




ORIGINAL ARTICLE

Validity and reliability of the Colorado Adult Joint Assessment Scale in adults with moderate-severe hemophilia A

Sharon M. Funk¹  | Sylvia Engelen² | Katy Benjamin³ | Olga Moshkovich³ |
Brittany Gentile⁴ | Nikki Church²  | Walter Hong² | Dianne Thornhill¹ |
Marilyn J. Manco-Johnson¹ 

¹University of Colorado Anschutz Medical Campus, Aurora, Colorado

²Bayer, Whippany, New Jersey

³ICON, Gaithersburg, Maryland

⁴ICON, San Francisco, California

Correspondence

Sharon M. Funk, Hemophilia and Thrombosis Center, University of Colorado Anschutz Medical Campus, 13199 E Montview Blvd, Suite 100, Aurora, CO 80045.
Email: sharon.funk@cuanschutz.edu

Funding information

Bayer

Abstract

Background: The Colorado Adult Joint Assessment Scale (CAJAS) is designed to assess joint health in adults with hemophilia. The CAJAS comprises nine items (swelling, muscle atrophy, axial deformity, crepitus, range of motion, contracture, instability, strength, gait) and assesses six joints.

Objective: To assess CAJAS content validity and psychometric properties.

Patients/Methods: Data were obtained from the Trial to Evaluate the Effect of Secondary Prophylaxis With rFVIII Therapy in Severe Hemophilia A Adult and/or Adolescent Subjects Compared to That of Episodic Treatment (SPINART) study and a separate CAJAS validation study. CAJAS assessments in SPINART were performed by physical therapists (PTs) from the United States, Romania, Bulgaria, and Argentina. In the validation study, content validity was assessed from interviews with six PTs at three US hemophilia centers; cultural equivalence was assessed with seven non-US PTs from SPINART. Reliability data were collected from 30 subjects at four US centers. Test-retest reliability was evaluated by having the same PT perform CAJAS examinations at two visits, 7-10 days apart. Inter-rater reliability was assessed by comparing CAJAS scores of two different PTs performing separate examinations of the same patient several hours apart at the same visit. Psychometric properties were assessed using SPINART and validation study data.

Results: The CAJAS demonstrated good content validity. Test-retest reliability was high (intraclass correlation coefficient, 0.98), as was inter-rater reliability (intraclass correlation coefficient, 0.88). Internal consistency reliability was strong ($\alpha = .90$). The CAJAS demonstrated good convergent/divergent validity, known-groups validity, and ability to detect change.

Conclusions: The CAJAS is a valid and reliable measure of joint health in adults with moderate-severe hemophilia and is appropriate for use in clinical practice.

Sylvia Engelen and Walter Hong were employed at Bayer at the time this study was conducted.

Katy Benjamin and Brittany Gentile were employed at ICON at the time this study was conducted.

Manuscript handled by: Donna DiMichele

Final decision: Donna DiMichele, 19 September 2019

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2019 The Authors. *Journal of Thrombosis and Haemostasis* published by Wiley Periodicals, Inc. on behalf of International Society on Thrombosis and Haemostasis

KEYWORDS

adult, hemophilia, joints, psychometrics, validity and reliability

1 | INTRODUCTION

Hemophilia is a congenital bleeding disorder resulting from a deficiency of clotting factor (factor VIII in hemophilia A; factor IX in hemophilia B).¹ Patients with severe hemophilia have factor levels <1% of normal² and are at risk for spontaneous bleeding episodes, most commonly into joints and muscles.¹ Repeated joint bleeding can lead to irreversible hemophilic arthropathy and consequent disability.^{3,4} Joint bleeding and hemarthropathy can be minimized or prevented with prophylactic factor infusions.¹ The efficacy of prophylaxis for preventing bleeding episodes and joint damage in pediatric patients is well-established.⁵ Prophylaxis can prevent bleeding episodes in adults.^{6,7} Whether it can also prevent progression of hemarthropathy is less well studied.

Validated assessment measures are needed to gauge the natural progression of hemophilic joint disease and the effects of prophylaxis on joint structure and function in adults with hemophilia. The Hemophilia Joint Health Score (HJHS), based on the Gilbert score, was developed to identify the earliest signs of joint degeneration, but has only been validated in children. Currently, the Gilbert score (or World Federation of Hemophilia physical examination score) is the only available hemophilia-specific measure for physical assessment of joint impairment in adults.⁸

Joint health varies considerably in adults with hemophilia,⁹⁻¹¹ many of whom are treated on-demand or are not currently receiving prophylaxis.¹² The Colorado Adult Joint Assessment Scale (CAJAS) is a clinician-reported outcome measure of joint structure and function developed specifically to assess the range of joint health in adults with hemophilia. Based on physical examination of six joints (knees, elbows, ankles), the CAJAS evolved from the Colorado Physical Examination full-point and half-point scales (Colorado PE-1 and PE-0.5),¹³ both modifications of the Gilbert score.^{8,13} The CAJAS includes all seven domains of the Gilbert score, with axial deformity, contracture, and instability specifically included because these domains are consistent with the greater progression of joint disease typically seen in adults. Six of the domains preserved from the Gilbert score (swelling, muscle atrophy, axial deformity, crepitus, range of motion and contracture) were modified to allow assessment of subtle changes in joint structure and function seen in early stages of joint arthropathy. Two additional domains, strength and gait, were added to complete the scale.

The CAJAS was initially developed as an 11-domain instrument to measure joint health in adults with hemophilia. This version was used in Trial to Evaluate the Effect of Secondary Prophylaxis With rFVIII Therapy in Severe Hemophilia A Adult and/or Adolescent Subjects Compared to That of Episodic Treatment (SPINART; clinicaltrials.gov identifier NCT00623480), a phase 3 randomized, controlled, parallel group clinical trial designed to compare prophylaxis versus on-demand treatment in

Essentials

- Colorado Adult Joint Assessment Scale (CAJAS) assesses joint health in adults with hemophilia.
- CAJAS validity and reliability were assessed in two separate multinational studies.
- CAJAS demonstrated good validity and reliability.
- CAJAS is appropriate for clinical use by hemophilia clinicians trained in hemophilia musculoskeletal examination and CAJAS evaluation.

adolescents and adults with severe hemophilia A. The HJHS was not considered as an outcome measure for this study because it had not been validated in adults and differed more substantively from the Gilbert score. Following use of the CAJAS in SPINART, two domains (pain and splint/orthotic) were removed based on US Food and Drug Administration guidelines regarding outcome measures for pain evaluation and statistical analysis of the splint/orthotic domain. Splint/orthotic demonstrated low ability to discriminate between healthy and unhealthy joints and physical therapists (PTs) had difficulty scoring it. The pain item included both subjective and objective pain assessments (e.g., stiffness, activity level, use of pain medications), which could be better evaluated using a patient-reported outcome (PRO) measure.

The CAJAS is designed for evaluation of knees, elbows, and ankles in adults with hemophilia, by trained PTs (CAJAS Worksheet and Instruction Guidelines; Appendix S1, Appendix S2). The scale comprises nine domains for assessing joint status: swelling, muscle atrophy, axial deformity, crepitus, range of motion, contracture, instability, strength, and gait (Figure 1). Each item is assessed using a Likert-type scale and scored from 0 to 2 (for axial deformity, instability, and gait), 0 to 3 (for swelling, muscle atrophy, crepitus, range of motion, and contracture), or 0 to 4 (for strength). Higher scores indicate more severe joint dysfunction. The CAJAS score for each joint consists of a single unweighted summary score across the nine domains. The maximum scores for knees and ankles (combined) and elbows are 25 and 21, respectively. The total six-joint score is 142.

The goal of this study was to assess the content validity and psychometric properties of the CAJAS in adults with moderate or severe hemophilia A.

2 | METHODS

The CAJAS was administered in SPINART⁶ at baseline, year 1, year 2, and year 3. The CAJAS psychometric analyses included data from patients aged ≥ 17 years who had completed all PRO and

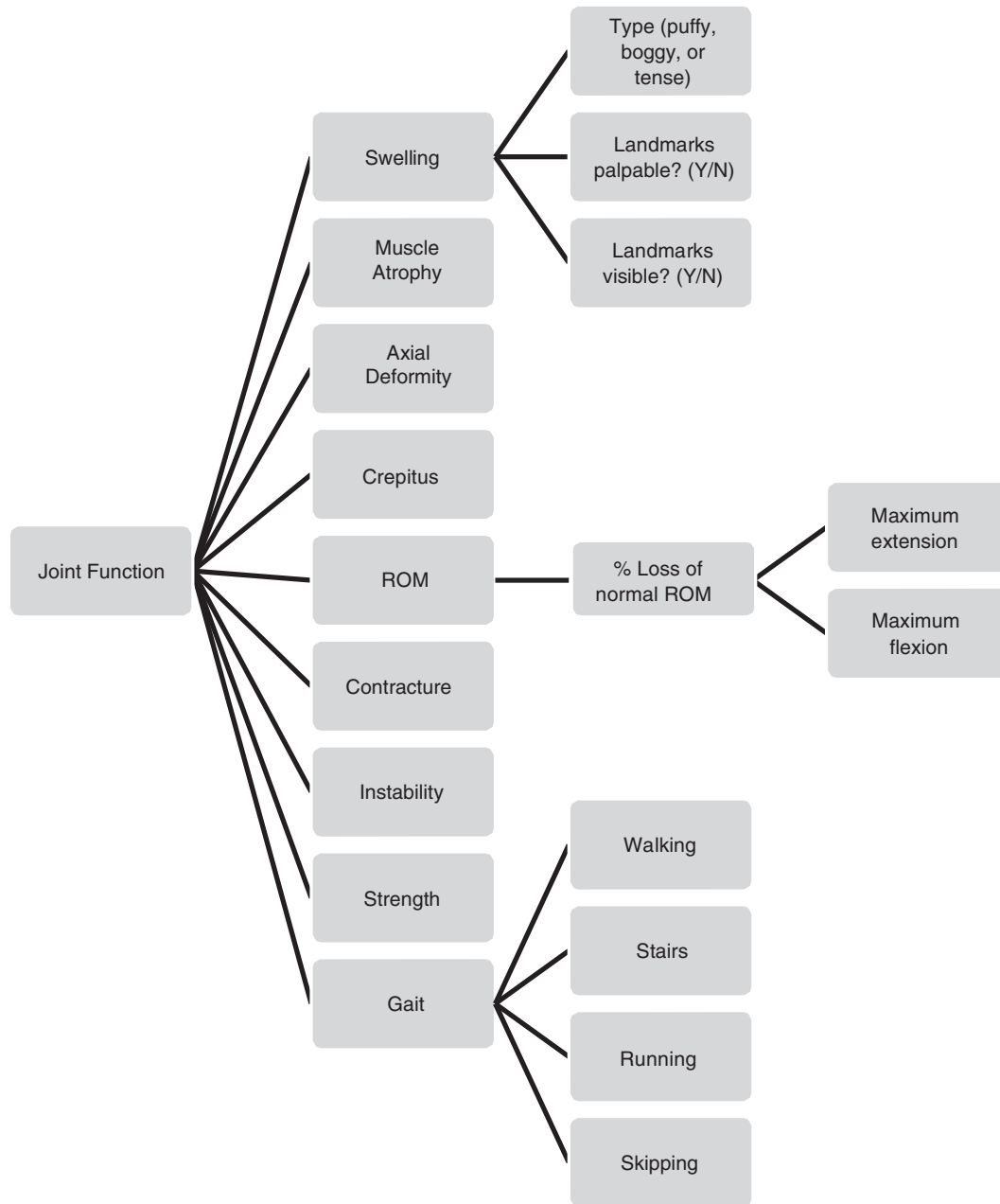


FIGURE 1 Conceptual framework of the Colorado Adult Joint Assessment Scale. All items are measured for the left and right elbows, knees, and ankles with the exception of Axial Deformity and Gait, which are measured for the left and right knees and ankles only. ROM, range of motion

clinician-reported outcome assessments. Patients with very poor joint status, defined as routine need for a wheelchair or inability to walk without assistance of a brace, cane, or crutches or having three or more ankylosed joints, were excluded from SPINART.

In addition to the SPINART data, a separate validation study using the nine-item CAJAS was conducted to assess the scale's content validity, inter-rater and test-retest reliability. Male patients aged 18 to 50 years from four United States treatment centers were eligible if they had moderate or severe hemophilia A (factor VIII:C, <1%-2%), could read and understand English, and did not routinely need a wheelchair, two crutches, or canes. For the psychometric analyses, the CAJAS total score was calculated as the average of the mean scores for each joint type (elbows,

knees, ankles). For the reliability analyses, the summed CAJAS by joint and total scores were used.

For the validation study, eight PTs (two per site) were trained to administer and score the CAJAS in a live training session and received a training manual and video demonstrating a CAJAS patient evaluation.

2.1 | Content validity

Content validity was assessed using standard qualitative research methods.¹⁴ Cognitive debriefing interviews were conducted with six PTs who evaluated 12 patients from three US clinical sites. Interviews were audio-recorded and transcribed; transcripts were qualitatively analyzed on an item-by-item basis using MAXQDA software (Verbi GmbH).

2.2 | Cultural validity

In SPINART, the CAJAS was administered in the United States, Romania, Bulgaria, and Argentina. The CAJAS was not translated into other languages; all PTs used the English-language version and were trained on the measure in English. To determine whether PTs in Romania, Bulgaria, and Argentina correctly interpreted the CAJAS items and found the concepts relevant, a cultural equivalence study was conducted with seven PTs at seven sites located outside the United States (two in Romania, three in Bulgaria, two in Argentina). Each of the PTs had participated in SPINART and were interviewed about their experience administering the CAJAS. PTs were interviewed in their native language regarding the meaning of terms used in the CAJAS and their ability to understand and score CAJAS items. To assess any differences in joint-related symptoms between US and non-US patients that could affect CAJAS results, both US and non-US PTs were asked about each CAJAS domain with respect to how common it was among their patients and how it affected patients' daily lives. Interviews were audio recorded and transcribed into English; transcripts were analyzed in comparison with results from the US content validity study to assess the cultural equivalence of the CAJAS.

2.3 | Psychometric properties

Psychometric properties assessed included instrument reliability (e.g., internal consistency, test-retest, inter-rater), construct validity (e.g., convergent/divergent, known-groups validity), and ability to detect change.

Reliability analyses were conducted using data from the validation study previously described. Test-retest reliability was evaluated by having the same PT perform the same examination on two patient visits 7 to 10 days apart. Before the second assessment, the PT confirmed that no muscle or joint-bleeding episodes had occurred since the last evaluation, using information documented in the patient's journal combined with a musculoskeletal health assessment form completed immediately before each evaluation (noting bleeding history, increases in pain and swelling, and need for additional adaptive equipment). Inter-rater reliability was assessed by comparing the CAJAS scores of two PTs performing separate examinations of the same patient, at the same visit, 4 hours apart. A total of 30 patients were evaluated.

Baseline and final study assessment data from SPINART were used to assess internal consistency reliability (using Cronbach α), construct validity, ability to detect change (including calculation of effect size), minimally important difference (MID; clinically meaningful change at the group level), and responder definition (clinically meaningful change for an individual patient). To evaluate construct validity, the relationships between the CAJAS and other instruments administered in SPINART were assessed. These included the Haemo-QoL-A Physical Functioning subscale (a subscale of a PRO measure used to assess health-related quality of life [HRQoL] in adults with hemophilia), the EQ-5D-3L visual analog scale (EQ-5D VAS; a PRO

of current health state), the Short-Form McGill Pain Questionnaire (SF-MPQ; a PRO assessment of pain severity), the Activity List (a PRO to assess changes in physical activity over time), the Extended Magnetic Resonance Imaging (MRI) Scale (used by experienced radiologists to determine the extent of joint damage in adults), and the International Prophylaxis Study Group MRI 17 (IPSG MRI 17) scale. The MID was estimated as the mean CAJAS change scores for joints deteriorating in each outcome measure.

Ability to detect change was examined by correlating change in the CAJAS total score from baseline to final evaluation 3 years later with changes during the same timeframe in Haemo-QoL-A Physical Functioning; EQ-5D VAS; current, recent, and overall pain (SF-MPQ); current activity level; and Extended MRI scores. One-way analyses of variance were used to assess ability of change in the CAJAS total score to differentiate between patients classified into three change groups (e.g., improvement, no change, deterioration) by the degree (at the tertiles) and direction of change from baseline in Haemo-QoL-A Physical Functioning, current activity level, and Extended MRI scores.

Receiver operating characteristic curves were used to estimate responder definitions for the CAJAS total score in terms of moving from one change category to another (from no change/improvement to deterioration) in Haemo-QoL-A Physical Functioning and Extended MRI scores. The best cut-off point was the value of the change in the CAJAS score that best differentiated (e.g., gave the best balance between sensitivity and specificity) between patients deteriorating or not deteriorating on each measure.

2.4 | Statistical analysis

Statistical analyses were performed using SAS 9.3.

3 | RESULTS

Data from 80 of the 84 patients enrolled in SPINART were included in the analyses. Two patients were excluded because they did not complete the Haemo-QoL-A. Two other patients were evaluated by a non-US clinician who did not demonstrate sufficient understanding of the CAJAS; their data were excluded from psychometric analyses, but the clinician's responses in the cultural validity study were included to document these challenges. The patients enrolled in the validation study ($n = 30$) were somewhat more ethnically diverse but otherwise demographically similar to those in SPINART (Table 1).

3.1 | Content validity

Overall, the six US PTs interviewed understood the terms and assessment methods used for each of the nine CAJAS domains and had minimal difficulty performing and scoring the joint assessments. The CAJAS methods for assessing muscle atrophy, axial deformity, crepitus, instability, and strength were considered similar to those used in clinical practice. Two items, swelling and gait, presented

TABLE 1 Demographics and treatment regimen of the SPINART and validation study patient populations

	SPINART ^a (n = 80)	Validation study (n = 30)
Age, y		
Mean	31.0	30.9
Range	17-50	18-47
Race, n (%)		
White	73 (91.2)	20 (66.7)
Black	0	1 (3.3)
Asian	2 (2.5)	2 (6.7)
Mixed ^b	0	3 (10.0)
Ethnicity/Hispanic n (%)	5 (6.2)	4 (13.3)
Treatment regimen, n (%)		
On-demand	41 (51.3) ^c	2 (6.7)
Prophylaxis	39 (48.8) ^c	22 (73.3) ^d
Other	0	6 (20.0)

^aSPINART, Trial to Evaluate the Effect of Secondary Prophylaxis With rFVIII Therapy in Severe Hemophilia A Adult and/or Adolescent Subjects Compared to That of Episodic Treatment.²⁰

^bWhite/Hispanic, Native American/Hispanic, White/Hawaiian.

^cPatients were randomized 1:1 to prophylaxis (3x/week) or on-demand treatment regimens.

^dProphylaxis dosing regimens included 1x/week, 2x/week, 3x/week, or every other day.

challenges for the US PTs. The CAJAS emphasis on identification of very minimal swelling was a change from what some PTs previously considered to be swelling, initially causing confusion in identifying mild versus moderate swelling. Definitions of mild, moderate, and severe swelling were clarified during training sessions. Questions arose regarding potential confounding with joint surgery and swelling for reasons unrelated to bleeding. However, no consensus was reached on how to improve assessment of this domain. Issues in scoring the gait domain (e.g., lack of means to account for other body parts that could affect gait, such as the hip) were reported, but overall the PTs understood the concepts in the domain and could appropriately score each joint (knee and ankle), on each side.

3.2 | Cultural validity

Seven non-US clinicians were interviewed to assess the cultural equivalence of the CAJAS concepts, terms, and clarity of instructions. Results were compared with cognitive interview results from US PTs to ensure common knowledge and comprehension across all clinicians. Overall, the US and non-US PTs demonstrated a common understanding of CAJAS domains and items as well as the instructions for assessment and scoring. They perceived the constructs included in the CAJAS to be common problems for adult hemophilia patients with important consequences in terms of patients' ability to perform usual activities of daily living and to lead full, productive lives. For the most part, all PTs could perform the CAJAS evaluations using the procedures outlined in the training manual. Both US and non-US PTs generally agreed that

the signs of arthropathy measured in the CAJAS were relevant, common, and affected patients' daily lives. Instability was considered less common by most US PTs (five of six), whereas five of the seven non-US therapists indicated it was a common problem for adult patients with hemophilia. True ligamentous instability, resulting from repeated hemarthroses, is observed in very severe end-stage joint disease. Although still observed in older patients with hemophilia in the United States, it is much less common in younger adult patients because of access to more effective treatment regimens at younger ages. This is likely not the case in the non-US study patients, where even on-demand treatment was delayed compared with US standards. Swelling was reported to be more common by the US (six of six) than the non-US PTs (four of seven). US PTs have been evaluating earliest signs of swelling in patients for more than 20 years as part of prophylaxis monitoring. Because non-US patients had limited access to prophylaxis or even early aggressive on-demand treatment, it is likely that early identification of subtle swelling may not have been recognized in these countries. Although there were some differences in the degree to which various signs of arthropathy affected patient functioning, these differences did not appear culturally related but rather geographically related. Most PTs reported that swelling, muscle atrophy, range of motion, and contractures had a major effect on patients' functioning. Crepitus was considered the least likely to disrupt everyday activities.

3.3 | Psychometric properties

By-joint and total CAJAS summary statistics for the CAJAS validation and SPINART (entry and exit) studies are provided (Table 2). Results for psychometric properties are reported for evaluations of the CAJAS total score, a secondary efficacy endpoint in SPINART.⁶

3.3.1 | Reliability

Individual by-joint and total CAJAS intraclass correlation coefficients for inter-rater reliability and test-retest reliability all indicated good (>0.75) to excellent (>0.90) reliability (Table 3).¹⁵ Internal consistency reliability was strong (Cronbach $\alpha = .90$).

3.3.2 | Construct validity

Convergent/divergent validity was shown by correlating the CAJAS total score and other outcome measures used in SPINART (Table 4). The negative correlations of the CAJAS total score with Haemo-QoL-A scores demonstrated convergent validity, given that higher CAJAS scores indicate worse joint health and higher Haemo-QoL-A scores indicate better functioning. The CAJAS total score also correlated significantly with the EQ-5D VAS ($P < .001$). The CAJAS total score was moderately positively correlated with pain within the past 4 weeks on the SF-MPQ, and with current activity level (both $P < .05$), and strongly correlated with Extended MRI Scale scores ($P < .001$). All correlations were in the expected direction, with higher CAJAS total scores associated with worse physical functioning and HRQoL, more pain, and greater activity and joint impairment.

TABLE 2 By-joint and total CAJAS summary statistics for the validation and SPINART (entry and exit) study populations

	N	Mean ± SD	Median (range)
CAJAS validation study			
CAJAS total	30	36.87 ± 20.68	35 (3-98)
Left ankle total	30	8.80 ± 3.60	9 (1-18)
Right ankle total	30	8.80 ± 4.46	9 (1-18)
Left elbow total	30	5.07 ± 4.96	4 (0-15)
Right elbow total	30	4.30 ± 4.60	2 (0-14)
Left knee total	30	5.07 ± 4.98	3 (0-18)
Right knee total	30	4.83 ± 4.55	4 (0-17)
SPINART study entry			
CAJAS total	80	45.66 ± 19.41	47 (7-97)
Left ankle	80	8.36 ± 4.12	8 (0-18)
Right ankle	80	7.65 ± 4.29	7 (0-17)
Left elbow	79	7.15 ± 4.41	7 (0-18)
Right elbow	78	6.01 ± 4.06	6 (0-16)
Left knee	80	8.11 ± 5.92	8 (0-22)
Right knee	80	8.61 ± 5.97	7 (0-21)
SPINART study exit			
CAJAS total	70	46.67 ± 21.62	49 (7-114)
Left ankle	70	8.26 ± 4.12	8 (0-18)
Right ankle	70	8.11 ± 4.45	8 (0-19)
Left elbow	70	6.93 ± 4.66	7 (0-18)
Right elbow	70	6.11 ± 4.37	6 (0-18)
Left knee	70	8.71 ± 6.14	9 (0-24)
Right knee	70	8.54 ± 5.33	9 (0-21)

Note: Elbow scores do not contain "Axial Deformity" and "Gait" sub-scores and thus cannot have as high a range of values as the knees and ankles.

Abbreviations: CAJAS, Colorado Adult Joint Assessment Scale; SD, standard deviation; SPINART, Trial to Evaluate the Effect of Secondary Prophylaxis With rFVIII Therapy in Severe Hemophilia A Adult and/or Adolescent Subjects Compared to That of Episodic Treatment.

Divergent validity was demonstrated by the smaller, nonsignificant correlations between the CAJAS total score and more distal concepts (e.g., worry, consequences of bleeding).

For known-groups validity, significant differences in CAJAS total scores at baseline were observed between patients aged 35 to 50 years and 17 to 26 years (mean difference, 4.01 [95% confidence interval, 1.90-6.12]; $P < .001$), with older patients having higher CAJAS scores, indicating worse joint status. The CAJAS total scores at baseline were also higher for patients with moderate or severe joint damage (defined as an IPSPG MRI 17 score >4) versus patients with little or no joint damage (defined as an IPSPG MRI 17 ≤ 4 ; mean difference, 4.37 [95% confidence interval, 2.93-5.81]; $P < .001$).

3.3.3 | Ability to detect change

Changes in CAJAS total scores were significantly, moderately correlated with changes in Haemo-QoL-A Physical Functioning, EQ-5D

TABLE 3 Intraclass correlation coefficients of inter-rater and test-retest reliability by the summed total six joint CAJAS score and individual joint scores

	Inter-rater ICC (95% CI)	Test-retest ICC (95% CI)
CAJAS total	0.88 (0.77-0.94)	0.98 (0.95-0.99)
Left ankle total	0.76 (0.56-0.88)	0.92 (0.84-0.96)
Right ankle total	0.86 (0.73-0.93)	0.93 (0.86-0.97)
Left elbow total	0.84 (0.69-0.92)	0.97 (0.94-0.99)
Right elbow total	0.91 (0.82-0.96)	0.98 (0.97-0.99)
Left knee total	0.84 (0.69-0.92)	0.93 (0.85-0.96)
Right knee total	0.82 (0.66-0.91)	0.95 (0.89-0.97)

Note: Inter-rater reliability was analyzed by a two-way random-effects model of consistency; test-retest reliability was analyzed by a two-way mixed-effects model of absolute agreement.

Abbreviations: CAJAS, Colorado Adult Joint Assessment Scale; CI, confidence interval; ICC, intraclass correlation coefficients.

VAS, current activity level, and the SF-MPQ pain summary score and pain in the past 4 weeks, but not the Extended MRI Scale scores or SF-MPQ current pain score. All changes were in the expected direction; decreasing CAJAS total scores were associated with improved physical functioning and HRQoL, decreased pain over the past 4 weeks, and decreased activity impairment.

Table 5 presents the relationship between changes in CAJAS total score and changes in Haemo-QoL-A physical functioning, current activity level, and Extended MRI Scale scores. CAJAS total score changes generally were small; effect sizes within each change group (improvement, no change, deterioration) were all <0.5 points, the cut-off for a medium-sized effect.¹⁶ Overall effect sizes (η^2) for change in CAJAS total score by change in Haemo-QoL-A physical functioning and current activity level indicated a moderate to large effect. Correlations between the CAJAS total score and these measures were statistically significant ($P < .05$; Table 5).

3.3.4 | MID and responder definition

The Haemo-QoL-A and Activity List current activity level produced the most consistent MID estimates, resulting in a recommended MID of 0.78 points for the CAJAS total score, and 0.30, 0.62, and 1.29 for elbows, knees, and ankles, respectively (Table 6). For the responder definition, a 1.0- or 1.5-point difference for the CAJAS total score is estimated to be indicative of a true individual change that is beyond the measurement error of the instrument; responder definitions for individual joints are slightly higher (Table 7).

4 | DISCUSSION

Using data from SPINART and a separate content validation study, the CAJAS was shown to be a reliable and valid measure of joint structure and function in adults with moderate-severe hemophilia A. The CAJAS showed excellent test-retest reliability and very good inter-rater

TABLE 4 Correlation between CAJAS total score and SPINART outcome measures

Haemo-QoL-A	
Physical functioning	-0.51 ^a
Role functioning	-0.23 ^c
Worry	-0.24 ^c
Consequences of bleeding	-0.20
Emotional impact	-0.06
Treatment concern	-0.31 ^b
Haemo-QoL-A total	-0.34 ^b
EQ-5D	
Visual analog scale	-0.43 ^a
Short-Form McGill Pain Questionnaire	
Pain summary score	0.19
Pain (past 4 wk)	0.28 ^c
Pain (current)	0.07
Activity List	
Change in physical activities or lifestyle	0.10
Current activity level	0.25 ^c
Extended MRI Scale	
Total score	0.78 ^a
Ankle	0.53 ^a
Elbow	0.60 ^a
Knee	0.65 ^a

Note: All coefficients were calculated using Pearson correlations except for the Extended MRI Scale, which was calculated using Spearman correlations.

Abbreviations: CAJAS, Colorado Adult Joint Assessment Scale; MRI, magnetic resonance imaging.

^a $P < .001$.

^b $P < .01$.

^c $P < .05$.

reliability, internal consistency reliability, construct validity, and ability to detect change. Content and cultural validation showed a common understanding of CAJAS instructions and scoring among US and non-US PTs.

Evaluating the effects of prophylaxis on joint health in adults with hemophilia requires an assessment measure developed with sensitivity and specificity over a wide range of joint findings, given the wide range in manifestation of joint damage seen between older and younger patients and relative to the effectiveness of early therapy. Before the CAJAS, the only direct measure of joint health in adults was the Gilbert score.⁸ Although widely used, the Gilbert score was developed before the widespread availability of prophylaxis, especially in adults, and may be a better measure of joint health in patients with severe arthropathy than in those with mild joint damage.¹⁷ The validity, reliability, and ability of the Gilbert score to detect change have not been confirmed.¹⁸ The HJHS was developed to address the limitations of the Gilbert score and is more sensitive for detecting the earliest signs of joint damage. It has demonstrated reliability and validity in pediatric populations. Validation in adult populations over a wide range of joint pathologies has not been established.^{17,19}

The CAJAS, which preserves domains of the Gilbert frequently seen in more advanced joint arthropathy (i.e., axial deformity, contracture, and instability), also focuses on the earliest signs of joint change (i.e., swelling and gait changes), addressing the continuum of progressive joint degeneration. The HJHS did not retain axial deformity and instability, although contracture is recognized through the Extension Loss domain. The CAJAS is currently the only validated measure for physical assessment of joint structure and function designed specifically for adults with moderate to severe hemophilia. It has the potential to provide accurate, reliable assessment of joint function across a wide spectrum of joint disease. The CAJAS was successfully used in SPINART to assess joint health over a 3-year period in adolescent and adult patients with severe hemophilia A, demonstrating improvement in joint structure and function associated with prophylaxis.²⁰

4.1 | Limitations

A limitation of the validation study is that CAJAS assessments were conducted by a small number of PTs. However, the construct validity and cultural validation performed with PTs from three continents highlights the ability of the CAJAS to be applied to diverse populations, provided that PTs have adequate CAJAS training and understanding of English. Patients with very poor joint status were excluded from SPINART and the separate validation study. It is likely the CAJAS would be an appropriate measure in such patients but requires further testing.

Further validation of the CAJAS is needed for patients with hemophilia B and mild hemophilia A because the SPINART and validation studies included only patients with moderate-severe hemophilia A. However, joint bleeding, the underlying cause of the joint disease measured by the CAJAS, occurs at all levels of severity in patients with hemophilia A, hemophilia B, or other bleeding disorders. The progression of joint disease in patients with mild hemophilia, and in children who have been treated with prophylaxis or aggressive on-demand treatment protocols, differs with a smaller number of affected joints but not by qualitative changes in arthropathy. Evidence from in vitro animal and clinical studies indicates that very few hemarthroses into the same joint, an annualized joint bleed rate of 2 to 3, may cause irreversible, progressive structural joint changes.⁴ Joint damage noted in these populations indicates that some bleeding has occurred.⁵ The CAJAS domains of axial deformity, contracture, and instability reflect the joint changes seen in adults. For this reason, they were retained in the CAJAS. The CAJAS focus on identifying early joint degeneration suggests that it would be expected to perform similarly in mild patients and children who have only experienced minimal joint bleeding, although additional psychometric testing is needed.

In the assessment of the CAJAS ability to detect change, the lack of association between change in CAJAS total scores and change in Extended MRI Scale scores possibly resulted from the small changes seen in the Extended MRI Scale score over the 3-year observation period. Although prophylaxis can prevent joint bleeding and potentially slow hemarthropathy progression, it

Score difference from baseline, defined at the tertiles	CAJAS total score change		
	N	Mean change	Effect size
Haemo-QoL-A – physical functioning			
Improvement	24	–0.75	–0.20
No change	23	–0.13	–0.04
Deterioration	23	0.76	0.25
P value (linear trend)		.002 (.004)	
η^2		0.176	
r_s (P value)		–.39 (<.001)	
Activity List – current activity level			
Improvement	22	–0.52	–0.21
No change	31	–0.20	–0.06
Deterioration	16	0.79	0.20
P value (linear trend)		.020 (.009)	
η^2		0.111	
r_s (P value)		.26 (.028)	
Extended MRI Scale			
Improvement	20	–0.26	–0.06
No change	21	–0.30	–0.12
Deterioration	18	0.11	0.03
P value (linear trend)		.652 (.790)	
η^2		0.015	
r_s (P value)		.06 (.639)	

Note: Effect size thresholds: small, ≥ 0.2 ; medium, ≥ 0.5 ; large, ≥ 0.8 .¹⁶

Abbreviations: CAJAS, Colorado Adult Joint Assessment Scale; MRI, magnetic resonance imaging.

	Haemo-QoL-A physical functioning	Current activity level	Haemo-QoL-A physical functioning & current activity level, mean	Extended MRI score
Total	0.76	0.79	0.775	0.11
Ankle	1.18	1.39	1.285	0.10
Elbow	0.22	0.38	0.300	0.11
Knee	0.79	0.45	0.620	0.38

Abbreviations: CAJAS, Colorado Adult Joint Assessment Scale; MRI, magnetic resonance imaging.

	Baseline SD	α coefficient	0.5 SD	SEM	MDC ₉₀	RD 1 ^a (Lower)	RD 2 ^a (Higher)
Total	3.57	0.91	1.79	1.07	2.41	1.0	1.5
Ankle	3.83	0.85	1.92	1.48	2.84	1.5	2.0
Elbow	3.61	0.82	1.81	1.53	2.89	1.5	2.0
Knee	5.10	0.89	2.55	1.69	3.03	1.5	2.0

Abbreviations: CAJAS, Colorado Adults Joint Assessment Scale; MDC₉₀, minimal detectable change based on 90% confidence level; RD, responder definition; SD, standard deviation; SEM, standard error of mean.

^aRD 1 obtained with reference to SEM; RD 2 with reference to 0.5 SD.

likely cannot reverse established bone and cartilage damage, as reflected in the minimal change seen in Extended MRI Scale scores.²⁰ Prophylaxis, however, can improve some physical/

structural attributes measured by the CAJAS, including swelling, range of motion, muscle atrophy, and strength, all of which can affect the functional item of gait. This may be the cause of the

TABLE 5 Changes from baseline to final visit in CAJAS total scores by changes in patient health status

TABLE 6 Minimally important difference estimated for CAJAS total and joint subscale scores

TABLE 7 Distribution-based measures of responder definitions for the CAJAS total and joint subscale scores

discrepancy between the change scores of these two outcome measures.

Data on how the CAJAS could model joint change over longer periods (>3 years) are lacking. The strong association of changes in CAJAS with changes in other relevant hemophilia outcomes, including HRQoL, pain, and daily activities, supports the usefulness of CAJAS in assessing joint health and function in adults with hemophilia.

An additional limitation is the small sample size used to calculate the MID and responder definitions for the CAJAS total and joint scores. These calculations provide guidance for clinicians in determining whether treatment resulted in a clinically meaningful change in joint status. Although the results were fairly consistent, further analysis with a larger sample would provide additional support for the robustness of these estimates.

4.2 | Use of the CAJAS in clinical practice

To simplify the psychometric analyses, a single total score was calculated as the average of the mean scores for each joint type. This represents a potential study limitation, considering that in clinical practice, clinicians score each joint separately, so individual CAJAS joint scores can be compared over time. This allows clinicians to determine whether changes in the CAJAS total score result from substantial impairment in a single joint or from milder impairment in several joints. The calculation of a single total score was necessary for the purpose of CAJAS validation given that inclusion of multiple joint scores from single individuals would bias the statistical analysis.

Evaluating joint health in patients with hemophilia is a specialized area of physiotherapy. Use of the CAJAS by non-PT hemophilia clinicians or therapists with a general PT background does not guarantee the ability to properly conduct a CAJAS assessment and highlights the need for training in proper administration of this evaluation tool.

4.3 | Conclusion

Analysis of CAJAS data from SPINART and a separate content validation study demonstrated that CAJAS is a valid and reliable measure that can be used by trained PTs for assessing joint structure and function in adults with moderate-severe hemophilia. With increasing use of prophylaxis in adults with hemophilia, CAJAS can provide a useful means of assessing changes in joint health over time.

ACKNOWLEDGMENTS

We thank Helen Doll for assistance with data analysis. This study was funded by Bayer (Leverkusen, Germany). Medical writing assistance was provided by Karen L. Zimmermann from Complete Healthcare Communications, LLC (West Chester, PA) and Rachel Price from Darwin Medical Communications (London, England) and was fully funded by Bayer.

CONFLICT OF INTEREST

Sharon Funk developed and Marilyn J. Manco-Johnson implemented the use of the CAJAS, which will be copyrighted. Sharon Funk was a paid consultant on the Bayer-sponsored SPINART study. Marilyn J. Manco-Johnson has received honoraria for advisory board participation from Bayer, Baxter BioScience, Biogen Idec, CSL Behring, and Novo Nordisk and has received research grant funding from Bayer. Nikki Church is an employee of Bayer. Sylvia Engelen and Walter Hong were employees of Bayer at the time this study was conducted. Olga Moshkovich is an employee of ICON, an external vendor contracted for this research by Bayer. Katy Benjamin was an employee of ICON at the time this study was conducted. Brittany Gentile was an employee of ICON at the time this study was conducted and is a current employee of Roche/Genentech; the views and opinions expressed in this article are the author's own and do not necessarily reflect those of Roche/Genentech. Dianne Thornhill has no disclosures to address.

AUTHOR CONTRIBUTION

S.M. Funk developed the CAJAS, which was based on the earlier Gilbert scale; developed the training manual and produced the training video used for both the SPINART and CAJAS validation studies; participated in the validation study concept and design; trained evaluators in the United States, Argentina, Bulgaria, and Romania; provided quality assurance for CAJAS data in SPINART; and was an evaluator for patients in both the CAJAS validation study and SPINART. She also contributed to interpretation of data. S. Engelen participated in development of the CAJAS validation study protocol and contributed to data analysis and interpretation. K. Benjamin was the project director for the content validation and reliability studies, directed the project design for these studies, and participated in the analysis of all content and psychometric validation data. O. Moshkovich participated in developing the CAJAS validation study protocol, managed the validation study, and conducted interviews and qualitative analysis for the assessment on content validity and cultural validity. B. Gentile participated in developing the CAJAS psychometric validation study statistical analysis plan, conducting psychometric validation analyses, and summarizing results. N. Church and S. Engelen contributed to analysis/interpretation of data. W. Hong was involved in study concept and design. D. Thornhill conducted reliability analyses and results summarization, and contributed to data interpretation. M.J. Manco-Johnson participated in the development of the CAJAS examination, served on the Steering Committee that advised on and guided outcome assessment protocols for SPINART, and helped guide the data analysis and interpretation for the CAJAS. All authors contributed to the development of the manuscript, reviewed and commented on each draft, and approved the final draft.

ORCID

Sharon M. Funk  <https://orcid.org/0000-0001-5592-6140>

Nikki Church  <https://orcid.org/0000-0001-7618-6137>

Marilyn J. Manco-Johnson  <https://orcid.org/0000-0001-5215-1482>

REFERENCES

1. Srivastava A, Brewer AK, Mauser-Bunschoten EP, et al. Treatment guidelines working group on behalf of the world federation of hemophilia. Guidelines for the management of hemophilia. *Haemophilia*. 2013;19:e1-e47.
2. White GC 2nd, Rosendaal F, Aledort LM, et al. Definitions in hemophilia. Recommendation of the scientific subcommittee on factor VIII and factor IX of the Scientific and Standardization Committee of the International Society on Thrombosis and Haemostasis. *Thromb Haemost*. 2001;85:560.
3. Valentino LA. Blood-induced joint disease: the pathophysiology of hemophilic arthropathy. *J Thromb Haemost*. 2010;8:1895-1902.
4. Gringeri A, Ewenstein B, Reininger A. The burden of bleeding in haemophilia: is one bleed too many? *Haemophilia*. 2014;20:459-463.
5. 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.
6. Manco-Johnson MJ, Kempton CL, Reding MT, et al. Randomized, controlled, parallel-group trial of routine prophylaxis vs. on-demand treatment with sucrose-formulated recombinant factor VIII in adults with severe hemophilia A (SPINART). [published correction appears in *J Thromb Haemost*. 2014;12:119-122]. *J Thromb Haemost*. 2013;11:1119-27.
7. Collins P, Faradji A, Morfini M, Enriquez MM, Schwartz L. Efficacy and safety of secondary prophylactic vs. on-demand sucrose-formulated recombinant factor VIII treatment in adults with severe hemophilia A: results from a 13-month crossover study. *J Thromb Haemost*. 2010;8:83-89.
8. Gilbert MS. Prophylaxis: musculoskeletal evaluation. *Semin Hematol*. 1993;30:3-6.
9. Jackson SC, Yang M, Minuk L, et al. Prophylaxis in older Canadian adults with hemophilia A: lessons and more questions. *BMC Hematol*. 2015;15:4.
10. Oldenburg J, Brackmann HH. Prophylaxis in adult patients with severe haemophilia A. *Thromb Res*. 2014;134:S33-S37.
11. Oldenburg J. Optimal treatment strategies for hemophilia: achievements and limitations of current prophylactic regimens. *Blood*. 2015;125:2038-2044.
12. Biss TT, Chan AK, Blanchette VS, et al. The use of prophylaxis in 2663 children and adults with haemophilia: results of the 2006 Canadian national haemophilia prophylaxis survey. *Haemophilia*. 2008;14(5):923-930.
13. Manco-Johnson MJ, Nuss R, Funk S, Murphy J. Joint evaluation instruments for children and adults with haemophilia. *Haemophilia*. 2000;6:649-57. hae439 [pii].
14. Rothman M, Burke L, Erickson P, Leidy NK, Patrick DL, Petrie CD. Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR good research practices for evaluating and documenting content validity for the use of existing instruments and their modification pro task force report. *Value Health*. 2009;12:1075-1083.
15. Nunnally JC, Bernstein IH. *Psychometric theory*. New York, NY: McGraw-Hill; 1994.
16. Cohen J. *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
17. 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.
18. De Kleijn P, Heijnen L, Van Meeteren NL. Clinimetric instruments to assess functional health status in patients with haemophilia: a literature review. *Haemophilia*. 2002;8:419-427.
19. Hilliard P, Funk S, Zourikian N, et al. Hemophilia joint health score reliability study. *Haemophilia*. 2006;12:518-525.
20. Manco-Johnson MJ, Lundin B, Funk S, et al. Effect of late prophylaxis in hemophilia on joint status: a randomized trial. *J Thromb Haemost*. 2017;15:2115-2124.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Funk SM, Engelen S, Benjamin K, et al. Validity and reliability of the Colorado Adult Joint Assessment Scale in adults with moderate-severe hemophilia *AJ Thromb Haemost* 2020; 18:285-294. <https://doi.org/10.1111/jth.14651>.

APPENDIX

The SPINART CAJAS evaluators were: Dr Zakarija (United States) and Drs Nachevska and Todorov (Bulgaria); A. Iliescu and O. Surau (Romania); J. Carlos Lopez and D. Magliaro (Argentina); and S. Akins, K. Baumann, M. Braun, T. Casey, A. Clark, L. Doring, K. Dusek, S. Funk, J. Gibbs Hinkle, G. Hernandez, E. Hockey, A. Janich, J. Koop, J. Lee, D. Maloney, J. Musick, D. Oldfield, L. Slovensky, D. Voss, and R. Young (United States). The CAJAS validation study investigators were: S. Funk, L. Fox, D. Oleson, N. Durben, G. Hernandez, S. Leedy, S. Akins, and T. Kaltenmark (United States).