Haemophilia (2006), **12**, 36–46 DOI: 10.1111/j.1365-2516.2006.01186.x

Measuring patients' perceptions on their functional abilities: validation of the Haemophilia Activities List

F. R. VAN GENDEREN,†* P. WESTERS,‡ L. HEIJNEN,§* P. DE KLEIJN,†* H. M. VAN DEN BERG*, P. J. M. HELDERS¶ and N. L. U. VAN MEETEREN†**

**Van Creveldkliniek, University Medical Centre Utrecht, Utrecht, The Netherlands; †Rudolf Magnus Institute of Neuroscience, Department Neurology and Neurosurgery, Section Rehabilitation Medicine, University Medical Centre Utrecht, Utrecht, The Netherlands; ‡Centre for Biostatistics, Utrecht University, Utrecht, The Netherlands; §Rehabilitation Centre De Trappenberg, Huizen, The Netherlands; ¶Department of Paediatric Physiotherapy and Exercise Physiology, Wilhelmina Children's Hospital, University Medical Centre Utrecht, Utrecht, The Netherlands; and **Department of Physiotherapy–Research, Academy of Health Sciences Utrecht, Utrecht, The Netherlands*

Summary. Recently, the Haemophilia Activities List (HAL), a haemophilia-specific self-assessment questionnaire to assess a patient's self-perceived functional ability, was introduced and a limited pilot study warranted its further development. The present study finalizes the HAL and assesses the convergent and construct validity, as well as the internal consistency of its definitive version. Three questionnaires (HAL, Dutch-Arthritis Impact Measurement Scales 2 and the Impact on Participation and Autonomy questionnaire) were completed by 127 patients with severe haemophilia (<1% clotting activity), as well as four performance tests (button test, 50 metre walking test, timed-up-and-go test and figure-8 walking test). After removal of 15 non-informative items from the provisional HAL, three components within the questionnaire were identified (upper extremity activities, basic lower extremity activities and complex lower extremity activities). The internal consistency of these components was high (Cronbach's $\alpha = 0.93$–0.95), as was internal consistency for the seven domains of the HAL ($\alpha = 0.61$–0.96). The convergent validity of the HAL when compared to the other two questionnaires was good ($r = 0.47$–0.84). The construct validity of the HAL when compared to the four performance tests was generally lower ($r = 0.23$–0.77). The final version of the HAL has good internal consistency and convergent validity and gives the clinician insight into a patient's self-perceived ability to perform activities of daily life. It is likely that self-assessment instruments (questionnaires) and performance tests consider different concepts of functional health status and it is therefore recommended that both types are included when clinicians assess a patient's functional abilities.

Keywords: activities, haemophilia, HAL, questionnaire, validation

Introduction

As a result of a hereditary deficiency in clotting factor, patients with haemophilia suffer from repeated bleeding episodes, most frequently located in the major joints (i.e. elbows, knees and ankles) and muscle tissue. Because of repeated joint bleeds, affected individuals experience progressive joint impairments that may have serious implications for their functional abilities. Appropriate clinimetric instruments are essential to monitor functional abilities in individual patients. Moreover, assessment of the effects of medical treatment (on demand or prophylactic factor substitution) and orthopaedic procedures on a patient's functional ability should be improved. Recently, De Kleijn *et al.* [1,2] reviewed the literature and found no haemophilia-specific clinimetric instruments suitable for measuring functional abilities in this patient group. The importance of measuring functional abilities as a part of a patient's health status becomes more and more recognized, and therefore the need for clinimetrics to do so is imperative [3,4]. Therefore, a haemophilia-specific

Correspondence: Frank R. van Genderen, University Medical Centre Utrecht, Department of Rehabilitation Medicine, Room STR 5.133, P.O. Box 85500, 3508 GA Utrecht, The Netherlands. Tel.: +31 30 253 8484; fax: +31 30 250 5450; e-mail: F.R.vanGenderen@umcutrecht.nl

Accepted after revision 27 October 2005

instrument to measure functional abilities, the Haemophilia Activities List (HAL), was developed [5]. Patients played a major role in the development of the provisional version of the questionnaire and a pilot study yielded results that granted further developmental steps. Moreover, the development of the HAL did not go unnoticed by the World Federation of Hemophilia [6], which is important for the further implementation of the HAL in a core-set of measurements for haemophilia [2,7,8]. Further development of the HAL should focus on eliminating non-informative items, and assessing of the psychometric properties of the final version in a larger group of patients. Additionally, validation of the HAL should not only consider questionnaires, but also actual performance instruments. Self-report instruments assess the perception of an individual regarding a functional task, whereas performance-based instruments assess the actual performance of that task [9].

The aim of this study was therefore twofold: (i) construct the definitive version of the HAL by removing non-informative items and (ii) assess the internal consistency and convergent and construct validity of that definitive version of the HAL in a large group of patients with severe haemophilia, using both self-report as well as performance-based instruments.

Methods

Consecutive patients with severe (<1% clotting activity) haemophilia A or B visiting the Van Creveldkliniek for their regular check-up were invited to participate in the study. After obtaining written informed consent, three questionnaires [the Haemophilia Activities List (HAL), the Dutch Arthritis Impact Measurements Scales 2 (Dutch-AIMS2) and the Impact on Participation and Autonomy questionnaire (IPA)] were administered. As a part of the check-up, each participant was also required to perform four performance tests [button test, 50 metre walking test, timed-up-and-go (TUG) test and figure-8 walking test] under the supervision of a physical therapist (P.d.K.). The study was approved by the Ethical Review Board of the University Medical Centre, Utrecht.

Haemophilia Activities List

The HAL is a haemophilia-specific questionnaire, evaluating the self-perceived functional abilities of patients with haemophilia [5]. The provisional version of the HAL (used here) was developed in close collaboration with the patients. All items included in the questionnaire were the result of over 160 interviews based on the McMaster-Toronto Arthritis Patient Preference Disability Questionnaire (MACTAR) [10]. During this semi-structured interview, patients were allowed to mention up to five problematic activities that they would like to be able to perform without pain and/or difficulty. The mentioned activities served as items in the questionnaire. Subsequently, the HAL was presented to two focus groups (care givers and patients) to address the expert validity of the questionnaire. Moreover, a small pilot study in 50 patients with haemophilia addressed the convergent validity and internal consistency of the provisional version of the HAL [5].

This provisional HAL encompasses 57 items, distributed over eight domains. The eight domains were based on activity-classes of daily life and are related to the International Classification of Functioning, Disability and Health (ICF) [11]: 'lying down/sitting/kneeling/standing', 'functions of the legs', 'functions of the arms', 'use of transportation', 'self-care', 'household tasks', 'leisure activities and sports' and 'other'.

All items are preceded by a standard question: 'In the previous month, did you encounter any difficulty due to haemophilia with …'. The response options were a five-point Likert scale, complemented by the option of 'impossible'; this response ought to be chosen by patients who felt unable to perform the requested activity. For each domain a composite score was calculated, as well as an overall summarized score for the HAL (HAL_{sum}). Moreover, three component scores were calculated: upper extremity activities (HAL_{upper}; e.g. reaching, carrying heavy objects), basic lower extremity activities (HAL_{lowbas}; e.g. walking) and complex lower extremity activities (HAL_{lowcom}; e.g. running, jumping). These component scores relate to three underlying constructs within the HAL questionnaire itself. For all scores of the HAL, a higher score represents more self-perceived difficulty in performing activities.

Dutch Arthritis Impact Measurement Scales 2

The Dutch-AIMS2, originally designed for use in patients with rheumatoid arthritis [12], evaluates physical, psychological and social aspects of patients; its application in haemophilia was recently established [13,14], but the Dutch-AIMS2 has only a limited scope on activities [2]. This self-administered questionnaire encompasses 81 items, distributed over 12 health status scales and five components. In this study, only the scores for the first six scales and the first component were calculated, as they relate to activities: 'mobility level', 'walking and bending', 'hand and finger function', 'arm function', 'self-care', 'household tasks' and the 'physical component' ($AIMS2_{physical}$). Moreover, a sum scores for the entire questionnaire is also calculated ($AIMS2_{sum}$). As in the HAL, a higher score represents more self-assessed activity limitations. The validity of the Dutch-AIMS2 for use in patients with haemophilia has been established [13,14].

Impact on Participation and Autonomy questionnaire

The IPA is a generic questionnaire addressing the personal impact of illness on participation and autonomy and related experience of problems [15,16]. This self-administered questionnaire consists of 31 items, distributed over five domains: autonomy indoors ($IPA_{auto\text{-}in}$), family role ($IPA_{famrole}$), autonomy outdoors ($IPA_{auto\text{-}out}$), social relations (IPA_{socrel}) and work and educational opportunities ($IPA_{workedu}$). A higher score represents greater restrictions in participation or worse autonomy [17]. Because the IPA has not yet been validated for use in patients with haemophilia, the internal consistency of the questionnaire was evaluated. Cardol *et al.* [15] obtained the following reliability coefficients (Cronbach's α coefficient) for each of the subscales: autonomy indoors (0.91), family role (0.90), autonomy outdoors (0.81), social relations (0.86) and work and educational opportunities (0.91). In our sample, we obtained similar coefficients (0.93, 0.93, 0.88, 0.89 and 0.92, respectively), which led us to consider the IPA an instrument with satisfactory reliability indices in patients with severe haemophilia.

Performance tests

Four performance tests were incorporated into the study design. The choice of instruments was based on (i) expert opinion regarding a match between activity problems in haemophilia and suitable instruments, and (ii) results from an earlier investigation into problematic activities as mentioned by patients with haemophilia [5]. Moreover, it is generally known that most problems in haemophilia are found in relation to activities of the lower extremities, as the ankles and knees are very frequently impaired by haemorrhages and resulting arthropathy [18]. Therefore, only one of the four performance tests is aimed at activities involving the upper extremities, which is the button test. The button test is incorporated because in this activity the use of the elbows and, to a lesser extent, the shoulders is needed. The patient was required to close and open the three top most buttons (except the top button) of a standard shirt. The button test used in this study was an adaptation of the original button test in which a button board is used to assess the functional abilities of the rheumatic hand [19]. The reliability of the original button test was established in patients with rheumatoid arthritis [19]. In the 50 metre walking test [20,21], the patient walks in a straight line down a corridor of the hospital towards a designated point 50 m away from the starting point. A comfortable walking speed is instructed. The TUG test [22–24] required the patient to rise from a chair with armrests, walk forward for 3 m, turn around, walk back to the chair and sit down again. The reliability and validity of the TUG has been established [22–24]. The figure-8 walking test is part of the validated Timed Movement Battery [25]. The figure-8 walking test required the patient to perform a walking test around two points separated 8 m, in an eight-shaped manner. For all tests, the outcome was the amount of time (in seconds) needed to perform the specific task.

Data analysis

Data analysis was done with SPSS 11.5, using a one-sided significance level of $\alpha = 0.05$. All data were assessed for normal distribution using *Q–Q* plots. The data analysis was divided into three consecutive parts. In the first part, non-informative or redundant items were removed from the questionnaire, hereafter called the final HAL. The second part tested whether the three underlying constructs that were established in the pilot study [5] were distinct in the final HAL. In the third step, the internal consistency and convergent validity of the final HAL were assessed. Both the questionnaires and the performance tests were used in the assessment of the convergent validity.

Part 1: item reduction In order to reduce the number of items in the HAL, three steps were taken. First, to identify possible non-informative items in the HAL, floor and ceiling effects were evaluated by the percentage of scores at the extremes [i.e. 'impossible' or 'never' (…a problem)] of the response options [26]. When at least 50% of all respondents indicated being unable to perform a certain activity, that item was addressed as showing a ceiling effect, whereas a floor effect was found when at least 50% of all respondents indicated having no problems with that activity. Items with either floor or ceiling effects were eligible for removal from the questionnaire.

Second, data from the MACTAR interviews on which the items in the provisional HAL were based [5] were considered. If an item was identified in the first step as showing either a floor or ceiling effect, the MACTAR data were searched to identify the number of patients who originally mentioned that specific activity. These MACTAR data originated from a study in which 162 patients mentioned problematic activities and – after categorization – all mentioned activities were included as items in the provisional version of the HAL. If the identified item in the current was mentioned by less than 5% of patients in that previous investigation, it was considered redundant.

Third, in order not to drop items with a low response, but a high impact for the individual patient (i.e. a clinically relevant item), the items eligible for removal after the second step were subsequently presented to two haemophilia care experts (a physical therapist and a rehabilitation physician, both with over 20 years of experience in the field of haemophilia). The final decision on whether or not to remove an item from the questionnaire was based on their independent judgement.

Part 2: identifying underlying constructs After removal of redundant items, an explorative factor analysis with varimax rotation was used to assess whether the components could still be identified in the finalized HAL, thereby addressing the construct validity. Only factors with an eigenvalue of 1.0 or higher were considered as a construct and separate items with a factor-loading of at least 0.4 were considered as representative of that construct.

Part 3: assessing internal validity and convergent validity Descriptive statistics for the three questionnaires (final HAL, Dutch-AIMS2 and IPA) were calculated.

The internal consistency of the final version of the HAL was determined by calculating Cronbach's α for the whole questionnaire, as well as for each separate domain and the three components of the HAL (as established in the factor analysis in part 2) that resulted from the above-mentioned factor analysis. For each item, the 'α if item deleted' was also determined.

The convergent validity of the final HAL was determined by correlating the scores for all three questionnaires, using the Spearman rank order correlation coefficient. Two sum scores were used for the Dutch-AIMS2: the overall sum score ($AIMS2_{sum}$) and the score for the physical component ($AIMS2_{physical}$). Additionally, the scale scores for the first six scales of the Dutch-AIMS2 encompassing the physical component were correlated with the scores of the HAL (all seven domains, three components and the sum score). Additionally, Spearman correlations between the HAL scores and the performance tests were assessed.

Because a large number of patients were to be included in the analysis, even relatively low correlations will be statistically significant [27]. For the sake of interpretation of the data, only correlations >0.60 were considered in substantiating the convergent validity. Correlation coefficients <0.60 were omitted, correlations in the range of 0.60–0.80 were addressed as good and correlations >0.80 were addressed as excellent.

Results

Patient characteristics

One hundred and twenty-seven patients with severe haemophilia agreed to participate in this study (110 with haemophilia A and 17 with haemophilia B). Their median age was 42.0 years (p25–p75, 31.0–51.0). None of the patients had acute pathology (e.g. bleeding) or recent bleedings. Patients required a median of 8 min (5.0–12.0) to complete the provisional HAL (57 items were administered).

Part 1: item reduction Twenty-nine items (50.8%) of the provisional HAL showed a floor effect, whereas no items with ceiling effects were found. According to the '50%-limit', most redundant items were found in the domain of self-care, seven of the 11 items were eligible for removal. Additionally, the two-item domain 'other' could also be eliminated, as over 70% of the respondents reported having no problems at all with writing or using a personal computer (75.6% and 70.9%, respectively).

The 29 items identified as showing floor effects were originally mentioned by a maximum of seven patients during the MACTAR interviews (4.3% of the 162 participating patients [5]), with a median of two patients (1.2%). Thus, all 29 items were considered redundant.

Two experts subsequently addressed the 29 items on their clinical relevance (overall agreement was 75% on 29 items; $\kappa = 0.459$, $P = 0.011$). Their discussion led to consensus on the removal of 15 items (Table 1); the remaining 14 were considered too important from a clinical perspective by one or both of the experts to be removed from the HAL. Eventually, 42 items in seven domains remained (see Appendix).

Table 1. Items removed from the questionnaire.

Item	Domain	Floor effect ($n = 127$)	MACTAR [5] ($n = 167$)
Lying down	LSKS	90 (70.9%)	2 (1.2%)
Getting back up after lying down	LSKS	77 (60.6%)	2 (1.2%)
Lying down for a longer period of time	LSKS	92 (72.4%)	1 (0.6%)
Holding or carrying light objects	ARMS	87 (68.5%)	1 (0.6%)
Driving a car yourself	TRANSP	86 (67.7%)	7 (4.3%)
Washing your hair	SELFCA	93 (72.2%)	2 (1.2%)
Taking a shower	SELFCA	90 (70.9%)	5 (3.1%)
Brushing your teeth	SELFCA	100 (78.4%)	1 (0.6%)
Shaving	SELFCA	102 (80.3%)	2 (1.2%)
Putting on trousers	SELFCA	86 (67.7%)	3 (1.9%)
Eating (cutting, using cutlery)	SELFCA	97 (76.4%)	3 (1.9%)
Preparing a meal	HOUSEH	77 (60.6%)	5 (3.1%)
Playing games indoors	LEISPO	106 (83.5%)	4 (2.5%)
Writing	OTHER	96 (75.6%)	1 (0.6%)
Using a computer	OTHER	90 (70.9%)	2 (1.2%)

Part 2: identifying underlying constructs An explorative factor analysis on the data of the final HAL revealed six factors that together accounted for 75.9% of the total variance. The three major factors that were identified corresponded with the components addressed in the provisional HAL and explained 50.6% of the total variance: 'upper extremity' (17.4% explained variance), 'basic lower extremity' (12.7%) and 'complex lower extremity' (20.5%). The remaining three factors were identified as 'household activities' (12.3%), 'leisure activities and sports' (9.5%) and 'use of public transportation' (3.6%).

Part 3: assessing internal validity and convergent validity All analyses in this part were based on the data of the final version of the HAL encompassing 42 items in seven remaining domains, with scores for all separate domains, three component scores and an overall sum score.

For all three questionnaires, the raw scores were calculated (see Table 2). In general, the data from the three questionnaires were not normally distributed. The median sum score for the HAL was 106.5, 117.0 for the overall Dutch-AIMS2, 1.0 for the physical score of the Dutch-AIMS2 and the median sum score for the IPA was 30.0. For the performance tests, the median time to complete the button test was 16.4 s, 50 metre walking test 33.8 s, TUG 9.0 s and 16.7 s for the figure-8 walking test.

For the final HAL as a whole, Cronbach's α was 0.97 (Table 3). Cronbach's α statistics for the internal consistency of the seven domains of the HAL ranged from 0.61 (domain 'use of transportation') to 0.96 (domain 'functions of the legs'). The internal consistency of the three components ranged from 0.93 to 0.95.

Spearman rank correlation coefficients between the several HAL scores and the scores for the Dutch-AIMS2 and IPA were calculated to assess the convergent validity of the HAL (Table 4a). All correlations ($n = 55$) were significant ($P < 0.001$) and ranged from 0.47 and 0.84; 89% ($n = 49$) of these correlations were considered in the analysis ($r > 0.60$). The highest correlation was found between the 'physical component' of the Dutch-AIMS2 and the 'leg functions' domain of the HAL ($r = 0.84$), followed by the correlations between the HAL sum score and the physical component and the sum score of the Dutch-AIMS2 followed by the correlations between the HAL sum score and the physical component and the sum score of the Dutch-AIMSZ ($r = 0.83$ and 0.81, respectively). An in-depth analysis (Table 4b) comparing the domains of the HAL with the first six scales of the Dutch-AIMS2 (encompassing the physical component) showed the highest correlations between the 'walking and bending' scale of the Dutch-AIMS2 and the domains for 'lying/sitting/kneeling/standing and leg' functions ($r = 0.83$ and 0.86, respectively) and between the 'arm and hand and finger' scales of the Dutch-AIMS2 and the 'self-care' domain of the HAL ($r = 0.83$ and 0.80, respectively).

The Spearman rank correlation coefficient between the several HAL scores and the performance tests were also calculated (Table 4c). Two correlation coefficients were not significant (button test and the domains of 'household tasks' and 'leisure and sports'), whereas all other correlations were significant ($P < 0.05$). The button test did not correlate higher than 0.39 with any of the domains of the HAL. The highest correlation was found with the 'self-care' domain ($r = 0.43$). The 50 metre walking test correlated highest with the domains on 'lying/sitting/kneeling/standing' and 'leg functions'. The best correlations for the TUG test were

Table 2. Results from the study.

Questionnaires	Possible scoring range*	Median	IQR
HAL			
Lying/sitting/kneeling/standing (LSKS)	8.0–48.0	26.0	15.0–33.0
Functions of the legs (LEGS)	9.0–54.0	33.0	20.0–41.0
Functions of the arms (ARMS)	4.0–24.0	8.0	4.0–12.0
Use of transportation (TRANS)	3.0–18.0	4.5	3.0–8.0
Self care (SELFCA)	5.0–30.0	5.5	5.0–10.0
Household tasks (HOUSEH)	6.0–36.0	10.0	6.0–15.0
Leisure activities and sport (LEISPO)	7.0–42.0	13.0	8.0–20.0
Upper extremity component (HAL_{upper})	9.0–54.0	14.0	9.0–22.0
Basic lower extremity component (HAL_{lowbas})	6.0–36.0	18.5	12.0–25.0
Complex lower extremity component (HAL_{lowcom})	9.0–54.0	37.0	19.0–45.0
SUMscore (HAL_{sum})	42.0–252.0	106.5	62.0–131.0
Dutch-AIMS2			
Physical component ($AIMS2_{phys}$)	0.0–10.0	1.0	0.4–2.6
Affect component ($AIMS2_{affect}$)	0.0–10.0	2.5	1.8–3.8
Symptom component ($AIMS2_{symptom}$)	0.0–10.0	3.0	2.0–5.0
Social/interaction component ($AIMS2_{social}$)	0.0–10.0	2.9	2.1–3.9
Role component ($AIMS2_{role}$)	0.0–10.0	0.6	0.0–2.5
SUMscore ($AIMS2_{sum}$)	70.0–349.0	117.0	104.0–138.5
IPA			
Autonomy indoors ($IPA_{auto\text{-}in}$)	0.0–28.0	3.0	0.0–7.5
Family role ($IPA_{famrole}$)	0.0–28.0	7.0	1.0–11.5
Autonomy outdoors ($IPA_{auto\text{-}out}$)	0.0–20.0	5.0	1.0–7.0
Social relations (IPA_{socrel})	0.0–24.0	4.0	0.5–6.5
Work and education ($IPA_{workedu}$)	0.0–24.0	8.0	3.0–13.0
SUMscore (IPA_{sum})	0.0–124.0	30.0	9.5–43.0
Performance tests			
Button test		16.4	13.5–19.0
50 Metre walking test		33.8	29.9–41.2
Timed Get-Up-and-Go test		9.0	7.7–11.4
Figure-8 walking test		16.7	14.7–20.4

IQR, interquartile range; n/a, not applicable.
*A lower score represents better health.

Table 3. Internal consistency of the HAL.

	Items (*n*=)	Cases (*n*=)	Cronbach's α
HAL overall	42	110	0.97
HAL domains			
Lying/sitting/kneeling/standing (LSKS)	8	113	0.92
Functions of the legs (LEGs)	9	113	0.96
Functions of the arms (ARMS)	4	117	0.89
Use of transportation (TRANS)	3	118	0.61
Self-care (SELFCA)	5	117	0.89
Household tasks (HOUSEH)	6	118	0.89
Leisure activities and sport (LEISPO)	7	118	0.84
HAL components			
Upper extremity (HAL_{upper})	9	116	0.93
Basic lower extremity (HAL_{lowbas})	6	113	0.95
Complex lower extremity (HAL_{lowcom})	9	113	0.95

found with the 'lying/sitting/kneeling/standing' domain and the 'complex lower extremity' component. The figure-8 walking test correlated well with the 'lying/sitting/kneeling/standing' domain.

Discussion

The aim of this study was to finalize the HAL and to assess the internal consistency and both convergent

Table 4a. Convergent validity; questionnaires.

	$AIMS2_{phys}$	$AIMS2_{sum}$	$IPA_{auto\text{-}in}$	$IPA_{auto\text{-}out}$	IPA_{sum}
Lying/sitting/kneeling/standing	*0.79*	*0.71*	*0.70*	*0.69*	*0.68*
Leg functions	**0.84**	*0.77*	*0.67*	*0.68*	*0.65*
Arm functions	*0.78*	*0.67*	*0.71*	*0.66*	*0.70*
Transportation	*0.64*	*0.66*	0.54	0.56	0.59
Self-care	*0.76*	*0.61*	*0.65*	*0.63*	*0.68*
Household tasks	*0.73*	*0.75*	*0.66*	*0.67*	*0.67*
Leisure and sports	0.55	*0.62*	0.50	0.50	0.47
Upper extremity (HAL_{upper})	*0.78*	*0.68*	*0.72*	*0.68*	*0.72*
Basic lower extremity (HAL_{baslow})	*0.79*	*0.71*	*0.65*	*0.68*	*0.64*
Complex lower extremity (HAL_{comlow})	*0.77*	*0.75*	*0.63*	*0.66*	*0.64*
SUMscore (HAL_{sum})	**0.83**	**0.81**	*0.72*	*0.73*	*0.71*

All Spearman correlations are significant ($P < 0.001$).
Values underlined are correlations not considered for evaluation; values in italics are good correlations; values in bold are excellent correlations.

Table 4b. Convergent validity; AIMS2 scales.

	Mobility	Walking and bending	Hand and finger	Arm	Self-care	Household
Lying/sitting/kneeling/standing	*0.64*	**0.83**	*0.60*	0.58	0.38	0.59
Leg functions	*0.69*	**0.86**	0.59	0.55	0.35	*0.66*
Arm functions	*0.65*	*0.71*	*0.73*	*0.73*	0.48	*0.67*
Transportation	0.58	0.57	0.56	0.49	0.31	0.48
Self-care	*0.66*	*0.66*	**0.80**	**0.83**	0.50	*0.63*
Household tasks	*0.65*	*0.66*	0.55	0.56	0.50	*0.62*
Leisure and sports	0.44	0.53	0.35	0.37	0.30*	0.41
Upper extremity (HAL_{upper})	*0.66*	*0.70*	*0.76*	*0.76*	0.50	*0.65*
Basic lower extremity (HAL_{baslow})	*0.69*	*0.80*	0.55	0.52	0.36	*0.64*
Complex lower extremity (HAL_{comlow})	*0.63*	*0.78*	0.59	0.51	0.33	0.59
SUMscore (HAL_{sum})	*0.70*	*0.80*	*0.65*	*0.64*	0.45	*0.67*

All Spearman correlations are significant for $P < 0.001$, unless specified $^{*}P = 0.001$.
Values underlined are correlations not considered for evaluation; values in italics are good correlations; values in bold are excellent correlations.

Table 4c. Construct validity; performance tests (Spearman; outliers omitted).

	Button	50 Metre walking	Timed up-and-go	Figure-8 walking
Lying/sitting/kneeling/standing	0.31**	*0.66*	*0.65*	*0.61*
Leg functions	0.27**	*0.65*	0.59	0.53
Arm functions	0.33	0.56	0.57	0.56
Transportation	0.23**	0.41	0.39	0.36
Self-care	0.39	0.52	0.50	0.48
Household tasks	0.19*	0.50	0.45	0.45
Leisure and sports	ns	0.38	0.34	0.29**
Upper extremity (HAL_{upper})	0.35	0.55	0.54	0.54
Basic lower extremity (HAL_{baslow})	0.25*	*0.62*	0.55	0.51
Complex lower extremity (HAL_{comlow})	0.27**	*0.61*	*0.62*	0.55
SUMscore (HAL_{sum})	0.30**	*0.60*	0.59	0.54

All Spearman correlations are significant for $P < 0.001$, unless specified $^{*}P < 0.05$; $^{**}P < 0.01$.
Values underlined are correlations not considered for evaluation; values in italics are good correlations; values in bold are excellent correlations.

and construct validity of the final version of the HAL using self-rating (questionnaires) and performance-based instruments (performance tests). Ceiling effects suggested removal of several items and eventually 15 items were rejected from the questionnaire. The subsequent results were very satisfactory: the internal consistency was addressed as high, whereas the convergent validity of the HAL when compared to the Dutch-AIMS2 and the IPA was also very good. The construct validity of the HAL when compared to the 50 metre walking test and the TUG test was also good, meaning that more reported 'problems' in performing activities also meant more time needed to perform the tests. On the other hand, the correlations between the HAL and the remaining performance tests (button test, figure-8 walking test) were rather low. Owing to the short time needed to complete the questionnaire, the clinical applicability was good: the original 57-item HAL required 9.5 min to complete; the final version of the HAL with 42 items will require patients even less time to complete.

Over half of the items in the HAL showed floor effects, which led us to consider removal of non-informative, redundant items. Because an item was defined as showing a floor effect when over half of all respondents scored 'no problems' for that specific item, which is a rather crude limit, we had to ascertain that no valuable, clinically important information would be lost when removing specific items. The data from which the original questionnaire was constructed supported removal of these items, as the number of patients who mentioned that specific activity was very small. Moreover, two experts in the field of haemophilia care reached a consensus on the removal of 15, clinically non-informative items, which led us to conclude that by removing the non-informative items from the questionnaire, no important information would be lost.

The three major underlying components within the HAL that were identified in our previous study were also present in the final version of the HAL. An explorative analysis was considered appropriate here, as the composition of the final HAL was altered because of the removal of redundant items. Therefore, some items that encompassed the initial components have been eliminated from the questionnaire, thereby obstructing the possibility to run a confirmative factor analysis. All of these components clearly point to activities relating to specific body regions (one for the upper and two for the lower extremities). The third component additionally encompasses activities that are likely to be firstly affected in severe patients with haemophilia, like running and jumping. Therefore, the final designation of the components is 'upper extremity activities', 'lower extremity basic activities' and 'lower extremity complex activities'. The component scores serves as an indication of a patient's self-perceived abilities to perform activities that are encompassed in each component; higher scores represent more problems.

The internal consistency of the 'use of transportation' domain of the HAL is the lowest of all domains ($\alpha = 0.64$), which is probably caused by diversity of the three items within this domain ('riding a bicycle', 'getting in and out of a car' and 'using public transportation') plus the fact that there are only three items [27]. The internal consistency of the other domains, as well as the three identified components and the questionnaire as a whole, is otherwise very satisfactory. Although a high α coefficient might suggest redundancy in the items [27], results from the statistical analyses indicated that if any other item was removed from the subsequent domain or component, the α would decrease, indicating a lower internal consistency for that specific domain/component. Therefore, further removal of items affects the internal consistency of the questionnaire, but also leaves out clinically important information and is thus not recommended. These statistics also give back-up to the choice of both experts to remove or conserve specific items.

The convergent validity based on the associations between the final HAL and the Dutch-AIMS2 and IPA questionnaires is good. The sum score of the HAL and the domain score for 'leg activities' correlated highly with the 'physical component' score of the Dutch-AIMS2. Partially based on the results from the pilot study using the provisional HAL [5], we hypothesized that the HAL domains and components relating to activities involving the upper extremities (arm functions, self-care and HAL_{upper}) were associated with the 'hand and finger function', 'arm function and 'self-care' scales of the Dutch-AIMS2 and the button test. The results confirm this hypothesis, as the correlations between the aforementioned scores were indeed the highest for those scores. The low correlation between the 'self-care' domains of both the HAL and the Dutch-AIMS2 is remarkable, but might be explained by the fact that the Dutch-AIMS2 assesses whether the patient *needs assistance* in performing self-care tasks, whereas the HAL assesses whether the patient is *able to perform those activities independently*. The high correlation between the 'self-care' domain of the HAL and the 'arm functions' domain of the Dutch-AIMS2 does confirm this finding, because most activities within the 'self-care' domain involve the upper extremities. The upper extremity component, however, also correlated highly with $AIMS2_{phys}$. This can be explained by the overrepresentation of activities involving the upper extremities in the Dutch-AIMS2. The Dutch-AIMS2 is originally designed for patients with rheumatoid arthritis and it is generally known that these patients – in contrast to patients with haemophilia – experience problems with hand and finger functions, hence the orientation of the

Dutch-AIMS2 towards these specific tasks. The domains of the HAL relating to activities involving the lower extremities ('lying down/sitting/kneeling/standing' and 'functions of the legs'), as well as both component scores (HAL_{lowbas} and HAL_{lowcom}) were hypothesized to be associated with the 'mobility level' and 'walking and bending' scales of the Dutch-AIMS2, the 'autonomy indoors' and 'outdoors' domains of the IPA ($IPA_{auto\text{-}in}$ and $IPA_{auto\text{-}out}$), and the 50 metre walking test, the TUG test and the figure-8 walking test. These hypothesized associations were also confirmed by the data.

The convergent validity of the final HAL is good when considering the correlations with the performance tests, but lower than the convergent validity when compared to the other two questionnaires. The first two domains of the HAL correlated well with the 50 metre walking test and the TUG test, whereas the button test, although generally correlating poorly with the different domains, showed the best associations with the domains of 'arm functions' and 'self-care' and the 'upper extremity' component. These low correlations might question the validity of the HAL, but the results on the comparison between the HAL and the other two questionnaires contradict this. Moreover, no negative correlations were found, which means that a good perception of functional abilities goes together with 'fast' performances and vice versa. One might subsequently question the choice of the performance tests themselves. The tests were selected by the two haemophilia care experts to match the most common problems in patients with haemophilia. Two tests (50 metre walking test and the TUG test), were also suggested to be eligible for a future core-set in measuring patients with haemophilia [2]. We, therefore, feel confident that the performance tests are indeed suitable for our patient group. In addition, low to moderate correlations between self-report and performance measures, as found in our study, are more often reported in the literature [28–31], thereby questioning our choice for a cut-off point for considering correlations lower than 0.60. A study involving 93 patients with rheumatoid arthritis correlated the 50 metre test with several scales of a functional questionnaire [20] and found correlations ranging from −0.53 to −0.68. Explanations have been speculated on measurement errors, and the fact that performance tests are simplifications of the demands associated with real activities of daily life [29, 31]. Both self-report and performance measures contribute to the assessment of a patients' functional status on their own, and they might not measure the same concept [28, 31]; they can be considered complimentary. It is, therefore, recommended to use both types of instruments in clinical practice as well as for research purposes.

The fact that only patients with severe haemophilia were included in this study is a limitation of the generalization of the results. Additional validation of the HAL for use in moderate and mild haemophilia is needed, before the HAL can be used in all patient categories.

The final version of the HAL has good internal consistency and both construct and convergent validities. Moreover, its clinical applicability is good. Especially, this last finding is very important. Because correlations between the HAL and several performance tests were not very high, it is suggested that the HAL, other than the performance tests, assesses a different, but equally important concept of functional health status (i.e. perception). It is, therefore, recommended that both types of instruments are included when assessing a patient's functional health status. Future research regarding the HAL should focus on the reliability and responsiveness of the questionnaire and on generating reference values for the HAL.

References

1 De Kleijn P, Heijnen L, Van Meeteren NLU. Clinimetric instruments to assess functional health status in patients with haemophilia: a literature review. *Haemophilia* 2002; **8**: 419–27.

2 De Kleijn P, Van Genderen FR, Van Meeteren NLU. Assessing functional health status in adults with haemophilia: towards a preliminary core set of clinimetric instruments based on a literature search in rheumatoid arthritis and osteoarthritis. *Haemophilia* 2005; **11**: 308–18.

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

4 Dekker J, Dallmeijer AJ, Lankhorst GJ. Clinimetrics in rehabilitation medicine: current issues in developing and applying measurement instruments 1. *J Rehabil Med* 2005; **37**: 193–201.

5 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**: 565–71.

6 World Federation of Hemophilia. Extensive musculoskeletal program at Bangkok congress. *Hemophilia World* 2005; **12**: 10.

7 De Kleijn P, Heijnen L, Roosendaal G *et al.* Towards a new WFH standard for assessment of health status in haemophilia patients. *Haemophilia* 2000; **6**: 401.

8 De Kleijn P, Heijnen L, Van Meeteren NLU. From WFH-1 towards WFH-2, the kick off. In: *7th Musculoskeletal Congress*. 2001; **1**: 56–57. (abstract).

9 Hidding A, Van Santen M, De Klerk E *et al.* Comparison between self-report measures and clinical observations of functional disability in ankylosing spondylitis, rheumatoid arthritis and fibromyalgia. *J Rheumatol* 1994; **21**: 818–23.

10 Tugwell P, Bombardier C, Buchanan WW *et al.* The MACTAR Patient Preference Disability Questionnaire – an individualized functional priority approach for assessing improvement in physical disability in clinical trials in rheumatoid arthritis. *J Rheumatol* 1987; **14**: 446–51.

11 World Health Organization. *International Classification of Functioning, Disability and Health (ICF)*. Geneva: WHO, 2001.

12 Riemsma RP, Taal E, Rasker JJ *et al.* Evaluation of a Dutch version of the AIMS2 for patients with rheumatoid arthritis. *Br J Rheumatol* 1996; **35**: 755–60.

13 Van Meeteren NLU, Strato IHM, Van Veldhoven NHMJ *et al.* The utility of the Dutch Arthritis Impact Measurement Scales 2 for assessing health status in individuals with haemophilia: a pilot study. *Haemophilia* 2000; **6**: 664–71.

14 De Joode EW, Van Meeteren NLU, Van den Berg HM, De Kleijn P, Helders PJM. Validity of health status measurement with the Dutch Arthritis Impact Measurement Scale 2 in individuals with severe haemophilia. *Haemophilia* 2001; **7**: 190–7.

15 Cardol M, De Haan RJ, De Jong BA, Van den Bos GA, De Groot IJM. Psychometric properties of the Impact on Participation and Autonomy Questionnaire. *Arch Phys Med Rehabil* 2001; **82**: 210–16.

16 Cardol M, De Haan RJ, Van den Bos GA, De Jong BA, De Groot IJM. The development of a handicap assessment questionnaire: the Impact on Participation and Autonomy (IPA). *Clin Rehabil* 1999; **13**: 411–19.

17 Cardol M, De Jong BA, Van den Bos GA *et al.* Beyond disability: perceived participation in people with a chronic disabling condition. *Clin Rehabil* 2002; **16**: 27–35.

18 Klukowska A, Czyrny Z, Laguna P *et al.* Correlation between clinical, radiological and ultrasonographical image of knee joints in children with haemophilia. *Haemophilia* 2001; **7**: 286–92.

19 Pincus T, Brooks RH, Callahan LF. Reliability of grip strength, walking time and button test performed according to a standard protocol. *J Rheumatol* 1991; **18**: 997–1000.

20 Spiegel JS, Paulus HE, Ward NB *et al.* What are we measuring? An examination of walk time and grip strength. *J Rheumatol* 1987; **14**: 80–6.

21 Lindeman E, Leffers P, Reulen J, Spaans F, Drukker J. Reduction of knee torques and leg related functional abilities in hereditary motor and sensory neuropathy. *Arch Phys Med Rehabil* 1994; **75**: 1201–5.

22 Morris S, Morris ME, Iansek R. Reliability of measurement obtained with the timed 'up & go' test in people with Parkinson disease. *Phys Ther* 2001; **81**: 810–18.

23 Podsiadlo D, Richardson S. The Timed 'Up & Go' test: a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc* 1991; **39**: 142–8.

24 Hayes KW, Johnson ME. Measures of adult general performance tests. S28–S42. *Arthritis Care Res* 2003; **49**:

25 Creel GL, Light KE, Thigpen MT. Concurrent and construct validity of scores on the Timed Movement Battery. *Phys Ther* 2001; **81**: 789–98.

26 Guccione AA, Mielenz TJ, Devellis RF *et al.* Development and testing of a self-report instrument to measure actions: Outpatient Physical Therapy Improvement in Movement Assessment Log (OPTIMAL). *Phys Ther* 2005; **85**: 515–30.

27 Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford: Oxford Medical Publications, 1989.

28 Kennedy D, Stratford PW, Pagura SM, Walsh M, Woodhouse LJ. Comparison of gender and group differences in self-report and physical performance measures in total hip and knee arthroplasty candidates. *J Arthroplasty* 2002; **17**: 70–7.

29 Stratford PW, Kennedy D, Pagura SM, Gollish JD. The relationship between self-report and performance-related measures: questioning the content validity of timed tests. *Arthritis Rheum* 2003; **49**: 535–40.

30 Cress ME, Schechtman KB, Mulrow CD *et al.* Relationship between physical performance and self-perceived physical function. *J Am Geriatr Soc* 1995; **43**: 93–101.

31 Hoeymans N, Feskens EJ, Van den Bos GA, Kromhout D. Measuring functional status: cross-sectional and longitudinal associations between performance and self-report (Zutphen Elderly Study 1990–1993). *J Clin Epidemiol* 1996; **49**: 1103–10.

Appendix 1. Items of the Haemophilia Activities List (HAL). Each domain starts with the question: 'In the previous month, did you encounter any difficulty due to haemophilia with: '. Each item has a six-point Likert scale response option.

Domains *Items*	Upper Extremity Component	Lower Extremity Basic Component	Lower Extremity Complex Component
Lying/Sitting/Kneeling/Standing			
Sitting down (e.g. on a chair or couch)			
Rising from a chair with armrests			
Rising from a chair without armrests			✔
Kneeling/squatting			✔
Bending forward			
Kneeling for a longer period of time			✔
Squatting for a longer period of time			✔
Standing for a longer period of time		✔	
Leg Functions			
Walking short distances (less than 1 kilometre/15 minutes)		✔	
Walking long distances (more than 1 kilometre/15 minutes)		✔	
Walking on a soft surface (e.g. on the beach or through the woods)		✔	
Walking on an uneven surface (e.g. cobblestones, high sidewalks)		✔	
Strolling/(window-)shopping		✔	
Climbing up the stairs			✔
Climbing down the stairs			✔
Running (e.g. in order to catch the bus)			✔
Jumping			✔
Arm Functions			
Lifting heavy objects	✔		
Carrying heavy objects in the arms	✔		
Fine hand movements (e.g. sewing)	✔		
Reaching above your head (to pick something up from a high shelf)	✔		
Use of Transportation			
Riding a bicycle			✔
Getting in and out of a car			
Using public transportation (bus, train, subway)			
Self-Care			
Drying your whole body	✔		
Putting on a shirt, sweater etc.	✔		
Putting on sock and shoes	✔		
Putting on a tie or closing the upper button of a shirt	✔		
Toileting	✔		
Household Tasks			
Going out shopping (for food, drink etc.)			
Washing the dishes, cleaning the sink			
Cleaning the house			
Other household tasks (ironing, making the beds)			
Doing odd jobs (both in and around the house)			
Gardening			
Leisure Activities and Sports			
Playing games (outdoors, e.g. with your children)			
Sports			
Going out (theatre/museum/movie theatre/bar)			
Hobbies			
Dancing			
Going on a holiday (active)			
Going on a holiday ('passive'; beach-/hotel holiday)			

