Received: 18 February 2018 | Revised: 20 April 2018 | Accepted: 4 May 2018

DOI: 10.1111/hae.13631

ORIGINAL ARTICLE

Clinical haemophilia

WILEY Haemophilia

Measurement of joint health in persons with haemophilia: A systematic review of the measurement properties of haemophilia-specific instruments

Samantha C. Gouw[1,2] | Merel A. Timmer[3] | Alok Srivastava[4] | Piet de Kleijn[3,5] | Pamela Hilliard[6] | Marjolein Peters[1] | Victor Blanchette[7] | Kathelijn Fischer[3]

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

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

[3]Van Creveldkliniek, Department of Hematology, University Medical Center, Utrecht, The Netherlands

[4]Department of Hematology, Christian Medical College, Vellore, India

[5]Department of Rehabilitation, Physical Therapy Science and Sport, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, the Netherlands

[6]Department of Rehabilitation, Hospital for Sick Children, Toronto, Ontario, Canada

[7]Department of Paediatrics, Division of Hematology/Oncology, Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada

Correspondence
Samantha C. Gouw, Department of Paediatric Hematology, Academic Medical Center, Amsterdam, The Netherlands.
Email: s.c.gouw@amc.uva.nl

Introduction: Accurate assessment of joint health in persons with haemophilia is crucial. Several haemophilia-specific measurement tools are available, but an overview of the measurement properties is lacking.

Aim: To provide an overview of the measurement properties of haemophilia-specific measurement tools to assess clinical joint health.

Methods: MEDLINE and EMBASE were searched for reports on reliability, validity or responsiveness of the World Federation of Haemophilia Orthopedic Joint Score (WFH), Colorado Physical Examination Score (CPE), joint examination score by Petrini (PJS) and Hemophilia Joint Health Score (HJHS). Methodological quality of the studies was assessed using an adapted COSMIN checklist.

Results: The search yielded 2905 unique hits, and 98 papers were included. The methodological quality of the included studies was limited. The HJHS was studied most extensively, which yielded limited evidence for good internal consistency and structural validity, moderate evidence for hypothesis testing in adults and conflicting evidence for hypothesis testing in children. Reliability, measurement error and responsiveness were rated unknown due to low COSMIN scores. For the CPE and PJS, we found limited to moderate evidence for good responsiveness and conflicting evidence for hypothesis testing.

Conclusion: Only patchy evidence is available on the quality of measurement properties of all haemophilia-specific joint health scores. Although significant gaps in the evidence for all instruments remain, measurement properties of the HJHS were most extensively studied and show no drawbacks for use in clinical practice. This review forms the basis for further research aimed at the assessment of measurement properties of measurement tools to assess joint health.

KEYWORDS

arthropathy, haemophilia, reliability, responsiveness, validity

1 | INTRODUCTION

Recurrent haemorrhages into muscles and joints, particularly in ankles, knees and elbows, are the hallmark of haemophilia. Exposure to blood leads to degeneration of joint cartilage and bone, resulting in haemophilic arthropathy, that is characterized by chronic pain, loss of range of motion and muscle atrophy, leading to reduction in activities and social participation. These outcomes are dependent on the severity of haemophilia and inhibitor status of persons with haemophilia.

Joint health refers to the structural integrity and function of a joint. Within the World Health Organization's International Classification of Functioning, Disability and Health (ICF) model, joint health is classified in the domain of "body functions and structures." For a complete evaluation of health, the World Health Organization recommends to assess all domains of the organization's ICF model. The use of a set of high quality instruments for the measurement of all ICF domains is expected to optimize management of individual patients with haemophilia and improve comparability between outcome studies.

Several instruments have been developed that aim to measure joint health in persons with haemophilia. The Musculoskeletal Working Group of the World Federation of Haemophilia and other consensus groups have suggested core sets of outcome assessment instruments to measure functional and physical status of persons with haemophilia, including joint health. (https://elearning.wfh.org/resource/compendium-of-assessment-tools).[1,2] These choices were the result of consensus procedures, and all are experience based. Several reviews on the instruments' measurement properties have been published while systematic appraisal of available data, including quality assessment of the included studies, was lacking.[3-5] Recently, systematic reviews on measurement properties of health-related quality of life instruments and for instruments measuring activities and participation did include methodological quality appraisal using the COSMIN checklist.[6,7] However, a systematic review is still lacking for instruments that aim to assess joint health.

The current review aims to obtain a comprehensive overview of the measurement properties that is, validity, reliability and responsiveness of haemophilia-specific instruments used for the assessment of joint health by physical examination in adult and paediatric patients with all types of haemophilia and to identify gaps in the current knowledge on measurement properties in order to direct further research.

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). Study inclusion criteria and methodological quality criteria were prespecified in a protocol. The initial search included joint examination scores as well as instruments for the assessment of activities, participation and health-related quality of life. At an early stage, it was decided to report the results of the systematic review according to topic; measurement tools for assessment of joint health, activity and participation and health-related quality of life are reported separately.[6,7]

2.1 | Selection of joint assessment instruments

The joint assessment instruments were identified by a survey distributed among several comprehensive haemophilia treatment centres (Columbus, USA; Denver, USA; Milan, Italy; Utrecht, The Netherlands; Valencia, Spain; Vellore, India), supplemented by a survey during an expert meeting held in Toronto, October 2014.[2] Four joint assessment instruments were selected: the physical examination score as recommended by the Orthopedic Advisory Committee of the World Federation of Haemophilia (WFH score),[8] the Colorado Physical Examination score (CPE),[9] the adjusted physical examination score as reported by Petrini [Petrini joint score (PJS)],[10] and different versions of the Hemophilia Joint Health Score (HJHS).[11]

2.2 | Study selection criteria

Studies that reported on reliability, validity or responsiveness of one of the joint scores were included. Studies published as an article in a peer-reviewed journal in English without publication date restrictions were eligible for inclusion. In addition to studies that explicitly reported on the development or investigation of measurement properties of joint scores, studies that used joint scores as an outcome measure were included as these findings contribute to the evidence on validity. Studies including children and adults with severe, moderate or mild haemophilia A or B were considered for inclusion. Results of studies among healthy children or adults were only described but not included in the overall rating of measurement properties. The following measurement properties were evaluated: (i) Validity; structural validity and hypotheses testing. Cross-cultural validity, content validity and criterion validity were not applicable to joint scores and were therefore not evaluated, (ii) Reliability, internal consistency, reliability and measurement error and (iii) Responsiveness. (see Table S1 for definitions[12]).

2.3 | Information sources and literature search

Studies were identified by searching electronic databases, by screening the bibliographic references of retrieved studies and review papers, and by contacting experts in the field. MEDLINE and EMBASE were initially searched on 7 October 2014. A search update was performed on 30 May 2016. The full search for each database is listed in Appendix S1 (Search strategy). The search was designed and supervised by professional librarians (CN and R.S).

2.4 | Study selection

The study selection, data extraction and methodological quality appraisal were performed by two investigators independently (SG

and MT). In case of disagreement between the two reviewers, consensus was reached by discussion, with a third reviewer (KF).

Studies were selected by assessing titles and abstracts of studies retrieved by the literature search. All potentially relevant studies were retrieved as complete manuscripts and examined for compliance with the inclusion criteria.

2.5 | Data collection

Overlap in study populations between studies was checked by the authors' names, authors' affiliations, catchment areas, time of inclusion and inclusion criteria of the participants. A structured electronic data collection form was used to collect the following data: patient inclusion criteria, study design, number of included patients, setting, study period, number of patients with severe haemophilia A, number of patients receiving prophylaxis, number of patients with inhibitors, measurement properties.

2.6 | Quality assessment of individual studies

The methodological quality of the selected studies was evaluated according to the 4-point COSMIN checklist.[13] For each measurement property, the COSMIN includes items that assess design requirements and statistical methods. Every item is rated on a 4-point scale (poor, fair, good, excellent), and the lowest score determines the overall rating. We adjusted the minimal sample size for reliability, measurement error and hypotheses testing from 30 (as stated in the COSMIN checklist) to 20 to score "fair" instead of "poor." This was justified as haemophilia is a rare disease and reliability studies of performance-based measurement tools require a smaller sample size than that of self-reported tools for which the COSMIN is primarily designed.[14,15] Further, as missingness is not applicable in joint assessment tools, two COSMIN items on missingness were excluded.

2.7 | Data synthesis

Summary of the evidence was performed by two investigators independently (SG and K.F). First, results of the measurement properties reported in the individual studies were rated as positive, negative or indeterminate based on criteria as shown in Table S1.[12] Secondly, the level of evidence was summarized, combining results over all studies for each measurement property, taking into account the methodological quality of the studies, the number of available studies and the consistency of the results (Table 1).[16]

2.8 | Additional analyses

Subgroup analyses were performed for adult and paediatric patients.

3 | RESULTS

3.1 | Study selection

Figure 1 presents the flow chart of the study selection process. Using the above search strategies, 2905 unique references were identified. After screening of the titles and abstracts, 286 unique articles were selected for full-text inspection. After reviewing full-text papers, 96 studies were eligible for inclusion. Two additional studies were included after screening reference lists of selected studies.[8,9] Overlapping study populations were excluded when no additional data was provided.

3.2 | Study characteristics

In total, 98 studies were included. Study characteristics and the overall methodological quality assessment per measurement property are presented in Table S2. A total of 58 studies (17 paediatric, 16 adult and 25 combined populations) reported on the WFH joint score, seven studies (five paediatric, one adult and one combined study population) on the Colorado Physical Examination scale, six studies (five paediatric and one combined study populations) on the joint assessment score by Petrini and 32 (20 paediatric, five adult and seven combined study populations) studies reported on the HJHS. Measurement properties of more than one joint examination score were reported in five studies.[9,17-20] The COSMIN score of all studies was assessed as "fair" to "poor." Eight studies explicitly aimed at investigating the measurement properties of the used joint assessment instruments (one CPE and seven HJHS). [9,11,19,21-25] Most

TABLE 1 Levels of evidence for the overall quality of the measurement property

Level	Rating	Criteria
Strong	+++ or −−−	Consistent findings in multiple studies of good methodological quality OR in one study with excellent methodological quality
Moderate	++ or −−	Consistent findings in multiple studies of fair methodological quality OR in one 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.
Based on the GRADE approach by the Cochrane Back Review Group[16].

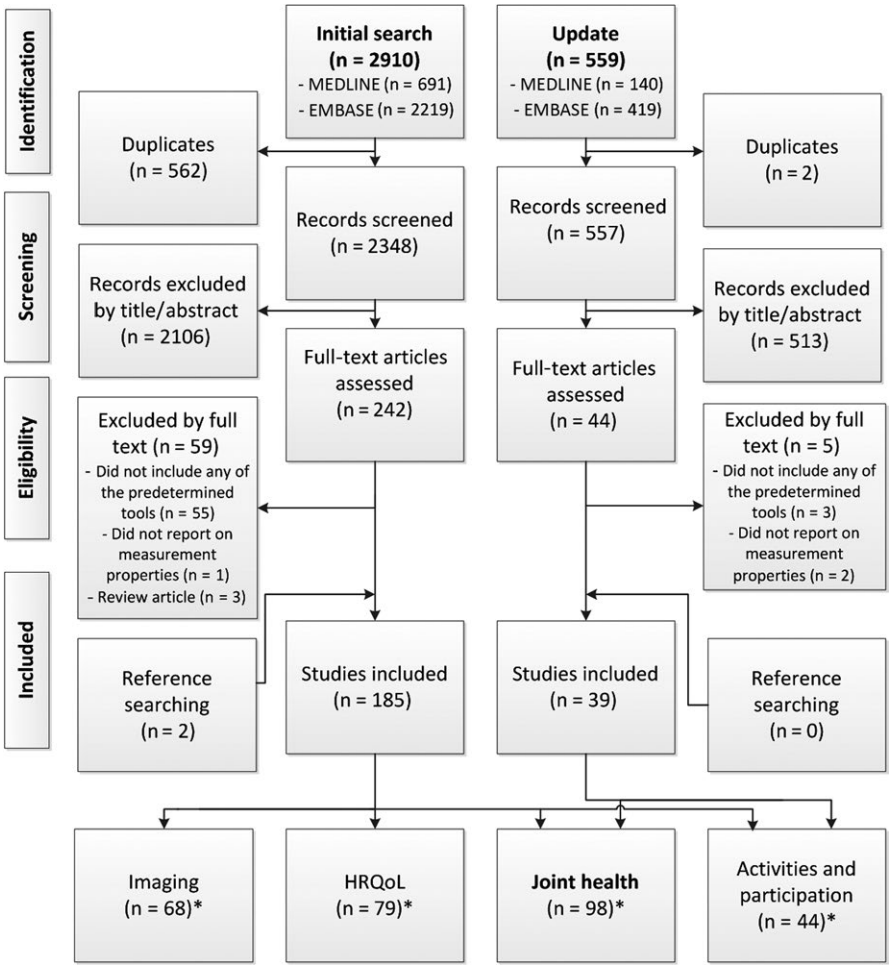

FIGURE 1 Flow diagram of study selection. Number of unique reports is indicated in parenthesis. In total, resulting from the initial literature search and search update, 2905 (2348 + 557) unique references were screened, 286 (242 + 44) unique full-text articles were assessed of which 96 studies were eligible for inclusion. After screening reference lists, we added two additional studies, so in total, 98 (96 + 2) studies were included. Numbers of studies in the systematic reviews exceed the number of studies eligible for inclusion in any of the three systematic reviews as individual studies may have included more than one outcome measurement instrument in different domains. HRQoL indicates Health-related quality of life

studies reported on hypotheses testing (convergent and discriminative validity). A limited number of studies reported on reliability and responsiveness. The measurement properties of 54 studies with COSMIN methodological quality scores of at least "fair" on any of the measurement properties are presented in Table S3. The other 44 studies with "poor" methodological quality on the COSMIN checklist are presented in the Appendix S1 (Table S4). Synthesis of the results for each joint examination score including the level of evidence is presented for the total patient group and for adults and children separately in Table 2.

3.3 | World federation of hemophilia orthopaedic joint score (WFH)

The WFH Orthopaedic Joint score (also known as the Gilbert Score) is a haemophilia-specific physical examination scoring system for the assessment of joint health of the knees, ankles and elbows.[8] The WFH score was described by the Orthopaedic Advisory Council of the WFH. The WFH score is an additive score ranging from 0 to 12 for knees and ankles and 0 to 10 for elbows, with 0 being an unaffected joint and 10 or 12 being most affected. For each joint, the following items are assessed: joint swelling (0-2), muscle atrophy (0-1), axial deformity of the knee and ankle (0-2), crepitus on motion (0-1), range of motion (0-2), flexion contracture (0-2) and instability (0-2). The maximum total score is 68 without the pain score and 86 with pain score (0 to 3 for each joint). A total of 58 studies reported on measurement properties of the WFH score, of which 17 studies reported on paediatric patients, 16 on adult patients and 25 on paediatric and adult patients combined. None of the studies explicitly aimed at evaluating measurement properties of the WFH score.

For hypotheses testing, 24 studies (six studies in children, seven in adults and 11 in a mixed population of children and adults) were assessed as "fair" on the COSMIN checklist. The other studies were assessed as "poor," mostly because of sample size <20 subjects. Conflicting results were found for hypotheses testing in all populations. The WFH score showed a good correlation with the CPE[9],

TABLE 2 Best evidence synthesis: Quality of the measurement properties of joint assessment instruments

	Reliability			Construct validity		
Measurement property	Internal consistency	Reliability	Measurement error	Structural validity	Hypothesis testing (correlations/ discriminative validity)	Responsiveness
Children						
WFH	na	na	na	na	+/−	−
CPE	na	na	na	na	+/−	+
PJS	na	na	na	na	++	na
HJHS	+	?	?	+	+/−	?
Adults						
WFH	na	na	na	na	+/−	na
CPE	na	na	na	na	na	+
PJS	na	na	na	na	na	na
HJHS	na	na	na	na	++	?
Adults and/or children						
WFH	na	na	na	na	+/−	+/−
CPE	na	na	na	na	+/−	++
PJS	na	na	na	na	+/−	+
HJHS	+	?	?	+	+/−	?

WFH, World Federation of Haemophilia orthopedic joint score; CPE, Colorado Physical Examination score; PJS, the joint examination score by Petrini; HJHS, the Hemophilia Joint Health Score.
Based on the GRADE approach by the Cochrane Back Review Group.[16]
Criterion validity and content validity were deemed irrelevant in the case of joint assessment scores and not evaluated.

a borderline correlation with a global joint physician health scale[19] but a low correlation with MRI scores[26]. Two studies assessed as having "fair" methodological quality on responsiveness found improvement in WFH scores after radiosynovectomy, which led to limited evidence for good responsiveness in mixed populations[27]. For paediatric patients, there was limited evidence for unsatisfactory responsiveness based on a lack of change in WFH joint scores after radiosynoviorthesis.[17] None of the studies addressed reliability or structural validity.

Significant floor effects of up to 88% were seen in European boys[26], whereas only 2% of high-titre inhibitor patients scored zero on the WFH joint score.[28] In one study in patients with severe haemophilia, the score distribution in haemophilia patients was similar to that in healthy controls, reflecting the relative insensitivity of the WFH score.[29] The WFH score was used as a starting point for the development of the following three physical examination scales.

3.4 | Colorado physical examination scale

The CPE score was developed specifically for children and aims to detect subtle abnormalities in joint health.[9] The original CPE (CPE-1) assesses the following items of the knee, ankle and elbow: swelling (0-3), muscle atrophy (0-3), axial deformity (0-2, knees/ankles), crepitus of motion (0-3), range of motion (0-3), flexion contracture (0-3, hips/knees/ankles/elbows), instability (0-2), strength (0-3), pain with activity (0-3), pain without activity (0-3) and gait abnormalities (0-2). Scores for the CPE-1 range from 0 to 31 for ankles/knees and 0 to 29 for elbows. Subsequently, the scoring system was adapted into the CPE-0.5 with maximum scores of 25 for ankles/knees and 23 for elbows. Higher scores indicate worse joint health. The Child CPE includes age-specific assessments and need for orthotics and/or splinting. The highest score achievable using the Child CPE is 31 for ankles/knees and 29 for elbows. None of the patients in the development study scored zero.[9] A total of seven studies reported on measurement properties of one of the CPE versions. These studies did not specify the version used. One study explicitly aimed at investigating measurement properties.[9]

Three studies (two studies including children and one including children and adolescents <20 years) were assessed as having "fair" methodological quality for hypotheses testing.[9,30,31] One study was excluded from the analyses because of duplication of study cohorts.[32] Overall, conflicting results were found for hypotheses testing. Understandably, correlations were high with the WFH Orthopaedic Joint Score and lower with WFH pain scale.[9] Correlations with MRI scores were contradictory. In one paediatric study, correlations between CPE scores and the additive and progressive MRI scores for elbows and knees were good, but unsatisfactory for ankles.[30] In another study, the overall CPE score did not correlate well with the overall MRI scores.[31]

Moderate evidence for good responsiveness was found. One study in a paediatric population and one in an adult population[17,33] showed improvement in CPE scores after radiosynoviorthesis and after resistance training with and without pulsed electromagnetic field therapy, respectively. However, the expected extent of improvement was not reported, and consequently, the study was scored as "fair".

Reliability of the CPE has not been studied in persons with haemophilia. In a study among 62 healthy boys, the interrater intraclass correlations (ICCs) were unsatisfactory, 0.34 for ankles, 0.67 for knees and 0.46 for elbows.[34] This study established normal values in 72 healthy boys, aged between 1 and 7 years (left ankle 1.14 ± 0.42, right ankle 1.10 ± 0.30, left knee 0.46 ± 0.69, right knee 0.50 ± 0.77 and left elbow 0.24 ± 0.72, right elbow 0.22 ± 0.65). None of the studies addressed internal consistency or structural validity.

3.5 | The petrini joint score

The PJS was adapted from the WFH scoring system for use in children.[10] It includes assessment of joint swelling (0-2), muscle atrophy (0-2), crepitus on motion (0-2), range of motion (0-2), flexion contracture (0-2), pain (0-3), gait (0-3) and strength against gravity (0-3). Three points can be added when the joint is considered a target joint and when chronic synovitis is present. Ankles, knees and elbows are assessed. The PJS ranges between 0 and 25 per joint with a maximum total score of 150; a higher score indicates worse joint health. Six studies reported on measurement properties of the PJS. None of the studies explicitly aimed at investigating measurement properties.

For hypotheses testing, four studies (three including children and one including adults) were assessed as having "fair" methodological quality.[18,35-37] In a follow-up study, a subset of patients were re-evaluated after changes in prophylactic treatment and therefore were excluded in this systematic review to avoid duplicate inclusion of patients.[36,37] Satisfactory correlation was found with the Pettersson score; however, the correlation with MRI was borderline at 0.48, with a predefined threshold of 0.50 for a good correlation.[36] Discriminative validity between severe and non-severe haemophilia, and clotting factor availability was assessed as satisfactory. In a study with a mixed adult and paediatric population, there was no correlation with the Pettersson score in patients awaiting surgical synovectomy.[35] Therefore, in children, moderate evidence for good hypotheses testing properties were found and for the overall population evidence was conflicting.

Responsiveness was reported in two studies. Mean PJS did not substantially change in one paediatric study after start or intensification of prophylaxis. As we do not expect that several years of regular prophylaxis in children with a mean score of 2.0 ± 3.6 would substantially decrease the score, responsiveness cannot be reliably assessed.[37] However, in a population of children and adults, there was a clear improvement in PJS score after surgical synovectomy.[35]

None of the studies addressed internal consistency, reliability, measurement error or structural validity.

3.6 | Hemophilia joint health score

The HJHS incorporates items of the WFH Orthopaedic Joint Score, the CPE and the PJS.[11] It was originally designed to detect mild joint

impairment in children with haemophilia ages 4-18 years. The current HJHS 2.1 version consists of assessment of swelling (0-3), duration of swelling (0-1), muscle atrophy (0-2), crepitus on motion (0-2), flexion loss (0-3), extension loss (0-3), joint pain (0-2) and strength (0-4) of the knees, ankles and elbows and a global gait score (0-4). The maximum score is 124, with a higher score indicating worse joint health.[25] Version 1.0 additionally included instability, gait per joint and axial alignment, with a maximum score of 148.[11] The 2.1 version was adapted to include updated normative tables for range of motion and gait assessment at a patient level rather than at joint level. A total of 32 studies reported on measurement properties of the HJHS. Seven studies explicitly aimed at investigating measurement properties of the HJHS, two of these studies were rated "poor" on the COSMIN methodological quality scale because of the small sample size.[11,25].

One paediatric study with "fair" methodological quality resulted in limited evidence for good internal consistency (Cronbach's alpha 0.83-0.84).[19] Although three paediatric studies reported good test-retest and interrater reliability with ICCs > 0.70, reliability was rated unknown due to insufficient numbers of included patients (<20)[11,22,25] In three studies, limits of agreement were reported, but measurement error was considered unknown due to insufficient number of included patients (<20).[11,22,25,38] Limited evidence is available for good structural validity. An explained variance of 52% in factor analyses was reported in a study of paediatric patients.[19] Responsiveness was considered unknown, because all studies included less than 20 patients.[39-41]

A total of 22 studies (13 paediatric, four adult and five mixed populations) were assessed as having "fair" methodological quality for hypotheses testing. Overall, for hypotheses testing, conflicting evidence was found for paediatric patients, whereas moderate evidence was found for good hypotheses testing properties in adult populations. In a paediatric study, HJHS correlated well with the HEAD-US Ultrasound Score.[42] In conflict with this finding, in other paediatric populations, the correlations with a global physician joint health scale, with Pettersson scores, Arnold-Hilgartner scores, MRI additive and progressive scores and HEAD-US scores were insufficient.[19,24,43] This results from a relatively high rate of abnormal findings with imaging studies without clinical symptoms in some studies,[43] but not in other studies.[42] Discrimination was good in paediatric patient groups comparing inhibitor status, onset of prophylaxis, severity of haemophilia, prophylaxis regimen, HEAD-US abnormalities, need for secondary prophylaxis and patients' age.[19,21,24,42,44,45] However, in other studies, HJHS scores did not discriminate well by severity of haemophilia in paediatric populations with access to primary prophylaxis or in children treated on demand only.[46,47] Overall, evidence for hypotheses testing in paediatric patients was rated as conflicting.

In adult populations, there was a clear discrimination by HJHS scores between groups according to age at start prophylaxis, the presence of synovitis and continuation of prophylaxis.[48-50] Overall, moderate evidence was found for good hypotheses testing properties in adult populations. In populations including both children and adults, correlations between HJHS and MRI scores were low ($r = 0.27$),[51] whereas reported correlations with Pettersson scores were high ($r = 0.67$-0.86).[22,52] Overall, for hypotheses testing, conflicting evidence was found for mixed paediatric and adult populations.

4 | DISCUSSION

In this overview of measurement properties of four haemophilia-specific measurement tools for the assessment of joint health, a total of 98 studies were included. For most of the included measurement tools, hypotheses testing and to a lesser extent responsiveness were investigated in multiple studies. Evidence on hypotheses testing was conflicting for most measurement tools. Yet, overall moderate evidence for good hypotheses testing properties was found in paediatric populations for the PJS and in adult populations for the HJHS. Limited to moderate evidence exists for responsiveness of the CPE. Internal consistency, reliability, measurement error and structural validity were only reported for the HJHS. However, test-retest reliability and measurement error were rated as unknown, because of small sample size.

Overall, the methodological quality of included studies was rated as "fair" to "poor" on the COSMIN scale. This was mainly due to limitations in study design, small patient numbers and the failure to meet the criterium of prespecified hypotheses for hypotheses testing (correlations/discriminative validity). The limited methodological quality can be explained by the fact that 92% of studies were not specifically designed to study the measurement properties of joint scores, but rather used one of the joint examination scores to measure outcomes. Unfortunately, the two studies that investigated measurement properties of HJHS were rated as "poor" on the COSMIN scale because of a small sample size (n = 8).[11,25]

Including all studies that used one of the included tools as a measurement instrument for the outcome rather than only including those that aimed to investigate measurement properties, allowed us to provide a good overview of the current evidence on hypotheses testing. Yet, including more studies leads to a higher chance of finding conflicting results. Therefore, well-designed clinimetric studies of higher methodological hierarchy (good to excellent) are needed to overcome the conflicting results of studies of "fair" methodological quality. At least 20 patients should be included, and hypotheses on the expected differences in scores between patient categories and size and direction of correlations should be formulated beforehand and explicitly reported.

Especially in some populations with significant floor or ceiling effects, underestimation of correlations with other constructs may have been caused by patient selection and may not reflect the true measurement properties of the measurement tools reported. Future studies on correlations with other constructs should be designed in a way such that sufficient variation in score results can be expected.

Strengths of the current study are the systematic literature search and independent study selection and methodological quality

assessment by two investigators. In the absence of alternatives, the COSMIN checklist was used to assess methodological quality of the included studies. This may be suboptimal because the COSMIN checklist was designed for quality assessment of studies on patient-reported outcomes rather than performance-based instruments. We adapted the COSMIN checklist to make it more suitable for our study by lowering the criterium of at least 30 participants for a "fair" score to 20, and dropping the requirement of reporting handling of missing items.

In the current study, we focused on measurement tools of joint health specific for haemophilia and did not include generic tools or disease-specific tools for related diseases. Since children and persons with access to early intensive treatment may have limited arthropathy, this necessitates sufficiently sensitive assessment by instruments designed specificly for haemophilic arthropathy. The included haemophilia-specific joint examination scores were identified by experts in the field. The Colorado Adult Joint Assessment Score (CAJAS) was not included in the current review as it was not yet published in full.[53] To the best of our knowledge, all published haemophilia-specific joint health assessment scores are included.

Previous reviews on physical examination tools also included haemophilia-specific instruments to assess joint health.[3-5] However, unlike the present study, these reviews did not formally include methodological quality criteria in their evaluation of measurement properties and are therefore considered incomplete.

Future studies should be well designed for investigation of measurement properties of instruments developed to measure joint health, particularly reliability, structural validity and responsiveness. Knowledge on limitations in measurement properties will guide improvements of the existing joint scores, including their feasibility. Currently, insufficient evidence on measurement properties exists to propose a hierarchy in joint scores. Measurement properties of the HJHS are most frequently reported, especially in children and intensively treated young adults and these show no significant pitfalls. Given the available evidence, there are no drawbacks for the use of the HJHS in clinical practice in these patient groups. Further knowledge is needed on the use in adults, especially those with more advanced arthropathy.

In conclusion, this review provides an overview of available evidence on measurement properties of haemophilia-specific instruments used for the assessment of clinical joint health. Although significant gaps in the evidence for all instruments remain, measurement properties of the HJHS were most extensively studied and show no drawbacks for use in clinical practice in children and intensively treated young adults.

ACKNOWLEDGEMENTS

The authors thank Cheri Nickel, Hospital Library and Archives, The Hospital for Sick Children, Toronto, Canada and René Spijker, Information Specialist, Medical Library Academic Medical Hospital Amsterdam, The Netherlands for expert support in designing the original literature search and the literature search update.

DISCLOSURES

SCG has received research funding from Bayer, Baxter, Novo Nordisk and CSL Behring. PH is a member of the International Prophylaxis Study Group (IPSG) Physical Health and Joint Function Expert Working Group. The IPSG is a not for profit collaborative study group supported by grants from Bayer Health care, Bioverativ, CSL-Behring, NovoNordisk, Pfizer and Shire. VB is Chair of the International Prophylaxis Study Group (IPSG). The IPSG is a not for profit collaborative study group whose mission is to facilitate the acquisition of new knowledge regarding regular replacement therapy (prophylaxis) with clotting factor concentrates in persons with inherited bleeding disorders (focus the hemophilias), and the dissemination of this knowledge globally. The IPSG is supported by grants from Bayer Health care, Bioverativ, CSL-Behring, NovoNordisk, Pfizer and Shire to the Hospital for Sick Children Foundation, Toronto, Canada. VB does not receive any personal remuneration for his role as Chair of the IPSG. VB has received speaker fees and fees for participation in Advisory Boards from Am, Bayer Healthcare, Novo Nordisk, Pfizer, Roche and Shire. VB is a Co-Principal Investigator for investigator initiated, industry supported research grants from Bioverativ and Shire. VB is a member of Data Safety Monitoring Boards for Octopharma and Shire. All other authors stated that they had no interests which might be perceived as posing a conflict or bias.

AUTHOR CONTRIBUTION

S.C. Gouw and V. Blanchette contributed to the concept and design. S.C. Gouw, M.A. Timmer and K. Fischer contributed to the analysis and/or interpretation of data. SCGouw drafted the manuscript. All authors contributed to the critical writing or revising the intellectual content and all authors approved the final version to be published.

REFERENCES

1. de Moerloose P, Fischer K, Lambert T, et al. Recommendations for assessment, monitoring and follow-up of patients with haemophilia. *Haemophilia*. 2012;18:319-325.
2. Fischer K, Poonnoose P, Dunn AL, et al. Choosing outcome assessment tools in haemophilia care and research: a multidisciplinary perspective. *Haemophilia*. 2017;23:11-24.
3. Stephensen D, Drechsler WI, Scott OM. Outcome measures monitoring physical function in children with haemophilia: a systematic review. *Haemophilia*. 2014;20:306-321.
4. De Kleijn P, Van Genderen FR, Van Meeteren N. 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;11:308-318.
5. De Kleijn P, Heijnen L, Van Meeteren N. Clinimetric instruments to assess functional health status in patients with haemophilia: a literature review. *Haemophilia*. 2002;8:419-427.
6. Limperg PF, Terwee CB, Young NL, et al. Health-related quality of life questionnaires in individuals with haemophilia: a systematic review of their measurement properties. *Haemophilia*. 2017;23:497-510.
7. Timmer M, Gouw SC, Feldman B, et al. Measuring limitations in activities and participation in persons with haemophilia:

a systematic review of commonly used measurement tools. *Haemophilia*. 2017;29-140.
8. Gilbert MS. Prophylaxis: musculoskeletal evaluation. *Semin Hematol*. 1993;30:3-6.
9. Manco-Johnson MJ, Nuss R, Funk S, Murphy J. Joint evaluation instruments for children and adults with haemophilia. *Haemophilia*. 2000;6:649-657.
10. Hill F, Ljung R. Third and fourth workshops of the European paediatric network for haemophilia management. *Haemophilia*. 2003;9:223-228.
11. Hilliard P, Funk S, Zourikian N, et al. Hemophilia joint health score reliability study. *Haemophilia*. 2006;12:518-525.
12. 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.
13. 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. *Quality Life Res*. 2012;21:651-657.
14. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther*. 2005;85:257-268.
15. 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.
16. 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.
17. Manco-Johnson MJ, Nuss R, Lear J, et al. 32P Radiosynoviorthesis in children with hemophilia. *J Pediatr Hematol Oncol*. 2002;24:534-539.
18. Schramm W, Gringeri A, Ljung R, et al. Haemophilia care in Europe: the ESCHQoL study. *Haemophilia*. 2012;18:729-737.
19. Feldman BM, Funk SM, Bergstrom BM, et al. Validation of a new pediatric joint scoring system from the International Hemophilia Prophylaxis Study Group: validity of the hemophilia joint health score. *Arthritis Care Res (Hoboken)*. 2011;63:223-230.
20. Khanum F, Bowen DJ, Kerr BC, Collins PW. Joint health scores in a haemophilia A cohort from Pakistan with minimal or no access to factor VIII concentrate: Correlation with thrombin generation and underlying mutation. *Haemophilia*. 2014;20:426-434.
21. Bladen M, Main E, Hubert N, Koutoumanou E, Liesner R, Khair K. Factors affecting the Haemophilia Joint Health Score in children with severe haemophilia. *Haemophilia*. 2013;19:626-631.
22. Fischer K, de Kleijn P. Using the Haemophilia Joint Health Score for assessment of teenagers and young adults: Exploring reliability and validity. *Haemophilia*. 2013;19:944-950.
23. Nijdam A, Bladen M, Hubert N, et al. Using routine Haemophilia Joint Health Score for international comparisons of haemophilia outcome: standardization is needed. *Haemophilia*. 2016;22:142-147.
24. Oymak Y, Yildirim AT, Yaman Y, et al. The effectiveness of tools for monitoring hemophilic arthropathy. *J Pediatr Hematol Oncol*. 2015;37:e80–e85.
25. Sun J, Hilliard PE, Feldman BM, et al. Hemophilia Joint Health Score 2.1 reliability study. *Haemophilia*. 2014;20:435-440.
26. Lundin B, Ljung R, Pettersson H. European Paediatric Network for Haemophilia M. MRI scores of ankle joints in children with haemophilia–comparison with clinical data. *Haemophilia*. 2005;11:116-122.
27. De la Corte-Rodriguez H, Rodriguez-Merchan EC, Jimenez-Yuste V. Radiosynovectomy in hemophilia: quantification of its effectiveness through the assessment of 10 articular parameters. *J Thromb Haemost*. 2011;9:928-935.
28. Stieltjes N, Torchet MF, Misrahi L, et al. Epidemiological survey of haemophiliacs with inhibitors in France: orthopaedic status, quality of life and cost–the 'Statut Orthopedique des Patients Hemophiles' avec Inhibiteur study. *Blood Coagul Fibrinolysis*. 2009;20:4-11.
29. Brunner A, Stauber F, Gohler S, et al. Impact of joint status on contraction steadiness of m. quadriceps femoris in people with severe haemophilia. *Haemophilia*. 2014;20:884-890.
30. Kraft J, Blanchette V, Babyn P, et al. Magnetic resonance imaging and joint outcomes in boys with severe hemophilia A treated with tailored primary prophylaxis in Canada. *J Thromb Haemost*. 2012;10:2494-2502.
31. Manco-Johnson MJ, Abshire TC, Shapiro AD, et al. Prophylaxis versus episodic treatment to prevent joint disease in boys with severe hemophilia. *N Engl J Med*. 2007;357:535-544.
32. Feldman BM, Pai M, Rivard GE, et al. Tailored prophylaxis in severe hemophilia A: interim results from the first 5 years of the Canadian Hemophilia Primary Prophylaxis Study. *J Thromb Haemost*. 2006;4:1228-1236.
33. Parhampour B, Torkaman G, Hoorfar H, Hedayati M, Ravanbod R. Effects of short-term resistance training and pulsed electromagnetic fields on bone metabolism and joint function in severe haemophilia A patients with osteoporosis: a randomized controlled trial. *Clinical rehabilitation*. 2014;28:440-450.
34. Hacker MR, Funk SM, Manco-Johnson MJ. The colorado haemophilia paediatric joint physical examination scale: normal values and interrater reliability. *Haemophilia*. 2007;13:71-78.
35. Rampal V, Odent T, Torchet MF, et al. Surgical synovectomy of the knee in young haemophiliacs: long-term results of a monocentric series of 23 patients. *J Child Orthop*. 2010;4:33-37.
36. Pergantou H, Matsinos G, Papadopoulos A, Platokouki H, Aronis S. Comparative study of validity of clinical, X-ray and magnetic resonance imaging scores in evaluation and management of haemophilic arthropathy in children. *Haemophilia*. 2006;12:241-247.
37. Pergantou H, Platokouki H, Matsinos G, et al. Assessment of the progression of haemophilic arthropathy in children. *Haemophilia*. 2010;16:124-129.
38. Feldman BM, Pullaneyagum E. Response to 'Limits of agreement between raters are required for use of HJHS 2.1 in clinical studies'. *Haemophilia*. 2015;21:e71.
39. 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.
40. Teyssler P, Taborska K, Kolostova K, Bobek V. Radiosynoviorthesis in hemophilic joints with yttrium-90 citrate and rhenium-186 sulfide and long term results. *Hellenic J Nucl Med*. 2013;16:44-49.
41. 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.
42. Foppen W, van der Schaaf IC, Fischer K. Value of routine ultrasound in detecting early joint changes in children with haemophilia using the 'Haemophilia Early Arthropathy Detection with UltraSound' protocol. *Haemophilia*. 2016;22:121-125.
43. Altisent C, Martorell M, Crespo A, Casas L, Torrents C, Parra R. Early prophylaxis in children with severe haemophilia A: clinical and ultrasound imaging outcomes. *Haemophilia*. 2016;22:218-224.
44. Saulyte Trakymiene S, Ingerslev J, Rageliene L. Utility of the Haemophilia Joint Health Score in study of episodically treated boys with severe haemophilia A and B in Lithuania. *Haemophilia*. 2010;16:479-486.
45. Saulyte Trakymiene S, Clausen N, Poulsen LH, Ingerslev J, Rageliene L. Progression of haemophilic arthropathy in children: a Lithuanian-Danish comparative study. *Haemophilia*. 2013;19:212-218.
46. Groen WG, Takken T, van der Net J, Helders PJ, Fischer K. Habitual physical activity in Dutch children and adolescents with haemophilia. *Haemophilia*. 2011;17:e906–e912.

47. Payal V, Sharma P, Chhangani NP, Janu Y, Singh Y, Sharma A. Joint health status of hemophilia patients in jodhpur region. *Indian J Hematol Blood Transfus*. 2015;31:362-366.
48. Khawaji M, Astermark J, Berntorp E. Lifelong prophylaxis in a large cohort of adult patients with severe haemophilia: a beneficial effect on orthopaedic outcome and quality of life. *Eur J Haematol*. 2012;88:329-335.
49. Kidder W, Nguyen S, Larios J, Bergstrom J, Ceponis A, von Drygalski A. Point-of-care musculoskeletal ultrasound is critical for the diagnosis of hemarthroses, inflammation and soft tissue abnormalities in adult patients with painful haemophilic arthropathy. *Haemophilia*. 2015;21:530-537.
50. 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.
51. Den Uijl IE, De Schepper AM, Camerlinck M, Grobbee DE, Fischer K. Magnetic resonance imaging in teenagers and young adults with limited haemophilic arthropathy: baseline results from a prospective study. *Haemophilia*. 2011;17:926-930.
52. 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:e80–e86.
53. Hong W, Raunig D, Funk S. Validation of the colorado adult joint assessment scale in patients with severe hemophilia A. *J Thromb Haemost*. 2013;11:461.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Gouw SC, Timmer MA, Srivastava A, et al. Measurement of joint health in persons with haemophilia: A systematic review of the measurement properties of haemophilia-specific instruments. *Haemophilia*. 2019;25:e1–e10. https://doi.org/10.1111/hae.13631