'crepitus' an important argument for dropping it was that crepitus is also reported frequently in healthy people [16,18]. The items considered as most important were swelling (100%: important), extension loss (100%), strength (95%) and global gait (84%). In addition, the experts discussed the item 'global gait': whether it should be part of the HJHS as a tool assessing structure and function, or whether 'global gait' should be scored separately.

Validation of $HJHS_{short}$

An $HJHS_{short}$ was created by deletion of the items 'duration swelling' and 'crepitus' (0-106). Abnormal HJHS joint scores based on 'crepitus' only were observed infrequently (elbows 1.0%; knees 3.6%; ankles 5.5%). Abnormal HJHS joint scores based on 'duration swelling' only were not observed. HJHS total scores for the $HJHS_{full}$ and the $HJHS_{short}$ and proportions of affected joints are shown in Table 4. The proportions of affected joints ($HJHS_{full}$ ≥4; $HJHS_{short}$ ≥3) were slightly higher for the $HJHS_{short}$. $HJHS_{short}$ was still able to discriminate between children vs. adults and PWH with less intensive treatment vs. intensive treatment. The normalized $HJHS_{full}$ and $HJHS_{short}$ correlated strongly in children and adults (r = 0.98) and in PWH with less (r = 0.99) and more intensive treatment (r = 0.97), with an ICC of 0.99 and LoA of -3.1 to 3.3 for the normalized scores.

Table 2. Proportions of zero scores in PWH with intensive treatment vs. less intensive treatment and children vs. adults

		Intensive treatment (n=220)			Less intensive treatment (n=279)		
		Elbow	Knee	Ankle	Elbow	Knee	Ankle
	% zero scores		**n=183**			**n=142**	
Children (n=275)	Swelling	97	98	86	79	74	83
	Duration swelling	98	98	92	87	85	91
	Atrophy	97	95	90	79	66	75
	Crepitus	98	92	85	88	78	87
	Flexion loss	95	96	93	72	74	80
	Extension loss	94	97	88	84	84	88
	Pain	98	97	97	78	74	90
	Strength	98	98	92	70	67	75
	Global gait		64			35	
	% zero scores		**n=37**			**n=137**	
Adults (n=174)	Swelling	99	95	92	86	77	72
	Duration swelling	99	95	92	91	86	78
	Atrophy	89	89	78	86	71	73
	Crepitus	91	92	78	80	69	70
	Flexion loss	66	89	74	74	73	72
	Extension loss	68	93	70	75	86	76
	Pain	97	99	97	80	77	78
	Strength	93	95	81	82	71	74
	Global gait		57			40	

In grey: <85% zero scores.

Table 3. Results of discussion session 1 aimed at identifying redundant HJHS items

	Voting 1	Voting 2
Item	**Redundant, n (%)**	**Redundant, n (%)**
Swelling	0 (0)	0 (0)
Duration swelling	11 (58)	**18 (95)**
Atrophy	7 (37)	8 (42)
Crepitus	13 (68)	**18 (95)**
Flexion loss	2 (11)	7 (37)
Extension loss	0 (0)	0 (0)
Pain	8 (42)	12 (63)
Strength	6 (33)[a]	1 (5)
Global gait	4 (21)	3 (16)

Question to experts: Is this item important or redundant? Answer options: important / redundant.
[a] 18 voters during first voting.

Table 4. Comparison of the $HJHS_{full}$ total score vs. the $HJHS_{short}$ total score, after item deletion of 'duration swelling' and 'crepitus'

	$HJHS_{full}$ (0-124)		$HJHS_{short}$ (0-106)	
	median (IQR)	**% affected (≥4)**	**median (IQR)**	**% affected (≥3)**
Intensive treatment (n=220)	2.0 (0.0-7.0)	43.2	2.0 (0.0-5.0)	44.5
Less intensive treatment (n=279)	12.0 (5.0-26.0)	78.5	10.0 (4.0-22.0)	80.3
Children (n=325)	5.0 (0.0-12.0)	55.7	4.0 (0.0-10.0)	57.8
Adults (n=174)	11.5 (4.0-23.0)	76.4	9.0 (3.0-20.0)	77.0

Proportions of PWH with affected joint were calculated with a cut-off point ≥4 for $HJHS_{full}$ and ≥3 for $HJHS_{short}$, according HJHS scores from 0-3 shown in healthy subjects with scores on crepitus and flexion loss [16].

Discussion session 2: scoring options

Frequency of endorsement for scoring options

Reduction of the number of scoring options for each item may be another way to reduce the time needed for HJHS assessment. Frequencies of endorsement for all scoring options are shown in Table 5. All items except 'duration swelling' and 'global gait' had scoring options which were scored in ≤5% of the PWH.

Table 5. Distribution (%) of the scoring options of all HJHS items, for all PWH (n=499)

Scoring option	Swelling	Duration swelling	Atrophy	Crepitus	Flexion loss	Extension loss	Pain	Strength	Global gait
0	85.4	90.5	82.8	84.1	81.9	85.8	87.5	82.8	48.9
1	9.4	9.5	14.3	12.5	9.5	7.7	9.8	11.4	15.4
2	***4.7***		***2.9***	***3.4***	***4.5***	***3.4***	***2.7***	***2.9***	13.7
3	***0.4***				***4.0***	***3.0***		***1.8***	8.7
4								***1.1***	13.3

In bold/italics: proportions <5%.

Expert voting 2

The results of the voting on scoring options are shown in the Supplementary material. For 'pain' 79% of the experts voted that the scoring options could be reduced from three

categories (*no pain through active range of motion/no pain through active range; only pain on gentle overpressure or palpation/pain through active range*) to a binominal scoring. An important argument against reduction of scoring options was that reducing the scoring options would only result in a minor reduction of the duration of HJHS assessment. It was decided that reduction of scoring options was not a feasible suggestion for shortening HJHS assessment.

Discussion session 3: joints

Joints

Reduction of the number of joints which needs assessment may be another way to reduce the time needed for HJHS assessment. The ankles were the most frequently affected joints, followed by the knees and elbows (see Figure 1). Ankles were most frequently affected in PWH on intensive treatment, while knees were most frequently affected in PWH on less intensive treatment.

Expert voting 3

The results of the voting on joints are shown in the Supplementary material, which proposed measuring a reduced number of joints. During the discussion experts suggested that screening of joints instead of a full HJHS assessment could be a way to reduce time of HJHS assessment: 'assess all joints that fail a screening examination of medical/bleeding history and a physical examination'. However, a decision regarding which items to screen and how was considered beyond the scope of this meeting. Another topic discussed by the experts was that the most affected joint is not always the joint which needs the most attention. The experts reached consensus (94%) that for clinical practice a way to reduce assessment time is that only joints that fail a screening examination should be assessed with the full HJHS. The experts did not reached consensus (74%) about the statements for a research setting which proposed measuring a reduced number of joints.

DISCUSSION

This study describes real-life HJHS data of 499 children and young adults with different treatment regimens combined with international expert opinion as a first step to develop a more convenient version of HJHS. The items 'duration swelling' and 'crepitus' were identified as candidates for item reduction. The resulting $HJHS_{short}$ was still able to discriminate between different ages and treatment regimens. Another way

of shortening HJHS assessment for clinical practice suggested by the experts was a screening examination to select joints which need full HJHS assessment.

Internal and external validity

This is the first study presenting HJHS data of 499 PWH from heterogeneous populations. The results showed different patterns of HJHS scoring in children vs. young adults and PWH with less intensive treatment vs. intensive treatment. We observed a wide variety of scores (range HJHS total: 0-63) with only 21% of the persons achieving a HJHS total score of 0 in this relatively young PWH with a wide range in treatment intensity. The real-life data in the present study were representative for clinical use and research purposes, while variation between raters was unavoidable in this study design. Some items show more variability in scoring between raters which is a limitation of the HJHS, despite the good overall interobserver reliability [20,21]. In absence of a cut-off score for affected joints according the HJHS, a cut-off score was chosen above scores (0-3) shown in healthy adults which was established in a single observer study [16].

In addition to the use of real-life data, international expert opinion of HJHS developers, clinical HJHS users and investigators using the HJHS increased the clinical value of the results. The experts who participated in the discussion sessions are representative of HJHS developers, users and researchers. This blended approach was used to compensate for the absence of a gold standard.

Comparison with other studies

According to the analyses of the pooled data, 'crepitus' was not a candidate for item reduction. However, the experts voted that 'crepitus' could be regarded redundant because crepitus is a sign which is also reported frequently (13-14%) for the knees in healthy children and young adults [16,18]. Despite 'flexion loss' being reported in ankles of healthy young adults (12%) [16] and ROM assessment is time consuming, only 37% of the experts voted that this item was redundant and should be eliminated.

Furthermore, the experts voted that 'duration swelling' was redundant. According to the analyses of the pooled data, 'duration swelling' was indeed a candidate for item reduction based on the floor effects. An additional argument of the experts was potential recall bias which could lower the reliability of this item. Although inter-observer and test-retest reliability of this item were reported in two studies (ICC=0.44-0.90)

[20,21], these findings do not support the experts' argument of recall bias since these studies did not address the risk on recall bias over six months.

Clinical implications and future research

The items 'duration swelling' and 'crepitus' were identified as candidates for item reduction. Dropping these two items will not lead to a substantial gain in time. Therefore, it is relevant to search for further ways to achieve shorter joint assessment in clinical practice. As suggested by the experts, a next step to explore is joint screening to select the joints which need full assessment.

Besides shortening the HJHS to make joint assessment more feasible in routine clinical practice, additional focus on standardization of items is needed.

CONCLUSION

This study in 499 PWH showed that the HJHS is able to discriminate between children and adults and different treatment regimens. Based on expert (n=19) consensus, reduction of the items 'duration swelling' and 'crepitus' resulted in the $HJHS_{short}$, which had the same discriminative ability. To achieve a shorter joint assessment in clinical practice, joint screening to select the joints which need full assessment was suggested.

REFERENCES

1 Srivastava A, Brewer AK, Mauser-Bunschoten EP, et al. Guidelines for the management of hemophilia. Haemophilia *2013*; 19:e1-47.

2 Fischer K, Poonnoose P, Dunn AL, et al. Choosing outcome assessment tools in haemophilia care and research: A multidisciplinary perspective. Haemophilia *2016* 1–14.

3 Feldman BM, Funk S, Lundin B, et al. Musculoskeletal measurement tools from the International Prophylaxis Study Group (IPSG). Haemophilia *2008*; 14: 162–9.

4 Feldman BM, Funk SM, Bergstrom B-MM, et al. Validation of a new pediatric joint scoring system from the international hemophilia prophylaxis study group: Validity of the hemophilia joint health score. Arthritis Care Res *2011*; 63: 223–30.

5 Fischer K, de Kleijn P. Using the Haemophilia Joint Health Score for assessment of teenagers and young adults: Exploring reliability and validity. Haemophilia *2013*; 19: 944–50.

6 Mokkink LB, Terwee CB, Patrick DL, et al. COSMIN checklist manual *2012*.

7 Groen W, Van der Net J, Lacatusu AM, et al. Functional limitations in Romanian children with haemophilia: Further testing of psychometric properties of the Paediatric Haemophilia Activities List. Haemophilia *2013*; 19: 116–25.

8 Fischer K, Carlsson KS, Petrini P, et al. Intermediate-dose versus high-dose prophylaxis for severe hemophilia: Comparing outcome and costs since the 1970s. Blood *2013*; 122: 1129–36.

9 Nijdam A, Bladen M, Hubert N, et al. Using routine Haemophilia Joint Health Score for international comparisons of haemophilia outcome: Standardization is needed. Haemophilia *2016*; 22: 142–7.

10 Khanum F, Bowen DJ, Kerr BC, Collins PW. Joint health scores in a haemophilia A cohort from Pakistan with minimal or no access to factor VIII concentrate: correlation with thrombin generation and underlying mutation. Haemophilia *2014*; 20: 426–34.

11 Carneiro JDA, Blanchette V, Ozelo MC, et al. Comparing the burden of illness of haemophilia between resource-constrained and unconstrained countries: the São Paulo-Toronto Hemophilia Study. Haemophilia *2017*; 23: 682–8.

12 Kempton CL, Recht M, Neff A, et al. Impact of pain and functional impairment in US adults with haemophilia: Patient-reported outcomes and musculoskeletal evaluation in the pain, functional impairment and quality of life (P-FiQ) study. Haemophilia *2018*; 24: 261–70.

13 Saulyte-Trakymiene S, Juodyte A, Jusinskaite V, Kulikauskaite R. Systematic evaluation of hemophilic arthropathy in Lithuania. J Med Sci *2019*; 7: 1–15.

14 International Prophylaxis Study Group. Hemophilia Joint Health Score 2.1 Instruction Manual n.d.

15 De Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine: A practical guide. New York: Cambridge University Press; *2011*.

16 Sluiter D, Foppen W, de Kleijn P, Fischer K. Haemophilia Joint Health Score in healthy adults playing sports. Haemophilia *2014*; 20: 282–6.
17 Ribeiro T, Abad A, Feldman BM. Developing a new scoring scheme for the Hemophilia Joint Health Score 2.1. Res Pract Thromb Haemost *2019*; 3: 405–11.
18 Hacker MR, Funk SM, Manco-Johnson MJ. The Colorado Haemophilia Paediatric Joint Physical Examination Scale: Normal values and interrater reliability. Haemophilia *2007*; 13: 71–8.
19 Kuijlaars IAR, Timmer MA, de Kleijn P, Pisters MF, Fischer K. Monitoring joint health in haemophilia: Factors associated with deterioration. Haemophilia *2017*; 23: 934–40.
20 Sun J, Hilliard PE, Feldman BM, et al. Chinese Hemophilia Joint Health Score 2.1 reliability study. Haemophilia *2014*; 20: 435–40.
21 Hilliard P, Funk S, Zourikins N, et al. Hemophilia joint health score reliability study. Haemophilia *2006*; 12: 518–25.

SUPPLEMENTARY MATERIAL

Supplemental table 1. Factor analysis loadings of the three factor model

Items	Factor 1 - Elbow	Factor 2 - Knee	Factor 3 - Ankle
Elbow swelling	**0.788**	0.218	0.245
Elbow duration swelling	**0.782**	0.150	0.221
Elbow atrophy	**0.830**	0.308	0.113
Elbow crepitus	**0.805**	0.004	0.236
Elbow flexion loss	**0.788**	0.132	0.097
Elbow extension loss	**0.824**	0.073	0.121
Elbow pain	**0.716**	0.375	0.061
Elbow strength	**0.768**	0.487	-0.021
Knee swelling	0.221	**0.886**	0.190
Knee duration swelling	0.176	**0.806**	0.211
Knee atrophy	0.233	**0.888**	0.233
Knee crepitus	0.184	**0.693**	0.238
Knee flexion loss	0.143	**0.786**	0.072
Knee extension loss	0.110	**0.826**	0.059
Knee pain	0.189	**0.915**	0.022
Knee strength	0.254	**0.947**	0.029
Ankle swelling	0.081	0.071	**0.905**
Ankle duration swelling	0.130	0.035	**0.873**
Ankle atrophy	0.098	0.524	**0.646**
Ankle crepitus	0.079	-0.042	**0.728**
Ankle flexion loss	0.249	0.177	**0.514**
Ankle extension loss	0.120	0.078	**0.557**
Ankle pain	0.132	0.371	**0.632**
Ankle strength	0.181	0.523	**0.549**

In bold: highest factor loadings for each item.
Root Mean Square Error of Approximation (RMSEA): 0.048

Supplemental table 2. Factor analysis loadings of the one factor models of each joint separately

Items	Elbow	Knee	Ankle
Swelling	0.867	0.945	0.844
Duration swelling	0.852	0.881	0.817
Atrophy	0.891	0.930	0.755
Crepitus	0.797	0.758	0.650
Flexion loss	0.790	0.799	0.616
Extension loss	0.812	0.835	0.584
Pain	0.794	0.925	0.779
Strength	0.858	0.970	0.786
Global gait		0.680	0.752

Supplemental table 3. Results of discussion session 2 regarding reduction of scoring options

	Voting 1	Voting 2
Item	**Reduction, n (%)**	**Reduction, n (%)**
Swelling	8 (42)	10 (53)
Atrophy	12 (63)	11 (58)
Crepitus	**18 (95)**	**19 (100)**
Flexion loss	12 (63)	9 (47)
Extension loss	6 (32)	4 (21)
Pain	14 (74)	15 (79)
Strength	11 (58)	7 (37)
Global gait	3 (16)	6 (32)

Question to experts: Should we measure all scoring options? Answer options: No/Yes. Voting 'No' means reduction of scoring options in some way.

Supplemental table 4. Results of discussion session 3 about which joints to assess

		Voting 1	Voting 2
Question		**n (%)**	**n (%)**
Is it necessary to measure all joints in clinical practice?	Yes	9 (47)	9 (47)
	No	10 (53)	10 (53)
Is it necessary to measure all joints in clinical practice? [a]	Full HJHS		1 (6)
	Screened, not necessarily a full HJHS		**17 (94)**
	Limited number of joints		
Which joint(s) should we measure in clinical practice?	All joints	10 (53)	2 (11)
	Only the most affected joint	9 (47)	
	Only the elbows		
	Only the knees		
	Only the ankles		
	Elbow and knee		
	Elbow and ankle		
	Knee and ankle		
	Joints that fail on screening		**17 (89)**
Is it necessary to measure all joints in a research setting?	Yes	**16 (84)**	14 (74)
	No	3 (16)	5 (26)
Which joint(s) should we measure in a research setting? [b]	All joints	**15 (83)**	14 (74)
	Only the most affected joint	1 (6)	2 (11)
	Only the elbows		
	Only the knees		
	Only the ankles		
	Elbow and knee		
	Elbow and ankle		
	Knee and ankle	2 (11)	
	Joints that fail on screening		3 (16)

[a] 18 voters during the voting.

[b] 18 voters during the first voting.

pedHAL

CHAPTER 4

The Paediatric Haemophilia Activities List (pedHAL) in routine assessment: changes over time, child-parent agreement and informative domains

Isolde A.R. Kuijlaars;[1]
Janjaap van der Net;[2]
Roger E.G. Schutgens;[1]
Kathelijn Fischer;[1]

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

[2] Center for Child Development, Exercise and Physical Literacy, University Medical Center Utrecht, Utrecht University, University Children's Hospital, Utrecht, The Netherlands

Haemophilia. 2019;25: 953-959.

ABSTRACT

Introduction

The Paediatric Haemophilia Activities List (pedHAL) assesses self-reported limitations in activities and participation in children with haemophilia.

Aim

To assess longitudinal changes, child-parent agreement and to identify which pedHAL domains yielded most information in boys with access to early prophylaxis.

Methods

The pedHAL (53 items, 7 domains, optimum 100) was completed annually at the Van Creveldkliniek by boys aged 4-18 years with moderate/severe haemophilia and their parents. Development of the pedHAL in relation to bleeds, changes per domain over 3-5 years, child-parent agreement (% difference child-parent≤|5|) per domain and domain scores (limitations defined as ≤95) were determined.

Results

Seventy-three patients and their parents (92% severe haemophilia, median age 13.1 years [range 5.4;18.0]) completed ≥1 pedHAL. Median (IQR) pedHAL sum score was 99.5 (95.2;100.0) for children and 99.6 (95.8;100.0) for parents. If patients scored >95 and had no joint and/or muscle bleed, 90.9% of the patients scored >95 at the next assessment. The median change in sum score was 0.0 for both the 3- and 5-year interval. Child-parent agreement varied between domains from 92% ('self-care') to 71% ('sitting/kneeling/standing'). Most limitations were reported in the domains 'sitting/kneeling/standing', 'functions of the legs' and 'leisure activities and sports'.

Conclusion

In routine clinical practice in Dutch children on prophylaxis pedHAL scores were high and remained stable in 3-5 years at group level. In individual patients without joint and/or muscle bleeds, pedHAL scores remained high after 1 year. Child-parent agreement was not optimal which indicated that both child report and parent proxy should be reported.

INTRODUCTION

The Paediatric Haemophilia Activities List (pedHAL) assesses self-reported limitations that children and youth with haemophilia (4-18 years) experience in various activities [1]. The pedHAL measures activities and participation according to the International Classification of Functioning, Disability and Health for Children and Youth (ICF-CY) [2]. It was directly derived from the Haemophilia Activities List for adults [3,4] and subsequently validated [1,5]. The pedHAL includes 53 items, distributed over seven domains: 'sitting/kneeling/standing', 'functions of the legs', 'functions of the arms', 'use of transportation', 'self-care', 'household tasks' and 'leisure activities and sports'. The pedHAL is recommended for both research purposes and clinical management of individual patients [6].

Regular outcome assessment should be part of regular care. However, the optimal frequency of measuring activities and participation is unknown. In addition, some clinicians and researchers reported on items which seemed redundant and some patients experienced the pedHAL as a long questionnaire. Vigorous testing should follow development of a new outcome measure and is not yet performed for the pedHAL. A pilot test in a small sample (n=32) was performed to describe the score distributions, child-parent agreement, test-retest reproducibility and construct validity of the pedHAL [1]. This first report showed disagreement between patients and their parents in the domain of household tasks and ceiling effects in all pedHAL domains. Field testing plays a role in defining a definitive version of the questionnaire with the minimum number of meaningful items [7]. As a first step meaningful domains can be explored.

In this field test in Dutch patients with access to early prophylaxis we explored the optimal frequency of administering the pedHAL and assessed child-parent agreement. In addition, we identified which pedHAL domains yielded most information in Dutch patients.

MATERIALS AND METHODS

Study design and study population

This study was a single-centre observational field test of the pedHAL. Since 2010, completion of this list is part of the routine assessment of severe/moderate haemophilia patients aged 4-18 treated at the Van Creveldkliniek in Utrecht, the Netherlands. Children and their parents who completed ≥1 pedHAL during routine assessment were included. For children aged 4-7 years only parents completed the pedHAL. All pedHAL data collected between October 2010 and June 2017 were included, except for those forms in which more than half of the items was missing.

All the data, including patient characteristics and joint and muscle bleeds, were obtained from the electronic medical files. The Medical Research Ethical Committee (MREC) of the University Medical Center Utrecht reviewed the study (protocol number 17-591/C).

Measurements

The pedHAL assesses self-reported limitations in activities and participation in children and youth with haemophilia and their parents. It consists of 53 items, distributed over seven domains, including a patient version (8-18 years) and parent version (4-18 years). The pedHAL was administered as a paper questionnaire.

Patients score the items on a 6-point Likert scale ('impossible', 'always', 'usually', 'sometimes', 'almost never', 'never'), with a 'not applicable (N/A)' scoring option. Domain scores and sum scores are converted to a normalized domain score ranging from 0 (worst possible functional abilities) to 100 (best possible functional abilities). If more than half of the items of a domain were missing or scored 'N/A', no valid domain score was calculated. If more than half of the items were missing or scored 'N/A', no valid total score was calculated.

Patient characteristics analyzed included age at pedHAL assessment, type of haemophilia (A or B), severity of the disease (moderate [factor VIII/IX activity 0.01–0.05 IU/ml] or severe [factor VIII/IX activity <0.01 IU/ml]) and clotting factor regimens (prophylaxis yes/no and age start prophylaxis), as well as the results of routine annual Haemophilia Joint Health Score ($HJHS_{2.1}$) for elbows, knees and ankles (optimum total score 0, worst 124 points) [8]. Joint and muscle bleeds documented by clinicians were counted during the periods that pedHAL data were available.

Statistical analyses

Patient characteristics were presented as proportions or medians (interquartile ranges [IQR:P25;P75]). Descriptive analyses (median, IQR, range, mean and standard deviation [SD]) were performed for the sum score for the most recently completed pedHAL by patients and parents. Differences in age and HJHS total score for children with pedHAL scores ≤95 points vs. pedHAL scores >95 points were tested with a Mann-Whitney U test.

Change in pedHAL between the first and second pedHAL sum score (time between pedHAL: 6-18 months) was shown for patients without limitations in activities and participation (>95) and with/without joint and/or muscle bleeds. In addition, proportions of scores which deteriorated by more than 5 points were evaluated after three (2.5-3.5) and five (4.5-5.5) years of follow-up. The number of joint and/or muscle bleeds were compared according deterioration of the pedHAL with descriptive analyses and Mann-Whitney U-tests. On group level, change scores of the pedHAL sum and domain scores (*follow-up – baseline*) were evaluated after three (2.5-3.5) and five (4.5-5.5) years of follow-up. Wilcoxon signed-rank tests were used to test differences in scores between baseline and follow-up. Because most data were available from questionnaires completed by the parents, follow-up analyses were performed on the parent data.

Patient–parent agreement was evaluated per domain and for the sum score with descriptive statistics due to non-normally distributed data. Scores were divided in three categories (>95, 90-95 and <90 points) and for every category the numbers and proportions of patients and parents were presented to show agreement in scores. In addition, the proportion of the patients and parents with similar scores (difference child-parent: ≤|5|) was assessed. Wilcoxon signed-rank tests were used to test differences in scores between patients and parents [9].

Descriptive analyses (median, IQR, range, mean and standard deviation [SD]) were performed per domain of the pedHAL for the most recently completed pedHAL by patients and parents. Based on reported limits of agreement (LoA) of test-retest data [1], limitations in activities and participation were defined as ≤95 points for domain and sum scores.

RESULTS

Patient characteristics and pedHAL sum scores

Seventy-three children with haemophilia A or B were included in this study. All eligible patients have completed at least one pedHAL, except from one patient with severe haemophilia and comorbidities that limited his capacity to complete the pedHAL. Patient characteristics are shown in Table 1. Their median age at the time of completing the last pedHAL was 13.1 years (range 5.4;18.0). Most patients had severe haemophilia (91.8%). Median annual bleeding rate was 1.3 (range 0.0;4.8). The median age (IQR) of start with prophylaxis of clotting factor was 1.9 (1.3;2.8) years. One of 67 severe haemophilia patients had a positive inhibitor titre and received immune tolerance induction (ITI) without prophylaxis with bypassing agents, and 19 were ex-inhibitor patients. None of the six moderate haemophilia patients had a history of inhibitors. Joint health was excellent with a median HJHS score of 0.0 points (range 0.0;11.0) at the time of completing the last pedHAL.

At the time of the last completed pedHAL, children had a median (IQR) sum score of 99.5 (95.2;100.0), and their parents had a median (IQR) sum score of 99.6 (95.8;100.0) (Table 2). 'Positive' pedHAL sum scores (≤95 points) were observed in a quarter of the children (23.8%) and parents (23.3%). Age was similar between patients with pedHAL scores ≤95 points and >95 points. HJHS scores were higher in patients with pedHAL scores ≤95 points (median 2.0 [IQR: 0.0;4.0] vs median 0.0 [IQR: 0.0;1.0], P=0.044). All patients with moderate haemophilia had a pedHAL sum score >95 points.

Table 1. Patient characteristics

Patient characteristics (n=73)	Median (IQR) or n (%)
Age (years)	13.1 (10.3 ; 15.6)
Haemophilia A	64 (88)
Severe haemophilia	67 (92)
Annual bleeding rate[a]	1.3 (0.8 ; 2.5)
On prophylactic replacement therapy	68 (93)
Inhibitor	
Current	1 (1)
Former	19 (26)
Never	53 (73)
HJHS total score (version 2.1)	0.0 (0.0 ; 2.0)

[a] n = 68 (with ≥1 pedHAL assessment)

Abbreviation: HJHS = Haemophilia Joint Health Score

Changes of the PedHAL over time

Short term changes according to reported bleeds

90.9% of patients with a pedHAL sum score >95 at first assessment and no joint and/or muscle bleeds during follow-up maintained a pedHAL sum score > 95 over median (IQR) 1.0 (0.9-1.2) years (see Figure 1). The pedHAL sum score after 6-18 months for the patient who had a pedHAL ≤95 in the absence of joint and/or muscle bleeds (n=1) was 93.7; the median (IQR) pedHAL sum score after 6-18 months for the patients who had a pedHAL ≤95 and reported joint and/or muscle bleeds (n=6) was 85.5 (74.3;93.0).

Changes over three years

Changes over three years were assessed in 49 parents. Median (IQR) age of the children was 9.7 (7.0;12.1) at the first assessment. After three years pedHAL scores deteriorated in 18.4%. There was a trend of more joint and/or muscle bleeds in children with a deteriorated pedHAL score than in children without a deteriorated pedHAL score (median number of joint and/or muscle bleeds [IQR] 7.0 [3.0-12.5] vs. 3.0 [2.0-6.0], p=0.065). The median (IQR) sum scores were similar at baseline (100.0 [96.4;100.0]) and three years later (99.6 [94.9;100.0]): median (IQR) change sum score after three year follow-up was 0.0 (-0.9;+0.9).

Table 2. Score distribution pedHAL per domain

Domain	Children (n=63)				Parents (n=73)			
	Median (IQR)	Min	Max	Score ≤95 (%)	Median (IQR)	Min	Max	Score ≤95 (%)
Sitting/kneeling/ standing	100 (94.0;100)	66.0	100	29	100 (96.0;100)	56.0	100	22
Functions of the legs	100 (94.5;100)	62.0	100	29	100 (96.4;100)	14.6	100	23
Functions of the arms	100 (96.0;100)	68.0	100	19	100 (100;100)	60.0	100	15
Use of transport	100 (100.0;100)	60.0	100	10	100 (100;100)	60.0	100	16
Self-care	100 (100.0;100)	75.6	100	13	100 (100;100)	64.4	100	14
Household tasks	100 (100.0;100)	60.0	100	14	100 (100;100)	60.0	100	13
Leisure activities and sports	100 (96.4;100)	54.3	100	20	100 (95.1;100)	7.3	100	25
Sum score	99.5 (95.2;100)	76.3	100	24	99.6 (95.8;100)	49.6	100	23

Note: Descriptive analyses were performed according complete case analysis.

Changes over five years

Changes over five years were assessed in 35 parents. Median (IQR) age of the children was 9.3 (6.9;10.6) at the first assessment. After five years pedHAL scores deteriorated in 14.3%. The number of joint and/or muscle bleeds was higher in children with a deteriorated pedHAL score than in children without a deteriorated pedHAL score (median number of joint and/or muscle bleeds [IQR] 5.0 [4.0-12.3] vs. 15.0 [8.0-24.5], p=0.024). The median (IQR) sum scores were similar at baseline (100.0 [97.3;100.0]) and five years later (100.0 [98.8;100.0]): median (IQR) change sum score after five year follow-up was 0.0 (-1.0;+0.9).

For all seven domains, median change scores were also 0.0 after three and five years. Repeat analyses of the pedHAL scores completed by children yielded similar results.

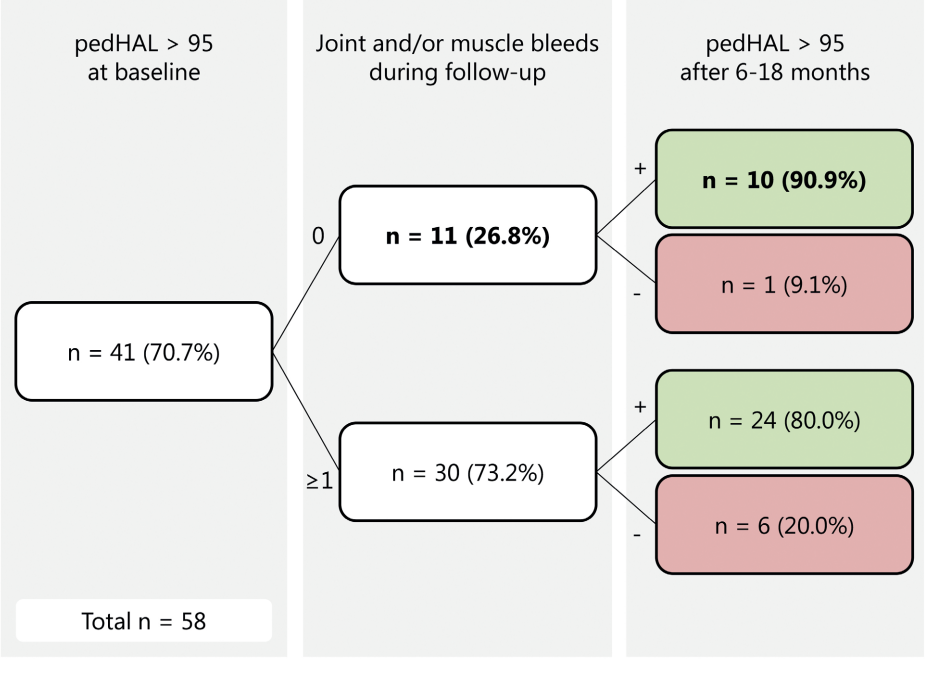

Figure 1. Changes of pedHAL score within one year (6-18 months) of follow-up according to bleeding. + present; - absent

Child-parent agreement

Table 3 shows descriptive analyses of agreement of the domain scores and sum score of the pedHAL in 63 children and parents. At domain level, child-parent agreement (difference child-parent ≤|5|) varied across domains: agreement was highest in the domain 'self-care' (92%) and lowest in the domain 'sitting/kneeling/standing' (71%). In addition, at domain level children and parents scored >95 points in 79% ('sitting/kneeling/standing') to 98% ('self-care'). For the sum score, child-parent agreement (difference child-parent ≤|5|) was 81%. If children scored >95 points on the pedHAL, parents scored >95 points in 96%.

Score distribution of the pedHAL domains

Table 2 shows the median (IQR) and the range of the domain scores and proportions of sum and domain scores ≤95 points of the last completed pedHALs. Mean scores (SD) are shown in the Supplemental table 1. All domain scores were median 100.0. The domain 'functions of the legs' showed most 'positive' scores (28.6% of children and 23.3% of parents). In children, the domain of 'use of transport' showed least 'positive' scores: in 9.8%. In parents, the domain of 'household tasks' showed least 'positive' scores: in 12.9%.

DISCUSSION

This study showed that Dutch children on early prophylaxis and their parents reported almost no limitations in activities and participation. On group level, after three or five years follow-up the sum scores and domain scores remained stable. On patient level, in patients without limitations in activities and participation (pedHAL sum score >95) and without joint and/or muscle bleeds, pedHAL scores remained high till the next assessment after median (IQR) 1.0 (0.9-1.2) years. If patients did report limitations, they reported most limitations in the domains 'sitting/kneeling/standing', 'functions of the legs' and 'leisure activities and sports'. Almost no limitations were reported on the pedHAL items in the domains 'functions of the arms', 'use of transportation', 'self-care' and 'household activities'. In addition, child-parent agreement (difference child-parent ≤5) varied across domains from 71% agreement for 'functions of the legs' up to 92% agreement for 'self-care'. The differences indicated that both child report and parent proxy should be reported.

Table 3. Score distribution pedHAL per domain

Children	Parents	Sitting/kneeling/ standing	Functions of the legs	Functions of the arms	Use of transport *	Self-care	Household tasks	Leisure activities and sports	Sum score
	>95	**38 (79)**	**41 (87)**	**48 (92)**	**50 (96)**	**55 (98)**	**44 (92)**	**33 (83)**	**46 (96)**
>95	90 ≤ 95	6 (13)	2 (4)	1 (2)		1 (2)	1 (2)	1 (3)	2 (4)
n (%)	<90	4 (8)	4 (9)	3 (6)	1 (2)		1 (2)	1 (3)	
	N/A				1 (2)		2 (4)	5 (13)	
	>95	2 (40)	1 (25)	1 (50)	2 (100)		1 (100)	1 (100)	
90 ≤ 95	**90 ≤ 95**		**2 (50)**	**1 (50)**					**1 (33)**
n (%)	<90	3 (60)	1 (25)						2 (67)
	N/A								
	>95	3 (30)	2 (17)	1 (11)	3 (33)		1 (14)	3 (25)	1 (8)
<90	90 ≤ 95		1 (8)	1 (11)		2 (29)	2 (29)	1 (8)	1 (8)
n (%)	**<90**	**7 (70)**	**9 (75)**	**6 (67)**	**5 (56)**	**5 (71)**	**4 (57)**	**5 (42)**	**10 (83)**
	N/A			1 (11)	1 (11)			3 (25)	
	>95						3 (43)	2 (20)	
N/A	90 ≤ 95							1 (10)	
n (%)	<90							2 (20)	
	N/A						**4 (57)**	**5 (50)**	
Pairs with similar scores, difference child-parent ≤5 (%)		**71**	**73**	**84**	**84**	**92**	**81**	**78**	**81**

Number and proportion n (%) of children and parents who scored >95 points, 90 ≤ 95 points, <90 points and N/A for all domains and the sum score. Concordant scores between children and parents are marked in grey.

*p-value of the Wilcoxon signed-rank order test is significant (p<0.05) signifying statistically significant discordant scores between children and their parents.

Internal and external validity

So far, studies on the psychometric properties of the pedHAL have been limited [10]. The results of the present report were dependent on the choice of cut-off points. The three categories (>95, 90-95 and <90 points), cut-off score of ≤95 (limitations in activities and participation) and cut-off score of more than 5 points for changes in

pedHAL were based on reported limits of agreement (LoA) of test-retest data [1], but are arbitrary. To determine exact changes in limitations measured with the pedHAL, Smallest Detectable Change and Minimal Important Change should be known.

Results show high pedHAL scores, no changes on group level and stable pedHAL scores in patients without limitations and joint and/or muscle bleeds during 1 year follow-up. Although this field test of routine assessments included data of an unselected cohort with a full age range and long follow-up, the results are only applicable to intensively treated patients as patients in the Van Creveldkliniek receive early prophylaxis and intensive treatment.

Children with a deteriorated pedHAL score after three or five years showed more joint and/or muscle bleeds, which is in line with the hypothesis that bleeds results in limitations of activities and participation. This needs to be confirmed in another population.

Comparison with other studies

Neither long-term follow-up data of self-reported limitations in activities and participation measured by the pedHAL, nor child-parent agreement has been published until now. This is the first study which showed three to five years follow-up data and the development of the pedHAL in relation to joint and/or muscle bleeds. Only in a pilot test child-parent agreement was studied in a small sample of 15 patients with LoA (change in mean score ± 95% confidence interval) varying from 0.7 ± 3.4 to 0.7 ± 28.2 [1]. Child- parent agreement in other patient reported outcome measures (PROMs) used in children with haemophilia show different scores on self-reported health related quality of life between children and their parents for the Canadian Hemophilia Outcome – Kids' Life Assessment Tool (CHO-KLAT). Differences between child and parent proxy scores were also reported in other paediatric PROMs and the recommendation is to measure both perspectives [11,12].

The domain and sum scores as reported in the present study in Dutch children were higher than previous reported data in UK, Romania and Lithuania [5,13,14]. Comparable sum scores were reported in a Canadian study [15]. Similar to the present study, the domains 'sitting/kneeling/standing', 'functions of the legs' and 'leisure activities and sports' were the most informative domains and most ceiling effects were observed in the domains of 'functions of the arms', 'use of transportation', 'self-care' and 'household

activities' in English, Romanian and Lithuanian children [5,13,14]. No studies were available about pedHAL scores in children without haemophilia, to compare the limitations to healthy children.

In the adult version of the HAL the same seven domains are represented and reported HAL data were mostly comparable to data of children. Adults with mild to severe haemophilia from the United States and adults with severe haemophilia from the UK reported most complaints in the domains 'lying down/sitting/kneeling/standing' and 'functions of the legs' and least complaints in the domain 'self-care' [16,17].

Clinical implications and future research

Results suggest that annual pedHAL assessment in clinical care is not necessary in children who have a pedHAL sum score >95 and no joint and/or muscle bleeds during follow-up. In children with lower pedHAL scores and/or bleeds, annual pedHAL assessment is recommended to monitor limitations in activities and participation. At group level changes were clinically insignificant after 3-5 years. The present data suggest that the frequency of administrating the pedHAL can be lowered to every 3-5 years when studying groups of children with low bleeding rates receiving intensive prophylaxis. However, both child report and parent proxy should be reported because scores differed between children and their parents. These discrepancies are important to discuss with the child and their parents.

Shortening the questionnaire may enhance the feasibility of the pedHAL within the context of multiple outcome assessment in haemophilia care. The present study identified several pedHAL domains which were less informative; especially 'functions of the arms', 'use of transportation', 'self-care' and 'household activities' may be candidates for shortening. However, before deciding on the removal of items, this study should be repeated including data from other paediatric haemophilia populations, including patients with more frequent bleeding and/or extensive arthropathy.

CONCLUSION

This explorative clinical study suggests that annual pedHAL assessment has limited clinical value in patients without limitations in activities and participation (pedHAL sum score >95) and without joint and/or muscle bleeds. Furthermore, both child report and parent proxy should be reported since scores between children and their parents differ

(child-parent agreement: 71-92%). This study revealed little limitations in activities and participation (pedHAL >95: 76%) in Dutch children on prophylaxis and pedHAL scores remained stable over three to five years at group level. Domain analyses showed that the domains 'sitting/kneeling/standing', 'functions of the legs' and 'leisure activities and sports' are most informative domains in patients receiving early prophylaxis.

REFERENCES

1 Groen WG, van der net J, Helders PJM, Fischer K. Development and preliminary testing of a Paediatric Version of the Haemophilia Activities List (pedhal). Haemophilia *2010*; 16: 281–289.

2 World Health Organization. International Classification of Functioning, Disability and Health: Children & Youth Version: ICF-CY *2007*.

3 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: 565–571.

4 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: 36–46.

5 Groen W, van der Net J, Lacatusu AM, et al. Functional limitations in Romanian children with haemophilia: Further testing of psychometric properties of the ophilia Activities List. Haemophilia *2013*; 19: 116–125.

6 Fischer K, Poonnoose P, Dunn AL, et al. Choosing outcome assessment tools in haemophilia care and research: A multidisciplinary perspective. Haemophilia *2016* 1–14.

7 de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine: A practical guide. New York: Cambridge University Press; *2011*.

8 Feldman BM, Funk SM, Bergstrom B-MM, et al. Validation of a new pediatric joint scoring system from the international hemophilia prophylaxis study group: Validity of the hemophilia joint health score. Arthritis Care Res *2011*; 63: 223–230.

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

10 Timmer MA, Gouw SC, Feldman BM, et al. Measuring activities and participation in persons with haemophilia: A systematic review of commonly used instruments. Haemophilia 2018; 24(2): e33-e49.

11 Schulte F, Wurz A, Reynolds K, Strother D, Dewey D. Quality of Life in Survivors of Pediatric Cancer and Their Siblings: The Consensus Between Parent-Proxy and Self-Reports. Pediatr Blood Cancer *2016*; 63: 677–683.

12 Brandon TG, Becker BD, Bevans KB, Weiss PF. Patient-Reported Outcomes Measurement Information System Tools for Collecting Patient-Reported Outcomes in Children With Juvenile Arthritis. Arthritis Care Res *2017*; 69: 393–402.

13 Khair K, Holland M, Bladen M, et al. Study of physical function in adolescents with haemophilia: The SO-FIT study. Haemophilia *2017*; 23: 918–925.

14 Radzevič V, Raistenskis J, Ragelienė L, Kowalski IM. Relationship between physical activity and functional ability in school-aged children with hemophilia. Polish Ann Med *2013*; 20: 13–18.

15 Bouskill V, Hilliard P, Stephens S, et al. An institutional pilot study to investigate physical activity patterns in boys with haemophilia. Haemophilia *2016*; 22: e383–389.

16 Kempton CL, Recht M, Neff A, et al. Impact of pain and functional impairment in US adults with haemophilia: Patient-reported outcomes and musculoskeletal evaluation in the pain, functional impairment and quality of life (P-FiQ) study. Haemophilia *2018*; 24: 261–270.

17 McLaughlin P, Morris R, Chowdary P. Investigating the relationship between the HJHS and HAL in routine clinical practice: A retrospective review. Haemophilia *2018;* 24(6): 988-994.

SUPPLEMENTARY MATERIAL

Supplemental table 1. Mean (standard deviation [SD]) scores of pedHAL per domain

Domain	Children		Parents	
	Mean	(SD)	Mean	(SD)
Sitting/kneeling/standing	95.3	(8.0)	96.1	(8.3)
Functions of the legs	94.7	(9.1)	94.9	(12.8)
Functions of the arms	96.3	(7.8)	96.7	(7.8)
Use of transport	97.8	(7.5)	96.4	(9.5)
Self-care	98.1	(5.0)	97.5	(6.6)
Household tasks	97.7	(7.1)	96.6	(9.8)
Leisure activities and sports	94.8	(10.9)	93.1	(16.0)
Sum score	95.9	(6.7)	95.6	(8.9)

CHAPTER 5

Shortening the pediatric Haemophilia Activities List (pedHAL) based on pooled data from international studies

Isolde A.R. Kuijlaars;[1]
Janjaap van der Net;[2]
Vanessa Bouskill;[3]
Pamela Hilliard;[4]
Agne Juodyte;[5]
Kate Khair;[6]
Sonata Saulyte Trakymiene;[5]
Kathelijn Fischer[1]

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

[2] Center for Child Development, Exercise and Physical Literacy, University Medical Center Utrecht, Utrecht University, University Children's Hospital, Utrecht, The Netherlands

[3] Department of Nursing, Division of Hematology/Oncology, The Hospital for Sick Children, Toronto, Canada

[4] Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, Canada

[5] Clinic of Children's diseases, Faculty of Medicine, Vilnius University, Vilnius, Lithuania

[6] Centre for Outcomes and Experience Research in Children's Health, Illness and Disability (ORCHID), Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK

Haemophilia. 2021;27: 305-313.

ABSTRACT

Introduction

The pediatric Haemophilia Activities List (pedHAL) was developed to measure activities and participation in children and youth with haemophilia. Results from international studies provide an opportunity to determine which items are universally important.

Aim

The aim of this study was to determine which items of the pedHAL are redundant to construct a shorter version of the pedHAL.

Methods

This study is a cross-sectional multicenter secondary analysis on pooled data of published studies using the pedHAL (7 domains, 53 items, optimum score: 100) in children with haemophilia A/B aged 4-18 years. To identify redundant items the following aspects were evaluated: floor and ceiling effects, proportions of missing and 'not applicable' responses, inter-item correlations, component loadings in an exploratory factor analysis, internal consistency and item-total correlations.

Results

Data on 315 patients with haemophilia from 6 studies were evaluated. Median age was 12.2 years [range 4.0-18.0], 87.3% had severe haemophilia and 80.3% received prophylaxis. Median (IQR) pedHAL sum score was 96.7 (88.0-100). After a stepwise procedure 31 items were removed, resulting in a $pedHAL_{short}$ of 22 items, representing all original 7 domains. Most remaining items belonged to the domains 'sitting/kneeling/standing' and 'functions of the legs'. The $pedHAL_{short}$ sum score was similar to the original pedHAL sum score, with small differences in 5 domains.

Conclusion

This clinimetric study resulted in >50% reduction of the length of the pedHAL. The 22-item $pedHAL_{short}$ reduces patient burden and is expected to capture the information on activities and participation. The $pedHAL_{short}$ needs validation in other populations.

INTRODUCTION

The paediatric Haemophilia Activities List (pedHAL) assesses self-reported limitations in various activities of daily living, which are relevant to children and youth with haemophilia [1]. It was directly derived from the Haemophilia Activities List (HAL) for adults [2,3] and subsequently validated [1,4]. The pedHAL includes 53 items, distributed over seven domains similar to the HAL domains: 'sitting/kneeling/standing', 'functions of the legs', 'functions of the arms', 'use of transportation', 'self-care', 'household tasks' and 'leisure activities and sports'. All items belong to 'activities and participation', according the World Health Organization's International Classification of Functioning (ICF). 'Activity' is defined as 'the execution of a task or action by an individual' and 'participation' as 'involvement in a life situation' [5]. The pedHAL is recommended for both research purposes and clinical management of patients [6].

After introduction of the pedHAL to clinical care and research in 2010, clinicians and researchers reported some items to be non-informative. Most ceiling effects were observed in the domains of 'functions of the arms', 'use of transportation', 'self-care' and 'household activities' in Dutch, English, Romanian and Lithuanian children. The domains 'sitting/kneeling/standing', 'functions of the legs' and 'leisure activities and sports' were the most informative domains [4,7–9].

Shortening the questionnaire may enhance the feasibility of pedHAL use within the context of multiple outcome assessments in haemophilia care. With more studies using the pedHAL being conducted internationally over the past years, there is an opportunity to determine which items are universally important for different patient populations.

The aim of this study was to determine which items of the pedHAL are redundant in order to construct a shorter version of the pedHAL for the measurement of activities and participation in children and youth with haemophilia.

MATERIALS AND METHODS

Study design and study population

This study was a cross-sectional multicenter secondary analysis of pooled data of published studies using the pedHAL. The Medical Research Ethical Committee (MREC) of the University Medical Center Utrecht reviewed the study (protocol number 18-309/C).

Pooling of published pedHAL data

A literature search identified five studies published between 2010 and April 2018, which used the pedHAL in children with haemophilia. In addition, two studies in preparation for publication were identified and included. Inclusion criteria were children with haemophilia A (FVIII) and B (FIX) of all severities, aged 4-18 years. Patients were excluded if more than half of the pedHAL items were missing. If both the children and parent proxy pedHAL scores were available, only the pedHAL completed by the child was included in the analyses. Authors of all seven papers were invited to share the original pedHAL data (scores per item) and de-identified patient characteristics. All but one authors accepted our invitation. Only data of children with haemophilia A (FVIII) and B (FIX) aged 4-18 years who participated in the selected studies were included [1,4,7,9–11].

Measurements

The pedHAL assesses self-reported limitations in activities and participation in children with haemophilia. It consists of a patient version (8-18 years) and parent version (4-18 years) both with 53 items, distributed over seven domains. Patients score the items on a 6-point Likert scale ('impossible', 'always', 'usually', 'sometimes', 'almost never', 'never'), with a 'not applicable (N/A)' scoring option. Domain scores and sum scores are converted to a normalized domain score ranging from 0 (worst possible functional abilities) to 100 (best possible functional abilities) in the scoring tool (available at www.vancreveldkliniek.nl). According to the pedHAL scoring manual, domain scores were only calculated if half or more of the items of a domain were scored on the 6-point Likert scale.

Patient characteristics analyzed included age at pedHAL assessment, type of haemophilia (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 and start prophylaxis before age of 3 years yes/no) and current inhibitor status.

Statistical analyses

Patient characteristics were presented as proportions or medians (interquartile ranges [IQR:P25 - P75]). A Kruskal Wallis test was performed to compare age according to treatment regimen (prophylaxis start <3 years vs. prophylaxis start ≥3 years vs. no prophylaxis). Descriptive analyses (median, IQR, range, mean and standard deviation [SD]) were performed for the pedHAL domain and sum scores. Based on reported limits of agreement (LoA) of test-retest data [1], limitations in activities and participation were defined as ≤95 points for domain and sum scores.

Non-informative items were identified in a stepwise process (7 steps) according to the method of de Vet et al. (2011), from the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative [12]. After each step non-informative items were deleted, before proceeding with the following step.

Step 1: Per item the proportions of each scoring option ('impossible', 'always', 'usually', 'sometimes', 'almost never' or 'never') was determined, excluding the missing and 'N/A' scored questions. Proportions of minimum ('impossible') and maximum ('never' problems) scores were analyzed to detect floor and ceiling effects. Items with ≥85% minimum or maximum scores were removed.
Step 2: The missing data and scores with 'N/A' were examined. The authors removed items which were scored >15% as 'missing' or 'N/A'.
Step 3: *Inter-item* correlations were evaluated. Inter-item correlations calculated with Spearman's rho <0.2 indicated items which do not correlate with any of the others and >0.9 indicated item redundancy. Items with inter-item correlations <0.2 and >0.9 were not included in the factor analysis.
Step 4: Component loadings on exploratory factor analysis were evaluated. Items were analyzed on categorical level. Items with factor loadings <0.5 were removed. Model fit was evaluated with the Root Mean Square Error of Approximation (RMSEA); <0.08 indicates moderate model fit and <0.05 indicates good model fit.
Step 5: *Inter-item* correlations were evaluated for the second time. Inter-item correlations calculated with Spearman's rho >0.7 within one factor were indicators for item redundancy. Items which had a correlation >0.7 were reviewed by IK, KF and JJ and one of the items was removed.
Step 6: Internal consistency calculated with Cronbach's α and internal consistency after item deletion were compared. Cronbach's α should be 0.7 and 0.9; a higher Cronbach's α after item deletion was considered a reason to eliminate an item.

Step 7: Item-total correlations for the pedHAL total score were evaluated. Item-total correlations were calculated with Spearman's rho. Items with item-total correlations <0.3 were removed.

A sensitivity analysis was performed where only parent scores were analyzed in cases where both were available (n=72).

After removing non-informative items, a $pedHAL_{short}$ was created. Median (IQR) normalized sum and domain scores and percentages of scores ≤95 points were calculated for the $pedHAL_{short}$, similar to the calculation of sum and domain scores in the original scoring tool. The differences between the pedHAL and $pedHAL_{short}$ sum scores were calculated and shown in a boxplot. In addition, a one-way analysis of variance (ANOVA) was performed to compare differences according to treatment regimen (prophylaxis start <3 years vs. prophylaxis start ≥3 years vs. no prophylaxis). A secondary exploratory factor analysis was performed for the $pedHAL_{short}$ to detect possible underlying constructs.

SPSS (version 25, IBM) was used for data analyses. Mplus (version 6.12, Muthen & Muthen) was used for the exploratory factor analysis.

RESULTS

Patient characteristics

From the data of six studies, 315 children with haemophilia A or B were included. The data are from the Netherlands (n=84) [1,9], Romania (n=28) [4], United Kingdom (UK) (n=123) [7], Lithuania (n=15) [10] and Canada (n=65) [11]. Patient characteristics are shown in Table 1. Median age at the time of completing the last pedHAL was 12.2 years (range 4.0 - 18.0) and was similar for patients with different treatment regimens (p=0.22). The majority of the patients had severe haemophilia (87.3%). One patient was excluded because he completed less than half of the items of the questionnaire. The bulk of the questionnaires (81.3%) that were analysed were completed by the children, the others were completed by parents.

PedHAL domain and sum scores

Domain and sum scores are shown in Table 2. The median (IQR) pedHAL sum score was 96.7 (88.0 - 100.0). 'Positive' pedHAL sum scores (≤95 points) were observed in

43% of participants. The median (IQR) domain scores were lowest for the domains 'sitting/kneeling/standing' (97.8 [82.2 - 100.0]), 'functions of the legs' (97.8 [83.6 - 100.0]) and 'leisure activities and sports' (97.8 [80.0 - 100.0]). The other domains had median scores of 100.0. Domain scores were not calculated for 2 participants for the domains 'functions of the arms' and 'self-care', and for up to 60 participants for the domain 'leisure activities and sports', because more than half of the items were scored as missing or 'N/A'.

Table 1. Patient characteristics (n=315)

Patient characteristics (n=315)	**Median (IQR), % (n)**
Age (years)	12.2 (9.7 - 15.0)
Haemophilia A	87.0 (n=274)
Severity haemophilia	
Mild	7.6 (n=24)
Moderate	5.1 (n=16)
Severe	87.3 (n=275)
Prophylaxis	80.3 (n=253)
Early prophylaxis (<3 years)	51.9 (n=139)[a]
Inhibitor (current)	5.7 (n=18)

[a] missing data on prophylaxis (n=47).

Table 2. Domain and sum scores of the pedHAL (n=315)

Domain	**Domain scores**			**Score ≤95**	**Missing /NA**
	Median (IQR)	**Min**	**Max**	**(%)**	**N**
Sitting/kneeling/standing	97.8 (82.2-100)	10.0	100	45	0
Functions of the legs	97.8 (83.6-100)	0.0	100	44	3
Functions of the arms	100 (90.0-100)	10.0	100	34	2
Use of transport	100 (93.3-100)	0.0	100	26	17
Self-care	100 (96.8-100)	6.7	100	20	2
Household tasks	100 (93.3-100)	0.0	100	26	37
Leisure activities and sports	97.8 (80.0-100)	0.0	100	45	60
Sum score	96.7 (88.0-100)	19.2	100	43	0

Item reduction

The stepwise process to select non-informative items is shown in Table 3. The frequency tables generated for step 1 and 2 and the table with item-total correlations for step 7 in the item reduction process are shown in the Supplementary material.

1. Floor and ceiling effects
   Minimum and maximum scores were evaluated for all items. There was no floor effect in any pedHAL item. Ceiling effects were shown in 2/6 items of the domain 'functions of the arms' and in 6/9 items of the domain 'self-care'.
2. Missing data and scores with 'N/A'
   There were small numbers of missing responses (0-3) on the items. Missing and/or 'N/A' responses were scored in >15% of the children in 1/3 items of the domain 'use of transport', 2/3 items of the domain 'household tasks' and in 9/11 items of the domain 'leisure activities and sports'.
3. *Inter-item* correlations (1)
   None of the items had correlations with other items lower than 0.2 or higher than 0.9. All remaining items were used for the exploratory factor analysis.
4. Component loadings of the exploratory factor analysis
   Table 4 shows the component loadings of the exploratory factor analysis. The exploratory factor analysis suggested no items were eligible for item reduction. A 2-factor model was selected which included all remaining items and the two factors were identified as arm activities and leg activities. The model fit of the 2-factor model was 0.07 (RMSEA), indicating moderate model fit. The factor loadings were >0.5.
5. *Inter-item* correlations (2)
   Inter-item correlations were re-evaluated. In the domain 'sitting/kneeling/standing' 3/10 items, which had inter-item correlations >0.7 with other items, were removed. In the domain 'functions of the legs' 5/11 items were removed. The items 'running' and 'jumping' had a correlation of 0.73. The authors decided to remove the item 'jumping', which was considered less relevant in lifelong outcome assessment. The items 'walking upstairs' and 'walking downstairs' had a correlation of 0.81. The authors decided to remove the item 'walking downstairs', which was scored as less difficult than 'walking upstairs' by the participants. In the domain 'use of transport' the item 'using public transport' (1/3) was removed. In the domain 'self-care' 2/9 items were removed.
6. Internal consistency calculated with Cronbach's α
   The remaining 22 PedHAL items were strongly related (Cronbach's α of 0.97),

which indicates redundancy of items. Only complete cases (n=201, 63.8%) were included in the analysis. The Cronbach's α after deletion of separate items was equal or smaller, which did not identify candidate items for removal. Eventually, the authors decided to keep the remaining 22 items, because the Cronbach's alpha was already lowered by removing the 31 items.

7. Item-total correlations for pedHAL total scores
   All item-total correlations were high (Spearman's rho = 0.55 - 0.76), thus identifying no candidates for item reduction.

The sensitivity analysis with parent proxy (n=131) and child (n=184) forms resulted in a shorter $pedHAL_{short}$ (20 items). The items 'walking or riding up a small hill or slope without help', 'stretching to reach something above your head' and 'putting on pants' were removed in step 5 (inter-item correlations) and 'putting on shoes and socks' was not removed in step 5.

$PedHAL_{short}$ with 22 items

In Table 3 all items of the $pedHAL_{short}$ are shown. Domain and sum scores of the pedHAL and $pedHAL_{short}$ are shown in Table 5.

Twenty-two items remained after removing the items (n=31) according to the seven steps. All domains were still represented in the $pedHAL_{short.}$ Most items of the $pedHAL_{short}$ belonged to the domains 'sitting/kneeling/standing' (n=7) and 'functions of the legs' (n=6). For the domains 'use of transport', 'self-care' and 'household tasks' only one item remained in the $pedHAL_{short}$. The median (IQR) $pedHAL_{short}$ sum score was 97.3 (87.0 - 100.0), which was similar to the pedHAL sum score. The differences between the pedHAL and $pedHAL_{short}$ sum scores were similar in patients receiving prophylaxis started <3 years and started ≥3 years and patients receiving no prophylaxis (p=0.82) (see Figure 1). The domains had median scores of 100.0, in exception of a median domain score of 97.1 for 'sitting/kneeling/standing'. Domain scores for 'sitting/kneeling/standing', 'functions of the legs', 'use of transport' and 'self-care' were higher than the original pedHAL domain scores. Domain score for 'functions of the arms' was lower than the original pedHAL domain score. The largest discrepancy in the proportions of abnormal domain scores (≤95) was observed for the domain 'use of transport' (pedHAL: 26% vs. $pedHAL_{short}$: 16%), which was a result of removing the item 'cycling'.

The secondary exploratory factor analysis with the 22-item $pedHAL_{short}$ resulted in a 1-factor model, indicating that a sum score containing all 22 items needs to be used.

Table 3. Flow chart of steps to reduce the number of pedHAL items and the remaining 22 pedHAL items

Sitting/ kneeling/ standing	Functions of the legs	Functions of the arms	Use of transport	Self-care	Household tasks	Leisure activities and sports
<u>Step 1:</u> Floor and ceiling effects (≥85% maximum scores)						
		3 Fine hand movements 6 Shaking hands with someone		1 Drying off your entire body 5 Wiping your bottom after using the toilet 6 Fastening a hood or doing up the top button on your … 7 Buttering bread or making a sandwich 8 Unscrewing the lid from a bottle of water, juice, etc. 9 Brushing your teeth		
<u>Step 2:</u> Missing and/or N/A (>15% missing / N/A)						
			1 Cycling		2 Outside chores 3 Other household chores	1 Going out 4 School sports: athletics 5 School sports: ball sports 6 Playing non-contact team sports 7 Playing contact team sports 8 Individual non-contact sports 9 Individual contact sports 10 Taking part in a sports event 11 Going to school camp or summer camp

Table 3. (Continued)

Sitting/ kneeling/ standing	Functions of the legs	Functions of the arms	Use of transport	Self-care	Household tasks	Leisure activities and sports
Step 3: Inter-item correlations (r<0.2 and r>0.9)						
Step 4: Exploratory factor analysis (factor loadings <0.5)						
Step 5: Inter-item correlations (r>0.7)						
5 Kneeling/ squatting 8 Standing still for a short period 10 Standing still for a very long periods	1 Walking short distances 3 Walking long distances 4 Walking on an uneven surface 8 Jumping 10 Walking downstairs		3 Using public transport (bus, train, metro, tram)	2 Putting on a t-shirt or jumper 4 Putting on shoes and socks		
Step 6: Internal consistency						
Step 7: Item-total correlations (r<0.3)						
pedHAL$_{short}$: 22 items						
1 Sitting down 2 Sitting on the ground 3 Standing up from a chair with arm rests 4 Standing up from a chair without arm rests 6 Squatting for longer periods 7 Bending over forwards 9 Standing still for longer periods	2 Walking longer distances 5 Walking on a soft surface 6 Strolling 7 Running 9 Walking upstairs 11 Walking or riding up a small hill or slope without help	1 Carrying large or heavy objects with two hands 2 Stretching to reach something above your head 4 Writing 5 Leaning on your arms	2 Getting in and out the car	3 Putting on pants	1 Chores around the house	2 Playing outside, alone or with others 3 School sports: exercises and gymnastic equipment

Table 4. 2-factor model of the remaining pedHAL items

	Arm activities	Leg activities
Sitting/kneeling/standing		
Sitting down	**0.725**	0.507
Sitting on the ground	0.614	**0.658**
Standing up from a chair with arm rests	**0.720**	0.556
Standing up from a chair without arm rests	**0.654**	0.562
Kneeling/squatting	0.547	**0.709**
Squatting for long periods	0.521	**0.726**
Bending over forwards	0.612	**0.630**
Standing still for a short period	0.523	**0.726**
Standing still for longer periods	0.338	**0.905**
Standing still for very long periods	0.256	**0.925**
Functions of the legs		
Walking short distances	0.638	**0.657**
Walking longer distances	0.430	**0.851**
Walking long distances	0.347	**0.855**
Walking on an uneven surface	0.531	**0.727**
Walking on a soft surface	0.573	**0.710**
Strolling	0.501	**0.725**
Running	0.480	**0.782**
Jumping	0.514	**0.740**
Walking upstairs	**0.730**	0.606
Walking down-stairs	**0.766**	0.585
Walking or riding up a small hill or slope without help	**0.674**	0.634
Functions of the arms		
Carrying large or heavy objects with two hands	**0.824**	0.358
Stretching to reach something above your head	**0.790**	0.411
Writing	**0.650**	0.397
Leaning on your arms	**0.764**	0.389
Use of transport		
Getting in and out of the car	**0.845**	0.445
Using public transport	**0.763**	0.549
Self-care		
Putting on a t-shirt or jumper, etc.	**0.903**	0.312
Putting on pants	**0.905**	0.374
Putting on shoes and socks	**0.866**	0.385

Table 4. (Continued)

	Arm activities	Leg activities
Household tasks		
Chores around the house	**0.772**	0.497
Leisure activities and sports		
Playing outside, alone or with others	0.614	**0.621**
School sports: exercises and gymnastic equipment	0.610	**0.623**

Interpretation: Two factors were identified in the factor analysis; arm activities and leg activities. The grey highlighted factor loadings shows to which factor the items were allocated. The highest factor loading of each factor was shown in bold.

Table 5. Domain and sum scores of the pedHAL and pedHAL$_{short}$

	pedHAL		pedHAL$_{short}$	
Domain	**Median (IQR)**	**Score ≤95 (%)**	**Median (IQR)**	**Score ≤95 (%)**
Sitting/kneeling/ standing	97.8 (82.2-100)	45	97.1 (85.7-100)*	42
Functions of the legs	97.8 (83.6-100)	44	100 (86.7-100)*	39
Functions of the arms	100 (90.0-100)	34	100 (85.0-100)*	40
Use of transport	100 (93.3-100)	26	100 (100-100)*	16
Self-care	100 (96.4-100)	20	100 (100-100)*	18
Household tasks	100 (93.3-100)	26	100 (100-100)	19
Leisure activities and sports	97.8 (80.0-100)	45	100 (80.0-100)	44
Sum score	96.7 (88.0-100)	43	97.3 (87.0-100)	43

Only complete cases within each domain were included in the comparison of the pedHAL scores with pedHAL$_{short}$ scores. *p<0.05 Wilcoxon signed rank test.

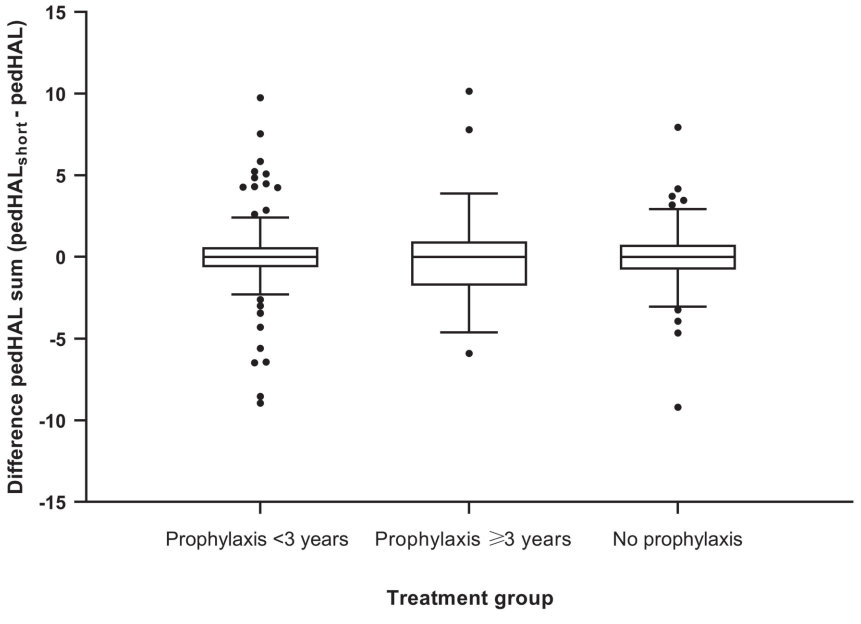

Figure 1. Boxplots of the differences in pedHAL and pedHAL$_{short}$ scores for patients with three different treatment regimens

DISCUSSION

This study analysed international pedHAL data in children with haemophilia with the aim of reducing the 53-item pedHAL questionnaire. A stepwise approach resulted in a pedHAL$_{short}$ of 22 items. The items of the pedHAL$_{short}$ belonged to the domains of the original pedHAL: 'sitting/kneeling/standing' (n=7), 'functions of the legs' (n=6), 'functions of the arms' (n=4), 'use of transportation' (n=1), 'self-care' (n=1), 'household tasks' (n=1) and 'leisure activities and sports' (n=2). Differences between the original pedHAL and pedHAL$_{short}$ sum score were similar between treatment regimens.

Internal and external validity

This is the first study presenting pedHAL data of 315 patients from heterogeneous populations. In these published pedHAL data, the majority of patients had received prophylactic treatment (80%) and half of the patients had early prophylaxis (51%). Ceiling effects are more likely to occur in intensively treated patients.

Despite the lack of cross-cultural validation studies of the pedHAL, the use of the pedHAL is recommended in international guidelines [13]. After development of the pedHAL in Dutch children, only one clinimetric study was performed in Romanian children [1,4]. In Romanian children high proportions of 'N/A' responses were recorded in the domains 'household tasks' and 'leisure activities and sports'. This was confirmed by the present 'pooling' study including more children from different populations. It seems that especially these domains are culturally dependent and removing these items is expected to results in a questionnaire that will perform better in a multicultural and global context. The shift towards higher domain scores in some domains (i.e. 'sitting/kneeling/standing', 'functions of the legs', 'use of transport' and 'self-care') and lower domain scores for 'functions of the arms' were a result of the different reasons for removing items. For example, in the domain 'functions of the arms' items were only removed for ceiling effects and in the domain 'use of transport' the most difficult item 'cycling' was removed for a high number of 'N/A' responses. However, the sum scores of the pedHAL and pedHAL$_{short}$ were similar.

For two items with a high item-total correlation rephrasing of the question may be considered. The items 'walking upstairs' and 'walking downstairs' had a high inter-item correlation of 0.81. 'Walking upstairs' was reported by the participants as being slightly more difficult. As both items are about walking stairs, 'walking stairs' may better capture the activity than choosing one of the two activities. For calculating the pedHAL$_{short}$ from the original pedHAL, any limitation reported on walking stairs could be scored as abnormal.

Internal consistency of the pedHAL$_{short}$ (Cronbach's $\alpha = 0.97$) is still higher than the recommended Cronbach's α between 0.7 and 0.9. As the internal consistency improved after reduction of the 31 items and there was no clear indication for removing any other specific items, it was decided to retain the remaining items.

To ensure that the pedHAL$_{short}$ contains all informative items, the 22-item pedHAL$_{short}$ was preferred above the 20-item pedHAL$_{short}$ resulting from the sensitivity analysis including all parent forms.

Comparison with other studies

Similar domains were important in two studies not included in this pooled data. In boys with haemophilia from Lithuania and Portugal, the most difficulties were reported in

the domains 'sitting/kneeling/standing', 'functions of the legs' and 'leisure activities and sports' [8,14]. The fewest difficulties were reported in the domain 'self-care' in the Lithuanian data and in the domains 'household tasks' and 'self-care' in the Portuguese data, which were both less informative domains in the pooled data [8,14]. Exact scores were difficult to compare because both studies reported mean scores. The highest proportions of 'N/A' responses were in the domain 'leisure activities and sports' in the Lithuanian data, which was similar in the pooled data [8]. No other studies were available to further compare our findings. In adults, similar scoring patterns were shown by domain level [15,16].

Clinical implications and future research

Within a context of multiple outcomes assessments in haemophilia care, a shorter assessment of limitations in activities and participation is desirable. This pooling study of international pedHAL data in children with mild to severe haemophilia with a wide range of treatment regimens suggested that 31 pedHAL items are redundant, resulting in a notable shortening of the questionnaire. The shorter version of the pedHAL includes the most relevant and informative items for children and youth with haemophilia. The $pedHAL_{short}$ can be derived from the original pedHAL, which allows for use in longitudinal studies. Only the sum score should be used for the $pedHAL_{short}$, since some domains only have 1 item in the $pedHAL_{short}$. Before introduction of the $pedHAL_{short}$ construct validity and reliability of the questionnaire should be investigated in diverse populations.

CONCLUSION

This clinimetric study resulted in a reduction of the pedHAL by more than half after a stepwise procedure of removing items. This short version of the pedHAL (22 items) is expected to retain the most relevant and informative items on activities and participation for children with haemophilia, representing all domains of the original pedHAL. It detects similar proportions of abnormal sum scores.

REFERENCES

1 Groen WG, van der net J, Helders PJM, Fischer K. Development and preliminary testing of a Paediatric Version of the Haemophilia Activities List (pedhal). Haemophilia *2010*; 16: 281–289.

2 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: 565–571.

3 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: 36–46.

4 Groen W, van der Net J, Lacatusu AM, et al. Functional limitations in Romanian children with haemophilia: Further testing of psychometric properties of the Paediatric Haemophilia Activities List. Haemophilia *2013*; 19: 116–125.

5 World Health Organization. International Classification of Functioning, Disability and Health: Children & Youth Version: ICF-CY; *2007.*

6 Fischer K, Poonnoose P, Dunn AL, et al. Choosing outcome assessment tools in haemophilia care and research: A multidisciplinary perspective. Haemophilia *2016*; 23*:* 1–14.

7 Khair K, Holland M, Bladen M, et al. Study of physical function in adolescents with haemophilia: The SO-FIT study. Haemophilia *2017*; 23: 918–925.

8 Radzevič V, Raistenskis J, Ragelienė L, Kowalski IM. Relationship between physical activity and functional ability in school-aged children with hemophilia. Polish Ann Med *2013*; 20: 13–8.

9 Kuijlaars IAR, van der Net J, Schutgens REG, Fischer K. The Paediatric Haemophilia Activities List (pedHAL) in routine assessment: changes over time, child-parent agreement and informative domains. Haemophilia *2019*; 25: 953–959.

10 Saulyte-Trakymiene S, Juodyte A, Jusinskaite V, Kulikauskaite R. Systematic evaluation of hemophilic arthropathy in Lithuania. J Med Sci *2019*; 7: 1–15.

11 Bouskill V, Hilliard P, Stephens S, et al. An institutional pilot study to investigate physical activity patterns in boys with haemophilia. Haemophilia *2016*; 22: e383–e389.

12 de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine: A practical guide. New York: Cambridge University Press; *2011.*

13 Srivastava A, Brewer AK, Mauser-Bunschoten EP, et al. Guidelines for the management of hemophilia. Haemophilia *2013;* 19:e1.

14 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: e54–67.

15 Kempton CL, Recht M, Neff A, et al. Impact of pain and functional impairment in US adults with haemophilia: Patient-reported outcomes and musculoskeletal evaluation in the pain, functional impairment and quality of life (P-FiQ) study. Haemophilia *2018*; 24: 261–270.

16 McLaughlin P, Morris R, Chowdary P. Investigating the relationship between the HJHS and HAL in routine clinical practice: A retrospective review. Haemophilia *2018; 24:* 1–7.

SUPPLEMENTARY MATERIAL

Supplemental table 1. Frequency tables of all domains with the distribution of responses for all items

Item	Missing and/or N/A scores (%)	Distribution of response options scored (%)					
		Impossible	Always	Usually	Sometimes	Almost never	Never
Sitting/kneeling/standing							
Sitting down	1.3	0.3	1.6	1.6	5.1	8.0	83.3
Sitting on the ground	2.9	1.0	2.6	1.6	8.5	10.1	76.1
Standing up from a chair with arm rests	2.2	0.6	1.9	2.9	2.9	7.8	83.8
Standing up from a chair without arm rests	5.4	1.3	2.3	1.3	4.0	8.1	82.9
Kneeling/squatting	0.3	2.9	4.1	3.8	9.6	13.7	65.9
Squatting for long periods	6.3	4.4	2.7	4.4	15.6	15.9	56.9
Bending over forwards	1.3	1.0	1.3	1.9	6.8	10.6	78.5
Standing still for a short period	1.3	0.6	1.9	5.1	7.1	11.9	73.3
Standing still for longer periods	3.8	3.0	3.0	6.6	11.9	14.9	60.7
Standing still for very long periods	3.8	3.0	3.0	6.6	11.9	14.9	60.7
Functions of the legs							
Walking short distances	1.0	1.0	2.2	2.6	5.8	11.5	76.9
Walking longer distances	1.9	1.6	2.3	4.2	15.9	12.6	63.4
Walking long distances	9.8	2.8	5.3	8.5	15.1	16.5	51.8
Walking on an uneven surface	1.6	0.6	2.6	2.9	11.3	14.5	68.1
Walking on a soft surface	5.7	0.3	2.4	1.3	2.7	9.8	83.5
Strolling	5.7	1.3	1.7	2.4	9.8	12.8	72.1
Running	2.9	2.3	2.6	4.6	6.5	13.7	70.3
Jumping	2.9	2.9	2.6	3.6	12.7	13.4	64.7
Walking upstairs	0.6	1.0	2.9	2.2	8.0	12.8	73.2
Walking down-stairs	0.6	0.6	2.9	2.2	6.1	9.6	78.6
Walking or riding up a small hill or slope without help	3.5	1.6	2.0	3.0	7.9	13.2	72.4
Functions of the arms							
Carrying large or heavy objects with two hands	4.1	0.7	2.0	2.6	7.6	14.9	72.2
Stretching to reach something above your head	2.5	0.7	1.6	2.3	6.2	12.1	77.2

Supplemental table 1. (Continued)

Item	Missing and/or N/A scores (%)	Distribution of response options scored (%)					
		Impossible	Always	Usually	Sometimes	Almost never	Never
Fine hand movements	1.0	0.3	2.6	0.3	3.8	3.5	89.4
Writing	1.3	0.3	1.9	2.3	6.4	10.6	78.5
Leaning on your arms	1.3	0.6	1.9	2.9	7.7	11.6	75.2
Shaking hands with someone	1.9	0.0	1.3	1.9	1.9	3.9	90.9
Use of transport							
Cycling	18.7	2.3	1.6	2.0	9.8	7.0	77.3
Getting in and out of the car	1.9	0.6	2.3	2.3	5.5	5.8	83.5
Using public transport	14.6	0.0	1.9	0.4	5.2	8.2	84.4
Self-care							
Drying off your entire body	1.0	0.6	2.2	1.3	2.9	4.8	88.1
Putting on a t-shirt or jumper, etc.	0.6	0.3	2.6	1.6	4.2	7.0	84.3
Putting on pants	0.6	0.6	3.8	1.0	5.1	7.7	81.8
Putting on shoes and socks	0.0	1.0	3.2	2.9	6.7	8.3	78.1
Wiping your bottom after using the toilet	1.9	0.3	1.6	1.9	2.3	4.2	89.6
Fastening a hood or doing up the top button on your jacket	1.6	0.3	2.6	1.6	3.2	4.8	87.4
Buttering bread or making a sandwich	6.7	0.3	2.0	1.4	1.0	4.1	91.2
Unscrewing the lid from a bottle of water, juice, etc.	2.5	0.3	2.3	1.3	4.6	6.2	85.3
Brushing your teeth	0.6	0.3	2.9	1.6	2.6	4.8	87.9
Household tasks							
Chores around the house	7.6	1.4	2.4	2.4	6.2	6.5	81.1
Outside chores	16.8	1.1	2.3	1.5	4.6	8.0	82.4
Other household chores	15.6	1.9	1.9	1.1	5.6	9.0	80.5
Leisure activities and sports							
Going out	18.7	1.2	1.6	2.3	5.5	9.4	80.1
Playing outside, alone or with others	4.4	0.7	1.3	5.3	8.6	12.6	71.4
School sports: exercises and gymnastic equipment	8.9	2.8	1.4	5.9	13.9	14.6	61.3
School sports: athletics	19.7	4.7	2.0	5.5	8.7	14.2	64.8
School sports: ball sports	16.5	2.7	2.3	5.3	8.0	13.3	68.4
Playing non-contact team sports	19.7	2.4	2.4	4.3	8.3	12.6	70.0

Supplemental table 1. (Continued)

Item	Missing and/or N/A scores (%)	Distribution of response options scored (%)					
		Impossible	Always	Usually	Sometimes	Almost never	Never
Playing contact team sports	25.7	5.6	3.4	6.4	11.1	12.8	60.7
Individual non-contact sports	21.6	1.6	2.4	3.2	7.7	6.9	78.1
Individual contact sports	57.1	17.0	5.2	5.2	4.4	6.7	61.5
Taking part in a sports event...	41.9	3.8	4.4	3.8	7.7	8.7	71.6
Going to school camp or summer camp	42.9	3.3	2.8	3.3	7.8	5.0	77.8

Supplemental table 2. Item-total correlations with the adjusted pedHAL sum score (22 items)

Item	**item-total correlation, r**
Sitting/kneeling/standing	
Sitting down	0.557
Sitting on the ground	0.655
Standing up from a chair with arm rests	0.557
Standing up from a chair without arm rests	0.552
Squatting for long periods	0.758
Bending over forwards	0.635
Standing still for longer periods	0.745
Functions of the legs	
Walking longer distances	0.738
Walking on a soft surface	0.565
Strolling	0.656
Running	0.669
Walking upstairs	0.697
Walking or riding up a small hill or slope without help	0.699
Functions of the arms	
Carrying large or heavy objects with two hands	0.636
Stretching to reach something above your head	0.632
Writing	0.547
Leaning on your arms	0.602
Use of transport	
Getting in and out of the car	0.594
Self-care	
Putting on pants	0.587
Chores around the house	0.608
Leisure activities and sports	
Playing outside, alone or with others	0.687

HAL

CHAPTER 6

Assessing the test retest reliability and smallest detectable change of the Haemophilia Activities List

Isolde A.R. Kuijlaars;[1]
Madelon van Emst;[1]
Janjaap van der Net;[2]
Merel A. Timmer;[1]
Kathelijn Fischer[1]

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

[2] Center for Child Development, Exercise and Physical Literacy, University Medical Center Utrecht, Utrecht University, University Children's Hospital, Utrecht, The Netherlands

Haemophilia. 2021;27: 108-112.

ABSTRACT

Introduction

The Haemophilia Activities List (HAL) is a preferred instrument to measure self-reported limitations in activities in persons with haemophilia (PWH). Information on reliability and interpretability of HAL scores is lacking.

Aim

To examine the test-retest reliability and smallest detectable change (SDC) of the HAL in adult PWH.

Methods

Fifty adult (≥18 years) persons with mild to severe haemophilia completed the HAL (42 items, 7 domains, optimum 100) at baseline (T0) and 3 to 4 weeks later (T1). The intraclass correlation coefficient (ICC) and SDC were calculated for sum and component scores.

Results

Fifty persons with haemophilia were included (median age 49 years; 92% haemophilia A; 70% severe haemophilia). The median (interquartile ranges) HAL sum score was 77 (62 to 99) at T0 and 81 (64 to 98) at T1. Reliability was good with ICCs for sum and component scores > 0.9. The SDC for the sum score was 10.2, for the upper extremity component score 9.2, for the basic lower extremity component score 16.7 and for the complex lower extremity component score 13.4.

Conclusion

The HAL has a good reliability for the sum and component scores. Score changes of the normalized sum HAL score greater than the SDC 10.2 indicate that the change was not a result of measurement error.

INTRODUCTION

Persons with haemophilia (PWH) suffer from recurrent joint bleeds that lead to synovial inflammation and blood related cartilage damage, eventually resulting in haemophilic arthropathy [1,2]. Joint impairment will result in limitations in functional abilities, daily activities and participation in society, and a reduction of quality of life [1].

In developed countries, treatment of haemophilia has greatly improved over the last decades and life expectancy of PWH has almost normalized [3]. Especially now, with gene therapy as a promising next step in haemophilia care [4], appropriate clinimetric instruments are essential to assess the effect of new (para)medical treatments and to monitor patients at individual level. Besides reporting bleeding episodes and joint assessment, measurement of the impact of haemophilia on activities and participation in relation with their society is important [1].

The Haemophilia Activities List (HAL) is recommended to measure self-reported activities and participation [5]. The HAL has been developed with patient interviews according to the World Health Organization (WHO) International Classification of Functioning, Disability and Health (ICF) and it measures self-reported limitations in activities and participation due to haemophilia in the previous month [6–8]. In addition to being clinically relevant, any instrument should be valid and reliable. Validity is the degree to which an instrument measures the construct which it aims to measure. Reliability is the degree to which the measurement is free from measurement error. Furthermore, interpretability is an important measurement property which is the degree to which one can assign qualitative meaning to an instrument's quantitative scores or change in scores [9,10]. The HAL was developed according the Classical Test Theory (CTT), which implies that the sum and component scores were a sum of all individual ordinal items of the questionnaire [8].

A recent systematic review performed according the CTT reported that the HAL had good content validity as it reflects daily activities which were based on interviews with PWH, while there was conflicting evidence for construct validity [5]. For example, the HAL discriminated well between patients on intensive and less intensive prophylaxis but not between patients who stopped or continued prophylaxis [5]. However, information on reliability including test-retest reliability and interpretability of scores is lacking, which is necessary to interpret HAL scores in clinical practice and research [5].

The aim of this study was to examine the test-retest reliability and the smallest detectable change (SDC) of the HAL in adult PWH. Furthermore, the measurement error needs to be considered to determine the SDC.

MATERIALS AND METHODS

Study design and study population

This study was a single-center prospective, psychometric study. Adult (≥18 years) persons with mild to severe haemophilia who visited the Van Creveldkliniek, Utrecht, The Netherlands, for routine assessment were asked to participate in the study. The first HAL (T0) was completed during a clinic visit. The second HAL was sent by mail three weeks later (T1) and PWH were asked to complete the questionnaire within one week. The time-interval between T0 and T1 was considered sufficiently long to prevent recall bias. Data were collected between September 2017 – September 2018. PWH were excluded if they had a recent bleed, synovitis or joint surgery at/between T0 or T1. We aimed for the inclusion of 50 PWH, according to the Consensus-based Standards for the development of Measurement Instruments (COSMIN) guidelines [11].

The Medical Research Ethical Committee (MREC) of the University Medical Center Utrecht reviewed and approved the study (17-591/C).

Measurements

The HAL contains 42 items across 7 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). Items are scored on a 6-point Likert scale ('impossible', 'always', 'usually', 'sometimes', 'almost never', 'never'), with a 'not applicable' option for some items. A summary score as well as component scores (upper extremity, basic lower extremity and complex lower extremity) can be calculated using the official scoring tool (available at www.vancreveldkliniek.nl) [12]. All these scores are converted to a normalized score from 0 to 100, where higher scores represent a better functional status. If more than half of the items were missing or scored 'not applicable', no valid domain, component and sum score were calculated [6,8].

Patient characteristics included age at baseline HAL assessment, type of haemophilia (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]), use of aids and time between test and retest.

Statistical analyses

Patient characteristics and time between T0 and T1 were presented as proportions or medians (interquartile ranges [IQR:P25;P75]). Descriptive analyses (median, IQR, range) were performed for the HAL sum score and component scores at T0 and T1. In addition, to assess an effect of delayed response (> 3-4 weeks) the time between T0 and T1 was plotted against the change of the HAL sum score of T0 and T1 and a linear regression analysis was performed.

Analyses were performed using IBM SPSS Statistics software version 26.

Reliability, measurement error and interpretability were evaluated and interpreted according to the definitions of COSMIN [9,10]. Both the development of the HAL and the analyses of the present study were performed according CTT. Using CTT, the standard error of measurement (SEM) is assumed to be stable over the total scale [9]. The SEM and SDC calculated in the present study should be interpreted as average SEM and SDC values for the HAL scores.

Reliability

Reliability is defined as the degree to which the measurement is free from measurement error and it expresses how well patients can be distinguished from each other despite the presence of the measurement error [9]. The intraclass correlation coefficient (ICC) ($ICC_{agreement}=\sigma_p^2/[\sigma_p^2 + \sigma_m^2 + \sigma_r^2]$) was calculated for test-retest reliability with a two-way random effects model for agreement, where each term refers to a variance component (σ^2): p = patient, m = measurement, r = residual [9,13]. The ICC represents the part of the variance between scores that can be attributed to 'true' differences between patients. ICC is expressed as a value between 0 and 1: a value of >0.70 is considered acceptable [9].

Measurement error

Measurement error is defined as the systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured [10]. The standard error of measurement (SEM) for agreement ($SEM_{agreement} = \sqrt{(\sigma^2_m + \sigma^2_r)}$) was calculated [9].

In addition, a Bland and Altman plot was shown for the HAL sum score to illustrate the measurement error, in relation to the mean HAL score. The 95% Limits of Agreement (LoA) (LoA = mean difference T0 – T1 ± 1.96 × SD difference T0 – T1) illustrates the variation in scores in stable patients [9,14].

Interpretability

Interpretability is defined as the degree to which one can assign qualitative meaning to an instrument's quantitative scores or change in scores [10]. The $SDC_{agreement}$ ($SDC_{agreement} = 1.96 \times \sqrt{2} \times SEM_{agreement}$) was calculated as a measure of interpretability and is the smallest change in score that you can detect above the measurement error [9].

The $ICC_{agreement}$, $SEM_{agreement}$, and $SDC_{agreement}$ were calculated for sum and component scores.

RESULTS

Patient characteristics

Sixty-nine PWH were invited, fifteen participants were excluded due to recent bleeding or synovitis and four participants did not return the HAL at T1. Eventually, 50 PWH were included and analyzed (Table 1). The sum and three component scores were available for all patients at both T0 and T1. Scoring was missing for 9/4200 items in total. The median age was 49.0 years (range 20 to 79) and 70.0% had severe haemophilia, 10.0% moderate haemophilia and 20.0% mild haemophilia. Nine PWH (18.0%) used aids when performing certain activities. Median (IQR) time between measurement at T0 and T1 was 3.4 weeks (3.0 ; 5.2), with a range of 1.4 ; 10.4 weeks.

HAL sum and component scores

Table 2 presents the sum and component scores at T0 and T1. At group level the median (IQR) HAL sum score was 77.1 (62.5 ; 98.6) at T0 and 81.2 (63.6 ; 98.5) at T1. The median (IQR) absolute difference for sum and component scores varied from |2.2| (0.0 ; 4.5) to |4.4| (0.0 ; 6.7). PWH scored highest on the upper extremity component and lowest on the complex lower extremity component. Scores were highest on the domain 'selfcare' and lowest on the domain 'functions of the legs' (Supplementary material). Maximum scores at T0 and T1 occurred frequently with maximum HAL sum in 18% and maximum component scores in 20-32%. The sum and component scores had left skewed distributions. The difference in HAL sum and component scores between T0

and T1 increased with increasing time between assessments (B=0.18, p=0.001); in 3/4 PWH who filled in the retest >50 days (7.1 weeks) scores varied >10.0 points.

Table 1. Patient characteristics at baseline (n=50)

Patient characteristics (n=50)	Median (IQR) or n (%)
Age (years)	49.0 (36.8 ; 61.3)
Haemophilia A	46 (92.0)
Severity of haemophilia	
Mild	10 (20.0)
Moderate	5 (10.0)
Severe	35 (70.0)
Using aids when performing certain activities [a]	7 (14.0)
One crutch/ cane	4 (8.0)
Two crutches	3 (6.0)
Wheelchair	1 (2.0)
Other aids [b]	2 (4.0)
Time (weeks) between T0 and T1	3.4 (3.0 ; 5.2)

[a] Three persons used two different aids
[b] Other aids: i.e. scooter or modified bicycle

Reliability, measurement error and interpretability

Table 3 presents the $ICC_{agreement}$, $SEM_{agreement}$, and $SDC_{agreement}$ for the sum and component scores. All ICC values exceeded 0.90. For the HAL sum score the SEM was 3.7 and the SDC was 10.2. The basic lower extremity component score had the highest variation with SEM (6.0) and SDC value (16.7), the upper extremity component score had the lowest variation with SEM (3.3) and SDC value (9.2). Figure 1 shows the Bland and Altman plot for the HAL sum score, with LoA of -0.92 ± 10.14. The differences between scores at T0 and T1 did not change with increasing mean HAL values, which was graphically checked.

After exclusion of PWH with a time between T0 and T1 >50 days, all ICC values increased and SEM and SDC values were smaller; the basic lower extremity component score had the highest variation with SEM (5.7) and SDC value (15.8), the HAL sum score had the lowest variation with SEM (2.8) and SDC value (7.8) (see Supplementary material).

Table 2. Characteristics of the test (T0) and retest (T1) for the HAL sum and component scores

HAL score	T0			T1			Absolute difference (T0 – T1)
	Median (IQR)	Min	Max	Median (IQR)	Min	Max	Median (IQR)
Sum	77.1 (62.5 ; 98.6)	23.4	100.0	81.2 (63.6 ; 98.5)	19.5	100.0	2.4 (0.5 ; 5.0)
Upper extremity	95.6 (82.8 ; 100.0)	37.8	100.0	95.6 (83.3 ; 100.0)	28.9	100.0	2.2 (0.0 ; 4.5)
Basic lower extremity	73.3 (55.0 ; 100.0)	6.7	100.0	78.3 (53.3 ; 100.0)	3.3	100.0	3.3 (0.0 ; 6.7)
Complex lower extremity	57.8 (30.6 ; 96.1)	6.7	100.0	56.7 (32.8 ; 100.0)	6.7	100.0	4.4 (0.0 ; 6.7)

Table 3. Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) and intra-class Correlation Coefficient (ICC) of the Haemophilia Activities List (n=50)

HAL score	$SEM_{agreement}$	$SDC_{agreement}$	$ICC_{agreement}$ (95% CI)
Sum	3.68	10.20	0.97 (0.95 ; 0.98)
Upper extremity	3.33	9.23	0.97 (0.94 ; 0.98)
Basic lower extremity	6.02	16.69	0.95 (0.91 ; 0.97)
Complex lower extremity	4.85	13.45	0.98 (0.96 ; 0.99)

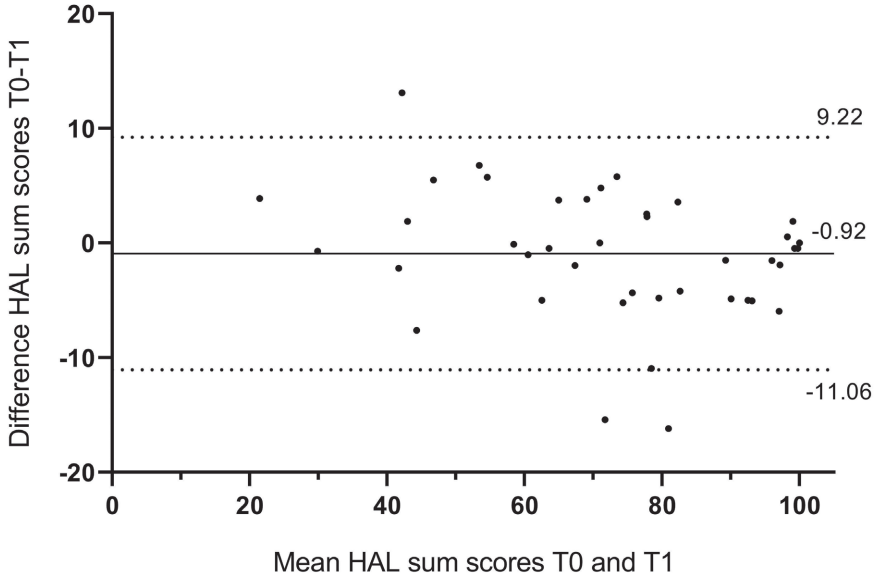

Figure 1. Bland and Altman plot of the HAL sum score, with Limits of Agreement of -0.92 ± 10.14

DISCUSSION

The present study aimed to determine the test-retest reliability and the SDC of the HAL. The HAL demonstrates a good test-retest reliability: the sum and components score had an ICC value >0.90. The average SDC value for the normalized HAL sum score was 10.2. This implies that a change in score of 10.2 signifies a true change in one patient and is not due to measurement error. For the upper extremity component score a change in score of 9.3 signifies a true change, for the basic lower extremity component score a change of 16.7 and for the complex lower extremity component score a change of 13.5. SDC values were smaller when excluding patients with a delayed response (> 50 days): the SDC for the sum score was 7.8, for the upper extremity component score 8.8, for the basic lower extremity component score 15.8 and for the complex lower extremity component score 11.7.

Comparison with other studies

Studies examining measurement properties of the HAL are limited. The ICC values of the present study are similar to previously reported ICC values of 0.87-0.97 in adult PWH in the USA (n=158-162), which reported on two questionnaires completed within

2 hours [15]. ICC values (0.66-0.90) were lower in Brazilian PWH (n=52) who completed the HAL during interviews (with an interval of 15 days), which is different to individually completing a paper questionnaire in the present study [16]. SEM and SDC values have not been published until now.

Strengths and limitations

A strength of this study was the follow up time of median 3.4 weeks, which is sufficiently long to prevent recall bias. In addition, the sample size of 50 patients was according to the recommendation of the COSMIN guideline.

A disadvantage of the CTT approach is that the calculated SEM and SDC values are stable over the whole continuum of the score. In the present study HAL sum scores were high, comparable to scores in studies in the United Kingdom (UK) and United States of America (USA) [17,18], indicating a ceiling effect of the HAL in Western countries. Therefore, the SEM and SDC values calculated in the present study best reflect the measurement error and SDC for the upper end of the HAL score (better functional status). Furthermore, the HAL sum and component scores (0-100) are a sum of the ordinal items and are not corrected for the difficulty of the separate items. For example, scoring 'impossible' on an easy item like 'sitting down' has the same weight for the sum score as scoring 'impossible' on a more difficult item like 'running'.

In addition, the skewed distribution affects the precision of the calculation of the ICCs, SEMs and SDCs which is based on variance components of analysis of variance (ANOVA). The ANOVA test assumes that the data is normally distributed, which was not the case in this study. Finally, the outliers with a delayed response time (T0 to T1 >50 days) increased the SEM and SDC, which implies that the assumption that patients do not change over time in this study design was not fully met.

Clinical implications and future research

Change scores of the normalized sum HAL score greater than the SDC 10.2 indicate that the change was not a result of measurement error. The SDC of the HAL helps to pick up real changes in activities and participation in clinical practice when patients were monitored with the HAL before and after an intervention or on an annual routine visit. In addition, the SDC score should be compared with the Minimal Important Change (MIC) which is the smallest change in score that is perceived as important by patients [9]. However, the MIC for the HAL still needs to be established.

CONCLUSION

The HAL is a reliable self-reported outcome measure for limitations in activities and participation in PWH. Average SDC values are 10.2 for the normalized HAL sum score, 9.2 for the upper extremity component score, 16.7 for the basic lower extremity component score, and 13.4 for the complex lower extremity component score, which signifies a true change in score that is not due to the measurement error. The difference in HAL scores between test and retest increases with larger time intervals between tests.

REFERENCES

1 Srivastava A, Santagostino E, Dougall A, et al. WFH Guidelines for the Management of Hemophilia, 3rd edition. Haemophilia *2020*; 26: 1–158.

2 van Vulpen, LFD; Holstein C; Martinoli K. Joint disease in haemophilia: Pathophysiology, pain and imaging. Haemophilia *2018*; 24: 44–49.

3 Shapiro S, Makris M. Haemophilia and ageing. Br J Haematol *2019*; 184: 712–720.

4 Miesbach W, O'Mahony B, Key NS, Makris M. How to discuss gene therapy for haemophilia? A patient and physician perspective. Haemophilia *2019*; 25: 545–557.

5 Timmer MA, Gouw SC, Feldman BM, et al. Measuring activities and participation in persons with haemophilia: A systematic review of commonly used instruments. Haemophilia *2017*; 24(2): e33-e49.

6 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: 36–46.

7 World Health Organization. International Classification of Functioning, Disability and Health: Children & Youth Version: ICF-CY; *2007*.

8 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: 565–571.

9 de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine: A practical guide. New York: Cambridge University Press; *2011*.

10 Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol *2010*; 63: 737–745.

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

12 van Genderen FR. Functional limitations in severe hemophilia; *2006*.

13 Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. Psychol Bull *1979*; 86: 420–428.

14 Bland J, Altman D. Measuring agreement in method comparison studies. Stat Methods Med Res *1999*; 8: 135–160.

15 Kempton CL, Wang M, Recht M, et al. Reliability of patient-reported outcome instruments in US adults with hemophilia: the Pain, Functional Impairment and Quality of life (P-FiQ) study. Patient Prefer Adherence *2017*; 11: 1603–1612.

16 Ramos AAT, Wolff ÁLP, Lorenzato CS, et al. Translation, validation and reliability of the functional capacity questionnaire Haemophilia Activities List for haemophilia patients in Brazil. Haemophilia *2019*; 25: e231–239.

17 McLaughlin P, Morris R, Chowdary P. Investigating the relationship between the HJHS and HAL in routine clinical practice: A retrospective review. Haemophilia *2018*; 24: 988-994.

18 Kempton CL, Recht M, Neff A, et al. Impact of pain and functional impairment in US adults with haemophilia: Patient-reported outcomes and musculoskeletal evaluation in the pain, functional impairment and quality of life (P-FiQ) study. Haemophilia *2018*; 24: 261–270.

SUPPLEMENTARY MATERIAL

Supplemental table 1. Characteristics of the test (T0) and retest (T1) for the HAL domains

HAL domain	T0			T1		
	Median, (IQR)	Min	Max	Median, (IQR)	Min	Max
Lying down/sitting/ kneeling/standing	72.5 (46.9 ; 100.0)	15.0	100.0	70.0 (47.5 ; 100.0)	22.5	100.0
Functions of the legs	61.1 (43.9 ; 100.0)	15.6	100.0	68.9 (43.3 ; 100.0)	11.1	100.0
Functions of the arms	90.0 (70.0 ; 100.0)	5.0	100.0	95.0 (80.0 ; 100.0)	20.0	100.0
Use of transportation	96.7 (80.0 ; 100.0)	20.0	100.0	100.0 (66.7 ; 100.0)	20.0	100.0
Self-care	100.0 (87.0 ; 100.0)	40.0	100.0	100.0 (88.0 ; 100.0)	36.0	100.0
Household tasks	92.0 (66.7 ; 100.0)	24.0	100.0	93.3 (65.8 ; 100.0)	16.7	100.0
Leisure activities and sports	76.7 (60.0 ; 100.0)	20.0	100.0	87.6 (60.0 ; 100.0)	14.3	100.0

Supplemental table 2. Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) and intra-class Correlation Coefficient (ICC) of the Haemophilia Activities List without outliers in test-retest time (n=46)

HAL score	$SEM_{agreement}$	$SDC_{agreement}$	$ICC_{agreement}$ (95% CI)
Sum	2.82	7.83	0.98 (0.97 ; 0.99)
Upper extremity	3.16	8.76	0.97 (0.95 ; 0.98)
Basic lower extremity	5.70	15.79	0.96 (0.92 ; 0.98)
Complex lower extremity	4.21	11.68	0.98 (0.97 ; 0.99)

CHAPTER 7

Shortening the Haemophilia Activities List (HAL) from 42 items to 18 items

Isolde A.R. Kuijlaars;[1]
Janjaap van der Net;[2]
Tyler W. Buckner;[3]
Christine L. Kempton;[4]
Roger E.G. Schutgens;[1]
Kathelijn Fischer[1]

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

[2] Center for Child Development, Exercise and Physical Literacy, University Medical Center Utrecht, Utrecht University, University Children's Hospital, Utrecht, The Netherlands

[3] University of Colorado School of Medicine, Aurora, CO, USA

[4] Hemophilia of Georgia Center for Bleeding & Clotting Disorders of Emory, Emory University School of Medicine, Atlanta, GA, USA

Submitted

ABSTRACT

Introduction

The Haemophilia Activities List (HAL) was developed to measure activities and participation in persons with haemophilia (PWH). Shortening the questionnaire may facilitate use of the HAL.

Aim

The aim of this study was to determine which items of the HAL are redundant, to construct a shorter version of the HAL, and to determine the construct validity of the HAL_{short}.

Methods

A secondary analysis was performed on pooled data of two published studies using the HAL (7 domains, 42 items, optimum score: 100) in adults with haemophilia A/B. Data were divided into a derivation (62%) and a validation set (38%). Redundant items were identified by evaluation of: floor and ceiling effects, proportions of missing and 'not applicable' responses, inter-item correlations, component loadings in an exploratory factor analysis, internal consistency, and item-total correlations. Correlations with the SF-36 and EQ-5D-5L were used to determine construct validity of the HAL_{short}.

Results

Data on 680 PWH were evaluated. In the derivation dataset (n=420), median age was 30 years [range 18-80], 43% had severe haemophilia and 61% received prophylaxis. Median (IQR) HAL sum score was 65.0 (55.7-88.8). The stepwise procedure resulted in a HAL_{short} of 18 items with a median sum score of 63.3 (54.4-86.7). Construct validity was similar for the HAL and HAL_{short} in the validation dataset (n=260).

Conclusion

This clinimetric study resulted in a >50% shortening of the HAL. The 18-item HAL_{short} reduces patient burden and is expected to capture the information on activities and participation. The HAL_{short} needs further validation.

INTRODUCTION

The Haemophilia Activities List (HAL) assesses self-reported limitations in various activities of daily living, which are relevant to persons with haemophilia (PWH) [1,2]. The HAL includes 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'. The HAL is recommended for both research purposes and clinical management of patients [3]. The questionnaire has been developed using patient interviews and classification according to the World Health Organization (WHO) International Classification of Functioning, Disability and Health (ICF) [4]. All items belong to the ICF 'activities and participation', with 'activity' defined as 'the execution of a task or action by an individual' and 'participation' as 'involvement in a life situation' [4].

After introduction of the HAL to clinical care and research in 2004, clinicians and researchers reported some items to be non-informative [5]. Shortening the questionnaire may enhance the feasibility of HAL use within the context of multiple outcome assessments in haemophilia care.

The aim of this study was to determine which items of the HAL are redundant to construct a shorter version of the HAL for the measurement of activities and participation in adults with haemophilia. In addition, construct validity of the HAL_{short} was determined in comparison to the SF-36 and EQ-5D-5L.

MATERIALS AND METHODS

Study design and study population

This study was a cross-sectional secondary analysis of pooled data of the *Pain, Functional Impairment, and Quality of Life* (P-FIQ) study and *Bridging Hemophilia B Experiences, Results, and Opportunities Into Solutions* (B-HERO-S) study, using the HAL in PWH in the United States [6,7]. The data of the P-FIQ and B-HERO-S studies were shared for this secondary analysis. Inclusion criteria were PWH A (FVIII) and B (FIX) of all severities, aged ≥18 years. Patients were excluded if >50% of the HAL items were missing, which results according to the HAL scoring tool in a 'not applicable' score.

The data were split in a derivation (n=420, 61.7%) and validation (n=260, 38.2%) dataset. The derivation set was used to identify non-informative items and the validation set was used to validate the HAL_{short}. To achieve equal representation, data were split according to the original study (P-FIQ vs. B-HERO-S) and treatment regimen and randomly assigned to the derivation and validation dataset in SPPS (version 25, IBM). The sample size for the stepwise process in the derivation dataset was set on 420 patients, needed for adequate field testing of measurement instruments and factor analysis [8].

The Medical Research Ethical Committee (MREC) of the University Medical Center Utrecht approved the study (protocol number 20-650/C).

Measurements

Haemophilia Activities List

The HAL assesses self-reported limitations in activities and participation in PWH. The questionnaire contains 42 items, distributed over seven domains. 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. The HAL was developed according the Classical Test Theory, which implies that the sum, domain and component scores are a sum of all individual ordinal items of the questionnaire. Domain, 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) in the scoring tool available at www.vancreveldkliniek.nl. Domain and component scores were only calculated if ≥50% of the items of a domain or component were scored on the 6-point Likert scale. The HAL demonstrates good test-retest reliability with an intraclass correlation coefficient value >0.90. The average SDC value for the normalized HAL sum score was 10.2 [9].

SF-36v2

The SF-36v2 measures health related quality of life across 8 domains: physical functioning, role physical, bodily pain, general health, vitality, social functioning, role emotional and mental health. In addition, physical and mental health summary scores are calculated. Scores range from 0 to 100, with higher scores indicating better health status [10].

EQ-5D-5L

The EQ-5D-5L measures overall health and covers five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression [11]. Each dimension has 5 levels, indicating 1 as 'no problems' up to 5 as 'extreme problems' [12]. In addition, a 100-point Visual Analogue Scale (VAS) records self-rated health on a 20-cm vertical scale with endpoints labelled as 'the worst health you can imagine' at 0 and 'the best health you can imagine' at 100 [12].

Patient characteristics were captured in all datasets. For the present analyses we extracted and analyzed age at HAL assessment, gender, type of haemophilia (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 regimen (prophylaxis yes/no) and inhibitor status (current/former or never).

Statistical analyses

Patient characteristics were presented as proportions or medians (interquartile ranges [IQR:P25-P75]). In the derivation dataset descriptive analyses (median, IQR, range, mean and standard deviation [SD]) were performed for the HAL domain, component and sum scores. Based on reported limits of agreement (LoA) of test-retest data [9], limitations in activities and participation were defined as ≤90 points for domain, component and sum scores. Normality of the data was checked visually using histograms.

Non-informative items were identified in the derivation dataset using a stepwise process (7 steps) according to the method of de Vet et al. (2011), from the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative [13]. After each step non-informative items were deleted, before proceeding with the following step. The 7 steps were described in detail in the Supplementary material and the publication about shortening the pediatric Haemophilia Activities List (pedHAL) [14].

Step 1 - Floor and ceiling effects: Items with ≥85% minimum or maximum scores were removed.

Step 2 - Missing data and scores with 'N/A': Items which were scored >15% as 'missing' or 'N/A' were removed.

Step 3 - Inter-item correlations (1): Items with inter-item correlations of <0.2 and >0.9 were not included in the factor analysis.

Step 4 - Component loadings of the exploratory factor analysis: Items with factor loadings <0.5 were removed.
Step 5 - Inter-item correlations (2): Items which had a correlation of >0.7 were reviewed by IK, KF and JJ and one of the items was removed. Of item-pairs with high correlation, those with least 'N/A' responses, most limitations and/or most variation belonging to ICF domains were manually selected.
Step 6 - Internal consistency: Cronbach's α should be between 0.7 and 0.9; a higher Cronbach's α after item deletion was considered a reason to eliminate an item.
Step 7 - Item-total correlations: Items with item-total correlations of <0.3 were removed.

After removing non-informative items, a HAL_{short} was created. Median (IQR) normalized domain, component and sum scores and percentages of scores <90 points were calculated for the HAL_{short}, similar to the calculation of domain, component and sum scores of the original scoring tool. Scores of the HAL and HAL_{short} were compared with a Wilcoxon signed rank test as the scores were not normally distributed. A Bland and Altman plot was generated to illustrate the differences between the HAL and HAL_{short} sum scores in relation to the mean HAL and HAL_{short} sum scores. The 95% LoA (LoA = mean difference HAL_{short}–HAL ± 1.96×SD difference HAL_{short}–HAL) illustrates the variation in scores [15]. A secondary exploratory factor analysis was performed for the HAL_{short} to detect any underlying constructs.

Hypotheses testing with a priori defined correlation thresholds and comparisons between subscores was performed in the validation dataset to determine the construct validity of the HAL_{short}. Hypotheses were defined a priori based on expert opinion (KF, JN, IK) and reported correlations of the P-FiQ study [16]. Spearman's correlations were calculated as the data were not normally distributed or on an ordinal scale. Correlation coefficients of ≥0.9 were considered as a very strong correlation, 0.7-0.89 as strong, 0.4-0.69 as moderate, 0.10-0.39 as weak and <0.10 as negligible [17].

SPSS (version 25, IBM) was used for data analyses. Mplus (version 6.12, Muthen & Muthen) was used for the exploratory factor analysis.

RESULTS

Patient characteristics

Data from all 381 PWH from the P-FIQ study and 299 PWH from the B-HERO-S study were included [6,7]. Patient characteristics for the derivation (n=420) and validation dataset (n=260) were similar and are shown in Table 1. In the derivation dataset, median age at the time of completing the HAL was 30.0 years (range 18-80). Most patients were male (88.3%) and a majority of the patients had moderate (36.0%) or severe (41.7%) haemophilia. Most patients were on prophylaxis (61.0%).

HAL domain, component and sum scores

Domain, component and sum scores of the derivation dataset are shown in Table 2. The median (IQR) HAL sum score was 65.0 (55.7-88.8), with a range of 11.7 to 100. 'Positive' HAL sum scores (<90 points) were observed in 76.0% of participants. For domain scores the median (IQR) scores were lowest for the 'sitting/kneeling/standing' (60.0 [52.5-85.0]) and 'functions of the legs' (60.0 [51.1-86.7]). The median score was highest for 'self-care' (88.0 [60.0-100]). For component scores, patients scored lowest on the 'complex lower extremity' component (60.0 [42.5-80.0]) and highest on the 'upper extremity' component (77.8 [60.0-93.3]). 'Not applicable' domain and component scores were rare, with ≤3 'not applicable' scores in 7/10 domain and components. Most 'not applicable' domain scores were reported for 'use of transport' in 59/420 (14%).

Table 1. Patient characteristics of the derivation- and validation datasets

Patient characteristics	Median (IQR), %	
	Derivation (n=420)	Validation (n=260)
Male	88.3	85.8
Age (years)	30.0 (26.0 - 36.5)	30.0 (26.0 - 41.9)
Haemophilia A	43.6	43.1
Severity haemophilia		
Mild	21.7	17.3
Moderate	36.0	33.8
Severe	41.7	48.1
Unknown	0.7	0.8
Prophylaxis	61.0	61.2
Inhibitor (current/former)	9.8	10.8

Table 2. Domain, component and sum scores of the HAL (derivation set, n=420)

				Score <90	Missing/ NA
	Median(IQR)	**Min**	**Max**	**(%)**	**N**
Domains					
Lying/sitting/kneeling/standing	60.0 (52.5-85.0)	10.0	100	76.7	2
Functions of the legs	60.0 (51.1-86.7)	0	100	77.1	1
Functions of the arms	65.0 (60.0-85.0)	0	100	75.2	0
Use of transport	66.7 (53.3-100)	0	100	57.4	59
Self-care	88.0 (60.0-100)	20	100	54.5	0
Household tasks	70.0 (56.7-100)	0	100	62.9	9
Leisure activities and sports	65.4 (54.3-90.0)	5.7	100	68.1	38
Components					
Upper extremity	77.8 (60.0-93.3)	13.3	100	69.0	0
Basic lower extremity	63.3 (56.7-93.3)	0	100	70.5	1
Complex lower extremity	60.0 (42.5-80.0)	0	100	81.6	3
Sum					
Sum score	65.0 (55.7-88.8)	11.7	100	76.0	0

Item reduction

The stepwise process to select non-informative items is shown in Table 3. Detailed information, including frequency tables generated for step 1 and 2, inter-item correlations of step 3 and 5, factor loading of step 4 and the table with item-total correlations for step 7 in the item reduction process are shown in the Supplementary material.

Step 1 - Floor and ceiling effects: Minimum and maximum scores were evaluated for all items. There was no floor or ceiling effect in any of the HAL items.

Step 2 - Missing data and scores with 'N/A': There were few missing responses (0-6) on the items. Missing and/or 'N/A' responses were scored in >15% of the PWH in 2/3 items of the domain 'use of transport', 1/5 items of the domain 'household tasks' and in 1/7 items of the domain 'leisure activities and sports'.

Step 3 - Inter-item correlations (1): 1/9 items of the domain 'functions of the legs' was removed, after evaluating the inter-item correlations (r>0.9). All remaining items were included in the exploratory factor analysis.

Step 4 - Component loadings of the exploratory factor analysis: The exploratory factor analysis suggested no items were eligible for item reduction. A 4-factor model was selected which included all remaining items and the four factors were identified as 'entire body – non-ambulatory activities of daily living', 'lower extremity – weight bearing', 'lower extremity – ambulation' and 'upper extremity – weight carrying'. The model fit of the 4-factor model was 0.07 (RMSEA), indicating moderate model fit. The factor loadings were >0.5.

Step 5 - Inter-item correlations (2): Inter-item correlations were re-evaluated. In the domain 'lying/sitting/kneeling/standing' 4/8 items, which had inter-item correlations >0.7 with other items, were removed. In the domain 'functions of the legs' 4/9 items were removed. The items 'walking upstairs' and 'walking downstairs' had a correlation of 0.86. The authors decided to remove the item 'walking downstairs', which was scored as less difficult than 'walking upstairs' by the participants. In the domain 'functions of the arms' 2/4 items were removed. In the domain 'self-care' 3/5 items were removed. In the domain 'household tasks' 3/6 items were removed. In the domain 'self-care' 3/7 items were removed.

Step 6 - Internal consistency: The remaining 18 HAL items were strongly related (Cronbach's α of 0.96), which indicates redundancy of items. Only complete cases (n=319, 76.0%) were included in this analysis. The Cronbach's α after deletion of separate items was equal or smaller, and therefore did not identify candidate items for removal. Eventually, the authors decided to keep the remaining 18 items, because the Cronbach's alpha was already lowered by removing the 24 items.

Step 7 - Item-total correlations: All item-total correlations were high (r=0.64-0.81), thus identifying no candidates for item reduction.

HAL_{short} with 18 items

Table 3 shows all items of the HAL_{short}. The original domain, component and sum scores of the HAL were calculated for both HAL and HAL_{short} and are shown in Table 4. Eighteen items remained after removing the items (n=24) according to the seven steps. All domains were still represented in the $HAL_{short.}$ Most items of the HAL_{short} belonged to the domains 'lying/sitting/kneeling/standing' (n=4) and 'functions of the legs' (n=4). For the domains 'use of transport' only one item remained in the HAL_{short}. The median (IQR) HAL_{short} sum score was 63.3 (54.4-86.7). The domain, component and sum scores were statistically different between the HAL and HAL_{short} ($p<0.05$). Figure 1 shows the Bland and Altman plot for the HAL vs. HAL_{short} sum score, with LoA of -1.2±4.7. The mean (SD) difference between the HAL and HAL_{short} was 1.2 (2.4) with a range from -5.8 to 10.3.

The differences between the sum scores did not change with increasing mean HAL sum scores, which was graphically checked. The largest discrepancy in the proportions of abnormal domain scores (<90) was observed for the domain 'lying/ sitting/kneeling/ standing' (HAL: 77.0% vs. HAL_{short}: 73.4%) and the component 'basic lower extremity' (HAL: 70.6% vs. HAL_{short}: 68.4%). The vast majority of PWH (90.1%) who scored ≥90 on the HAL, scored ≥90 on the HAL_{short}. The secondary exploratory factor analysis with the 18-item HAL_{short} resulted in a 2-factor model without good model fit (RSMEA=0.09) and clear underlying constructs could not be defined. Therefore, the HAL_{short} generated a single sum score.

Construct validity HAL_{short}

The pre-defined hypotheses which were tested to determine construct validity of the HAL and HAL_{short} are shown in Table 5. Correlations between the $HAL_{(short)}$ and the SF-36v2 and EQ-5D-5L are shown in Table 6. All calculated correlation coefficients met pre-defined cut-off values for both the HAL and the HAL_{short} in the validation datasets, confirming the hypotheses to determine construct validity. In addition, 'basic lower extremity' component scores were lower than 'complex lower extremity' component scores for both the HAL and HAL_{short} ($p<0.001$).

Table 3. Flow chart of steps to reduce the number of HAL items and the remaining 18 HAL items

Lying / sitting/ kneeling/ standing	Functions of the legs	Functions of the arms	Use of transport	Self-care	Household tasks	Leisure activities and sports
Step 1: Floor and ceiling effects (≥85% maximum scores) – remaining items: 42						
Step 2: Missing and/or N/A (>15% missing / N/A) – remaining items: 38						
			1 Riding a bicycle 3 Using public transportation (bus, train, subway)		6 Gardening	5 Dancing
Step 3: Inter-item correlations (r<0.2 and r>0.9) – remaining items: 37						
	9 Jumping					
Step 4: Exploratory factor analysis (factor loadings <0.5) – remaining items: 37						
Step 5: Inter-item correlations (r>0.7) – remaining items: 18						
2 Rising from a chair with armrests 4 Kneeling / squatting 6 Kneeling for a longer period of time 8 Standing for a longer period of time	1 Walking short distances (less than 1 kilometer / 15 minutes) 3 Walking on a soft surface (e.g. on the beach or through the woods) 5 Strolling / (window-) shopping 7 Climbing down the stairs	2 Carrying heavy objects in the arms 3 Fine hand movements (e.g. closing buttons)		2 Putting on a shirt, sweater etc. 4 Putting on a tie or closing the top button of a shirt 5 Going to the toilet	1 Going out shopping (for food, drink etc.) 2 Washing the dishes, cleaning the sink 4 Other household tasks (ironing, making the beds)	3 Going out (theatre / museum / movie theatre / bar) 4 Hobbies 7 Going on a holiday ("passive"; beach-/hotel holiday)

Table 3. (Continued)

Lying / sitting/ kneeling/ standing	Functions of the legs	Functions of the arms	Use of transport	Self-care	Household tasks	Leisure activities and sports
Step 6: Internal consistency – remaining items: 18						
Step 7: Item-total correlations (r<0.3) – remaining items: 18						
HAL_{short}: remaining 18 items						
1 Sitting down (e.g. on a chair or couch) 3 Rising from a chair without armrests 5 Bending forward 7 Squatting for a longer period of time	2 Walking long distances (more than 1 kilometer / 15 minutes) 4 Walking on an uneven surface (e.g. cobblestones, high sidewalks) 6 Climbing up the stairs 8 Running (e.g. in order to catch the bus)	1 Lifting heavy objects 4 Reaching above your head (to pick something up from a high shelf)	2 Getting in and out of a car	1 Drying your whole body 3 Putting on socks and shoes	3 Cleaning the house 5 Doing odd jobs (both in and around the house)	1 Playing games (outdoors, e.g. with your children) 2 Sports 6 Going on a holiday (active)

Table 4. Original domain and sum scores of the HAL and HAL_{short}

	HAL		HAL_{short}	
	Median (IQR)	**Score <90 (%)**	**Median (IQR)**	**Score <90 (%)**
Domains				
Lying/sitting/ kneeling/standing	60.0 (52.5-85.0)	77.0	65.0 (55.0-90.0)	73.4
Functions of the legs	60.0 (51.1-86.7)	77.3	60.0 (45.0-85.0)	75.8
Functions of the arms	65.0 (60.0-85.0)	75.2	70.0 (50.0-90.0)	73.8
Use of transport	66.7 (53.3-100)	66.8	80.0 (60.0-100)	65.1
Self-care	88.0 (60.0-100)	54.5	80.0 (60.0-100)	53.3
Household tasks	70.0 (56.7-100)	64.1	70.0 (60.0-100)	64.4
Leisure activities and sports	61.4 (54.3-88.6)	76.1	60.0 (53.3-80.0)	78.3
Components				
Upper extremity	77.8 (60.0-93.3)	69.0	75.0 (60.0-90.0)	68.8
Basic lower extremity	63.3 (56.7-93.3)	70.6	60.0 (50.0-90.0)	68.4
Complex lower extremity	60.0 (42.5-80.0)	81.8	60.0 (45.0-80.0)	79.1
Sum				
Sum score	65.0 (55.7-88.8)	76.0	63.3 (54.4-86.7)	78.3

Note 1: Only complete cases within each domain were included in the comparison of the HAL scores with HAL_{short} scores.
Note 2: Domain and component scores for the HAL_{short} are for comparison purpose in the developmental stage of the HAL_{short} only. Due to the low number of items in some domains and the results of a secondary exploratory factor analysis, only the sum score should be used for the HAL_{short}.

Table 5. A priori defined hypotheses to determine the construct validity of the HAL_{short}

Hypotheses – construct validity HAL vs. HAL_{short}	**Confirmed**
r $HAL_{(short)}$ – SF36v2 Physical health > r $HAL_{(short)}$ – SF36v2 Mental health	V
r $HAL_{(short)}$ basic lower extremity – SF36v2 physical functioning ≥ 0.6	V
r $HAL_{(short)}$ complex lower extremity – SF36v2 physical functioning ≥ 0.6	V
r domain leisure activities and sports $HAL_{(short)}$ – SF36v2 role physical ≥ 0.5	V
$HAL_{(short)}$ basic lower extremity scores are inferior to complex lower extremity scores	V
r $HAL_{(short)}$ basic lower extremity – EQ-5D-5L mobility ≥ -0.6	V
r $HAL_{(short)}$ complex lower extremity – EQ-5D-5L mobility ≥ -0.6	V
r $HAL_{(short)}$ domain household tasks – EQ-5D-5L usual activities ≥ -0.5	V
r $HAL_{(short)}$ domain leisure activities and sports – EQ-5D-5L usual activities ≥ -0.5	V

r = correlation
HAL: Haemophilia Activities List

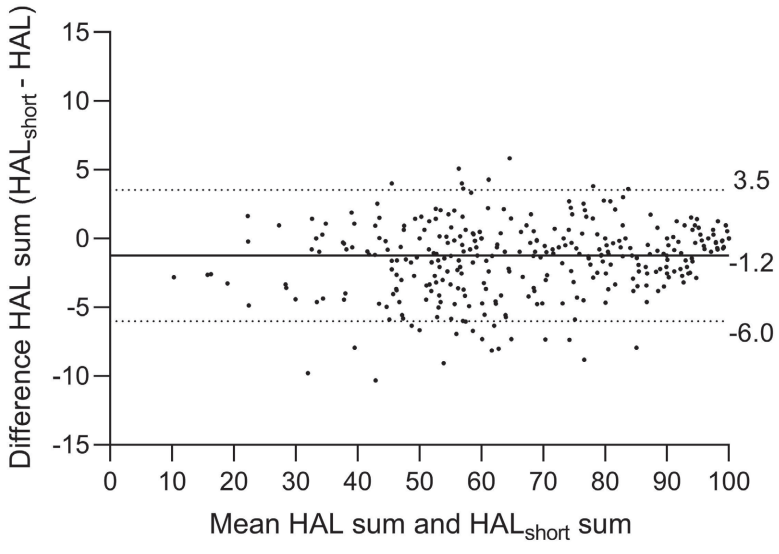

Figure 1. Bland Altman plot for the HAL and HAL_{short} scores in the derivation dataset

Table 6. Spearman correlations between HAL or HAL_{short} vs. SF-36v2 and EQ-5D-5L for the validation dataset

		SF36 PCS	SF36 MCS	SF36 PF	SF36 RP	EQ5D mobility	EQ5D usual activities
Validation data	**HAL / HAL_{short} sum**	0.77 / 0.77	0.32 / 0.32				
	HAL / HAL_{short} HOUSEH						-0.65 /-0.60
	HAL / HAL_{short} LEISPO				0.59 / 0.59		-0.60 /-0.55
	HAL / HAL_{short} LOWBAS			0.71 / 0.69		-0.74 /-0.73	
	HAL / HAL_{short} LOWCOMP			0.76 / 0.74		-0.65 /-0.66	

HAL: Haemophilia Activities List; HOUSEH: household tasks; LEISPO: leisure activities and sports; LOWBAS: basic lower extremity; LOWCOMP: complex lower extremity; SF36 PCS: SF-36 Physical component score; SF36 MCS: SF-36 Mental component score; SF36 PF: SF-36 Physical functioning; SF36 RP: SF-36 Role physical.

DISCUSSION

This study analysed HAL data in PWH with the aim of reducing the 42-item HAL questionnaire. A stepwise approach resulted in a HAL_{short} of 18 items. The items of the HAL_{short} represented all domains of the original HAL. Differences between the original HAL and HAL_{short} sum score were small (LoA: -1.3±4.7). The construct validity of the HAL and HAL_{short} was good as compared to the SF-36 physical health summary score and physical functioning domain and EQ-5D mobility and usual activities.

Internal and external validity

Data of the P-FIQ study were collected in PWH with a history of joint pain or bleeding and the B-HERO-S study was an online survey. Therefore, the data may not be representative for the entire US population. The HAL scores in the current data (median HAL sum: 65.0) were comparable to HAL scores in PWH from Jamaica (median: 66.1) and Brazil (weighted mean: 66.4), but lower than HAL scores in PWH from the United Kingdom (median: 80) and the Netherlands (median: 96) [18–21]. In Sweden, PWH with a later onset of treatment showed a median HAL sum score of 56, compared to a median of 98 in PWH with early treatment [22]. Therefore, the HAL_{short} may still include some items with ceiling effects when used in populations with less limitations in activities and participation.

In addition, some items in the domains 'self-care', 'household tasks' and 'leisure activities and sports' have been reported as inappropriate in Jamaican and Indian studies [5,18]. After the stepwise procedure some of these culturally dependent items were removed, while others were still included in the HAL_{short} (playing games, sports, putting on socks and shoes, going on a holiday active) because the items were appropriate for most populations [2,19,23–25]. Based on cross-cultural validation studies and the current study population, the HAL_{short} includes most relevant and informative items for PWH with access to intensive treatment [2,19,23,24]. However, as outcome monitoring will most likely be performed in patients with access to intensive treatment, the external validity of these findings is expected to be high.

For two items with a high inter-item correlation rephrasing of the question may be considered. The items 'walking upstairs' and 'walking downstairs' had a high inter-item correlation of 0.86. 'Walking upstairs' was reported by the participants as being slightly more difficult. As both items are about walking stairs, the descriptor 'walking stairs'

may better capture the activity than choosing one of the two activities. They will be combined into a new item 'walking stairs' for the HAL_{short}. For calculating the HAL_{short} from the original HAL, the worst score reported on walking upstairs or downstairs should be scored as abnormal for the new item 'walking stairs'.

Like the HAL, the HAL_{short} suffers from the limitations of Classical Test Theory. The HAL_{short} sum score (0-100) is a sum of the ordinal items and not corrected for the difficulty of the separate items. For example, scoring 'impossible' on an easy item like 'sitting down' has the same weight for the sum score as scoring 'impossible' on a more difficult item like 'running'.

When comparing the HAL and HAL_{short}, the domain-, component- and sum scores were considered to be similar despite significant p-values, as the variation was well below the smallest detectable change of 10.2. The statistical significance of these small differences may be attributed to the large sample size [9]. Only for the domain 'use of transport' scores of the HAL and HAL_{short} differed, because the most difficult item 'cycling' was removed as a result of a high number of 'N/A' responses.

Finally, the internal consistency of the HAL_{short} (Cronbach's $\alpha = 0.96$) is still higher than the recommended Cronbach's α between 0.7 and 0.9. As the internal consistency improved after reduction of the 24 items and there was no clear indication for removing additional items, it was decided to retain the remaining items.

Comparison with other studies

In contrast to the strong correlations between the HAL (domains) and the SF-36 physical health summary score and physical functioning domains observed in the present study, a recent systematic review reported conflicting evidence for construct validity of the HAL [26]. For example, the HAL correlated strongly with the Impact on Participation and Autonomy questionnaire and Arthritis Impact Measurement Scale which was reported in three studies, but correlated only moderately with the SF-36 domain of physical functioning, reported in one study [26]. The correlations in the current paper may be higher because the score distributions were better than in the Dutch study which had high scores on both the HAL and SF-36. The ceiling effects in some populations will potentially affected the convergent validity of the HAL_{short}.

Clinical implications and future research

Within a context of multiple outcomes assessments in haemophilia care, a shorter assessment and an easier way to quantify limitations in activities and participation is desirable. The shorter version of the HAL includes the most relevant and informative items for PWH in Western countries. However, before introduction of the HAL_{short} construct validity and reliability of the questionnaire should be established in diverse populations. The HAL_{short} can be derived from the original HAL, which allows for longitudinal studies that use the HAL to switch to the HAL_{short}. Only the sum score should be used for the HAL_{short}, since some domains only have one or two items in the HAL_{short}.

CONCLUSION

This clinimetric study resulted in a 52% reduction of the number of items in the HAL following a stepwise procedure of removing items. The short version of the HAL (18 items) is expected to capture the most relevant and informative items on activities and participation for PWH, represent all domains of the original HAL and result in similar proportions of abnormal sum scores.

REFERENCES

1 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: 565–571.

2 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: 36–46.

3 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.

4 World Health Organization. International Classification of Functioning, Disability and Health: Children & Youth Version: ICF-CY *2007*.

5 Poonnoose PM, Thomas R, Keshava SN, et al. Psychometric analysis of the Functional Independence Score in Haemophilia (FISH). Haemophilia *2007*; 13: 620–626.

6 Kempton CL, Recht M, Neff A, et al. Impact of pain and functional impairment in US adults with haemophilia: Patient-reported outcomes and musculoskeletal evaluation in the pain, functional impairment and quality of life (P-FiQ) study. Haemophilia *2018*; 24: 261–270.

7 Buckner TW, Witkop M, Guelcher C, et al. Impact of hemophilia B on quality of life in affected men, women, and caregivers-Assessment of patient-reported outcomes in the B-HERO-S study. Eur J Haematol *2018* 592–602.

8 Mokkink LB, Terwee CB, Patrick DL, et al. COSMIN checklist manual *2012*.

9 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: 108–12.

10 Ware JE, Snow KK, Kosinski M, Gandek B. SF-36 Physical and Mental Summary Scales: A User's Manual. *1993*.

11 EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. Health Policy (New York) *1990*; 16: 199–208.

12 Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res *2011*; 20: 1727–1736.

13 De Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine: A practical guide. New York: Cambridge University Press; *2011*.

14 Kuijlaars IAR, van der Net J, Bouskill V, et al. Shortening the paediatric Haemophilia Activities List (pedHAL) based on pooled data from international studies. Haemophilia *2021;* 27: 305-313.

15 Bland J, Altman D. Measuring agreement in method comparison studies. Stat Methods Med Res *1999*; 8: 135–160.

16 Batt K, Recht M, Cooper DL, Iyer NN. Construct validity of patient-reported outcome instruments in us adult people with hemophilia (PWH): Results from the pain, functional impairment, and quality of life (P-FIQ) study. Value Heal *2017*; 19: A93.

17 Schober P, Schwarte LA. Correlation coefficients: Appropriate use and interpretation. Anesth Analg *2018*; 126: 1763–1768.

18 Wharfe G, Buchner-Daley L, Gibson T, et al. The Jamaican Haemophilia Registry: Describing the burden of disease. Haemophilia *2018;* 24: e179-e186.

19 Ramos AAT, Wolff ÁLP, Lorenzato CS, et al. Translation, validation and reliability of the functional capacity questionnaire Haemophilia Activities List for haemophilia patients in Brazil. Haemophilia *2019*; 25: e231–239.

20 McLaughlin P, Morris R, Chowdary P. Investigating the relationship between the HJHS and HAL in routine clinical practice: A retrospective review. Haemophilia *2018*; 24: 988-994.

21 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: e80–86.

22 Brodin E, Hadzibajramovic E, Baghaei F, Sunnerhagen KS, Lundgren Nilsson A. Self-reported activity of swedish persons with haemophilia: Change over 2.5 years. J Rehabil Med *2018*; 50: 643–651.

23 Brodin E, Baghaei F, Elfvinger P, Lindvall K, Sunnerhagen KS. The Swedish version of the Haemophilia Activity List. Haemophilia *2011*; 17: 662–668.

24 Balestri E, Villafañe JH, Bertozzi L, et al. Validation of the Italian Version of the Haemophilia Activities List. Acta Haematol *2016*; 136: 152–156.

25 Srivastava A, Santagostino E, Dougall A, et al. WFH Guidelines for the Management of Hemophilia, 3rd edition. Haemophilia *2020*; 26: 1–158.

26 Timmer MA, Gouw SC, Feldman BM, et al. Measuring activities and participation in persons with haemophilia: A systematic review of commonly used instruments. Haemophilia *2017*; 24(2): e33-e49.

SUPPLEMEMTARY MATERIAL

Stepwise process to identify non-informative items

Non-informative items were identified in a stepwise process (7 steps) according to the method of de Vet et al. (2011), from the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative [12]. After each step non-informative items were deleted, before proceeding with the following step.

1. Floor and ceiling effects: Per item the proportions of each scoring option ('impossible', 'always', 'mostly', 'sometimes', 'rarely' or 'never') was determined, excluding the missing and 'N/A' scored questions. Proportions of minimum ('impossible') and maximum ('never' problems) scores were analyzed to detect floor and ceiling effects. Items with ≥85% minimum or maximum scores were removed.
2. Missing data and scores with 'N/A': The missing data and scores with 'N/A' were examined. The authors removed items which were scored >15% as 'missing' or 'N/A'.
3. Inter-item correlations (1): *Inter-item* correlations were evaluated. Inter-item correlations calculated with Spearman's rho <0.2 indicated items which do not correlate with any of the others and >0.9 indicated item redundancy. This step is a preparation for the factor analysis in which items with inter-item correlations of <0.2 and >0.9 were not included.
4 Component loadings of the exploratory factor analysis: Component loadings on exploratory factor analysis were evaluated. Items were analyzed on categorical level. Items with factor loadings <0.5 were removed. Model fit was evaluated with the Root Mean Square Error of Approximation (RMSEA); <0.08 indicates moderate model fit and <0.05 indicates good model fit.
5. Inter-item correlations (2): *Inter-item* correlations were evaluated for the second time. Inter-item correlations calculated with Spearman's rho >0.7 within one factor were indicators for item redundancy. Items which had a correlation of >0.7 were reviewed by IK, KF and JJ and one of the items was removed. When items had high inter-item correlations with more than one item, the authors considered which item was most relevant based on low proportions of 'N/A' responses, more limitations reported by PWH and more variation within the HAL domains according the ICF.
6. Internal consistency: Internal consistency calculated with Cronbach's α and internal consistency after item deletion were compared. Cronbach's α should be between 0.7 and 0.9; a higher Cronbach's α after item deletion was considered a reason to eliminate an item.

7. Item-total correlations: Item-total correlations for the pedHAL total score were evaluated. Item-total correlations were calculated with Spearman's rho. Items with item-total correlations of <0.3 were removed.

Supplemental table 1. Frequency tables of all HAL items

Item	Missing and/or N/A scores (%)	Distribution of response options scored (%)					
		Impossible	Always	Mostly	Sometimes	Rarely	Never
Lying/sitting/kneeling/standing							
Sitting down (e.g. on a chair or couch)	0.7	0.2	3.1	10.8	29.3	14.9	41.7
Rising from a chair with armrests	0.2	0.2	2.4	11.0	35.1	14.8	36.5
Rising from a chair without armrests	1.0	0.7	4.3	11.8	31.5	19.0	32.7
Kneeling / squatting	0.7	7.2	11.8	12.7	29.7	14.9	23.7
Bending forward	1.4	1.0	4.1	9.7	30.4	19.6	35.3
Kneeling for a longer period of time	0.7	11.5	12.0	12.9	31.7	11.0	20.9
Squatting for a longer period of time	0.5	13.2	13.2	12.9	29.7	11.5	19.6
Standing for a longer period of time	0.5	3.1	9.3	15.1	35.6	16.3	20.6
Functions of the legs							
Walking short distances (less than 1 kilometer / 15 minutes)	0.5	1.9	6.2	9.3	33.3	13.9	35.4
Walking long distances (more than 1 kilometer / 15 minutes)	0.7	4.8	8.4	13.4	29.0	14.1	30.2
Walking on a soft surface (e.g. on the beach or through the woods)	0.2	3.3	9.1	8.6	28.9	14.6	35.6
Walking on an uneven surface (e.g. cobblestones, high sidewalks)	0.5	2.9	11.0	11.2	31.6	15.1	28.2
Strolling / (window-) shopping	0.2	2.9	5.3	9.5	31.0	17.9	33.4
Climbing up the stairs	0.5	2.2	12.4	13.4	31.1	16.5	24.4

Supplemental table 1. (Continued)

Item	Missing and/or N/A scores (%)	Distribution of response options scored (%)					
		Impossible	Always	Mostly	Sometimes	Rarely	Never
Climbing down the stairs	0.2	2.1	10.7	12.4	32.5	17.2	25.1
Running (e.g. in order to catch the bus)	0.2	17.9	12.4	11.9	26.3	11.0	20.5
Jumping	0.2	17.4	10.3	13.4	27.0	13.1	18.9
Functions of the arms							
Lifting heavy objects		6.2	6.0	17.6	33.3	18.1	18.8
Carrying heavy objects in the arms		5.5	6.7	12.9	32.9	21.2	21.0
Fine hand movements (e.g. closing buttons)	0.2	1.0	2.9	5.0	30.3	22.2	39.6
Reaching above your head (to pick something up from a high shelf)	0.2	1.2	3.3	7.9	33.7	19.6	34.4
Use of transportation							
Riding a bicycle	18.8	8.6	3.6	9.3	24.8	9.3	25.8
Getting in and out of a car	1.4	1.0	7.4	10.7	29.1	17.4	33.2
Using public transportation (bus, train, subway)	24.5	1.0	2.9	7.6	25.5	9.1	29.6
Self-care							
Drying your whole body		0.2	1.9	9.5	26.7	14.0	47.6
Putting on a shirt, sweater etc.	0.2	0.2	2.1	6.0	25.1	15.8	50.8
Putting on sock and shoes			3.3	7.9	28.1	19.8	41.0
Putting on a tie or closing the top button of a shirt		0.7	2.4	7.1	23.8	16.0	50.0
Going to the toilet	0.2		1.7	5.3	27.0	12.9	53.2
Household tasks							
Going out shopping (for food, drink etc.)	0.5	0.5	4.3	7.9	34.0	15.2	37.6
Washing the dishes, cleaning the sink	2.6	0.5	3.6	10.0	28.6	13.3	41.4
Cleaning the house	3.8	0.7	3.8	11.2	31.5	16.2	32.9
Other household tasks (ironing, making the beds)	5.0	0.2	3.3	9.1	27.0	19.8	35.8
Doing odd jobs (both in and around the house)	5.5	0.7	4.0	7.9	36.7	16.2	29.0
Gardening	17.1	1.9	3.6	13.1	31.0	9.3	24.0

Supplemental table 1. (Continued)

Item	Missing and/or N/A scores (%)	Distribution of response options scored (%)					
		Impossible	Always	Mostly	Sometimes	Rarely	Never
Leisure activities and sports							
Playing games (outdoors, e.g. with your children)	10.7	3.1	4.5	13.1	35.6	10.5	22.7
Sports	12.9	7.4	10.5	11.5	29.8	10.5	17.7
Going out (theatre / museum / movie theatre / bar)	4.5	0.5	3.3	6.9	32.1	15.6	37.6
Hobbies	2.6	1.0	3.8	7.2	32.3	18.1	35.3
Dancing	19.8	5.3	6.9	9.1	25.8	11.0	22.5
Going on a holiday (active)	14.0	1.7	3.8	8.4	30.8	13.8	27.7
Going on a holiday ("passive"; beach-/hotel holiday)	14.5	1.2	2.6	6.7	32.5	14.4	28.5

Supplemental table 2. Inter-item correlations per domain

Lying/sitting/kneeling/standing								
	1. Sitting down	2. Rising from a chair with armrests	3. Rising from a chair without armrests	4. Kneeling / squatting	5. Bending forward	6. Kneeling for a longer period of time	7. Squatting for a longer period of time	8. Standing for a longer period of time
1. Sitting down		0.77	0.70	0.43	0.69	0.33	0.31	0.50
2. Rising from a chair with armrests			0.85	0.57	0.71	0.51	0.51	0.60
3. Rising from a chair without armrests				0.68	0.68	0.61	0.61	0.68
4. Kneeling / squatting					0.59	0.81	0.81	0.70
5. Bending forward						0.52	0.50	0.61
6. Kneeling for a longer period of time							0.88	0.69
7. Squatting for a longer period of time								0.71
8. Standing for a longer period of time								

Supplemental table 2. (Continued)

Functions of the legs

	1. Walking short distances	2. Walking long distances	3. Walking on a soft surface	4. Walking on an uneven surface	5. Strolling / (window-)shopping	6. Climbing up the stairs	7. Climbing down the stairs	8. Running	9. Jumping
1. Walking short distances		0.82	0.75	0.73	0.77	0.71	0.69	0.51	0.46
2. Walking long distances			0.79	0.80	0.77	0.74	0.72	0.65	0.62
3. Walking on a soft surface				0.83	0.73	0.72	0.74	0.64	0.63
4. Walking on an uneven surface					0.76	0.74	0.76	0.68	0.66
5. Strolling / (window-)shopping						0.73	0.71	0.56	0.56
6. Climbing up the stairs							0.86	0.68	0.68
7. Climbing down the stairs								0.70	0.70
8. Running									0.91
9. Jumping									

Functions of the arms

	1. Lifting heavy objects	2. Carrying heavy objects in the arms	3. Fine hand movements	4. Reaching above your head
1. Lifting heavy objects		0.85	0.50	0.62
2. Carrying heavy objects in the arms			0.44	0.59
3. Fine hand movements				0.72
4. Reaching above your head				

Supplemental table 2. (Continued)

Self-care

	1. Drying your whole body	2. Putting on a shirt, sweater etc.	3. Putting on sock and shoes	4. Putting on a tie or closing the top ...	5. Going to the toilet
1. Drying your whole body		0.88	0.78	0.80	0.82
2. Putting on a shirt, sweater etc.			0.76	0.80	0.81
3. Putting on sock and shoes				0.70	0.71
4. Putting on a tie or closing the top ...					0.79
5. Going to the toilet					

Household tasks

	1. Going out shopping	2. Washing the dishes, cleaning the sink	3. Cleaning the house	4. Other household tasks	5. Doing odd jobs
1. Going out shopping		0.84	0.82	0.82	0.80
2. Washing the dishes, cleaning the sink			0.84	0.85	0.77
3. Cleaning the house				0.83	0.83
4. Other household tasks					0.81
5. Doing odd jobs					

Supplemental table 2. (Continued)

Leisure activities and sports						
	1. Playing games	2. Sports	3. Going out	4. Hobbies	6. Going on a holiday (active)	7. Going on a holiday (passive)
1. Playing games		0.69	0.67	0.71	0.77	0.71
2. Sports			0.52	0.57	0.59	0.55
3. Going out				0.79	0.81	0.82
4. Hobbies					0.74	0.77
6. Going on a holiday (active)						0.86
7. Going on a holiday (passive)						

Supplemental table 3. Exploratory factor analysis - step 4

	Entire body - non-ambulatory activities of daily living	Lower extremity - weight bearing	Lower extremity - ambulation	Upper extremity - weight carrying
Lying/sitting/kneeling/standing				
Sitting down (e.g. on a chair or couch)	**0.732**	0.269	0.312	0.032
Rising from a chair with armrests	**0.661**	0.541	0.281	0.013
Rising from a chair without armrests	0.535	**0.644**	0.348	0.040
Kneeling / squatting	0.234	**0.833**	0.262	0.104
Bending forward	**0.591**	0.461	0.263	0.181
Kneeling for a longer period of time	0.188	**0.902**	0.171	0.175
Squatting for a longer period of time	0.146	**0.903**	0.210	0.178
Standing for a longer period of time	0.282	**0.651**	0.455	0.195
Functions of the legs				
Walking short distances (less than 1 kilometer / 15 minutes)	0.481	0.352	**0.680**	0.046
Walking long distances (more than 1 kilometer / 15 minutes)	0.386	0.475	**0.671**	0.120
Walking on a soft surface (e.g. on the beach or through the woods)	0.329	0.489	**0.687**	0.088
Walking on an uneven surface (e.g. cobblestones, high sidewalks)	0.283	0.515	**0.697**	0.127
Strolling / (window-)shopping	0.464	0.378	**0.627**	0.185
Climbing up the stairs	0.342	0.575	**0.610**	0.130
Climbing down the stairs	0.317	0.574	**0.624**	0.121
Running (e.g. in order to catch the bus)	0.083	**0.676**	0.485	0.290
Functions of the arms				
Lifting heavy objects	0.433	0.344	0.243	**0.734**
Carrying heavy objects in the arms	0.334	0.393	0.243	**0.758**
Fine hand movements (e.g. closing buttons)	**0.774**	0.189	0.126	0.193
Reaching above your head (to pick something up from a high shelf)	**0.693**	0.294	0.179	0.300
Use of transport				
Getting in and out of a car	**0.717**	0.358	0.358	0.125
Self-care				
Drying your whole body	**0.886**	0.148	0.199	0.126

Supplemental table 3. (Continued)

	Entire body - non-ambulatory activities of daily living	Lower extremity - weight bearing	Lower extremity - ambulation	Upper extremity - weight carrying
Putting on a shirt, sweater etc.	**0.891**	0.144	0.166	0.088
Putting on sock and shoes	**0.775**	0.314	0.218	0.059
Putting on a tie or closing the top button of a shirt	**0.821**	0.142	0.130	0.134
Going to the toilet	**0.848**	0.106	0.191	0.053
Household tasks				
Going out shopping (for food, drink etc.)	**0.724**	0.217	0.508	0.144
Washing the dishes, cleaning the sink	**0.793**	0.118	0.403	0.195
Cleaning the house	**0.706**	0.225	0.455	0.286
Other household tasks (ironing, making the beds)	**0.718**	0.216	0.412	0.274
Doing odd jobs (both in and around the house)	**0.651**	0.277	0.452	0.278
Leisure activities and sports				
Playing games (outdoors, e.g. with your children)	0.446	0.421	**0.563**	0.278
Sports	0.185	0.547	**0.548**	0.331
Going out (theatre / museum / movie theatre / bar)	**0.700**	0.192	0.488	0.209
Hobbies	**0.635**	0.247	0.493	0.174
Going on a holiday (active)	0.554	0.236	**0.605**	0.318
Going on a holiday ("passive"; beach-/hotel holiday)	**0.631**	0.192	0.577	0.284

Interpretation: Four factors were identified in the factor analysis. The grey highlighted factor loadings shows to which factor the items were allocated. The highest factor loading of each factor was shown in bold.

Supplemental table 4. Item-total correlations

Item	item-total correlation, r
Lying/sitting/kneeling/standing	
Sitting down (e.g. on a chair or couch)	0.697
Rising from a chair without armrests	0.804
Bending forward	0.731
Squatting for a longer period of time	0.635
Functions of the legs	
Walking long distances (more than 1 kilometer / 15 minutes)	0.815
Walking on an uneven surface (e.g. cobblestones, high sidewalks)	0.788
Climbing up the stairs	0.795
Running (e.g. in order to catch the bus)	0.662
Functions of the arms	
Lifting heavy objects	0.687
Reaching above your head (to pick something up from a high shelf)	0.686
Use of transport	
Getting in and out of the car	0.813
Self-care	
Drying your whole body	0.674
Putting on sock and shoes	0.722
Household tasks	
Cleaning the house	0.806
Doing odd jobs (both in and around the house)	0.804
Leisure activities and sports	
Playing games (outdoors, e.g. with your children)	0.809
Sports	0.710
Going on a holiday (active)	0.805

PROMIS

CHAPTER 8

Feasibility, measurement properties and relevance of generic PROMIS item banks for patient reported outcome assessment in adult persons with hemophilia

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

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

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

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

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

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

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

Submitted

ABSTRACT

Background

Legacy haemophilia-specific outcome measures 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 persons with haemophilia (PWH).

Patients/methods

In this cross-sectional study Dutch adult PWH 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, 142/373 of invited PWH (mean age 47 [range 18-79], 49% severe haemophilia, 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 in PWH were good for 'physical function', 'pain interference', 'satisfaction with social roles and activities' and 'fatigue'. The CAT 'physical function' showed 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 PWH, with few items and low floor-/ceiling effects.

INTRODUCTION

Clinical management and therapeutic options for haemophilia have greatly improved the last decades in resource rich countries. From prophylactic clotting factor replacement therapy to prevent bleeding [1], introduced in the Netherlands in 1968, to current ongoing haemophilia gene therapy trials [2] and upcoming non-replacement 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 (PROs) are essential to evaluate these and novel interventions in individual patients and should cover the wide range of consequences of haemophilia [6].

Patient Reported Outcomes Measurement Information System (PROMIS®) is a set of universal, person centred item banks that evaluates and monitors physical, mental, and social health in adults and children [7]. Other than Classical Test Theory (CTT) based legacy instruments like the Haemophilia Activities List (HAL) and RAND-36, PROMIS item banks were developed according the Item Response Theory (IRT). An important advantage of IRT is the application of Computer Adaptive Testing (CAT), where the next item presented to the patient depends on the response to earlier items and therefore it is not necessary to answer all items of the Patient Reported Outcome Measure (PROM). 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 4-6 PROMIS physical functioning CAT items [9]. When CATs are not available or 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 persons with haemophilia (PWH) who experience an increasing life expectancy and as a result acquire more comorbidities associated with ageing [5]. Furthermore, the occurrence of floor- and ceiling effects, a frequent limitation of the HAL and SF-36, are minimalized 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 adult

PWH and were to date seldomly applied in haemophilia research projects [19–21]. Therefore, this study aimed to evaluate (1) feasibility, (2) measurement properties and (3) relevance of nine PROMIS CATs and short forms for Dutch adult PWH. We hypothesized that PROMIS CATs and SFs are feasible alternatives to legacy instruments for PWH.

MATERIALS AND METHODS

Study design and study population

This study was a cross-sectional multicenter study in three Dutch Haemophilia Treatment Centers (HTCs): 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 PWH who participated in the Hemophilia in the Netherlands-6 (HiN-6) nationwide survey study, for which PWH 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 haemophilia 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, joint surgery and other major surgeries. PWH were invited by email and received a personal link to the research website to sign online informed consent and to complete PROMIS questionnaires. PWH were informed about the project by the Netherlands Haemophilia Patient Society (NVHP). 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 Haemophilia Regimen Treatment Adherence Scale – Prophylaxis [VERITAS-Pro], Patient Activation Measure-13 [PAM-13]) were extracted from the HiN-6 study.

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

Measurements

PROMIS item banks

Nine Dutch PROMIS item banks were selected by nine members of the 'PROMIS in haemophilia 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 SD of 10. For all item banks, higher scores represent more presence of the construct (e.g., more pain interference or better physical function). The scores of the short forms were calculated in the PROMIS Assessment Center Scoring Service. All items banks cover a seven-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.

Legacy instruments from HiN-6

The HAL is a validated instrument for assessment of self-reported limitations in activities and participation in PWH. 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 only calculated if a minimum of 50% of items of a domain or component were scored on the 6-point Likert scale. The HAL utilizes a recall period of one 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 PWH [27]. In 6/8 domains patients score the items on a 3- to 6-point Likert scale and in 2/8 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 PWH. 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 utilized a recall period of four weeks, except from 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 PWH. The 24-items 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 and an optimum score of 0. The VERITAS-Pro has a recall period of three months [32]. The VERITAS-Pro was only available 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].

Patient characteristics

Patient characteristics analyzed included age at HiN-6 participation, type of haemophilia (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).

Statistical analyses

SPSS (version 25, IBM) was used for data analyses. Complete case analyses were performed in case of missing data. Patient characteristics were presented as proportions or means (standard deviation [SD]).

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 PWH 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 was not available.

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 literature [13,14,31,36–39] and expert opinion (KF, MT, MP, MC, SG, MHC, MK) and are presented in supplementary material.

To test hypotheses regarding convergent validity, correlations between PROMIS items 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, 0.7-0.89 as strong, 0.4-0.69 as moderate, 0.10-0.39 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. non-severe [mild and moderate] haemophilia 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 PROMs in individual patients [41]. This SE cut-off point deviates from the stopping rule of ≤2.2 as described for the PROMIS CATs. To assess reliability of the legacy instruments of the PWH, internal consistency estimates (Cronbach's α) were calculated.

3. Relevance

To determine which items banks were relevant to adult PWH, descriptive analyses (mean T-scores and standard deviation [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- (0.5SD), moderately (1SD), or severely (2SD) deviant. Reference data from the general Dutch male population were used to determine the score cut-off points for these six PROMIS item banks, according data from the Dutch-Flemish PROMIS National center (personal communication CB Terwee, April 29, 2021). In the general population, 84.1% scored within normal limits or the mildly deviant categories. For the PROMIS CAT 'satisfaction with participation' and PROMIS short forms 'self-efficacy medications' and 'self-efficacy symptoms' T-scores were categorized in very high (+2SD), high (+1SD), average, low (-1SD) and very low (-2SD) of the construct being measured. In absence of Dutch reference data, score cut-off values from the general US population were used to categorize these three PROMIS scores. For these short forms, 84.1% of the general population scored within the average, high or very high categories. [42].

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

RESULTS

Patient characteristics

In total, 373 adult PWH were invited to participate in the study, 162 PWH signed informed consent, but six 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 PWH were included and started to complete the PROMIS item banks resulting in a response rate of 38%. Of these 142 PWH, 133 (94%) completed all nine PROMIS item banks. Patient characteristics are shown in Table 1. The median age was 47.3 years (range 18 to 79) and 49% had severe haemophilia. One-third (34%) of the PWH reported no comorbidities. Most common reported comorbidities were hepatitis C (51%), hypertension (20%), HIV (8%), hypercholesterolemia (8%) and 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.

Table 1. Patient characteristics

Patient characteristics (n=142)	Mean (SD) or %
Age (years)	47.3 (17.1)
Haemophilia A	86.5
Severity haemophilia	
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	
1	22.5
>1	20.4

PROMIS items banks and legacy instruments

1. Feasibility

Table 2 presents data on feasibility 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 Supplementary material.

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				Reliability (SE ≤3.2)	n
	%	%	Mean	(SD)	min	max	%	
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
Ability to participate in social roles and activities	-	9.4	6.4	2.9	3	12	90.6	139
Satisfaction with social roles and activities	-	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

CATs: Computer Adaptive Tests, PROMIS: Patient Reported Outcomes Measurement Information System, SD: standard deviation, SE: standard error, self-efficacy medications: self-efficacy for managing medications and treatment, self-efficacy symptoms: self-efficacy for managing symptoms.

Floor effects were observed in two PROMIS item banks; in the CATs 'pain interference' (26% minimum scores) and 'depression' (18% minimum scores). PWH had to administer the maximum of 12 CAT items when reporting minimum scores. Ceiling effects were observed in two PROMIS item banks; in 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 the legacy instruments are shown in

the Supplementary material. 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 %).

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' correlations with the legacy instruments were strong and met the predefined criteria for convergent validity. For the PROMIS CAT 'satisfaction with participation' correlations were moderate and met the predefined criteria for convergent validity. The correlation between the PROMIS CAT 'fatigue' and the RAND-36 'energy/fatigue' 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 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' and the legacy instruments RAND-36 'social functioning', RAND-36 '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 '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 'physical function' and 'participation' (Figure 1 and Table 3). Compared to patients with severe haemophilia, patients with non-severe haemophilia had better physical function (53.0 vs. 45.0, $p<0.001$) and better ability to participate in social roles and activities (54.4 vs. 50.5, $p<0.001$). Compared to PWH ≥30 years, PWH aged 18-29 years had better physical function than (57.9 vs. 46.4, $p<0.001$) and better ability to participate in social roles and activities (56.3 vs. 51.3, $p<0.001$). For PROMIS 'physical function' the minimal important change is 2 to 8 and the differences were considered to be clinically relevant [43]. For PROMIS 'participation' data on the minimal important change were not available.

Table 3. Predefined hypotheses and results of validity testing according to PROMIS item banks show that convergent validity was confirmed for the PROMIS CATs 'physical function', 'pain interference' and 'satisfaction with social roles and activities' and known-group validity was confirmed for the PROMIS CATs 'physical function' and 'ability to participate in social roles and activities'

PROMIS item bank	Legacy instrument	Predefined correlation	Spearman's correlation	Confirmed (Yes/N)
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	-.72	Yes
Depression	RAND-36 Emotional well-being	> -0.6	-.52	N
Anxiety	RAND-36 Emotional well-being	> -0.6	-.46	N
Ability to participate in social roles and activities	RAND-36 Social functioning	> 0.6	0.39	N
	RAND-36 Role limitations due to physical health problems	> 0.6	0.44	N
	HAL complex lower extremity	> 0.6	0.44	N
Satisfaction with social roles and activities	RAND-36 Social functioning	> 0.4	0.46	Yes
Fatigue	RAND-36 Energy/fatigue	> -0.6	-.59	N
Self-efficacy medications	VERITAS-Pro Time	> -0.4	-.08	N
	VERITAS-Pro Remember	> -0.4	0.01	N
Self-efficacy symptoms	PAM-13	> 0.4	0.37	N
PROMIS item bank	**Differences between:**			
Physical function	Severe and non-severe haemophilia Young adults (18-29 years) and adults (≥30 years)			Yes
Ability to participate in social roles and activities	Severe and non-severe haemophilia Young adults (18-29 years) and adults (≥30 years)			Yes

Note: in the non-severe categories all persons with mild and moderate haemophilia were included.
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, self-efficacy medications: self-efficacy for managing medications and treatment, self-efficacy symptoms: self-efficacy for managing symptoms, VERITAS-Pro: Validated Haemophilia Regimen Treatment Adherence Scale – Prophylaxis.

The reliability varied between the different PROMIS item banks (Table 2). For all PROMIS CATs and short forms >70% of the T-scores was reliable (SE ≤3.2, 90% reliable), except for the PROMIS short form '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 Supplementary material.

3. Relevance

Table 4 presents the T-scores for the PROMIS item banks. In addition, Figure 2 and Figure 3 show the distribution of the scores according to the score cut-off values. The PROMIS CAT 'physical function' (38%) was most frequently scored as limited and adult PWH reported lower scores than the general Dutch male population for 'physical function'. For all other PROMIS item banks adult PWH scored similar or better, compared to the general population.

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 'physical function'. Minimum and maximum scores occurred equally or less frequent in the PROMIS CATs than in the legacy instruments, except for the PROMIS CAT 'depression'. Convergent validity of the PROMIS CATs 'physical function', 'pain interference', 'satisfaction with participation' and 'fatigue' was confirmed by hypothesis testing. Convergent validity of the PROMIS CATs 'depression', 'anxiety' and 'participation' was, in this study, not confirmed. For the PROMIS CAT 'participation' as well as '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 'physical function' was considered to be most relevant for adult PWH, as most limitations were reported in this domain. The PROMIS short form '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 'self-efficacy medications' was not a feasible and reliable alternative to the VERITAS-Pro and measured a different construct.

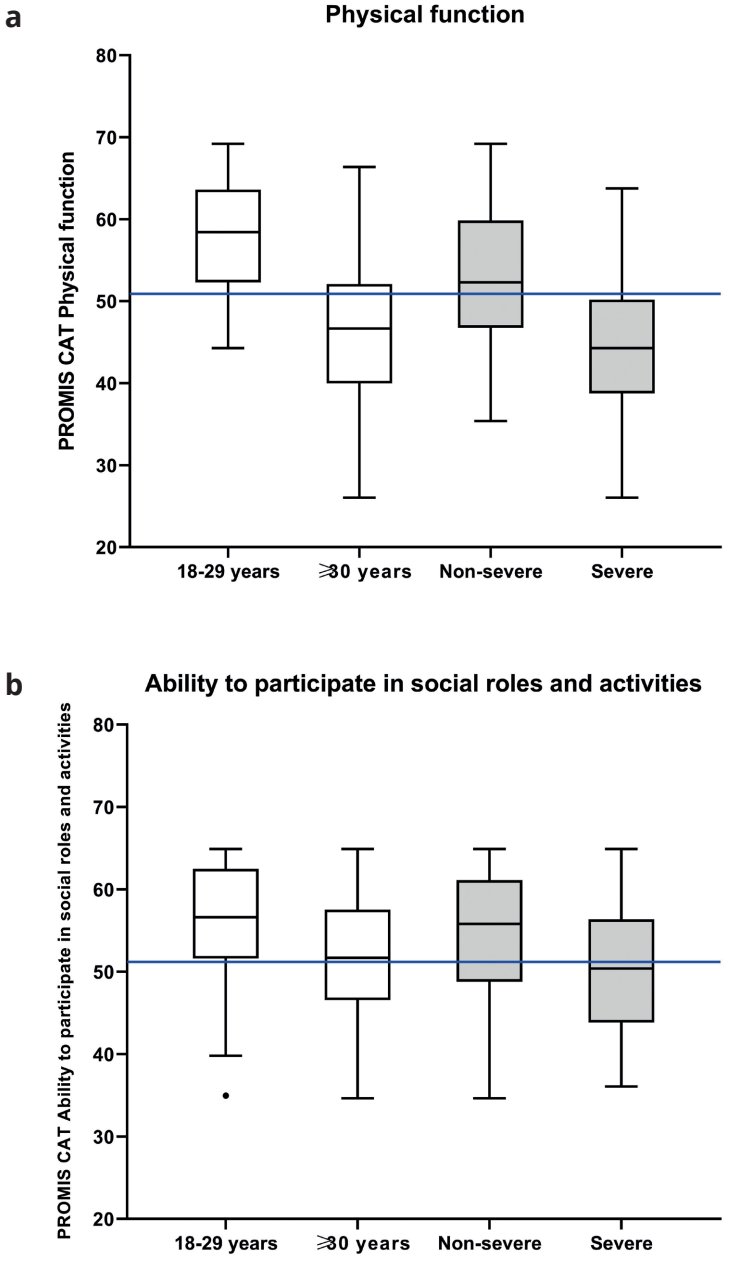

Figure 1. T-scores on the PROMIS CAT (a) 'physical function' and (b) 'ability to participate in social roles and activities' according to age and hemophilia severity.
The blue lines shows the mean score of the general adult Dutch male population on the PROMIS CAT 'physical function' (50.9) and 'ability to participate in social roles and activities' (51.2).

Table 4. T-scores on the PROMIS CATs and short forms

PROMIS item bank	PROMIS T-score				n
	Mean (SD)		min	max	
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
Ability to participate in social roles and activities	52.5	8.2	34.7	64.9	139
Satisfaction with social roles and activities	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

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 (e.g., more pain interference or better physical function).

CATs: computer adaptive tests; PROMIS: Patient Reported Outcomes Measurement Information System, SD: standard deviation, self-efficacy medications: self-efficacy for managing medications and treatment, self-efficacy symptoms: self-efficacy for managing symptoms.

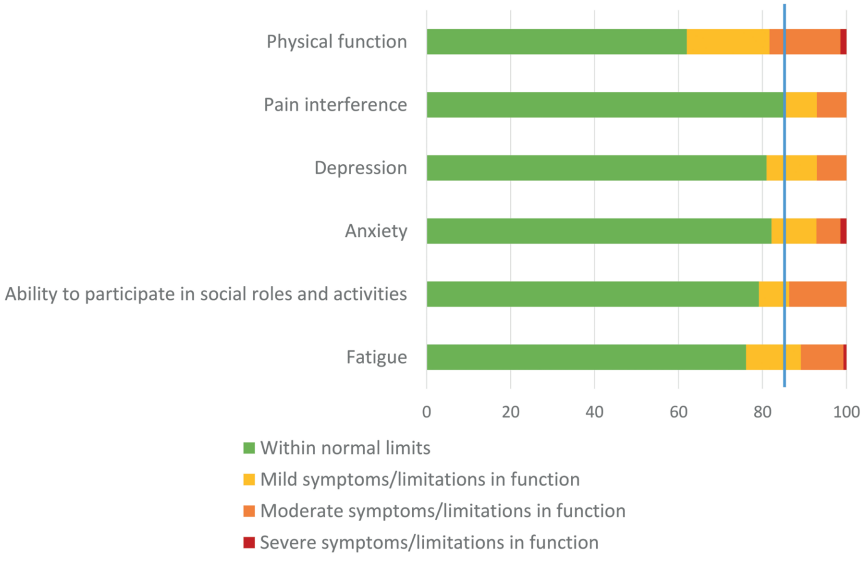

Figure 2. Scores on six PROMIS CATs
PROMIS T-scores were presented in four categories according to score cut-off points: within normal limits (green), mild (0.5SD) (yellow), moderate (1SD) (orange) and severe (2SD) (red) symptoms/limitations in function. As depicted, 84% (blue line) of the general adult Dutch male population scores within normal limits or mild symptoms.

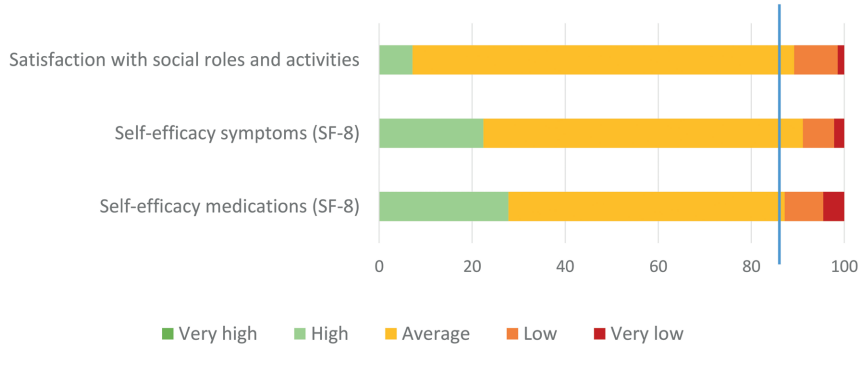

Figure 3. Scores on three PROMIS CATs and short forms
PROMIS T-scores were presented in five categories according to score cut-off points: very high (+2SD) (green), high (+1SD) (light green), average, low (-1SD) (orange), very low (-2SD) (red). As depicted, 84% (blue line) of the general US population scores within the very high, high or average categories.

Table 5. Synthesis of the results on feasibility, measurement properties and relevance of the PROMIS CATs and short forms shows that the PROMIS CATs 'physical function', 'pain interference', 'satisfaction with social roles and activities' and 'fatigue' were feasible alternatives to the legacy instruments

PROMIS item bank	Legacy instrument (n items)	Feasibility		Measurement properties			Relevance
		N items	Floor/ ceiling	Convergent validity	Known-group validity	Reliability	
Computer Adaptive Tests (CATs)							
Physical function	RAND-36 Physical functioning (10) HAL (42) HEP-test-Q (25)	6.0	+	+	+	++	+
Pain interference	RAND-36 Pain (2)	6.1	±	+	n.a.	+	-
Depression	RAND-36 Emotional well-being (5)	8.6	-	-	n.a.	+	-
Anxiety	RAND-36 Emotional well-being (5)	8.7	+	-	n.a.	+	-
Ability to participate in social roles and activities	RAND-36 Social functioning (2) RAND-36 RP (4) HAL complex lower extremity (9)	6.4	+	-	+	++	-
Satisfaction with social roles and activities	RAND-36 Social functioning (2)	5.2	+	+	n.a.	++	-
Fatigue	RAND-36 Energy/fatigue (4)	6.4	+	+	n.a.	++	-
Short forms							
Self-efficacy medications	VERITAS-Pro Time (4) VERITAS-Pro Remember (4)	8	-	-	n.a.	-	-
Self-efficacy symptoms	PAM-13 (13)	8	-	-	n.a.	+	-

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, self-efficacy medications: self-efficacy for managing medications and treatment, self-efficacy symptoms: self-efficacy for managing symptoms, VERITAS-Pro: Validated Haemophilia 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: + PWH had more limited scores than the general population, - PWH had fewer or similar limited scores than the general population

DISCUSSION

This study aimed to determine the feasibility, validity and relevance of nine PROMIS item banks in 142 adult Dutch adults with haemophilia. 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 to 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 'physical function' was more feasible than the legacy instruments and was most relevant for adult PWH. In addition, the PROMIS CATs 'pain interference', 'satisfaction with participation' and 'fatigue' were feasible alternatives to the legacy instruments. The PROMIS CAT '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 do not support the use of the PROMIS short forms on self-efficacy in adult PWH.

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 PWH aged 18-79 years with all severities included. However, 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 'self-efficacy medications' higher correlations with the VERITAS-Pro were expected, although the focus of the PROMIS short form 'self-efficacy medications' is more on confidence in managing medication schedules and the VERITAS-Pro on adherence to prophylactic treatment in haemophilia [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 PWH compared to other diseases as well as by the multifactorial character of adherence to prophylaxis [39,44]. The correlation between the PROMIS short form '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 'self-efficacy symptoms' focuses on confidence in managing symptoms, the PAM-13 has on 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 to 2.6 years) between the assessments for the HiN-6 and PROMIS studies. However, this is contradicted by the very high correlations observed for physical function. Moreover PWH with major changes in their health status were identified by an anchor question focused on physical health, and excluded. The facts that this anchor question did not cover changes in mental health and that assessment was done during COVID-19, may have affected the scores on the PROMIS CATs 'anxiety' and 'depression' [45]. However, the lack of correlation may also be attributed to the high scores, as these domains are generally less affected in haemophilia [40,46]. The lack of correlation between the PROMIS CAT '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'. This may be explained by a tendency that patients with lifelong conditions like haemophilia report higher health states than the general population, known as the 'disability paradox', suggesting the impact of haemophilia may be underestimated if general population references are used [48]. In absence of Dutch male reference data, reference data of the general US population were used for the PROMIS item banks 'satisfaction with participation', 'self-efficacy medications' and 'self-efficacy symptoms' which may have affected the results [49].

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 PWH [11–13]. The PROMIS T-scores in the current study were also comparable to T-scores in American PWH, although in American PWH higher correlations were reported between mental health domains and the EQ-5D-5L 'anxiety/depression' [21].

In the current study the PROMIS CAT 'pain interference' was limited in only 15% of the PWH, 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 PWH (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 [46,50]. Using US population references would have resulted in a score for increased pain interference in 33% of adult PWH.

The current results partly support the recommendations of the recent HaemoValue initiative. Based on expert opinion only, the core outcome set for haemophilia 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' [51].

Clinical implications and future research

Why should we use PROMIS above the legacy instruments in PWH? As 66% of the PWH 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 'physical function' is more feasible than the legacy instruments and is relevant to PWH when assessing disabilities at group level. In addition, the PROMIS CAT 'participation' should be considered for research purposes to compare PWH on group level. However, in day-to-day care for individual PWH all health domains may be of interest in a comprehensive care setting. The PROMIS CATs '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, the PROMIS short forms on self-efficacy for managing chronic

conditions are not recommended. 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' as an alternative to the RAND-36 'emotional well-being' domain (5 items, Cronbach's α = 0.87).

What work should be done before implementation of PROMIS CATs in day-to-day care and research? Several issues need to be addressed. Firstly, 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. Secondly, 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.

CONCLUSION

PROMIS CATs are feasible and may lower burden of outcome assessment by reducing the number of questions needed to assess various aspects of health compared to legacy instruments. The PROMIS CATs 'physical function', 'pain interference', 'satisfaction with participation' and 'fatigue' are feasible, reliable and valid alternatives to legacy instruments for PWH, with a low number of items and low floor- and ceiling effects. For the implementation of PROMIS CATs in haemophilia care with lifelong routine assessment, data on smallest detectable changes and minimal important changes and validation in children and young adults are essential.

REFERENCES

1 Srivastava A, Santagostino E, Dougall A, et al. WFH Guidelines for the Management of Hemophilia, 3rd edition. Haemophilia 2020; 26: 1–158.

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: 545–557.

3 Oldenburg J, Mahlangu JN, Kim B, et al. Emicizumab Prophylaxis in Hemophilia A with Inhibitors. N Engl J Med 2017; 377: 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: 1256–1264.

5 Shapiro S, Makris M. Haemophilia and ageing. Br J Haematol 2019; 184: 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 1–9.

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 (PROMISTM) Network: 2005–2008. J Clin Epidemiol 2010; 63: 1179–1194.

8 Fries JF, Witter J, Rose M, et al. Item response theory, computerized adaptive testing, and promis: Assessment of physical function. J Rheumatol 2014; 41: 153–158.

9 Kollmorgen RC, Hutyra CA, Green C, et al. 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: 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: 537-544.

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: 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: e80–86.

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: 1–9.

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: 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: 211–220.
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 1–18.
18 Crins MHP, van der Wees PJ, Klausch T, et al. Psychometric properties of the PROMIS Physical Function item bank in patients receiving physical therapy. PLoS One 2018; 13: 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: 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: 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 1–8.
22 Hassan S et al. Fifty years of hemophilia treatment in the netherlands. J Thromb Haemost 2021; accepted for publication.
23 Mokkink LB, Patrick DL, Alonso J, Bouter LM, Terwee CB. COSMIN Study Design checklist for Patient-reported outcome measurement instruments 2019.
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: 36–46.
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: 565–571.
26 Brodin E, Baghaei F, Elfvinger P, Lindvall K, Sunnerhagen KS. The Swedish version of the Haemophilia Activity List. Haemophilia 2011; 17: 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: 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: 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: 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: 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: 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: 1918–1930.
34 Rademakers J, Maindal HT, Steinsbekk A, et al. 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: 1–7.
35 McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res 1995; 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: 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 2019; 246: 437–443.
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: 1–9.
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; 3:14.
40 Schober P, Schwarte LA. Correlation coefficients: Appropriate use and interpretation. Anesth Analg 2018; 126: 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 n.d.
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 ou. World J Orthop 2018; 9: 41–49.
44 Schrijvers LH, Kars MC, Beijlevelt-van der Zande M, et al. Unravelling adherence to prophylaxis in haemophilia: A patients' perspective. Haemophilia 2015; 21: 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: 531–542.
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: 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: 2851–2861.
48 O'Hara J, Martin AP, Nugent D, et al. Evidence of a disability paradox in patient-reported outcomes in haemophilia. Haemophilia 2021 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 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: 13.

51 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 2021 1–13.

SUPPLEMENTARY MATERIAL

Supplemental table 1. Predefined hypotheses of validity testing according to PROMIS item banks

PROMIS item bank	Legacy instrument	Predefined correlation
Physical function	RAND-36 Physical functioning	> 0.6
	HAL	> 0.4
	HEP-test-Q	> 0.6
Pain interference	RAND-36 Pain	> -0.6
Depression	RAND-36 Emotional well-being	> -0.6
Anxiety	RAND-36 Emotional well-being	> -0.6
Ability to participate in social roles and activities	RAND-36 Social functioning	> 0.6
	RAND-36 Role limitations due to physical health problems	> 0.6
	HAL complex lower extremity	> 0.6
Satisfaction with social roles and activities	RAND-36 Social functioning	> 0.4
Fatigue	RAND-36 Energy/fatigue	> -0.6
Self-efficacy medications	VERITAS-Pro Time	> -0.4
	VERITAS-Pro Remember	> -0.4
Self-efficacy symptoms	PAM-13	> 0.4
PROMIS item bank	**Differences between:**	
Physical function	Severe and non-severe haemophilia Young adults (18-29 years) and adults (≥30 years)	
Ability to Participate in Social Roles and Activities	Severe and non-severe haemophilia Young adults (18-29 years) and adults (≥30 years)	

Note: in the non-severe categories all persons with mild and moderate haemophilia were included.
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, self-efficacy medications: self-efficacy for managing medications and treatment, self-efficacy symptoms: self-efficacy for managing symptoms, VERITAS-Pro: Validated Haemophilia Regimen Treatment Adherence Scale – Prophylaxis.

Supplemental table 2. Number of items, floor- and ceiling effects and internal consistency of the legacy instruments

Legacy instrument	n items	Floor	Ceiling	Cronbach's α	n
		n (%)	n (%)		
RAND-36 PF	10	1 (0.7)	35 (25.5)	0.93	137
RAND-36 SF	2	1 (0.8)	63 (49.2)	0.88	128
RAND-36 EW	5	-	3 (2.4)	0.87	126
RAND-36 Pain	2	-	36 (28.3)	0.81	127
RAND-36 RP	4	17 (13.0)	84 (64.1)	0.89	131
RAND-36 Energy/fatigue	4	1 (0.8)	2 (1.6)	0.76	126
PAM-13	13	1 (0.8)	4 (3.0)	0.80	133
VERITAS-Pro sum	24	2 (4.1)	-	0.89	49
VERITAS-Pro Time	4	11 (19.6)	1 (1.8)	0.88	56
VERITAS-Pro Remember	4	12 (21.8)	-	0.86	55
HAL sum	42	-	29 (21.8)	0.97	133
HAL LOWCOMP	9	1 (0.7)	42 (31.3)	0.97	135
HEP-test-Q	25	-	1 (0.7)	0.95	135

HAL: Haemophilia Activities List, HEP-test-Q: Haemophilia & Exercise Project-Test-Questionnaire, LOWCOMP: lower extremity complex, EW: emotional well-being, PAM-13: Patient Activation Measure-13, PF: physical functioning, RP: role limitations due to physical health problems, SF: social functioning, VERITAS-Pro: Validated Haemophilia Regimen Treatment Adherence Scale – Prophylaxis.

GENERAL DISCUSSION AND SUMMARY

CHAPTER 9

General discussion and summary

The general aim of this thesis was to optimize outcome assessment of functions, activities and participation in persons with haemophilia (PWH) through (1) improving the interpretation of legacy instruments and (2) reducing the time-investment of completing outcome assessment for PWH, caregivers and researchers by shortening legacy instruments and exploring the advantages of Patient Reported Outcomes Measurement Information System (PROMIS®).

Nowadays many haemophilia-specific and general tools are available, covering many health domains. The problem in clinical care and research is not the question whether an outcome measure exists for a certain domain, but more how to select one and use the outcome measure, while considering all relevant aspects. To describe all relevant aspects of the (1) Hemophilia Joint Health Score (HJHS), (2) paediatic Haemophilia Activities List (pedHAL) and Haemophilia Activities List (HAL) and (3) PROMIS® we used the 'Framework Clinimetrics' from the Royal Dutch Society for Physical Therapy (KNGF) which includes [1]:

1. Health domain of interest
2. Goal of outcome assessment
3. Choice of outcome measure
4. Availability
5. Feasibility
6. Measurement properties
7. Reference values
8. Calculation and interpretation

Within haemophilia care a broad range of generic and haemophilia-specific outcome measures are recommended to evaluate the impact of haemophilia on functions, activities and participation, as shown in the *Introduction – Figure 1*. Several initiatives, including different settings and goals, have provided reviews on tools to use for outcome assessment in haemophilia care and research [2–4].

As a next step to improve outcome assessment in haemophilia care, the results of this thesis focus on *(5) feasibility and (6)* measurement properties. Within a context of routine outcome assessment, feasibility is crucial. Especially within the constraints of a clinical practice, it may not take too much time to administer questionnaires or to do a test. Preferably, completing a questionnaire or test takes maximum ten minutes

[1]. Time was mentioned as a barrier for clinical use of outcome measures in general as well as within the field of haemophilia [2,5,6]. For the HJHS, international experts reported assessment time (30-60 minutes per patient) as an important disadvantage of the tool, especially in a busy clinical setting [7]. After introduction of the pedHAL (53 items) and HAL (42 items), clinicians and researchers reported some domains/items to be non-informative [8–10]. Furthermore, there is a significant gap in the evidence on measurement properties for traditionally used legacy instruments on joint health, activities and participation, including the reliability and responsiveness of the HJHS and HAL [11,12].

To solve issues like lengthy questionnaires and limited evidence on measurement properties, generic PROMIS item banks were recently introduced in the field of haemophilia [3,13,14]. In contrast to the disease-specific HJHS, pedHAL and HAL, generic outcome measures have the advantage that patients do not have to complete different questionnaires for every comorbidity. This is more and more relevant to adult PWH with an increasing life expectancy and more comorbidities of ageing [15].

The next paragraphs present the results of this thesis and describe all relevant aspects of the HJHS, pedHAL, HAL and PROMIS for clinical care and research purposes.

1. HJHS

Health domain of interest: The HJHS assesses joint structure and function in PWH [16].
Goal of outcome assessment: The HJHS is recommended to evaluate joint health in research and clinical care for children as well as adults in non-acute bleeding joints [2,17]. Although the HJHS was developed to detect early joint damage in boys (4-18 years) with haemophilia [18], current evidence showed that ultrasound is more sensitive to early joint damage [19]. The HJHS is not a direct measure of osteochondral changes but should be used as a structured clinical examination of the joints in PWH.
Choice of outcome measure: The HJHS is a physical examination of six joints, combined with a performance test assessing global gait.
Availability: The HJHS and instructions are available free of charge at the website of the International Prophylaxis Study Group (IPSG) (www.ipsg.ca). Training is recommended before use of the HJHS.
Feasibility: HJHS assessment is currently recommended every 1-2 years in children and adults [2]. As the assessment is time-consuming (30-60 minutes) [2,17], the optimal frequency should be established while considering factors like age, treatment regimen,

joint status and bleeding rates. In *Chapter 2* we investigated the long-term changes in joint health to determine the optimal frequency of HJHS assessment in adult PWH. Instead of annual joint assessment, we recommend to monitor all six joints every five years in adult PWH on long-term prophylaxis with low bleeding rates. Single joints with multiple bleeding episodes and/or the presence of synovitis should be assessed more frequently.

In addition, in *Chapter 3* we took the first step to develop a shorter version of the HJHS by combining real-life HJHS data of 499 PWH and expert opinion. The items 'duration swelling' and 'crepitus' were identified as candidates for item reduction, which resulted in the $HJHS_{short}$. The resulting $HJHS_{short}$ was able to discriminate between different ages and treatment regimens. Dropping the items 'duration swelling' and 'crepitus' will not lead to a substantial gain in time and the experts suggested a screening examination to select joints which needs full HJHS assessment. The screening examination suggested by the experts should be further explored, and was not included in the current data.

Measurement properties: A recent systematic review reported on measurement properties of the HJHS in children and adults [11]. According the review, studies in children have reported conflicting evidence regarding construct validity. The HJHS correlated well with Haemophilia Early Arthropathy Detection with Ultrasound (HEAD-US) scores in Dutch boys but not in in Spanish boys [19,20]. In addition, the HJHS discriminates well between age, inhibitor status, severity, (onset of) prophylaxis and HEAD-US abnormalities [16,19,21–24], but in children with access to primary prophylaxis or treated on demand only the HJHS did not discriminate between severities [25,26]. In contrast, in adults the HJHS discriminated well between different prophylactic regimens and the presence of synovitis [27–29]. Test-retest and interrater reliability were good in children and young adults [30–32], but was not studied in adults [11]. The HJHS was responsive to a short rehabilitation program in Romanian children and radiosynovectomy in Czech children and young adults but not to a Nordic walking program in Swedish adults [33–35]. As these studies were in small samples (<20), responsiveness is not formally established [11].

Reference values: Although the HJHS was developed and validated in children with haemophilia, reference values are only available in healthy young adults and adults. In healthy young adults, HJHS total scores range from 0 to 3 [36]. The reference range of the HJHS total score in healthy adult males aged <50 extends up to 12 [37].

Calculation and interpretation: A paper-based summary sheet is available to calculate the joint scores and total score. Scores range from 0 to 20 per joint and the global gait

score ranges from 0 to 4, resulting in a total HJHS score from 0 to 124 points [17]. A higher score indicates worse joint health.

pedHAL and HAL

Health domain of interest: The pedHAL and HAL assess limitations in activities and participation in children (4-18 years) and adults (≥18 years) with haemophilia [38,39].
Goal of outcome assessment: Both the pedHAL and HAL are recommended for evaluation of self-reported limitations in activities and participation in PWH in research and clinical care [2].
Choice of outcome measure: The pedHAL an HAL are questionnaires which can be administered on paper or electronically.
Availability: The pedHAL and HAL are available free of charge for clinical use and research by non-commercial parties at the Van Creveldkliniek website in 46 languages (www.vancreveldkliniek.nl).
Feasibility: Assessment of pedHAL or HAL is currently recommended in children and adults with an interval of ≥1 year, based on expert opinion [2]. In *Chapter 4*, we explored the optimal frequency of administering the pedHAL in data from routine clinical practice in Dutch parents and children with haemophilia. This study showed that Dutch children on early prophylaxis and their parents reported almost no limitations in activities and participation. At group level, the sum scores and domain scores remained stable during three and/or five years follow-up. At patient level, in patients without limitations in activities and participation (pedHAL sum score >95) and without joint and/or muscle bleeds, pedHAL scores remained high until the next assessment after median (IQR) 1.0 (0.9-1.2) years. The results suggested that annual pedHAL assessment in clinical care is not necessary in these children. In children with lower pedHAL scores and/or bleeds, annual pedHAL assessment is recommended to monitor limitations in activities and participation. In addition, the study identified ceiling effects in the domains 'functions of the arms', 'use of transportation', 'self-care' and 'household activities' and identified these as candidates for shortening the pedHAL.
In *Chapter 5* we studied international pedHAL data in 315 children with haemophilia with the aim to shorten the 53-item pedHAL questionnaire. A stepwise approach resulted in a $pedHAL_{short}$ of 22 items. The $pedHAL_{short}$ was expected to retain the most relevant and informative items on activities and participation for children with haemophilia, representing all domains of the original pedHAL.
In *Chapter 7*, a similar stepwise procedure was performed in 420 adult PWH to shorten the 42-item HAL. This resulted in an 18-item HAL_{short}, with comparable construct validity

as the original HAL. For both the $pedHAL_{short}$ and HAL_{short} only a sum score should be used. In addition, both short versions can be derived from the original HAL, which allows for use in longitudinal studies. The item reduction of >50% for both questionnaires reduces the burden for completing the questionnaires substantially.

Measurement properties: In a recent study, measurement properties of the pedHAL and HAL were reviewed [12]. The pedHAL and HAL had good content validity as their items reflect daily activities which were based on interviews with PWH. Evidence on construct validity of the pedHAL was conflicting [12]. The HAL was able to discriminate between patients on intensive and less intensive prophylaxis but not between patients who stopped or continued prophylaxis [27,40]. The pedHAL showed good test-retest reliability, although a substantial variability was reported in Romanian children [34]. For the HAL, information on reliability including test-retest reliability was lacking. In addition, information on interpretability of scores was lacking for both the pedHAL and HAL [12]. Data on the smallest detectable change (SDC) and minimally important change are necessary to interpret the scores in clinical practice and research [12]. For the HAL, we investigated the test-retest reliability and the smallest detectable change in 50 adult PWH in *Chapter 6*. The HAL showed good reliability for the sum and component scores (intraclass correlation coefficient >0.9). Average SDC values, which signify a true change in score that is not due to the measurement error, are 10.2 for the normalized HAL sum score, 9.2 for the upper extremity component score, 16.7 for the basic lower extremity component score and 13.4 for the complex lower extremity component score. In addition to the measurement properties on reliability, validity and responsiveness, patient-parent agreement is of interest for the pedHAL. In *Chapter 4* we studied child-parent agreement of the pedHAL in data from routine clinical practice in 63 Dutch parents and children with haemophilia. Child-parent agreement varied across pedHAL domains from 71% agreement for 'functions of the legs' up to 92% agreement for 'self-care'. The differences indicated that both child report and parent proxy should be reported.

Reference values: Reference values are not available for the pedHAL and HAL.

Calculation and interpretation: Electronic scoring sheets are available at the Van Creveldkliniek website. Normalized scores range from 0 to 100, where 100 represents no limitations in activities [39].

PROMIS

Health domain of interest: PROMIS is a set of universal, person centred item banks that evaluates and monitors physical, mental, and social health in adults and children [41].
Goal of outcome assessment: PROMIS item banks focus on measuring universally relevant domains of health to allow diagnostic assessments across diseases, clinical settings and research [42].
Choice of outcome measure: PROMIS measures are available as full item bank, short form and Computer Adaptive Tests (CATs) (www.dutchflemishpromis.nl, www.healthmeasures.net). In CATs the selection of the next item depends on the response on the earlier items and it is not necessary to answer all items.
Availability: PROMIS item banks and short forms are available free of charge in many languages. There are costs for the use of PROMIS CATs.
Feasibility: Until now, PROMIS item banks were seldomly used in haemophilia research projects [14,43,44]. In *Chapter 8* we investigated the feasibility of seven generic PROMIS CATs and two short forms in 142 Dutch adult PWH. The PROMIS CATs were considered to be feasible with a low number of items (mean number of CAT items: 5 to 9) and low floor- and ceiling effects. Especially the PROMIS CAT 'physical function' was more feasible than the legacy instruments and was the most relevant health domain for PWH. The PROMIS short forms with 8 items on 'self-efficacy for managing medications and treatment' and 'self-efficacy for managing symptoms' showed ceiling effects.
Measurement properties: PROMIS item banks have been validated in the general population and numerous paediatric and adult patient populations [45–49]. In PWH, only the PROMIS-29 Profile Instrument, representing seven health domains, was validated and considered to be a potentially valuable tool to study the impact of haemophilia [14]. PROMIS CATs and short forms have not yet been validated in adult PWH. In *Chapter 8* we investigated the measurement properties of seven generic PROMIS CATs and two short forms in Dutch adult PWH. The PROMIS CATs 'physical function', 'pain interference', 'satisfaction with participation' and 'fatigue' correlated well with the legacy instruments, but the PROMIS CATs 'depression', 'anxiety' and 'participation' did not meet the predefined correlation criteria. The PROMIS CATs 'physical function' and 'ability to participate in social roles and activities' were able discriminate between different age- and severity categories. The reliability of the PROMIS CATs and short forms was good, except for the short form on 'self-efficacy for managing medications and treatment'. The results did not support the use of PROMIS CATs 'depression' and 'anxiety' as an alternative to the RAND-36, which may have been a result of the study

design. In addition, the PROMIS short forms on self-efficacy for managing chronic conditions were not recommended.
Reference values: Reference values are available for the general US population. For some PROMIS item banks Dutch reference values are available.
Calculation and interpretation: 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 SD of 10. The scores of the short forms were calculated in the PROMIS Assessment Center Scoring Service.

Key findings regarding optimization outcome assessment

The key findings of this thesis are shown in Figure 1 and Figure 2.

In adult PWH with access to early prophylaxis, HJHS assessment is recommended every five years when there are no reported bleeds or synovitis. This interval is much longer than the recommended 1-2 years interval [2]. In addition, use of the $HJHS_{short}$ will lead to a small reduction in time. A screening examination of joints are also expected to reduce assessment time.

For the pedHAL, both child and parent proxy questionnaires should be reported. PedHAL assessment is recommended with an interval of ≥1 year [2]. However, annual pedHAL assessment has limited clinical value in patients without limitations in activities and participation and without joint and/or muscle bleeds and should be assessed less frequently in these patients. We suggest an interval of three years based on stable high scores in children on prophylaxis with low bleeding rates (Chapter 4). In addition, use of the $pedHAL_{short}$ with 22 items and HAL_{short} with 18 items will lower burden for the patient too. The HAL is considered to be a reliable self-reported outcome measure for limitations in activities and participation. Score changes of the HAL sum score greater than the SDC 10.2 indicate that the change was not a result of measurement error.

For PWH the PROMIS CATs 'physical function', 'pain interference', 'satisfaction with social roles and activities' and 'fatigue' are feasible, valid and attractive alternatives to legacy instruments. Especially the PROMIS CAT 'physical function' is more feasible than the legacy instruments and is relevant to PWH when assessing disabilities at group level.

Key findings HJHS, pedHAL and HAL

Outcome measure	Measurement properties	Original → short	Recommendations for the optimal frequency of monitoring in clinical care
HJHS	*Future research should focus on:* - *Reliability in adults* - *Responsiveness of scores*	9 → 7 items	- Adults on long-term prophylaxis and low bleeding rates: interval 5 years - Multiple bleeding episodes and/or the presence of synovitis: more frequent monitoring
pedHAL	- Both child and parent proxy should be reported	53 → 22 items	Children on intensive treatment: - PedHAL ≤95 and/or bleeds: interval 1 year - PedHAL >95, no joint and/or muscle bleeds during follow-up: no annual assessment, suggested interval 3 years
HAL	- Reliable outcome measure for limitations in activities and participation - Smallest detectable change HAL sum score: 10.2	42 → 18 items	- *Recommended interval of ≥1 year** - *Future research should focus on the optimal frequency*

*Based on expert opinion. Reference: 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.

Figure 1. Key findings of this thesis for the legacy instruments HJHS, pedHAL and HAL

Key findings PROMIS

PROMIS item bank	Legacy instrument	Findings
Physical function	HAL HEP-test-Q RAND-36	- More feasible than the legacy instruments - Good measurement properties - Discriminates well between different age- and severity categories - Most relevant health domain for PWH on group level
Pain interference **Satisfaction** **Fatigue**	RAND-36	- Feasible alternatives to the legacy instruments - Good measurement properties
Participation	RAND-36 and HAL	- Feasible and reliable - Convergent validity with the legacy instruments not confirmed - Discriminates well between different age- and severity categories
Anxiety **Depression**	RAND-36	- Feasible and reliable - Convergent validity with the legacy instruments not confirmed
Self-efficacy medications and Symptoms (SF)	VERITAS-Pro PAM-13	- Not feasible and/or reliable - Convergent validity with the legacy instruments not confirmed

Figure 2. Key findings of this thesis for the PROMIS item banks

METHODOLOGICAL AND PRACTICAL CONSIDERATIONS

Methodological issues

Consensus on terminology and definitions about measurement properties is relevant, since there is a variety in definitions in the international multidisciplinary field for health research. Therefore, we used the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) taxonomy [50]. The COSMIN taxonomy helped us to define the research purposes and methods and it helped us to interpret the results. Definitions about measurement properties should be standardized in haemophilia research to improve the interpretation of study results.

A strength of this thesis was the use of individual patient datasets to pool data from existing studies to shorten the HJHS, pedHAL and HAL and the use of the data from the Hemofilie in Nederland-6 (HiN-6) national survey to evaluate the PROMIS item banks in PWH. Secondary use of research data lowers the burden of participating in studies

for patients with rare disorders. The use of existing data from published studies was helpful to achieve sufficient sample sizes, which is a challenge in rare diseases. In our 'PROMIS in haemophilia' project the use of the HiN-6 contributed to include 142 PWH quickly with a limited time-investment for patients and caregivers. However, correlations between the PROMIS item banks and legacy instruments to test convergent validity were affected by our choice to use the best available legacy data and the extended interval (1.0 to 2.6 years) between the assessments for the HiN-6 and PROMIS studies. In my opinion, to improve the internal validity of the study we should have chosen to study only those PROMIS item banks for which the best legacy instruments were available in the HiN-6 study and for health domains expected to be stable over time and unaffected by the COVID-19 pandemic.

In the project to shorten the HJHS we have combined pooled data with expert opinion. A blended methodology was chosen since no criterion standard for the construct 'joint health' was available and consensus between key opinion leaders like HJHS developers, users and investigators was needed for adaptations of the HJHS, as well as implementation of recommended adaptations.

Practical issues

A secondary analysis of individual patient datasets requires dealing with practical issues like data transfer agreements between the centers, secured and encrypted data transfer, and anonymization. However, it took up to a year to arrange data sharing and the quality of some databases was limited. The FAIR (findability, accessibility, interoperability, and reusability) principles point the way forward for more systematic facilitation of data sharing, while respecting participants' rights [51]. For haemophilia the European Platform on Rare Diseases Registration (EU RD Platform), which aims to combine data on rare diseases across registries in the Europian Union, could be a promising initiative. To facilitate re-use of data the quality of databases is a point of attention and hospitals should provide professional support to researchers to create high quality datasets.

Another issue was the high costs of the HJHS expert meeting in 2019 for flights and hotel for the international experts. During the COVID-19 pandemic Zoom has become a popular tool for research purposes too. This could have been a feasible alternative for our in-person meeting.

Finally, to facilitate research in the Haemophilia Treatment Center broad consent for the use of retrospective data in studies is necessary to lower the burden of informed consent procedures for participants and researchers. In our center informed consent in still required for every single project with retrospective data. Ideally, a broad consent procedure for the use of retrospective data for research should be introduced for the whole university medical center.

WHAT ELSE IS NEEDED TO FURTHER IMPROVE OUTCOME ASSESSMENT IN HAEMOPHILIA?

PROMIS versus legacy instruments

Should we implement PROMIS in haemophilia care and research? For now, PROMIS is especially feasible for research purposes. PROMIS has advantages over legacy instruments like the RAND-36/SF-36 as it includes a wide spectrum of health domains and scores of CATs and short forms can be combined. In addition, PROMIS CATs will measure more precisely in the lower and upper ends of the score range while including more relevant items [52]. For example, the health domain physical function can be measured by long questionnaires such as the HAL (42 items) and the Haemophilia & Exercise Project-Test-Questionnaire (HEP-test-Q) (25 items) or very short questionnaires such as EQ-5D-5L (1 item on mobility) and PROMIS-29 (4 items on physical function) which are less precise. The PROMIS CAT with a mean of 6 items has the best of both worlds, as it selects the most relevant items for the individual patient from the full item bank with 121 items resulting in higher precision and less floor- and ceiling effects. In addition, the use of generic PROMIS item banks has the ability to consider effects of and/or comparison according to comorbidities, which is relevant as 66% of the PWH reported ≥1 comorbidity (Chapter 8). However, there are still some disease-specific health outcomes like bleeding episodes and adherence to prophylactic treatment which are not covered by generic instruments.

A successful implementation of PROMIS in daily clinical care is conditional on several issues. Firstly, the interpretation of PROMIS scores seems to be a barrier to clinicians. PROMIS T-scores are different from additive scores in legacy instruments, which were developed according the Classical Test Theory. In addition, data on SDC and minimal important change of PROMIS item banks are lacking, while these are needed to improve the interpretability of scores in a setting of routine follow-up assessment. Finally, using CATs in clinical care has the disadvantage that different items are administered during

follow-up. Therefore, we recommend making individual items and follow-up scores directly visible in a dashboard when using PROMIS in day-to-day care.

Practical issues which need attention are the limited number of data capture tools and electronic medical records that can be used to administer CATs. Good facilities for digital administration of CATs like a PROMs mobile app or routine data collection from electronic medical records are essential and will promote use of PROMs. In addition, the costs for using CATs could be a barrier for the implementation. Especially if IT facilities and budget for using CATs are limited, PROMIS SFs are an alternative.

Core outcome set

In haemophilia care broad implementation of a core outcome set should be the next step to reduce heterogeneity in research reports and improve the comparison of subjects between different populations and treatment centers [53]. Recently, two initatives developed a core outcome set with many similarities in the recommended outcome measures [2,13]. The HaemoValue initiative followed a stepwise standard procedure with patient and health care professional panels, resulting in a core outcome set with ten health outcomes for children and adults. The total HaemoValue core set is inclusive with the opportunity to expand from regular assessment to outcomes related to long-term consequences of the disease or treatment. The other initiative was based on a combination of a critical literature review plus a consensus conference between 48 health care experts. The consensus core outcome set described more practical issues like differences between various clinical settings, the differences between clinical care and research, and the optimal frequency of outcome assessment [2], which are important aspects for clinicians and researchers. Both core outcome sets will be useful for clinical care and research, with divergent recommendations regarding outcomes on life expectancy, pain and joint imaging. The legacy instruments described in this thesis are part of both outcome sets and the HaemoValue core set suggested PROMIS item banks too [13].

Future research

The development of outcome measures and evaluation of measurement properties are time consuming processes [54]. As many outcome measures are available the development of new ones is undesirable, while investment in studying measurement properties remains essential. Future research should focus on reliability of the HJHS in adult PWH as well as responsiveness and interpretability (SDC and minimal

important change) of the HJHS, pedHAL and some relevant PROMIS item banks. Data on responsiveness and interpretability are essential for evaluating the results of comprehensive care in PWH. The SDC of the HAL is already established (Chapter 6) and currently we are studying responsiveness and the minimal important change of the HAL in PWH undergoing lower extremity surgery.

Before widespread introduction of the pedHAL$_{short}$ and HAL$_{short}$, further validation is needed. The pedHAL$_{short}$ was developed in an international dataset but was not validated in other data (Chapter 5), which is the next step for the pedHAL$_{short}$. The HAL$_{short}$ was developed and validated in American PWH (Chapter 7) but needs validation in another population. Agreement between the original (ped)HAL and (ped)HAL$_{short}$ as well as convergent validity of the (ped)HAL$_{short}$ with legacy instruments (i.e. SF-36 and/or HEP-test-Q) should be assessed. In addition, the discriminative value of the short versions should be assessed.

For the HJHS, additional steps such as a screening examination are needed to achieve a substantially more time efficient HJHS assessment. A first step is to evaluate if any of the current HJHS items is predictive for a high total HJHS score in the dataset of the HJHS pooling project (Chapter 3).

Finally, lifelong monitoring is the standard in haemophilia care and the transition from outcome measures from childhood to adulthood should be investigated. On the one hand, paediatric PROMIS item banks should be validated in children with haemophilia. On the other hand, the transition from the pedHAL to HAL and PROMIS item banks from childhood to adulthood should be studied.

GENERAL CONCLUSION

Measurement properties help to select appropriate outcome measures and to interpret results in clinical care and research. This thesis contributed to the evidence on measurement properties of the legacy instruments pedHAL and HAL in PWH. Furthermore, this thesis contributed to shorter versions of the legacy instruments HJHS and (ped)HAL and to a more personalized frequency of monitoring joint health and limitations in activities and participation. Generic PROMIS CATs should be considered as a next step to lower the time-investment of outcome assessment, as these have the advantages of a low number of more relevant questions.

References

1 Swinkels R, Meerhoff G, Beekman E, Beurskens A. Raamwerk Klinimetrie voor evidence based products. KNGF 2016 27.

2 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.

3 Manco-Johnson MJ, Warren BB, Buckner TW, Funk SM, Wang M. Outcome measures in Haemophilia: Beyond ABR (Annualized Bleeding Rate). Haemophilia 2021;27 Suppl 3:87-95.

4 Dover S, Blanchette VS, Srivastava A, et al. Clinical outcomes in hemophilia: Towards development of a core set of standardized outcome measures for research. Res Pract Thromb Haemost 2020; 4: 652–658.

5 Swinkels RAHM, Van Peppen RPS, Wittink H, Custers JWH, Beurskens AJHM. Current use and barriers and facilitators for implementation of standardised measures in physical therapy in the Netherlands. BMC Musculoskelet Disord 2011; 12: 106.

6 Duncan EAS, Murray J. The barriers and facilitators to routine outcome measurement by allied health professionals in practice: A systematic review. BMC Health Serv Res 2012; 12: 96.

7 Brainstorm session, HJHS expert meeting, October 3rd 2019, Utrecht 2019.

8 Kuijlaars IAR, van der Net J, Schutgens REG, Fischer K. The Paediatric Haemophilia Activities List (pedHAL) in routine assessment: changes over time, child-parent agreement and informative domains. Haemophilia 2019; 25: 953-959.

9 Poonnoose PM, Thomas R, Keshava SN, et al. Psychometric analysis of the Functional Independence Score in Haemophilia (FISH). Haemophilia 2007; 13: 620–626.

10 Radzevič V, Raistenskis J, Ragelienė L, Kowalski IM. Relationship between physical activity and functional ability in school-aged children with hemophilia. Polish Ann Med 2013; 20: 13–18.

11 Gouw SC, Timmer MA, Srivastava A, et al. Measurement of joint health in persons with haemophilia: A systematic review of the measurement properties of haemophilia-specific instruments. Haemophilia 2019; 25: e1–10.

12 Timmer MA, Gouw SC, Feldman BM, et al. Measuring activities and participation in persons with haemophilia: A systematic review of commonly used instruments. Haemophilia 2018; 24: e33–49.

13 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 2021 1–13.

14 Barry V, Buckner TW, Lynch ME, et al. An evaluation of PROMIS health domains in adults with haemophilia: A cross-sectional study. Haemophilia 2021 1–8.

15 Shapiro S, Makris M. Haemophilia and ageing. Br J Haematol 2019; 184: 712–720.

16 Feldman BM, Funk SM, Bergstrom B-MM, et al. Validation of a new pediatric joint scoring system from the international hemophilia prophylaxis study group: Validity of the hemophilia joint health score. Arthritis Care Res 2011; 63: 223–230.

17 International Prophylaxis Study Group. Hemophilia Joint Health Score 2.1 Instruction Manual n.d.
18 Feldman BM, Funk S, Lundin B, et al. Musculoskeletal measurement tools from the International Prophylaxis Study Group (IPSG). Haemophilia 2008; 14: 162–9.
19 Foppen W, van der Schaaf IC, Fischer K. Value of routine ultrasound in detecting early joint changes in children with haemophilia using the "Haemophilia Early Arthropathy Detection with UltraSound" protocol. Haemophilia 2016; 22: 121–125.
20 Altisent C, Martorell M, Crespo A, et al. Early prophylaxis in children with severe haemophilia A: Clinical and ultrasound imaging outcomes. Haemophilia 2015; 22: 218–224.
21 Bladen M, Main E, Hubert N, et al. Factors affecting the Haemophilia Joint Health Score in children with severe haemophilia. Haemophilia 2013; 19: 626–631.
22 Oymak Y, Yildirim AT, Yaman Y, et al. The effectiveness of tools for monitoring hemophilic arthropathy. J Pediatr Hematol Oncol 2015; 37: e80-85.
23 Saulyte Trakymiene S, Ingerslev J, Rageliene L. Utility of the Haemophilia Joint Health Score in study of episodically treated boys with severe haemophilia A and B in Lithuania. Haemophilia 2010; 16: 479–486.
24 Saulyte Trakymiene S, Clausen N, Poulsen LH, Ingerslev J, Rageliene L. Progression of haemophilic arthropathy in children: a Lithuanian--Danish comparative study. Haemophilia 2013; 19: 212–218.
25 Groen WG, Takken T, van der Net J, Helders PJM, Fischer K. Habitual physical activity in Dutch children and adolescents with haemophilia. Haemophilia 2011; 17: e906-912.
26 Payal V, Sharma P, Chhangani NP, et al. Joint Health Status of Hemophilia Patients in Jodhpur Region. Indian J Hematol Blood Transfus 2015; 31: 362–366.
27 Nijdam A, Foppen W, De Kleijn P, et al. Discontinuing early prophylaxis in severe haemophilia leads to deterioration of joint status despite low bleeding rates. Thromb Haemost 2016; 115: 931–938.
28 Khawaji M, Astermark J, Berntorp E. Lifelong prophylaxis in a large cohort of adult patients with severe haemophilia: A beneficial effect on orthopaedic outcome and quality of life. Eur J Haematol 2012; 88: 329–335.
29 Kidder W, Nguyen S, Larios J, et al. Point-of-care musculoskeletal ultrasound is critical for the diagnosis of hemarthroses, inflammation and soft tissue abnormalities in adult patients with painful haemophilic arthropathy. Haemophilia 2015; 21: 530–537.
30 Sun J, Hilliard PE, Feldman BM, et al. Chinese Hemophilia Joint Health Score 2.1 reliability study. Haemophilia 2014; 20: 435–440.
31 Hilliard P, Funk S, Zourikins N, et al. Hemophilia joint health score reliability study. Haemophilia 2006; 12: 518–525.
32 Fischer K, de Kleijn P. Using the Haemophilia Joint Health Score for assessment of teenagers and young adults: Exploring reliability and validity. Haemophilia 2013; 19: 944–950.

33 Teyssler P, Taborska K, Kolostova K, Bobek V. Radiosynoviorthesis in hemophilic joints with yttrium-90 citrate and rhenium-186 sulfide and long term results. Hell J Nucl Med 2013; 16: 44–49.

34 Groen W, Van der Net J, Lacatusu AM, et al. Functional limitations in Romanian children with haemophilia: Further testing of psychometric properties of the Paediatric Haemophilia Activities List. Haemophilia 2013; 19: 116–125.

35 Salim M, Brodin E, Spaals-Abrahamsson Y, Berntorp E, Zetterberg E. The effect of Nordic Walking on joint status, quality of life, physical ability, exercise capacity and pain in adult persons with haemophilia. Blood Coagul Fibrinolysis 2016; 27: 467–472.

36 Sluiter D, Foppen W, de Kleijn P, Fischer K. Haemophilia Joint Health Score in healthy adults playing sports. Haemophilia 2014; 20: 282–286.

37 St-Louis J, Abad A, Atkins S, et al. Reference Ranges of HJHS Scores in Healthy Adult Males without Hemophilia. E-Poster Virtual ISTH Congr 2020.

38 Groen WG, van der net J, Helders PJM, Fischer K. Development and preliminary testing of a Paediatric Version of the Haemophilia Activities List (pedhal). Haemophilia 2010; 16: 281–289.

39 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: 36–46.

40 Fischer K, Carlsson KS, Petrini P, et al. Intermediate-dose versus high-dose prophylaxis for severe hemophilia: Comparing outcome and costs since the 1970s. Blood 2013; 122: 1129–1136.

41 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 (PROMISTM) Network: 2005–2008. J Clin Epidemiol 2010; 63: 1179–1194.

42 Witter JP. The Promise of Patient-Reported Outcomes Measurement Information System-Turning Theory into Reality. A Uniform Approach to Patient-Reported Outcomes Across Rheumatic Diseases. Rheum Dis Clin North Am 2016; 42: 377–394.

43 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: e54–67.

44 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: 2221–2229.

45 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: 211–220.

46 Crins MHP, van der Wees PJ, Klausch T, et al. Psychometric properties of the PROMIS Physical Function item bank in patients receiving physical therapy. PLoS One 2018; 13: 1–14.

47 Dampier C, Jaeger B, Gross HE, et al. Responsiveness of PROMIS® Pediatric Measures to Hospitalizations for Sickle Pain and Subsequent Recovery. Pediatr Blood Cancer 2016; 63: 1038–1045.

48 Bartlett SJ, Orbai AM, Duncan T, et al. Reliability and validity of selected PROMIS measures in people with rheumatoid arthritis. PLoS One 2015; 10: 1–14.

49 Irwin DE, Atwood CA, Hays RD, et al. Correlation of Promis Scales and Clinical Measures Among Chronic Obstructive Pulmonary Disease Patients With and Without Exacerbations. Qual Life Res 2015; 24: 999–1009.

50 Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010; 63: 737–745.

51 Boeckhout M, Zielhuis GA, Bredenoord AL. The FAIR guiding principles for data stewardship: Fair enough? Eur J Hum Genet 2018; 26: 931–936.

52 Fries JF, Cella D, Rose M, Krishnan E, Bruce B. Progress in assessing physical function in arthritis: PROMIS short forms and computerized adaptive testing. J Rheumatol 2009; 36: 2061–2066.

53 Feldman BM. The outcomes of haemophilia and its treatment: why we need a core set. Haemophilia 2017; 23: 485–487.

54 De Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine: A practical guide. New York: Cambridge University Press; 2011.

APPENDIX

NEDERLANDSE SAMENVATTING

Hemofilie is een zeldzame aangeboren stollingsaandoening, welke voorkomt bij 1 op de 5000 mannen. De ernst van de hemofilie hangt af van de hoeveelheid nog aanwezige stollingsfactoren in het bloed en varieert van mild tot ernstig. Het tekort aan stollingsfactoren in het bloed geeft een verhoogd risico op bloedingen in met name gewrichten en spieren. Gewrichtsbloedingen komen vooral voor in de enkels, knieën en ellebogen. De gewrichtsbloedingen resulteren in schade aan gewrichten waardoor mensen met hemofilie uiteindelijk moeite krijgen met dagelijkse fysieke activiteiten als lopen, fietsen en zelfzorg. Ook kan dit invloed hebben op school, werk of vrijetijdsbesteding.

De zorg voor mensen met hemofilie is de laatste decennia enorm verbeterd. In 1964 werd het eerste hemofiliebehandelcentrum opgericht in Nederland en in 1968 werd profylactische (regelmatige) behandeling met stollingsfactoren om bloedingen te voorkomen geïntroduceerd. In de loop der jaren is de medicatie verder ontwikkeld om het aantal bloedingen te verminderen en de kwaliteit van leven te verbeteren. Daarnaast wordt in de hemofiliebehandelcentra multidisciplinaire zorg aangeboden, door onder andere de arts, verpleegkundige, fysiotherapeut en maatschappelijk werker/psycholoog.

Het in kaart brengen van de gewrichten en beperkingen in dagelijkse bezigheden is belangrijk om het effect van de behandeling na te gaan, zowel in de dagelijkse zorg als onderzoek. Dit kan onder andere met meetinstrumenten zoals een gestandaardiseerd lichamelijk onderzoek en vragenlijsten die vragen naar functionele beperkingen. Het doel van dit proefschrift is om het meten van gewrichten (functies) en dagelijkse bezigheden (activiteiten en participatie) te verbeteren voor mensen met hemofilie. Enerzijds willen we meer kennis hebben over bestaande meetinstrumenten uit de hemofiliezorg, en anderzijds willen we de belasting en tijd die het kost om meetinstrumenten te gebruiken verminderen.

HJHS

De Hemophilia Joint Health Score (HJHS) is een gewrichtsscore van de enkels, knieën en ellebogen om de gezondheid van de gewrichten te meten. Een fysiotherapeut of arts beoordeelt de gewrichten tijdens een lichamelijk onderzoek op negen onderdelen, maar dit kost veel tijd (30-60 minuten).

In *hoofdstuk 2* hebben we de optimale frequentie voor het monitoren van gewrichten met behulp van de HJHS in kaart gebracht voor volwassenen met hemofilie. Na vijf tot tien jaar was de score van de HJHS in 37% van de mensen verslechterd en in 17% van de gewrichten verslechterd. De HJHS ging met name achteruit in gewrichten met een hoger aantal bloedingen en synovitis (kapselontsteking) van het gewricht. Nagenoeg alle gewrichten die bij aanvang geen beperkingen vertoonden en die geen bloeding of synovitis doormaakten bleven stabiel gedurende vijf tot tien jaar. Op basis van de resultaten adviseren we slechts eens per vijf jaar alle zes de gewrichten te meten met de HJHS. Indien een gewricht al beperkingen heeft of er is een bloeding en/of synovitis, adviseren we een gewricht vaker te monitoren.

Daarnaast hebben we in *hoofdstuk 3* gekeken of de HJHS korter gemaakt kan worden. Gewrichtsscores van 499 internationale patiënten zijn geanalyseerd en dit is vervolgens besproken in een bijeenkomst met 19 internationale experts. De conclusie is dat twee van de negen onderdelen van de HJHS weggelaten kunnen worden, namelijk 'hoe lang zwelling van het gewricht duurt' en 'crepitatie' (krakend geluid tijdens bewegen) van het gewricht. Een volgende stap zal zijn om te kijken of we kunnen screenen welke gewrichten in detail bekeken moeten worden.

PedHAL en HAL

De pediatrische Hemofilie Activiteiten Lijst (pedHAL) is een vragenlijst over beperkingen bij activiteiten en participatie voor kinderen met hemofilie. De vragenlijst bestaat uit 53 vragen en er is zowel een kinder- als oudervragenlijst beschikbaar. De Hemofilie Activiteiten Lijst (HAL) is een vragenlijst over beperkingen bij activiteiten en participatie voor volwassenen met hemofilie. De vragenlijst bestaat uit 42 vragen. Beide vragenlijsten worden zowel in de zorg als onderzoek gebruikt.

In *hoofdstuk 4* hebben we in Nederlandse jongens met hemofilie drie punten in kaart gebracht, namelijk (1) de veranderingen over de tijd gemeten met de pedHAL, (2) hoe goed de pedHAL scores van kinderen en ouders overeenkwamen en (3) welke domeinen van de pedHAL het meest informatief waren. Nagenoeg alle individuele patiënten met weinig tot geen beperkingen in activiteiten en zonder bloeding in de tussenliggende periode bleven stabiel na een jaar. Na drie tot vijf jaar bleven de pedHAL scores ook gelijk op groepsniveau. De pedHAL scores van kinderen en ouders kwamen niet voldoende overeen. Kinderen en hun ouders gaven de meeste beperkingen aan in de pedHAL domeinen 'zitten/knielen/staan' en 'functies van de benen'. Op basis van

de resultaten adviseren we niet standaard jaarlijks de pedHAL af te nemen bij kinderen met goede profylactische behandeling. Daarnaast adviseren we zowel de kinder- als de oudervragenlijst af te nemen.

Vervolgens hebben we in *hoofdstuk 5* onderzocht welke vragen uit de pedHAL mogelijk overbodig zijn met als doel een kortere versie van de vragenlijst te maken. In gegevens van de pedHAL uit vijf verschillende landen hebben we middels een stappenplan alle pedHAL vragen geëvalueerd. Dit heeft geleid tot een kortere versie van de pedHAL met 22 vragen in plaats van 53 vragen. De somscores van de volledige en de korte pedHAL kwamen overeen. De korte versie van de vragenlijst is efficiënter voor het uitvragen van ervaren beperkingen in activiteiten en participatie.

In *hoofdstuk 6* hebben we gekeken naar de betrouwbaarheid van de HAL. Vijftig volwassenen met hemofilie hebben tweemaal de vragenlijst ingevuld, met gemiddeld 3.4 weken tijd tussen de vragenlijsten. De vragenlijsten hebben we vervolgens met elkaar vergeleken. De HAL is betrouwbaar voor zowel de somscore als de componentscores. Veranderingen van de HAL somscore groter dan 10.2 punten worden gezien als verandering, buiten de meetfout van de vragenlijst om.

In *hoofdstuk 7* hebben we onderzocht welke vragen uit de HAL mogelijk overbodig zijn met als doel een kortere versie van de vragenlijst te maken. In gegevens van de HAL uit de Verenigde Staten hebben we middels een stappenplan alle HAL vragen geëvalueerd. Dit heeft geleid tot een kortere versie van de HAL met 18 vragen in plaats van 42 vragen. De somscores van de volledige en de korte HAL kwamen overeen. De korte HAL en de volledige HAL kwamen in gelijke mate overeen met andere vragenlijsten. De korte versie van de vragenlijst is efficiënter voor het uitvragen van ervaren beperkingen in activiteiten en participatie.

PROMIS

Patient Reported Outcomes Measurement Information System (PROMIS®) is een nieuwe ontwikkeling binnen onderzoek met vragenlijsten. PROMIS vragenlijsten kunnen worden afgenomen als verkorte lijst of via de computer met Computer Adaptief Testen (CAT). Welke vragen een patiënt moet beantwoorden wordt dan door de computer bepaald op basis van de antwoorden die iemand geeft. Bijvoorbeeld als iemand heeft geantwoord dat hij geen 100 meter kan lopen, dan is de volgende vraag of hij kan staan,

en wordt niet gevraagd of hij kan rennen. Dit zorgt ervoor dat iemand minder vragen hoeft te beantwoorden en meer relevante vragen krijgt.

In *hoofdstuk 8* hebben we de (1) toepasbaarheid, (2) meeteigenschappen en (3) relevantie van zeven PROMIS CAT vragenlijsten en twee verkorte PROMIS vragenlijsten onderzocht. De PROMIS vragenlijsten hebben we vergeleken met bekende vragenlijsten voor volwassenen met hemofilie, die al ingevuld waren in het kader van het Hemofilie in Nederland (HiN-6) onderzoek. Het gemiddelde aantal vragen voor de CAT vragenlijsten varieerden van vijf tot negen vragen. Bij de CAT vragenlijsten over 'belemmeringen door pijn' en 'depressie' scoorden mensen met hemofilie vaak de laagst mogelijke score. Volwassenen met hemofilie scoorden het vaakst beperkingen bij 'fysiek functioneren' (38%). De scores van de PROMIS CAT vragenlijsten over 'fysiek functioneren', 'belemmeringen door pijn', 'tevredenheid met sociale rollen en activiteiten' en 'vermoeidheid' kwamen voldoende overeen met de scores van de bekende hemofilie vragenlijsten. Ook waren de PROMIS CAT vragenlijsten betrouwbaar. We concluderen dat de PROMIS CAT vragenlijsten over 'fysiek functioneren', 'belemmeringen door pijn', 'tevredenheid met sociale rollen en activiteiten' en 'vermoeidheid' goed toepasbaar en betrouwbaar zijn. Deze vragenlijsten kunnen worden gebruikt als alternatief voor de bekende vragenlijsten uit de HiN-6 bij volwassenen met hemofilie.

DANKWOORD

Na vier jaar hard werken, ben ik erg blij met dit proefschrift als resultaat. Maar aan dit proefschrift hebben velen bijgedragen en zonder hen zou het niet zijn geworden wat het nu is. Daarom wil ik een aantal mensen in het bijzonder bedanken.

Allereerst dank aan alle deelnemers aan mijn onderzoeken. Zeker bij een zeldzame aandoening als hemofilie worden jullie regelmatig gevraagd voor onderzoek met als doel om telkens weer een stapje verder te komen. Het wordt enorm gewaardeerd dat jullie ook nu weer mee wilden werken!

Prof. dr. Roger Schutgens, beste Roger, veel dank voor de kans die je mij hebt gegeven om bij de Van Creveldkliniek onderzoek te doen. Niet alleen de zorg is multidisciplinair, ook het onderzoek, en dat maakt de Van Creveldkliniek een hele bijzondere plek. Het was prettig hoe je op de hoofdlijnen mee kon denken over de projecten en kaders kon stellen.

Dr. Kathelijn Fischer en dr. Janjaap van der Net, jullie zijn een mooi team samen. Zoals jullie zelf zeggen 'papa en mama'. Ik kwam in een warm bad terecht bij jullie. Beste Kathelijn, mede dankzij onze verschillen was het een perfecte match! Dank voor je enthousiasme voor dit onderzoek, waarbij we telkens weer hebben geprobeerd het zo dicht mogelijk bij de dagelijkse praktijk te houden. Het was elke keer weer een feestje als we nieuwe resultaten konden bekijken en bediscussiëren. Naast je enorme interesse in de studies, was er altijd ruimte voor een praatje over 'thuis'. Hopelijk kunnen we dit jaar nog mooi de puntjes op de i zetten. Beste Janjaap, alweer acht jaar geleden begon ik aan de opleiding Fysiotherapiewetenschap waar ik jou als docent heb leren kennen. Jouw ervaring, brede kijk en kennis hielpen mij om met een frisse blik naar ons werk te kijken. Daarnaast was jouw netwerk zeer waardevol voor onze projecten. Af en toe schakelde ik je in om mee te denken over mijn loopbaan, waar je dan ook echt de tijd voor nam. Dank voor de prettige begeleiding! We hadden dezelfde datum om naar toe te werken, ik voor het indienen van mijn proefschrift, jij voor je welverdiende pensioen! Hopelijk kun je daar nog vele jaren in goede gezondheid van genieten.

Graag bedank ik de beoordelingscommissie, prof. dr. C. Veenhof, prof. dr. E.M. van de Putte, prof. dr. M.A. Grootenhuis, dr. M.H. Cnossen, prof. dr. P.J. van der Wees, en de oppositie. Hartelijk dank dat jullie de tijd hebben genomen om mijn manuscript te lezen en te beoordelen.

Many thanks to all co-authors for your help and feedback on our work. Especially I would like to thank prof. dr. Brian Feldman and all the experts who participated in the HJHS expert meeting in Utrecht in 2019 for the inspiring meeting!

Beste collega's van de Van Creveldkliniek, dank voor jullie interesse en betrokkenheid en het heerlijke (zelfgemaakte) gebak. Het is mooi om te zien hoe iedereen zich inzet voor de patiënten! In het bijzonder wil ik Merel bedanken, bij jou mocht ik in 2015/2016 mijn afstudeeronderzoek voor de opleiding Fysiotherapiewetenschap doen, wat uiteindelijk een belangrijke aanleiding was voor dit proefschrift.

Lieve mede PhD studenten van de Van Creveldkliniek, bedankt voor alle (virtuele) kopjes koffie, gezelligheid en inspiratie! Ik heb genoten van de afgelopen jaren en langzaam gaan we nu allemaal weer onze eigen weg verder. Heel veel succes allemaal en hopelijk tot snel! Olav, jou wil ik in het bijzonder bedanken omdat je mijn paranimf wil zijn. Met hetzelfde promotieteam en dezelfde startdatum van ons PhD traject konden we elkaar vaak ondersteunen. Ik waardeer jouw enthousiasme in je werk!

Beste collega's van de research meeting van het WKZ kinderbewegingscentrum, dank voor alle inspirerende bijeenkomsten en jullie frisse blik!

Beste collega's van Centrum voor Fysiotherapie Ramaekers, dank voor jullie interesse in mijn onderzoek en de nodige afleiding op mijn werkdagen in Weert. Beste André en Liesbeth, hartelijk dank voor de mogelijkheid om mijn werk in de fysiotherapie praktijk te kunnen combineren met onderzoek.

Beste docenten van de opleiding Fysiotherapiewetenschap, dank voor alle inspirerende vrijdagen op de Uithof! Zonder de opleiding was dit boekje er nooit geweest. Het heeft me erg veel gebracht zowel in mijn werk als op persoonlijk vlak.

Studiegenoten van de opleiding Fysiotherapiewetenschap, cohort 2013, wat hebben we er drie mooie jaren van weten te maken in Utrecht! Ik heb genoten van jullie enthousiasme en samen hebben we er alles uitgehaald wat erin zat, zowel inhoudelijk als op sociaal vlak. De borrels en weekendjes weg waren top en hopelijk komen we elkaar nog vaak tegen!

Vrienden uit Luyksgestel, ook wel Go2, het boekje is er! Bedankt voor jullie belangstelling. Maar vooral heel erg bedankt voor de goede afleiding op alle feestjes en borrels. Dat er nog maar vele feestjes mogen volgen in ons mooie dorpje!

Lieve vrienden en vriendinnen, Nicky & Yvonne, Lisa & Marieke, Roy & Evy, Twan & Laura en Stefan & Anne in het bijzonder, heel erg bedankt voor jullie interesse en alle goede gesprekken! Bij jullie kan ik altijd terecht als het even tegen zit, maar daarna kan ik ook weer vol goede moed verder. Tot het volgende kopje thee!

Lieve (schoon)familie, bedankt voor jullie steun en belangstelling. En bedankt voor de nodige ondersteuning wanneer het wel erg veel tegelijk werd. We genieten samen volop van elkaar en de kindjes en dat is geweldig! Papa en mama, dank voor een warm thuis waar ik altijd welkom ben! Jacobine, bedankt dat jij mijn paranimf wil zijn. Ik vind het zo fijn dat jij weet wat onderzoek doen en promoveren is.

Lieve Imme, ons meisje, wij zijn zo blij met jou! Wij kijken er naar uit samen nog veel mooie herinneringen te maken. In dit boek staan wat dieren verstopt, kun jij ze vinden?

Lieve Bart, bedankt dat jij me hebt gesteund om aan dit avontuur te beginnen en het tot een goed einde te brengen! Zie het maar als in compliment dat jij degene bent die het meeste mee heeft gekregen van de frustraties die onderzoek en twee banen soms met zich meebrengen. *Want de liefste, dat ben jij!*

ABOUT THE AUTHOR

Isolde Kuijlaars was born on June 29, 1991 in Veldhoven, the Netherlands. She completed her secondary school in 2009 at Were Di in Valkenswaard and studied Physical therapy at the Fontys University of Applied Sciences in Eindhoven between 2009-2013. Her interest in research arose during working on her thesis for the bachelor Physical therapy. After graduation, she started the (pre)master Clinical health sciences, direction Physical therapy sciences, at the University Utrecht, which she obtained in 2016. She did an internship at the department of Rehabilitation, Physiotherapy and Sport and the van Creveldkliniek of the University Medical Center Utrecht.

During her premaster she started working as a physical therapist at Vital Fysiotherapie in Eersel and in 2014 she switched to Centrum voor Fysiotherapie Ramaekers in Weert. She is involved in the rehabilitation of patients with chronic obstructive pulmonary disease, symptomatic peripheral arterial disease and total knee/hip arthroplasty and is a member of Chronisch ZorgNet, a nationwide network of specialized physical therapists, providing high quality supervised exercise therapy and lifestyle counseling to patients with non-communicable diseases.

Besides her work as a physical therapist, Isolde started with her PhD project at the van Creveldkliniek under supervision of prof. dr. R.E.G. Schutgens, dr. K. Fischer and dr. J. van der Net in 2017. The results of this PhD project are presented in this thesis. Next year she will continue her research at the van Creveldkliniek.

Currently, she is living with her daughter Imme and her husband Bart in Luyksgestel, the Netherlands.

LIST OF PUBLICATIONS

Kuijlaars IAR, van der Net J, Bouskill V, Hilliard P, Juodyte A, Khair K, Trakymiene SS, Fischer K. Shortening the paediatric Haemophilia Activities List (pedHAL) based on pooled data from international studies. Haemophilia 2021; 27: 305–13.

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: 108–12.

Kuijlaars IAR, van der Net J, Feldman BM, Aspdahl M, Bladen M, de Boer W, Cuesta-Barriuso R, Matlary RED, Funk SM, Hilliard P, John JA, Kempton CL, de Kleijn P, Manco-Johnson M, Petrini P, Poonnoose P, St-Louis J, Thomas S, Timmer MA, Trakymiene SS, van Vlimmeren L, Fischer K. Evaluating international Haemophilia Joint Health Score (HJHS) results combined with expert opinion: Options for a shorter HJHS. Haemophilia 2020; 26: 1072–80.

Timmer M, Kloek C, de Kleijn P, **Kuijlaars IAR**, Schutgens REG, Veenhof C, Pisters MFA. Blended Physiotherapy Intervention for Persons With Hemophilic Arthropathy: Development Study. J Med Internet Res 2020; 22: e16631.

Kuijlaars IAR, van der Net J, Schutgens REG, Fischer K. The Paediatric Haemophilia Activities List (pedHAL) in routine assessment: changes over time, child-parent agreement and informative domains. Haemophilia 2019; 25: 953–9.

Kuijlaars IAR, Sweerts L, Nijhuis-van der Sanden MWG, van Balen R, Staal JB, van Meeteren NLU, Hoogeboom TJ. Effectiveness of Supervised Home-Based Exercise Therapy Compared to a Control Intervention on Functions, Activities, and Participation in Older Patients After Hip Fracture: A Systematic Review and Meta-analysis. Arch Phys Med Rehabil 2019; 100: 101-114.e6.

Kuijlaars IAR, Timmer MA, de Kleijn P, Pisters MF, Fischer K. Monitoring joint health in haemophilia: Factors associated with deterioration. Haemophilia 2017; 23: 934–40.

Submitted for publication

Kuijlaars IAR, van der Net J, Buckner TW, Kempton CL, Schutgens REG, Fischer K. Shortening the Haemophilia Activities List (HAL) from 42 items to 18 items. Submitted.

Kuijlaars IAR, Teela L, van Vulpen LFD, Timmer MA, Coppens M, Gouw SC, Peters M, Kruip MJHA, Cnossen MH, Muis JJ, van Hoorn ES, Haverman L, Fischer K. Feasibility, measurement properties and relevance of generic PROMIS item banks for patient reported outcome assessment in adult persons with haemophilia. Submitted.