Accepted: 2 October 2017

DOI: 10.1111/hae.13367

REVIEW ARTICLE

WILEY Haemophilia

Measuring activities and participation in persons with haemophilia: A systematic review of commonly used instruments

M. A. Timmer[1,2] | S. C. Gouw[3] | B. M. Feldman[4] | A. Zwagemaker[3] | P. de Kleijn[1,2] | M. F. Pisters[2,5] | R. E. G. Schutgens[1] | V. Blanchette[6] | A. Srivastava[7] | J. A. David[8] | K. Fischer[1] | J. van der Net[9]

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

[2]Physical Therapy Research, Department of Rehabilitation, Physical Therapy Science and Sport, Brain Center Rudolf Magnus, Utrecht University, The Netherlands

[3]Department of Pediatric Hematology, Academic Medical Center, Amsterdam, The Netherlands

[4]Division of Rheumatology, Department of Paediatrics, The Hospital for Sick Children, Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[5]Center for Physical Therapy Research and Innovation in Primary Care, Julius Health Care Centers, Utrecht, The Netherlands

[6]Department of Pediatrics, Division of Hematology/Oncology Hospital for Sick Children, University of Toronto, Toronto, ON, Canada

[7]Department of Haematology, Christian Medical College, Vellore, India

[8]Department of Physical Medicine and Rehabilitation, Christian Medical College, Vellore, India

[9]Department of Child Development and Exercise, University Medical Center Utrecht and Children's Hospital, Utrecht University, The Netherlands

Correspondence
Merel Timmer, Van Creveldkliniek, University Medical Center Utrecht, Utrecht, The Netherlands.
Email: m.a.timmer@umcutrecht.nl

Introduction: Monitoring clinical outcome in persons with haemophilia (PWH) is essential in order to provide optimal treatment for individual patients and compare effectiveness of treatment strategies. Experience with measurement of activities and participation in haemophilia is limited and consensus on preferred tools is lacking.

Aim: The aim of this study was to give a comprehensive overview of the measurement properties of a selection of commonly used tools developed to assess activities and participation in PWH.

Methods: Electronic databases were searched for articles that reported on reliability, validity or responsiveness of predetermined measurement tools (5 self-reported and 4 performance based measurement tools). Methodological quality of the studies was assessed according to the COSMIN checklist. Best evidence synthesis was used to summarize evidence on the measurement properties.

Results: The search resulted in 3453 unique hits. Forty-two articles were included. The self-reported Haemophilia Acitivity List (HAL), Pediatric HAL (PedHAL) and the performance based Functional Independence Score in Haemophilia (FISH) were studied most extensively. Methodological quality of the studies was limited. Measurement error, cross-cultural validity and responsiveness have been insufficiently evaluated.

Conclusion: Albeit based on limited evidence, the measurement properties of the PedHAL, HAL and FISH are currently considered most satisfactory. Further research needs to focus on measurement error, responsiveness, interpretability and cross-cultural validity of the self-reported tools and validity of performance based tools which are able to assess limitations in sports and leisure activities.

KEYWORDS

activities, haemophilia, measurement properties, outcome measure, participation

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2018 The Authors. *Haemophilia* Published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Bleeding into muscles and synovial joints is the main characteristic of haemophilia. Recurrent joint bleeds cause degenerative cartilage and bone changes (haemophilic arthropathy) through synovial inflammation and blood related cartilage damage.[1] This results in short- and long term limitations in joint function, performance of activities and participation in society. The available treatment includes on demand or prophylactic clotting factor replacement and functional rehabilitation after bleeding.[2]

Monitoring clinical outcome is essential in order to provide optimal individual treatment of persons with haemophilia (PWH) and to compare patient groups in scientific research. Self-reported bleeding and clinical and radiologic joint assessment used to be the main outcome measures in haemophilia treatment. However, according to the World Health Organization's International Classification of Functioning (ICF),[3] health is considered the result of an interaction between body structure and function, activities, participation and personal and environmental factors. In order to obtain a representative impression of a person's health assessment of all ICF domains, combining objective and self-reported measurement tools, is recommended.[4] According to the ICF "activity" is defined as "the execution of a task or action by an individual" and participation is defined as "involvement in a life situation". Consistent monitoring outcome at activity and participation level still needs to find its way into haemophilia care and research.

Various haemophilia specific and generic tools for assessing activity and participation are available, but consensus on a preferred set of measurement tools is not yet reached. Differences in access to expensive clotting factor concentrates cause significant differences in joint status between PWH in different parts of the world. Differences between age groups, severity and cultural differences also contribute to heterogeneity in the haemophilia population. Moreover, measurement tools might be used for different purposes, eg. monitoring clinical outcome vs comparing treatment groups and evaluation of chronic complaints vs acute bleeds. A core set of measurement tools should account for these different purposes and patient groups. Recently, a first recommendation for a core set based on expert opinion was published.[5] However, quality assessment of the outcome measures, considered an important step in the development of a core outcome set,[6] was incomplete.

The aim of the current study is to provide a systematic overview of the measurement properties (ie, validity, reliability and responsiveness) of a selection of measurement tools commonly used in PWH and developed to measure activities and participation. This information may promote the identification of core-set outcome tools and direct further research on these measurement tools.

2 | METHODS

This systematic review is reported in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement (www.prisma-statement.org). The research group aimed to perform systematic reviews on (i) imaging techniques to assess haemophilia arthropathy, (ii) tools to measure joint structure and function, (iii) tools to measure activity and participation and (iv) measurement tools to assess health related quality of life (HRQoL). For clarity of reporting, separate reviews are drafted for the different domains. The current manuscript reports on tools used to assess activities and participation.

Measurement tools included in the search were predetermined. The selection of measurement tools was made based on a survey among comprehensive haemophilia treatment centers (Columbus, USA; Denver, USA; Milan, Italy; Utrecht, The Netherlands; Valencia, Spain; Vellore, India; 2014) combined with an expert meeting (Toronto, October 2014)[5] and by the authors. Included measurement tools were: the Canadian Occupation Performance Measure (COPM), the Haemophilia Activity List (HAL), the Impact on Participation and Autonomy questionnaire (IPA), the International Physical Activity Questionnaire (IPAQ), the Pediatric HAL (PedHAL), accelerometry, the enhanced Functional Independence Score in Haemophilia (eFISH), the Functional Independence Score in Haemophilia (FISH), the 6 Minute Walk Test (6MWT) and the Timed Up and Go test (TUG).

2.1 | Search

Medline and Embase were searched until May 30th 2016. The full search, designed and supervised by a professional librarian (C. Nickel, Hospital Library and Archives, The Hospital for Sick Children, Toronto, Canada), is listed in Appendix S1. For the purpose of reviews on all domains, the initial search (October 2014) included measurement instruments on joint scores, imaging, health related quality of life, as well as activities and participation. The search update (May 30th 2016) was focused at joint scores and activities and participation. Reference lists were screened and experts in the field were contacted to identify studies not retrieved by the literature search.

2.2 | Study selection

Studies that reported on reliability, validity or responsiveness of one of the predetermined measurement tools used in pediatric and adult patients with severe, moderate or mild haemophilia A or B were eligible for inclusion. It was decided to also consider studies that used one of the selected measurement instruments as an outcome as these would include information on discriminant validity. Only peer-reviewed original articles written in English were considered.

Titles and abstracts of retrieved studies were independently examined by 2 investigators (S.G. and M.T.). Complete manuscripts of potentially relevant studies were examined for eligibility for inclusion. In case of disagreement between the 2 reviewers, consensus was reached by discussion with a third reviewer (K.F).

2.3 | Data extraction

Data extraction from manuscripts was performed independently by 2 investigators (S.G. and A.Z). In the case of disagreement between the 2

reviewers consensus was reached by discussion with a third reviewer (M.T.). The following data were extracted: sample size, proportion of severe PWH, proportion of patients receiving prophylactic therapy, age, spread the results in the population being studied and information about measurement properties. Measurement properties are defined according to COSMIN, as shown in Table 1. Interpretability, including Minimal Clinical Important Difference (MCID) and floor- and ceiling effects, was considered an important requirement for the suitability of an instrument, but not a measurement property. Therefore it was only described and not scored.

2.4 | Quality assessment

First, the methodological quality of the selected studies was evaluated according to the COSMIN checklist, which offers standards to assess each measurement property.[7] For each measurement property the COSMIN includes items that assess design requirements and statistical methods, the actual content of the items is specific for each measurement property. Every item is rated on a 4-point scale (poor, fair, good, excellent) and the lowest score determines the rating. Quality assessment of the studies was performed by 2 investigators independently (S.G. and A.Z.). In the case of disagreement between the 2 reviewers consensus was reached by discussion with a third reviewer (M.T.). Since haemophilia is a rare disease and reliability studies of performance-based measurement tools require a smaller sample size compared to self-reported tools[8] (as the COSMIN is initially designed for), it was decided to adjust the COSMIN criterion of minimum sample size to score fair from 30 to 20 for reliability, measurement error and hypothesis testing.[9,10]

Secondly, results of the measurement properties were assessed as positive, negative or indeterminate based on criteria of Terwee et al,[11] as shown in Table 1. Thirdly, the results of the studies were combined and adjusted for the methodological quality of the studies according to the methods of the Cochrane Back Review Group,[12] shown in Table 2. Overlap in study population between studies was checked for by comparing source population, inclusion period and inclusion criteria. In case of uncertainty, authors were contacted. Studies using the same population were considered as 1 study in the overall quality assessment.

3 | RESULTS

The study selection process is shown in a flowchart in Figure 1. Using the above search strategy, **3453** unique references were identified. After screening of the titles and abstracts, **288** articles were selected for full text inspection and **44** studies were included in the current paper. Included studies reported on the COPM, the HAL, the IPA, the IPAQ, the PedHAL, accelerometry the FISH, the 6MWT and the TUG. No studies were found that reported on the eFISH. The included studies comprised 1398 PWH, of which 40.5% were adults and were 30.3% children (<18 years). In 29.2% of the cases it remained unclear if it considered children or adults. In total 71.5% of the included PWH had severe haemophilia; the remaining patients had mild or moderate haemophilia. In less than 1% of the cases severity was not described.

Study characteristics and measurement properties are presented in Table 3, for studies that scored at least "fair" on the COSMIN checklist. Most studies reported on hypothesis testing (convergent and discriminant validity). Minimal clinical important differences (MCID) were not reported. Internal consistency and structural validity were not applicable for the included performance based tools since the (sub) scales are not based on a reflective model but on a formative model (which means that the items together form the construct and are therefore not interchangeable and not necessarily correlated). Cross-cultural validity was also considered not applicable for performance based tools, although we are aware that performance on certain activities can be culturally influenced. All studies assessed functioning in a steady state condition, 12 studies[13-24] explicitly excluded patients with acute bleeding; none reported assessment of patients after (sub) acute bleeding. Methodological quality ratings of each study are reported per measurement property in Appendix S2 and were generally fair to poor. Synthesis of the results for each measurement tool, including the level of evidence, is presented for the total patient group in Table 4. Separate reports for adults and children are presented in Appendix S3 A, B.

It was decided not to differentiate between tools measuring activities and tools measuring participation as this requires comprehensive research and is therefore beyond the scope of the current study.

3.1 | Self-reported measurement tools

3.1.1 | Canadian Occupation Performance Measure (COPM)

Occupation Performance Measure is a standardized semi-structured interview measuring self-perception of performance and satisfaction with the performance of tasks of daily living, leisure and productivity.[25] Patients identify problems with tasks and prioritize them according to the level of importance. The 5 most important problems become goals for treatment planning. For these 5 problems individuals rate their current level of performance and their satisfaction of their performance on a scale from 1 to 10 (lower scores represent more difficulties or less satisfaction). An average score for performance and satisfaction of the 5 problems is computed at the end of the interview process.

The COPM was only reported once in PWH,[18] in a study aiming to describe how the COPM was applied, along with other measures of assessment. Hypothesis testing was considered "unknown" due to poor methodological quality.

3.1.2 | Haemophilia Activity List (HAL)

The HAL is a haemophilia-specific questionnaire assessing self-perceived limitations in activities in adults due to haemophilia, in the previous month.[24] It contains 42 items across 7 domains. A summary

TABLE 1 Definitions and quality criteria (Based on Terwee et al[10]) for measurement properties

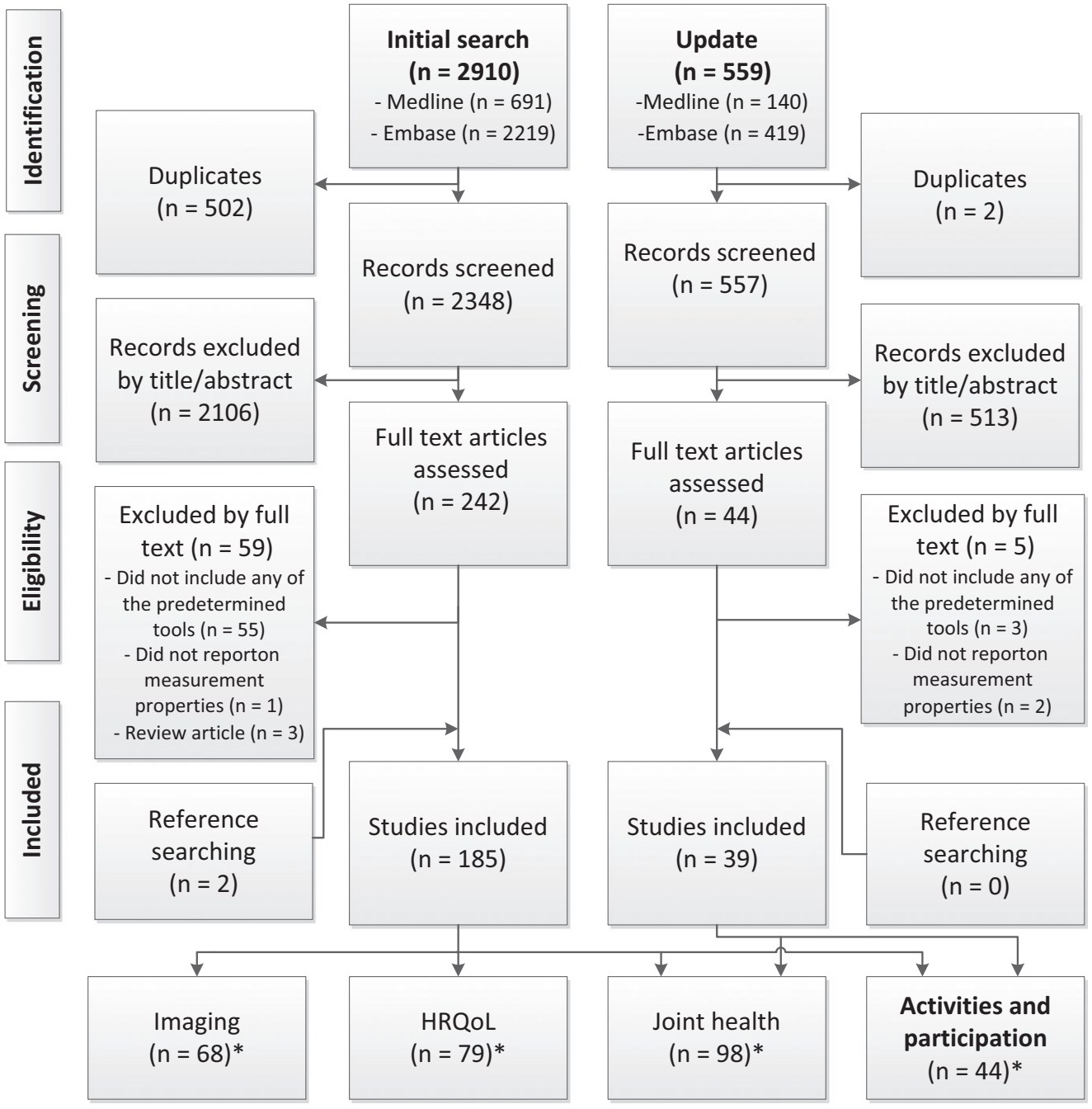
Property	Definition	Rating	Quality criteria
Reliability			
Internal consistency	The degree of interrelatedness among items and is generally calculated by Cronbach's alpha	+	(Sub)scale unidimensional AND Cronbach's alpha(s) > 0.70
		?	Dimensionality not known OR Cronbach's alpha not determined
		−	(Sub)scale not unidimensional OR Cronbach's alpha(s) < 0.70
Measurement error	Systematic and random error of a person's score that is not attributed to true changes in the construct and can be calculated bythe standard error of measurement (SEM), limits of agreements (LoA) or smallest detectable change (SDC)	+	MIC > SDC OR MIC outside the LOA
		?	MIC not defined
		−	MIC < SDC OR equals or inside LOA
Reliability	The proportion of variance due to true differences between persons and includes test-retest reliability, inter-rater reliability and intra-rater reliability	+	ICC/weighted Kappa > 0.70 OR Pearson's r > .80
		?	Neither ICC/weighted Kappa, nor Pearson's determined
		−	ICC/weighted Kappa < 0.70 OR Pearson's r < .80
Validity			
Content validity	The degree to which an instrument is an adequate reflection of the construct to be measured. This could be done by asking experts (eg. the target population) to judge the relevance of the items	+	The target population considers all items in the questionnaire to be relevant AND considers the questionnaire to be complete
		?	No target population involvement
		−	The target population considers items in the questionnaire to be irrelevant OR considers the questionnaire to be incomplete
Construct validity			
Cross- cultural validity	The degree to which the performance of items on a translated or culturally adapted instrument is an adequate reflection of the original version	+	Original factor structure confirmed OR no important DIF
		?	Confirmation original factor structure AND DIF not mentioned
		−	Original factor structure not confirmed OR important DIF
		+	Factors should explain at least 50% of the variance
Structural	The degree to which the scores of an instrument are an adequate reflection of the dimensionality of the construct to be measured	?	Explained variance not mentioned
		−	Factors explain <50% of the variance
Validity			
Hypothesis testing	The degree to which scores are consistent with pre-specified hypotheses and is evaluated in terms of convergent validity (correlation with demographic variables or instruments measuring the same construct) and discriminant validity (differences between known groups)	+	(Correlation with an instrument measuring the same construct > .50 OR at least 75% of the results are in accordance with the hypotheses) AND correlation with related constructs is higher than with unrelated constructs
		?	Solely correlations determined with unrelated construct
		−	Correlation with an instrument measuring the same construct <.50 OR <75% of the results are in accordance with the hypotheses OR correlation with related constructs is lower than with unrelated constructs
Responsiveness	The ability to detect change over time	+	(Correlation with an instrument measuring the same construct > .50 OR at least 75% of the results are in accordance with the hypotheses OR AUC > 0.70) AND correlation with related constructs is higher than with unrelated constructs
		?	Solely correlations determined with unrelated constructs
		−	Correlation with an instrument measuring the same construct <.50 OR >75% of the results are in accordance with the hypotheses OR AUC < 0.70 OR correlation with related constructs is lower than with unrelated constructs

MIC, minimal important change; SDC, smallest detectable change; LOA, limits of agreement; ICC, intraclass correlation coefficient; DIF, differential item functioning; AUC, area under the curve; +, positive rating; ?, indeterminate rating; −, negative rating.

TABLE 2 Levels of evidence for the quality of the measurement property[11]

Level	Rating	Criteria
Strong	+++ or −	Consistent findings in multiple studies of good methodological quality OR in 1 study with excellent methodological quality
Moderate	++ or −	Consistent findings in multiple studies of fair methodological quality OR in 1 study of good methodological quality
Limited	+ or −	One study of fair methodological quality
Conflicting	±	Conflicting findings
Unknown	?	Only studies with poor methodological quality

+, positive result; −, negative result.



*Studies could be assigned to multiple domains; HRQoL = Health Related Quality of Life

FIGURE 1 Flowchart of studyselection

TABLE 3 Characteristics and measurement properties of selected studies, scored as at least "fair" on the COSMIN checklist, per measurement tool

Author	N/% severe/% PR	Age in years	Distribution scores	Reliability (reliability, measurement error, internal consistency)	Validity (structural validity, discriminant validity, convergent validity, content validity, cross- cultural validity)
HAL					
Van Genderen et al 2004[23] (provisional version)	50/86/na	Mean 44.9 (SD 14.2)	Median 83.3 (IQR 77.5-98.2)	n.a	Content validity: Patients considered it to reflect daily activities better than dutch-AIMS2 and IPA. Convergent validity: Strong correlation with AIMS2 (r = .89; P < .001) and IPA(r = .75; P < .001)
Van Genderen et al 2005[22] (A)	43/100/na	Mean 45 (SD 14)	Median 68.7 (55.3-79.3)	n.a	Convergent validity: adjusted for age and psychological health: Pettersson score and pain explained 52.3% of the variance in HAL. HAL explained 61.5% of the variance on IPA. (all P < .05)
Van Genderen et al 2006[31] (A)	127/100/na	Median 42.0 (IQR 31.0-51.0)	Median 69.29 (IQR 57.6-90.5)	n.a	Convergent validity: strong positive correlation with AIMS2 (r = .81), IPA (r = .71) and 50 m walking (r = .60). Moderate positive correlation with TUG (r = .59) and figure 8 (r = .54) (all P < .001)
Van Genderen et al 2006[20] (A)	78/100/na	Median 40.35 (IQR 32-52)	Median 70.2 (IQR 54.8-90)	n.a	Convergent validity: Age, PR and pain explained 65.7% of the variance in HAL
Brodin et al 2011[27] (Swedish version) (B)	84/72/81	Mean 45.0 (SD 17.7)	Moderate: Median 91 (range 78) Severe: median 72 (range 89)	n.a	Convergent validity: strong positive correlation with AIMS2 physical (r = .84; P < .01), IPA indoor (r = .83; P < .01) and IPA mobility (r = .89; P < .01). Discriminant validity: No difference between moderate (median 9 range 79) and severe (median 28 range 89) (P > .05)
Biere-Rafi et al 2011[25]	30/13/na	Mean 55 (range 22-74)	Normal weight: median 97.9 Obese: median 88.4	n.a	Discriminant validity: Obese patients scored lower than normal weight patients (mean 88.4 vs. 97.9; P = .024). BMI explained 21% of the variance in HAL In only non-severe patients obese patients scored also lower than normal weight (mean 90.1 vs. 99.5; P = .017)
Van Genderen et al 2012[21] (A)	115/100/74	Median 42 (range 18-72)	Median 69.5 (IQR 57.1-90.5)	n.a	Convergent validity: After adjusting for age and psychological health, pain did not explain variance in HAL. 44% of the variance in HAL was explained by age, 6% by joint mobility (6%), and 2% by psychological health. HAL explained 25% of variation in IPA
Nijdam et al 2016[34] (C)	56/100/74	Median 32.4	n.a	n.a	Discriminant validity: No differences between patients that stopped PR (mean 84, IQR 67-98) and that continued PR (mean 84, IQR 78-97), P = .99. Patients that interrupted PR scored lower (mean 67, IQR 49-79) than the other groups, but too small the sample size to perform statistical analysis (n = 6)

(Continues)

TABLE 3 (Continued)

Author	N/% severe/% PR	Age in years	Distribution scores	Reliability (reliability, measurement error, internal consistency)	Validity (structural validity, discriminant validity, convergent validity, content validity, cross- cultural validity)
Fischer et al 2016[4] (C)	90/100/89	Median 25.5(IQR 20.9- 30.4)	Median 96 (IQR 84–100) 31.1% optimum score	n.a.	Convergent validity: Moderate correlation with sf36 physical functioning (r = .43) and sf36 utility (r = .41). Weak (r < .4) correlation with joint bleeds, eq5d and other components of sf36
Fischer et al 2013[33]	81/100/85	Mean 24.5 (range 14-37)	Netherlands: median 93 (IQR 81-98), Sweden: median 99 (IQR 93-100),	n.a	Discriminant validity: The high dose treatment group showed less limitations in activities (median 99) than the intermediate dose group (median 93), P < .01
Sherlock et al 2010[30] (D)	61/49/na	Mean 38 (range 16-63)	n.a	n.a	Convergent validity: Weak positive correlation between upper extremity and IPAQ (r = .36; P < .001) and between lower extremity and IPAQ (r = .297; P < .001) Discriminant validity: PWH spend less time in moderate PA than controls (mean 152.7 ± 167.2 vs. 318.7 ± 115.5; CI −195.2 to 136.8). No difference in vigorous PA (mean 141.1 ± 145.6 vs. 281 ± 631.15; CI = −299.5 to 17.5), walking (mean 444.0 ± 156.5 vs. 600 ± 878.3; CI −53.5 to 511.5), and sitting (mean 2262.0 ± 1326.8 vs. mean 2033.0 ± 1115.7; CI −53.5 to 511.5) Total MET of PWH was lower than controls (mean 4319.7 ± 3491.2 vs. 6505.9 ± 7332.54; CI 4030.9 to 343.9)
Poonnoose et al 2007[29]	20/100/na	Mean 14 (range 7-40)	Mean 71 (range 31-97)	n.a	Convergent validity: moderate positive correlation with Gibert score (r = .48;**), no correlation with Pettersson (r = .08;**)
IPA					
Van Genderen et al 2004[23] (A)	50/86/na	Mean 44.9 (SD 14.2)	Mean 31.2 (SD 19.5; range 0-77)	n.a	Convergent validity: Strong positive correlation with HALprovisional version (r = .75; P < .001)
Van Genderen et al 2005[22] (A)	43/100/na	Mean 45 (SD 14)	Median 23.1 (IQR 8.4-38.5)	Good internal consistency: autonomy indoors 0.93, family role 0.93, autonomy outdoors 0.90, social relations 0.90, and work and educational opportunities 0.93, sum score 0.97	Convergent validity: Adjusted for age and psychological health HAL explained 61.5% variance in IPA (P < .001) and pain and Pettersson score explained 47.9% variance in IPA (P = .005)
Van Genderen et al 2006[31] (A)	127/100/na	Median 42.0 (IQR 31.0-51.0)	Median 30.0 (IQR 9.5-43.0)	n.a	Convergent validity: Strong positive correlation with HAL (r = .71; P < .001)

(Continues)

TABLE 3 (Continued)

Author	N/% severe/% PR	Age in years	Distribution scores	Reliability (reliability, measurement error, internal consistency)	Validity (structural validity, discriminant validity, convergent validity, content validity, cross- cultural validity)
Brodin et al 2011[27] (B)	84/72/81	Mean 45.0 (SD 17.7)	n.a	n.a	Convergent validity: Strong positive correlation IPA indoor with HAL (r = .83; P < .01) and IPA outdoor (r = .89; P < .01) with HAL. LOA*: IPA indoor with HAL -24/50; IPA outdoor with HAL -19/38; IPA family with HAL -21/35
Van Genderen et al 2012[21] (A)	87/100/74	Median 42 (range 18-72)	Median 28 (IQR 10.0-37.5)	n.a	Convergent validity: Body functions and structures explained none of the variation in IPA, age explained 31% and psychological health 4%. Performance-based activity measures explained 3% variation in IPA. HAL explained 25% of variation in IPA
IPAQ					
Sherlock et al 2010[30] (D)	61/49/na	Mean 38 (range 16-63)	46% high, 28% moderate, 16% low PA levels	n.a	Convergent validity: Weak positive correlation with HAL upper extremity (r = .362; P < .001) and HAL lower extremity (r = .297; P < .001). Weak negative correlation with age (r = −.266; P < .05). Discriminant validity: PWH spend less time in moderate PA than controls (mean 152.7 ± 167.2 vs. 318.7 ± 115.5; CI −195.2 to 136.8). No difference in vigorous PA (mean 141.1 ± 145.6 vs. 281 ± 631.15; CI = −299.5 to 17.5), walking (mean 444.0 ± 156.5 vs. 600 ± 878.3; CI −53.5 to 511.5), and sitting (mean 2262.0 ± 1326.8 vs. mean 2033.0 ± 1115.7; CI −53.5 to 511.5). Total MET of PWH was lower than controls (mean 4319.7 ± 3491.2 vs. 6505.9 ± 7332.54; CI 4030.9 to 343.9)
Fischer et al 2013[33]	75/100/85	Mean 24.5 (range 14-37)	Intermediate dose: Median IPAQ 4294 (IQR 1037-13740) High dose: Median IPAQ 3200 (IQR 1152-9292)	n.a	Discriminant validity: No difference in physical activity between intermediate (median 4294, IQR 1037-13740) and high dose (median 3200, IQR 1152-9292). P = .50
PedHAL					
Groen et al 2010[35]	Validity 22/75/75	Mean 8.9 (SD 3.1)	Parents: mean 95 (SD 9; range 69-100) Children: mean 97 (SD 7; range 73-100) >50% max score	n.a	Content validity: Items considered relevant and questionnaire considered complete by health professionals, patients and caregivers

(Continues)

TABLE 3 (Continued)

Author	N/% severe/% PR	Age in years	Distribution scores	Reliability (reliability, measurement error, internal consistency)	Validity (structural validity, discriminant validity, convergent validity, content validity, cross- cultural validity)
Groen et al 2013[36] (E)	Reliability 22 responsiveness 19 Validity 29/86/0	Mean 13.2 (SD 4.0)	Median 83.5 (IQR 47.9-90.5) 34% optimum score	No difference between test and retest scores (P = .72). LOA- 0.7/17.4 IC= 0.95 (CI 0.88-0.98)	Convergent validity: Strong positive correlation with FISH (r = .65; P < .05). Moderate negative correlation with HJHS (r = −.59; P < .05) and CHQ-50 (physical function-ing) (r = .40; P < .05). No correlation with CHQ-50(behaviour) (r = .36; P > .05), CHQ-50(mental health) (r = .16; P > .05) and 6MWT (r = .13; P > .05)
Radzevic et al 2013[37] (F)	24/88/88	Mean 12.6 (SD 3.01)	Mean 83.6 (SD 11.4; range 58.86-97.29)	n.a	Convergent validity: Strong negative correlation with HJHS (r = −.962; P < .0001) and 6MWT (r = .903; P < .0001)
Accerelometer					
Goto et al 2014[55] (Omron, triaxial)	32/27/na	Self- moniotoring group mean 41.8 (SD 8.6), control group mean 43.9 (SD 10.7)	Self-monitoring group: mean 2.9 (SD 2.0) Controls: mean 2.5 (SD 1.9)	n.a	Discriminant validity: No difference in Ex (MET × h) between self-monitoring group and controls in locomotor PA (mean 2.4 SD 1.8 vs. mean 1.8 SD 1.5; P = .263), non-locomotor PA (mean 0.5 SD 0.8 vs. mean 0.7 SD 0.5; P = .348), total ex (mean 2.9 SD 2.0 vs. mean 2.5 SD 1.9; P = .517), steps (mean 5805.6 SD 3384.0 vs. mean 4910.2 SD 2663.5; P = .412) and step times (mean 86 SD 45.6 vs. mean 75.0 SD 34.9; P = .453)
Gonzales et al 2014[13] (Actigraph, triaxial)	41/12/15 + 25 healthy controls	Range 8-18	n.a	n.a	Discriminant validity: PWH had higher total PA than controls (mean 652.63 SEM 33.74 vs. mean 430.82 SEM 30.63; P < .001). PWH had lower sedentary behaviour (mean 356.78 SEM 16.6 vs. 479.41; P < .001), lower light PA (mean 450.24 SEM 18.68 vs. mean 339.63 SEM 19.62; P < .001) and higher moderate PA (mean 8.48 SEM 1.15 vs. mean 3.36 SEM 0.86; P = .001) than controls. No difference in vigorous PA (mean patients SEM 0.06 vs. mean controls 0.41 SEM 0.10; P > .05)
Buxbaum et al 2010[52] (Axtritrac, biaxial)	17/9/na	mean 13.71 (SD 2.1)	n.a	n.a	Convergent validity: Moderate correlation between social influences and sedentary behaviour (r = .59), high PA (r = −.59) and vigorous PA (r = −.51) and between self-efficacy and sedentary behaviour (r = .57), high PA (r = −.68) and vigorous PA (r = −.78) (all P < .05). Severe patients spend more time in sedentary behaviour than mild (P = .039). Discriminant validity: There was no difference between severe and mild in moderate PA (P = .81), high PA (P = .18) and vigorous PA (P = .29)

(Continues)

TABLE 3 (Continued)

Author	N/% severe/% PR	Age in years	Distribution scores	Reliability (reliability, measurement error, internal consistency)	Validity (structural validity, discriminant validity, convergent validity, content validity, cross- cultural validity)
Takken et al 2010[53] (Actiheart, uniaxial and heartrate)	63 with chronic disease (10 PWH)/ na/na	mean PWH 12.5 (SD 3.31)	Mean energy expenditure 208.4 (SD 118.2)	LOA with indirect calorimetry -144.2/148.4	Criterion validity: With mixed linear regression accelerometer counts, heart rate above sleep and gender were significant predictors for Activity Energy Expenditure (indirect calorimetry)($P < .05$) Strong correlation between indirect calorimetry and total output ($r = .81$), no correlation with sitting behind computer ($r = .11$), weak with sweeping($r = .27$) and hallway walking ($r = .31$), moderate with steps ($r = .51$), slow walking ($r = .41$) and moderate walking ($r = .47$), strong with fast walking ($r = .69$)
FISH					
Tasbihi et al 2016[46]	25/100/na	Median 63 (range 11-70)	Median 20 (range 16-28)	n.a	Convergent validity: Moderate correlation with MRI additive scale ($r = -.537$), weak correlation with MRI progressive scale ($r = .240$) and ultrasound ($r = -.365$)
Kachooie et al 2014[42]	133/80/14	Mean 26.9 (SD 14.24)	n.a	n.a	Discriminant validity: The odds of being categorized as disabled increases with 7.34 for patients with severe haemophilia (CI 1.32-40.67), with 1.07 for every year in age (CI 1.03-1.11) and with 9.75 for patients with an inhibitor (CI 1.03-92.12)
Ingerslev et al 2014[41]	40/100/33	Danish (PR) mean 23.2 (SD 5.0) Age-matched Russian (OD): 22.8 (SD 4.8) All Russian mean 23.3 (SD 4.2)	n.a	n.a	Discriminant validity: Patients on PR scored better than age matched patients with OD treatment (mean 27.8 ± 0.6 vs, mean 4.8 ± 2.2; $P < .0001$). No difference with the total OD population (mean 24.5 ± 1.9; $P = .66$)
Poonnoose et al 2007[29]	63/100/na	Mean 14 (range 7-40)	Mean 25.6 (range 7-32)	Good Interrater reliability (ICC = 0.98)	Convergent validity: Strong correlation with Gilbert score ($r = -.61^{**}$), HAQ ($r = -.75^{**}$), HAL ($r = -.66^{**}$) and WOMAC ($r = -.75^{**}$), moderate correlation with Pettersson score ($r = -.38^{**}$)
Tlacuilo-parra 2010 [19]	60/38/0 + 30 controls	Mean 10.0 (SD 4.0)	Mean 25.8 (SD 3.6) 65% optimum score	n.a	Discriminant validity: Severe patients had worse FISH scores than mild patients (mean 24.08 ± 4.74 vs 28 ± 0; $P < .05$) and moderate patients had worse scores than mild patients (26.27 ± 2.54 vs 28 ± 0; $P < .05$). There was no difference between mild patients and controls ($P > .05$)

(Continues)

TABLE 3 (Continued)

Author	N/% severe/% PR	Age in years	Distribution scores	Reliability (reliability, measurement error, internal consistency)	Validity (structural validity, discriminant validity, convergent validity, content validity, cross- cultural validity)
Hassan et al 2011[40]	50/50/22	mean 16 (SD 1.1)	Mean 23.32 (SD 4.69; range 13-28)	n.a	Convergent validity: Strong negative correlation with Pettersson score knee (r = −.896), ankle (r = −.877) and elbow (r = −.848) (all P < .001). Strong negative correlation with MRI knee (r = −.973), ankle (r = −.941) and elbow (r = −.946) (all P < .001). Strong negative correlation with BDI-SF (r = −91; P < .001). Discriminant validity: Lower scores in severe patients compared to moderate and mild (mean 20.52 ± 3.97 vs. mean 24.37 ± 2.43 and mean 27.90 ± 0.32; P < .001). Lower scores in patients with weekly joint bleeds compared to every 2 wk (mean 16.2 ± 2.1 vs mean 20.3 ± 2.1; P < .001). Higher scores in patient with regular FVIII compared to OD (27 ± 0.8 vs. mean 18.4 ± 2.9; P < .001)
Groen et al 2013[36] (E)	29/86/0	Mean 13.2 (SD 4.0)	Median 32.0 (IQR 25.0-32.0)	n.a	Convergent validity: Strong positive correlation with PedHAL (r = .65)
6MWT					
Douma-van Riet et al 2009[12]	113/41/43	Mean 12.7 y (SD 2.9)	Mean z-score -0.02 (SD 0.95)	n.a	Convergent validity: Not correlated with the maximum exercise test on cyclo ergometer (r = .07). Mean score of patients with overweight was lower than of patients with normal weight (zscore -0.8 (1.4) vs. 0.1 (1.0); P = .01)
Hassan et al 2010[16]	113 with chronic disease (47 PWH)/89/na	PWH mean 12.5 (SD 2.9)	Mean 628 m	n.a	Convergent validity: Moderate correlation with age (r = .49), height (r = .49) and weight (0.34). Multiple regression analysis showed height to be the best predictor of 6MWD, explaining 24% of the variance. PWH achieved a greater distance than JIA and SB (628 vs 459 and 391 m; P < .01)
Radzevic et al 2013[37] (F)	24/88/88	Mean 12.6 (SD 3.01)	Mean 408.46 (SD 68)	n.a	Convergent validity: Strong negative correlation with PedHAL (r = .903; P < .0001)
TUG					
Van Genderen et al 2006[31] (A)	127/100/na	Median 42.0 (IQR 31.0-51.0)	Median 16.7 (IQR 14.7–20.4)	n.a	Convergent validity: Moderate correlation with HAL sum score (r = .59) and HAL domains Basic Lower Extremity (r = .55) and Complex Lower Extremity (r = .62)
Fearn et al 2010[49]	20/70/35 + 20 healthy controls	Mean 39.4 (range 22-58) controls ages matched	Mean 11.61 (range 9.84–13.38)	n.a	Convergent validity: Moderate correlations between TUG and laboratory measures of walking speed (r = .424), step width (r = .311) and stepping and turning sway (r = .530). PWH were slower on TUG than healthy controls (11.61 sec. vs 7.78 sec; P < .001

PR, prophylaxis; OD, on demand; TUG, Timed Up & Go; IPA, Impact on Participation and Autonomy; AIMS2, Arthritis Impact Measurement 2; HAL, Haemophilia Activity List; PA, physical activity; LOA, Limits Of Agreement; n.a, no information of sufficient quality (at least fair at COSMIN checklist) available; FISH, Functional Independence Score in Haemophilia; HAL, Haemophilia Acitivity List; PWH, persons with haemophilia.

Studies that report on multiple measurement tools are described for every tool separately; (A), (B), (C), (D), (E), (F) studies used (partly) the same patients.

TABLE 4 Quality of measurement properties (best evidence synthesis)

Measu rement tool	Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Hypothesis testing (convergent/ discriminant validity)	Cross-cultural validity	Criterion Validity	Responsiveness
HAL	?		?	++	?	±	?		
PedHAL	?	+		+		±			
IPAQ						±			
COPM	?								
IPA	+					++			
FISH	n.a	+			n.a	++	n.a		?
TUG	n.a				n.a	++	n.a		
6MWT	n.a				n.a	±	n.a		?
Accelerometer	n.a				n.a	±	n.a	±	

+++ or −, strong evidence positive/negative result; ++ or −, moderate evidence positive/negative result; + or −, limited evidence positive/negative result; ±, conflicting evidence; ?, unknown; due to poor methodological quality; n.a., measurement property not applicable for measurement tool; empty entries, no information available; 6MWT, Minute Walk Test; COPM, Occupation Performance Measure; FISH, Functional Independence Score in Haemophilia; HAL, Haemophilia Acitivity List; IPA, Impact on Participation and Autonomy questionnaire; IPAQ, International Physical Activity Questionnaire; MET, metabolic equivalents of task; PedHAL, Pediatric HAL; TUG, Timed Up and Go test.

score as well as component scores involving upper extremity, basic lower extremity and complex lower extremity activities can be calculated. Normalized scores range from 0 to 100, with higher scores representing better functional status.

The HAL was used in 11 studies.[4,21-24,26-36] Three [31,32,35] included both children and adults. Another 3 studies explicitly aimed at investigating measurement properties.[24,29,33] Moderate positive evidence was found for content validity of the HAL. Conflicting results were found for hypothesis testing due to a good correlation with IPA, Arthritis Impact Measurement Scale (AIMS), TUG and 50 m walking speed but a low correlation with IPAQ and the physical functioning domain of the SF36. Furthermore, the HAL discriminates well between treatment groups[35] but not between patients who stopped or continued prophylaxis.[36] Internal consistency, structural validity and cross-cultural validity were considered "unknown" due to poor quality of the selected studies. When used as an outcome measure in an Indian study[31] most patients did not answer all questions as they considered them inappropriate. Reported proportions of persons achieving an optimal score vary according to age and intensity of treatment, ranging from 7% (only severe haemophilia, 76% prophylaxis, median 38 years) to 31% (only severe haemophilia, 89% prophylaxis, median 41 years).[4,28]

3.1.3 | Impact on Participation and Autonomy questionnaire (IPA)

The IPA is a generic questionnaire for adults, addressing personal impact of illness on participation and autonomy and related experience of problems.[37] This self-administered questionnaire consists of 31 items, distributed over 5 domains: autonomy indoors, family role, autonomy out-doors, social relations, and work and educational opportunities. Scores range from 0 to 120 with a higher score representing more restrictions in participation or worse autonomy.

Seven studies reported on the IPA, all included adult patients.[22-24,28,29,33,34] None of the studies explicitly aimed to investigate measurement properties. In 3 of them, the IPA was used to explore validity of the HAL.[21,24,29] Limited positive evidence was found for the internal consistency of the IPA in this population. Moderate positive evidence was found for hypothesis testing due to good correlations with HAL in multiple studies.

3.1.4 | International Physical Activity Questionnaire (IPAQ)

The IPAQ is a 7-day recall questionnaire, developed for adults, measuring current levels of physical activity.[38] Walking, moderate-intensity and vigorous-intensity activities are assessed for 4 different domains – leisure time, domestic and gardening, work-related and transport-related physical activity. Scoring of the IPAQ results in a continuous variable in the form of total metabolic equivalents of task (MET)-min per week, as well as categorization into high, moderate or low physical activity level.

Four studies reported on the IPAQ in PWH,[32,35,39,40] none of the studies explicitly aimed to investigate measurement properties. Three studies included adults as well as teenagers (age 16 year and above),[32,35,39] the other only included adults.[40] Conflicting evidence was found for hypothesis testing due to a weak correlation with the HAL and differences in moderate physical activity and total METs between PWH and controls but not in vigorous physical activity, light physical activity and sedentary behaviour.[32] Ceiling effects are not applicable for the IPAQ.

3.1.5 | Pediatric Haemophilia Activity List (PedHAL)

The PedHAL is a haemophilia specific measure assessing self-perceived limitations in activities in children between 8-18 years old.[41] It contains 53 items across 7 domains and can be completed by both parents and children. A summary score as well as component scores involving upper extremity, basic lower extremity and complex lower extremity activities can be calculated. Normalized scores, ranging from 0 to 100 with higher scores representing better functional status, can be obtained for the summary score and the component scores.

A total of 3 studies reported on the PedHAL.[41-43] Two studies explicitly aimed to investigate measurement properties.[41,42] Limited positive evidence was found for content validity of the PedHAL. Internal consistency, reliability, measurement error and hypothesis testing are considered "unknown" due to poor quality of the selected studies. Two studies reported a proportion of optimum scores of 34% (81% severe haemophilia, no prophylaxis, 13.2 years) and >50% (75% severe haemophilia, all prophylaxis, 8.9 years).[41,42]

3.2 | Performance based measurement tools

3.2.1 | Accelerometer

The accelerometer is an objective tool to assess physical activity. Currently a broad range of accelerometers is available.[44] The primary unit of measurement is counts or m/s^2, which can be converted into MET using an algorithm. One MET is equal to the resting metabolic rate, obtained during quite sitting.

Five studies reported on accelerometers; 4 only included children,[14,45-47] one only adults.[48] One study explicitly aimed to validate an accelerometer (Actiheart, uniaxial accelerometer combined with heart-rate) for the measurement of energy expenditure in children with different chronic diseases. In this study conflicting evidence was found for criterion validity due to a good correlation between Actiheart and indirect calorimetry for total activities and fast walking and a weak correlation between Actiheart and indirect calorimetry for sitting, slow walking and moderate walking.[46] The remaining studies showed conflicting evidence for hypothesis testing in different age groups. Ceiling effects are not applicable.

3.2.2 | Functional Independence Score in Haemophilia (FISH)

The FISH is a performance based test measuring the patient's independence in performing activities of daily living, transfers and mobility.[19] It was designed for adults and children above 7 years. Each function is graded from 1 to 4 depending on the amount of assistance the patient needs in performing the function. Total score ranges from 0 (functionally fully dependent) to 32 (functionally fully independent).

Fifteen studies reported on the FISH.[16,18-20,31,42,49-57] Two studies explicitly aimed to investigate measurement properties of the FISH.[19,31] Five studies included both children and adults,[18,19,31,51,55] 7 studies included only children,[16,20,42,49,52,54,56] 3 included only adults.[50,53,57] Moderate positive evidence was found for hypothesis testing in the total group, due to a good correlation with HAL, Western Ontario and McMaster Universities Arthritis Index (WOMAC) and Health Assessment Questionnaire (HAQ). In studies including children only, evidence for hypothesis testing was also considered moderately positive. Responsiveness and internal consistency were considered unknown due to poor methodological quality. Proportions of optimum scores reported range from 8% (only severe haemophilia, no prophylaxis, median 32 years)[16] to >50% (86% severe, no prophylaxis, 13.2 years).[42] One study showed optimum scores >50% of the FISH and optimum scores of 34% of the pedHAL in the same population.[42]

3.2.3 | Six Minute Walk Test (6MWT)

The 6MWT is a performance based test that measures walking speed and sub-maximal exercise capacity.[58] It was developed for patients with respiratory diseases and heart failure, but has been used in children and adults with a variety of chronic conditions. Patients are instructed to walk up and down a 30-50 meter track for 6 minutes and try to cover the largest possible distance without running. The covered distance in 6 minutes is recorded. The use of a walking aid or orthosis is allowed.

Six studies reported on the 6MWT,[13,17,42,43,59,60] 1 included adults[59,60] the others children only. The studies did not explicitly aim to investigate measurement properties. Conflicting evidence was found for hypothesis testing due to the ability of the 6MWT to discriminate between patients with and without overweight and between PWH, juvenile idiopathic arthritis and spina bifida vs the weak correlation between 6MWT and a maximum exercise test (cyclo-ergometer).[13,17,43] Responsiveness was only reported in studies with poor methodological quality and therefore considered unknown.

3.2.4 | Timed Up and Go Test (TUG)

The TUG is a performance based test developed to assess functional mobility in elderly persons.[61] Participants are required to stand up from a chair, walk 3 meters, return and sit in the chair. The time taken to perform the test is administered.

Three studies reported on the TUG, all included adult PWH.[26,33,62] These studies did not explicitly aim to validate the TUG in this population. Limited positive evidence was found for hypothesis testing due to 1 study with fair methodological quality showing a good correlation with the HAL.

4 | DISCUSSION

Five self-reported and 4 performance based measurement tools in 44 studies were investigated. Eight studies explicitly aimed at investigating measurement properties. For most measurement tools only hypothesis testing was investigated sufficiently. Two self-reported (the HAL and PedHAL) and 1 performance-based tool FISH were studied somewhat more extensively but with low COSMIN scores. Measurement error, cross-cultural validity, responsiveness and interpretability were rarely reported. For this review, no difference was made between tools measuring activities or participation.

4.1 | Self-reported measurement tools

The PedHAL is the only tool specifically developed for children. Given the measurement properties studied, the PedHAL seems a convenient tool to measure self-reported limitation in activities in children. The IPA is a validated tool for adults in several disease groups [63] and was used to validate the HAL.[24,29] The HAL and the IPA aim to measure a different construct, limitations in activities and impact on participation respectively. However, most participation questionnaires also include items on limitations in activities[64] and the HAL includes items on participation (eg. sports and going out). Cross-cultural validity of the HAL was not formally assessed, however when used as an outcome measure in an Indian study most patients did not answer all questions as they considered them inappropriate.[31] The IPAQ is considered a validated tool to measure physical activity in healthy adults[38] but correlates only weakly with the HAL.[32] This can be explained by the difference in construct between the HAL (limitations in activities) and IPAQ (level of physical activity). Differences between PWH and controls are found in some, but not all, aspects of physical activity. This might be due to large inter-individual differences, even within controls.[32]

4.2 | Performance based measurement tools

Given the measurement properties studied, the FISH seems a convenient tool to measure independence in performing activities of daily living, transfers and mobility in both children and adults. However, ceiling effects >50% were found in a population with 34% optimum scores on the PedHAL. Those differences can be explained by the differences in construct (self-perceived limitation measured by the PedHAL and observed limitations measured by the FISH) and by differences in the range of activities that is covered by the instruments; the FISH covers activities of daily living, transfers and mobility whereas the PedHAL also covers leisure and sports activities. The TUG is validated for frail elderly, patients with arthritis and stroke patients.[65] In middle aged patients with severe haemophilia it correlates moderately with HAL.[33,62] Measurement properties of the 6MWT in children vary among chronic conditions.[66] The reported poor correlation with maximum exercise tests in children with haemophilia is in accordance with results in children with juvenile idiopathic arthritis.[67] Possibly the 6MWT is not sensitive enough for patients with minimal joint complaints, indicating a ceiling effect. In populations with more severe joint complaints, like adults awaiting total knee or hip arthroplasty, reliability and responsiveness of the 6MWT is good.[68] Measurement properties of accelerometers vary among brands, but a wide range is validated for the assessment of physical activity in healthy children.[44] However, when accelerometers are used to predict energy expenditure significant over- and underestimation is found.[44] Similar results were found in children with haemophilia.[46]

4.3 | Strengths and limitations

Strengths of the current study are the systematic literature search and the independent study selection and independent methodological quality assessment by 2 investigators. Furthermore, we decided to include also studies that used one of the selected measurement instruments as an outcome. This enabled us to include more information about the measurement tools (on hypothesis testing), which was needed since studies that aimed at investigating measurement properties were limited. The different aim of these studies required different priorities in reporting, leading to low COSMIN scores. Nevertheless, consisting findings in studies scoring "fair" on the COSMIN checklist are also considered moderated evidence.

The main limitation of this systematic review is the lack of a suitable checklist for methodological quality assessment of studies investigation measurement properties of performance based tools. The COSMIN checklist was initially developed for self-reported outcome measures and is therefore less appropriate for performance-based tools. However, the COSMIN checklist seemed to be the most appropriate solution. By adjusting the minimal sample size of the COSMIN for hypothesis testing and reliability we attempted to make it more suitable for performance-based tools.[8] A second limitation of this study is that measurement tools included in the search were predetermined by expert opinion. Although we were only interested in commonly used tools, we could have missed a measurement tool. Given our broad international survey among comprehensive haemophilia treatment centers and expert meeting with an international and multidisciplinary group of professionals we are confident that all relevant tools are included.

4.4 | Clinical implications and future research

Based on this systematic review, the PedHAL and HAL are the recommended self-reported measurement tools. The FISH is the recommended performance based tool to measure independence in the performance of activities of daily living, transfers and mobility. When developing a core set of measurements, including the same instruments for a wide population of PWH is preferable in order to enable comparison between groups. Both in patient care and in clinical research 1 might wish to additionally add specific measurements for specific patients or purposes. Further research needs to focus on measurement

error, responsiveness, interpretability, cross-cultural validity of the self-reported tools and validity of performance based tools that are able to assess limitations in leisure and sports activities. Furthermore, weak to moderate correlations between self-reported and performance based measures of activities and participation suggest that different constructs are being measured and that there is a need to use both for full assessment.[4] Although in this study no difference was made between (items within) tools measuring activities or participation, both aspects should be represented in a core outcome set. Future studies should investigate to what extent the included tools measure activities or participation. Finally, in all selected studies functioning in a steady state condition was measured. Further research towards measurement tools which assess and monitor recovery after an acute bleed is needed in order to guide treatment of an (sub)acute bleed.

5 | CONCLUSION

Several self-reported and performance based measurement tools are available to assess activities and participation in PWH. Albeit based on limited evidence, the measurement properties of the PedHAL, HAL and FISH are currently considered most satisfactory. Based on the review, we recommend the PedHAL and HAL as self-reported measurement tools to measure self-perceived limitations in activities and the FISH as performance based tool to measure independence in daily activities, transfers and mobility.

DISCLOSURES

SCG has received research funding from Bayer, Baxter, Novo Nordisk and CSL Behring. BF has received research funding from Novo Nordisk, Bayer and Baxalta. All other authors stated that they had no interests which might be perceived as posing a conflict or bias.

ORCID

M. A. Timmer http://orcid.org/0000-0003-1910-999X
A. Zwagemaker http://orcid.org/0000-0002-8848-1229
A. Srivastava http://orcid.org/0000-0001-5032-5020

REFERENCES

1. Jansen NW, Roosendaal G, Lafeber FP. Understanding haemophilic arthropathy: an exploration of current open issues. *Br J Haematol*. 2008;143:632-640.
2. Hermans C, De MP, Fischer K, et al. Management of acute haemarthrosis in haemophilia A without inhibitors: literature review European survey and recommendations. *Haemophilia*. 2011;17:383-392.
3. WHO. International classification of functioning, disability, and health. World Health Organization (WHO), ICF Geneva 2001.
4. Fischer K, Nijdam A, Holmstrom M, et al. Evaluating outcome of prophylaxis in haemophilia: objective and self-reported instruments should be combined. *Haemophilia*. 2016;22:80-86.
5. Fischer K, Poonnoose P, Dunn AL, et al. Choosing outcome assessment tools in haemophilia care and research: a multidisciplinary perspective. *Haemophilia*. 2016;23:11-24.
6. Prinsen CA, Vohra S, Rose MR, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - a practical guideline. *Trials*. 2016;17:449.
7. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*. 2012;21:651-657.
8. Dobson F, Hinman RS, Hall M, Terwee CB, Roos EM, Bennell KL. Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a systematic review. *Osteoarthritis Cartilage*. 2012;20:1548-1562.
9. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther*. 2005;85:257-268.
10. Hscsbwgdn TB. *Designing clinical research: an epidemiologic approach*, 4th edn. Philadelphia: Lippincott Williams & Wilkins; 2013.
11. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60:34-42.
12. Furlan AD, Pennick V, Bombardier C, van Tulder M. 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. *Spine*. 2009;34:1929-1941.
13. Douma-van Riet DC, Engelbert RH, van Genderen FR, Ter Horst-De Ronde MT, de Goede-Bolder A, Hartman A. Physical fitness in children with haemophilia and the effect of overweight. *Haemophilia*. 2009;15:519-527.
14. Gonzalez LM, Peiro-Velert C, Devis-Devis J, et al. Comparison of physical activity and sedentary behaviours between young haemophilia A patients and healthy adolescents. *Haemophilia*. 2011;17:676-682.
15. 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-e912.
16. Gupta V, Sandeep HS, Rai A, Shukla J. Functional and radiological assessment of arthropathy in Indian children with haemophilia. *Haemophilia*. 2013;19:e373-e375.
17. Hassan J, van der Net J, Helders PJ, Prakken BJ, Takken T. Six-minute walk test in children with chronic conditions. *Br J Sports Med*. 2010;44:270-274.
18. Padankatti SM, Macaden AS, Cherian SM, et al. A patient-prioritized ability assessment in haemophilia: the Canadian Occupational Performance Measure. *Haemophilia*. 2011;17:605-611.
19. Poonnoose PM, Manigandan C, Thomas R, et al. Functional Independence Score in Haemophilia: a new performance-based instrument to measure disability. *Haemophilia*. 2005;11:598-602.
20. Tlacuilo-Parra A, Villela-Rodriguez J, Garibaldi-Covarrubias R, Soto-Padilla J, Orozco-Alcala J. Functional independence score in hemophilia: a cross-sectional study assessment of Mexican children. *Pediatr Blood Cancer*. 2010;54:394-397.
21. van Genderen FR, Fischer K, Heijnen L, et al. Pain and functional limitations in patients with severe haemophilia. *Haemophilia*. 2006;12:147-153.
22. van Genderen FR, van Meeteren NL, Fischer K, Helders PJ. Determinants of participation in patients with severe haemophilia. *Haemophilia*. 2012;18:860-867.
23. van Genderen FR, van Meeteren NL, Heijnen L, van den Berg HM, Helders PJ. The use of a disability model in haemophilia research. *Haemophilia*. 2005;11:472-480.
24. 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;10:565-571.
25. Law M, Baptiste S, McColl M, Opzoomer A, Polatajko H, Pollock N. The Canadian occupational performance measure: an outcome measure for occupational therapy. *Can J Occup Ther*. 1990;57:82-87.

26. Baumgardner J, Elon L, Antun A, et al. Physical activity and functional abilities in adult males with haemophilia: a cross-sectional survey from a single US haemophilia treatment centre. *Haemophilia*. 2013;19:551-557.
27. Biere-Rafi S, Haak BW, Peters M, Gerdes VE, Buller HR, Kamphuisen PW. The impairment in daily life of obese haemophiliacs. *Haemophilia*. 2011;17:204-208.
28. Binnema M, Schrijvers LH, Bos R, Schuurmans MJ, Fischer K. Coping in adult patients with severe haemophilia. *Haemophilia*. 2014;20:513-518.
29. Brodin E, Baghaei F, Elfvinger P, Lindvall K, Sunnerhagen KS. The Swedish version of the Haemophilia Activity List. *Haemophilia*. 2011;17:662-668.
30. De Kleijn P, Sluiter D, Vogely HC, Lindeman E, Fischer K. Long-term outcome of multiple joint procedures in haemophilia. *Haemophilia*. 2014;20:276-281.
31. Poonnoose PM, Thomas R, Keshava SN, et al. Psychometric analysis of the Functional Independence Score in Haemophilia (FISH). *Haemophilia*. 2007;13:620-626.
32. Sherlock E, O'Donnell JS, White B, Blake C. Physical activity levels and participation in sport in Irish people with haemophilia. *Haemophilia*. 2010;16:e202-e209.
33. 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.
34. Brodin E, Baghaei F, Sunnerhagen KS. Self-reported activity and functioning in daily life; the perspective of persons with haemophilia living in Sweden. *Eur J Haematol*. 2015;95:336-341.
35. Fischer K, Carlsson KS, Petrini P, et al. Intermediate-dose vs high-dose prophylaxis for severe hemophilia: comparing outcome and costs since the 1970s. *Blood*. 2013;122:1129-1136.
36. 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.
37. Cardol M, de Haan RJ, van den Bos GA, de Jong BA, de Groot IJ. The development of a handicap assessment questionnaire: the Impact on Participation and Autonomy (IPA). *Clin Rehabil*. 1999;13:411-419.
38. Craig CL, Marshall AL, Sjostrom M, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc*. 2003;35:1381-1395.
39. den Uijl I, Biesma D, Grobbee D, Fischer K. Turning severe into moderate haemophilia by prophylaxis: are we reaching our goal? *Blood Transfus*. 2013;11:364-369.
40. Niu X, Poon JL, Riske B, et al. Physical activity and health outcomes in persons with haemophilia B. *Haemophilia*. 2014;20:814-821.
41. Groen WG, van der Net J, Helders PJ, Fischer K. Development and preliminary testing of a Paediatric Version of the Haemophilia Activities List (pedhal). *Haemophilia*. 2010;16:281-289.
42. Groen W, van der Net J, Lacatusu AM, Serban M, Helders PJ, Fischer K. Functional limitations in Romanian children with haemophilia: further testing of psychometric properties of the Paediatric Haemophilia Activities List. *Haemophilia*. 2013;19:e116-e125.
43. Radzeui V, Raistenskis J, Ragelient L, Kowalski IM. Relationship between physical activity and functional ability in school-aged children with hemophilia. *Pol Ann Med*. 2013;20:13-18.
44. De Vries SI, Van Hirtum HW, Bakker I, Hopman-Rock M, Hirasing RA, Van Mechelen W. Validity and reproducibility of motion sensors in youth: a systematic update. *Med Sci Sports Exerc*. 2009;41:818-827.
45. Buxbaum NP, Ponce M, Saidi P, Michaels LA. Psychosocial correlates of physical activity in adolescents with haemophilia. *Haemophilia*. 2010;16:656-661.
46. Takken T, Stephens S, Balemans A, et al. Validation of the Actiheart activity monitor for measurement of activity energy expenditure in children and adolescents with chronic disease. *Eur J Clin Nutr*. 2010;64:1494-1500.
47. Walker RG, Obeid J, Nguyen T, et al. Sedentary Time and Screen-Based Sedentary Behaviors of Children With a Chronic Disease. *Pediatr Exerc Sci*. 2015;27:219-225.
48. Goto M, Takedani H, Haga N, et al. Self-monitoring has potential for home exercise programmes in patients with haemophilia. *Haemophilia*. 2014;20:e121-e127.
49. Hassan TH, Badr MA, El-Gerby KM. Correlation between musculoskeletal function and radiological joint scores in haemophilia A adolescents. *Haemophilia*. 2011;17:920-925.
50. Ingerslev J, Lethagen S, Hvitfeldt Poulsen L, et al. Long-standing prophylactic therapy vs. episodic treatment in young people with severe haemophilia: a comparison of age-matched Danish and Russian patients. *Haemophilia*. 2013;20:58-64.
51. Kachooei AR, Badiei Z, Zandinezhad ME, et al. Influencing factors on the functional level of haemophilic patients assessed by FISH. *Haemophilia*. 2014;20:185-189.
52. Gupta S, Garg K, Singh J. Assessment of Musculoskeletal Function and its Correlation with Radiological Joint Score in Children with Hemophilia A. *Indian J Pediatr*. 2015;82:1101-1106.
53. Hua B, Lian X, Li K, Lee A, Poon MC, Zhao Y. Low-dose tertiary prophylactic therapy reduces total number of bleeds and improves the ability to perform activities of daily living in adults with severe haemophilia A: a single-centre experience from Beijing. *Blood Coagul Fibrinolysis*. 2016;27:136-140.
54. Oymak Y, Yildirim AT, Yaman Y, et al. The effectiveness of tools for monitoring hemophilic arthropathy. *J Pediatr Hematol Oncol*. 2015;37:e80-e85.
55. Tasbihi M, Pishdad P, Haghpanah S, Ardeshiri R, Tavoosi H, Karimi M. A comparison between MRI, sonography and Functional Independence Score in Haemophilia methods in diagnosis, evaluation and classification of arthropathy in severe haemophilia A and B. *Blood Coagul Fibrinolysis*. 2016;27:131-135.
56. Hassan TH, Badr MA, Fattah NR, Badawy SM. Assessment of musculoskeletal function and mood in haemophilia A adolescents: a cross-sectional study. *Haemophilia*. 2011;17:683-688.
57. Ferreira AA, Bustamante-Teixeira MT, Leite IC, Correa CS, Rodrigues Dde O, da Cruz DT. Clinical and functional evaluation of the joint status of hemophiliac adults at a Brazilian blood center. *Rev Bras Hematol Hemoter*. 2013;35:23-28.
58. Balke B. A Simple Field Test for the Assessment of Physical Fitness. Rep 63-6. *Rep Civ Aeromed Res Inst US*. 1963;63:1-8.
59. 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.
60. Zlah WF. Effect of partial weight bearing program on functional ability and quadriceps muscle performance in hemofilic knee arthritis. *Egypt J Med Hum Genet*. 2013;14:413-418.
61. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc*. 1991;39:142-148.
62. Fearn M, Hill K, Williams S, et al. Balance dysfunction in adults with haemophilia. *Haemophilia*. 2010;16:606-614.
63. Cardol M, de Haan RJ, de Jong BA, van den Bos GA, de Groot IJ. Psychometric properties of the Impact on Participation and Autonomy Questionnaire. *Arch Phys Med Rehabil*. 2001;82:210-216.
64. Perenboom RJ, Chorus AM. Measuring participation according to the International Classification of Functioning, Disability and Health (ICF). *Disabil Rehabil*. 2003;25:577-587.
65. Hayes KW, Johnson ME. Measures of adult general performance tests. *Arthritis Care Res (Hoboken)*. 2013;49:28-42.
66. Bartels B, de Groot JF, Terwee CB. The six-minute walk test in chronic pediatric conditions: a systematic review of measurement properties. *Phys Ther*. 2013;93:529-541.

67. Lelieveld OT, Takken T, van der Net J, van Weert E. Validity of the 6-minute walking test in juvenile idiopathic arthritis. *Arthritis Rheum*. 2005;53:304-307.
68. Kennedy DM, Stratford PW, Wessel J, Gollish JD, Penney D. Assessing stability and change of 4 performance measures: a longitudinal study evaluating outcome following total hip and knee arthroplasty. *BMC Musculoskelet Disord*. 2005;6:3.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: 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–e49. https://doi.org/10.1111/hae.13367