ORIGINAL ARTICLE

Muskuloskeletal

Haemophilia WILEY

# 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 Centre Utrecht, Utrecht University, Utrecht, The Netherlands

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

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

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

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

## 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 (seven 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.

**KEYWORDS**
activities, haemophilia, participation, patient-reported outcome

# 1 | 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.

# 2 | MATERIALS AND METHODS

## 2.1 | 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 Haemophilia 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 Centre Utrecht approved the study (protocol number 20–650/C).

## 2.2 | Measurements

### 2.2.1 | 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 > .90. The average SDC value for the normalized HAL sum score was 10.2 [9].

### 2.2.2 | SF-36v2

The SF-36v2 measures health related quality of life across eight 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]

### 2.2.3 | 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 five 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 analysed age at HAL assessment, gender, type of haemophilia (A or B), severity of the disease (mild [factor VIII/IX activity .06–.40 IU/ml], moderate [factor VIII/IX activity .01–.05 IU/ml] or severe [factor VIII/IX activity < .01 IU/ml]), clotting factor regimen (prophylaxis yes/no) and inhibitor status (current/former or never).

## 2.3 | 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 (seven 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 seven steps were described in detail in the Supplementary material and the publication about shortening the paediatric 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 < .2 and > .9 were not included in the factor analysis.

Step 4 - Component loadings of the exploratory factor analysis: Items with factor loadings < .5 were removed.

Step 5 - Inter-item correlations (2): Items which had a correlation of > .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 $\alpha$ should be between .7 and .9; a higher Cronbach's $\alpha$ after item deletion was considered a reason to eliminate an item.

Step 7 - Item-total correlations: Items with item-total correlations of < .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 ≥.9 were considered as a very

**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 | .7 | .8 |
| Prophylaxis | 61.0 | 61.2 |
| Inhibitor (current/former) | 9.8 | 10.8 |

strong correlation, .70-.89 as strong, .40-.69 as moderate, .10-.39 as weak and < .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.

## 3 | RESULTS

### 3.1 | 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%).

### 3.2 | 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 11.7–100. '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 2** Domain, component and sum scores of the HAL (derivation set, n = 420)

| | Median (IQR) | Min | Max | Score < 90 (%) | Missing/NA 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 |

## 3.3 | 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 > .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 .07 (RMSEA), indicating moderate model fit. The factor loadings were > .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 > .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 .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 $\alpha$ of .96), which indicates redundancy of items. Only complete cases (n = 319, 76.0%) were included in this analysis. The Cronbach's $\alpha$ 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 = .64-.81), thus identifying no candidates for item reduction.

## 3.4 | 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'

**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<br>3 Using public transportation (bus, train, subway) | | 6 Gardening | 5 Dancing |
| Step 3: Inter-item correlations (r < .2 and r > .9) – remaining items: 37 | | | | | | |
| | 9 Jumping | | | | | |
| Step 4: Exploratory factor analysis (factor loadings < .5) – remaining items: 37 | | | | | | |
| Step 5: Inter-item correlations (r > .7) – remaining items: 18 | | | | | | |
| 2 Rising from a chair with armrests<br>4 Kneeling / squatting<br>6 Kneeling for a longer period of time<br>8 Standing for a longer period of time | 1 Walking short distances (less than 1 kilometer / 15 min)<br>3 Walking on a soft surface (e.g., on the beach or through the woods)<br>5 Strolling / (window-) shopping<br>7 Climbing down the stairs | 2 Carrying heavy objects in the arms<br>3 Fine hand movements (e.g., closing buttons) | | 2 Putting on a shirt, sweater etc.<br>4 Putting on a tie or closing the top button of a shirt<br>5 Going to the toilet | 1 Going out shopping (for food, drink etc.)<br>2 Washing the dishes, cleaning the sink<br>4 Other household tasks (ironing, making the beds) | 3 Going out (theatre / museum / movie theatre / bar)<br>4 Hobbies<br>7 Going on a holiday ("passive"; beach-/hotel holiday) |
| Step 6: Internal consistency – remaining items: 18 | | | | | | |
| Step 7: Item-total correlations (r < .3) – remaining items: 18 | | | | | | |
| HAL$_{short}$: remaining 18 items | | | | | | |
| 1 Sitting down (e.g., on a chair or couch)<br>3 Rising from a chair without armrests<br>5 Bending forward<br>7 Squatting for a longer period of time | 2 Walking long distances (more than 1 kilometer / 15 min)<br>4 Walking on an uneven surface (e.g., cobblestones, high sidewalks)<br>6 Climbing up the stairs<br>8 Running (e.g., in order to catch the bus) | 1 Lifting heavy objects<br>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<br>3 Putting on socks and shoes | 3 Cleaning the house<br>5 Doing odd jobs (both in and around the house) | 1 Playing games (outdoors, e.g., with your children)<br>2 Sports<br>6 Going on a holiday (active) |

**TABLE 4**   Original domain and sum scores of the HAL and HALshort

| | 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$_{short}$.
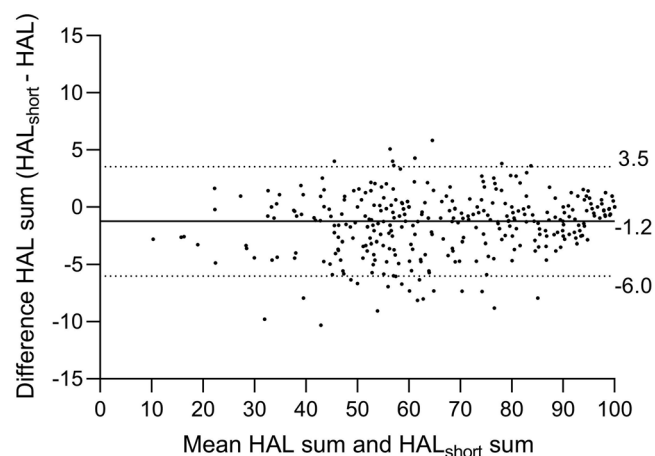


**FIGURE 1**   Bland Altman plot for the HAL and HAL$_{short}$ scores in the derivation dataset

($n = 4$). 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 < .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 = .09) and clear underlying constructs could not be defined. Therefore, the HAL$_{short}$ generated a single sum score.

## 3.5 | 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 < .001).

## 4 | 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

**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 ≥ .6 | V |
| r HAL$_{(short)}$ complex lower extremity – SF36v2 physical functioning ≥ .6 | V |
| r domain leisure activities and sports HAL$_{(short)}$ – SF36v2 role physical ≥ .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 ≥ -.6 | V |
| r HAL$_{(short)}$ complex lower extremity – EQ-5D-5L mobility ≥ -.6 | V |
| r HAL$_{(short)}$ domain household tasks – EQ-5D-5L usual activities ≥ -.5 | V |
| r HAL$_{(short)}$ domain leisure activities and sports – EQ-5D-5L usual activities ≥ -.5 | V |

r = correlation.
*Abbreviation*: HAL: Haemophilia Activities List.

**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 | .77 / .77 | .32 / .32 | | | | |
| | HAL / HAL$_{short}$ HOUSEH | | | | | | −.65 / −.60 |
| | HAL / HAL$_{short}$ LEISPO | | | | .59 / .59 | | −.60 / −.55 |
| | HAL / HAL$_{short}$ LOWBAS | | | .71 / .69 | | −.74 / −.73 | |
| | HAL / HAL$_{short}$ LOWCOMP | | | .76 / .74 | | −.65 / −.66 | |

*Abbreviations*: HAL, Haemophilia Activities List; HOUSEH, household tasks; LEISPO, leisure activities and sports; LOWBAS, basic lower extremity; LOW-COMP, 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.

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.

## 4.1 | 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 .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 $\alpha = .96$) is still higher than the recommended Cronbach's $\alpha$ between .7 and .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.

## 4.2 | 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$_{short}$.

## 4.3 | 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$_{short}$.

## 5 | 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.

## CONFLICT OF INTEREST

The Van Creveldkliniek has received speaker's fees from Bayer, Baxter/Shire, SOBI/Biogen, CSL Behring and NovoNordisk; has performed consultancy for Bayer, Biogen, CSL-Behring, Freeline, NovoNordisk, Roche and SOBI; and has received research support from Bayer, Baxter/Shire, Novo Nordisk, Pfizer and Biogen for work done by K. Fischer. K. Fischer is member of the group that developed the HAL. T.W. Buckner received honoraria for advisory board participation from Novo Nordisk, Takeda, CSL Behring, Bayer, Tremeau Pharmaceuticals, Spark, Pfizer; served as consultant to uniQure, BioMarin, and Tremeau Pharmaceuticals. T.W. Buckner is on the board of directors for the American Thrombosis and Hemostasis Network. C.L. Kempton received honoraria for Advisory Board Participation from Sanofi, Takeda, and Spark. I.A.R. Kuijlaars, J. van der Net and R.E.G. Schutgens do not have any conflict of interest regarding this manuscript other than membership of the group that developed the HAL. The other authors have no competing interests.

## AUTHOR CONTRIBUTION

I.A.R. Kuijlaars, J. van der Net, and K. Fischer contributed to the design of the study, C.L. Kempton and T.W. Buckner contributed to the data collection, I.A.R. Kuijlaars performed the statistical analyses, I.A.R. Kuijlaars wrote the first draft of the paper, all authors contributed to interpretation of the data, modification of statistical analyses and the writing of the manuscript.

## DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

## ORCID

*Isolde A. R. Kuijlaars* https://orcid.org/0000-0003-2920-2258
*Christine L. Kempton* https://orcid.org/0000-0002-2391-2016
*Roger E. G. Schutgens* https://orcid.org/0000-0002-2762-6033
*Kathelijn Fischer* https://orcid.org/0000-0001-7126-6613

## REFERENCES

1. Genderen FR, Meeteren NLU, Bom JG, et al. Functional consequences of haemophilia in adults: the development of the Haemophilia Activities List. *Haemophilia*. 2004;10:565-571.
2. 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, Emst M, 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.

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

11. Group EQ. EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy.* 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, 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 JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res.* 1999;8:135-160.

16. Batt K, Recht M, Cooper DL, Iyer NN, Kempton CL. 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, Boer C, 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.* 2018:1-8.

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: e231-9.

20. McLaughlin P, Morris R, Chowdary P. Investigating the relationship between the HJHS and HAL in routine clinical practice: a retrospective review. *Haemophilia.* 2018:1-7.

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-e86.

22. Brodin E, Hadzibajramovic E, Baghaei F, Sunnerhagen K, Nilsson Å. 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. *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.* 2018;24:e33-49.

## SUPPORTING INFORMATION

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