Title

Development and Reliability of the Joint Tissue Activity and Damage Examination for Quantitation of Structural Abnormalities by Musculoskeletal Ultrasound in Hemophilic Joints

Journal

Journal of Ultrasound in Medicine, 38(6)

ISSN

0278-4297

Authors

Volland, Lena M
Zhou, Jenny Y
Barnes, Richard FW
et al.

Publication Date

2019-06-01

DOI

10.1002/jum.14846

Development and Reliability of the Joint Tissue Activity and Damage Examination for Quantitation of Structural Abnormalities by Musculoskeletal Ultrasound in Hemophilic Joints

Lena M. Volland, PT, DPT, OCS, Jenny Y. Zhou, MD, Richard F. W. Barnes, PhD, Rebecca Kruse-Jarres, MD, Bruno Steiner, PT, DPT, RMSK, Doris V. Quon, MD, Cindy Bailey, PT, DPT, OCS, SCS, ATC, Tudor H. Hughes, MD, Randy E. Moore, DC, RDMS, RMSK, Eric Y. Chang, MD, Annette von Drygalski, MD, PharmD, RMSK

Supplemental material online at jultrasoundmed.org

Received July 25, 2018, from the Departments of Medicine, Division of Hematology/Oncology (L.M.V., J.Y.Z., R.F.W. B., A.v.D.); and Radiology (T.H.H.), University of California, San Diego, California, USA; Washington Center for Bleeding Disorders at Bloodworks Northwest, Seattle, Washington, USA (R.K.-J., B.S.); Orthopedic Hemophilia Treatment Center, Orthopedic Institute for Children, Los Angeles, California, USA (D.V.Q., C.B.); General Musculoskeletal Imaging, Inc, Cincinnati, Ohio, USA (R.E.M.); Radiology Service, VA San Diego Healthcare System, San Diego, California, USA (E.Y.C.); Department of Molecular Medicine, Scripps Research Institute, La Jolla, California, USA (A.v.D.). Manuscript accepted for publication September 24, 2018.

We thank Michelle Brackman, RN, for providing Joint Tissue Activity and Damage Examination (JADE) readings and Marlene Zepeda for creative editing of the JADE atlas. This study was supported by Health Resources and Services Administration service grant H30MC24045, Bioverativ (now Sanofi), and Pfizer. Dr Chang acknowledges support from the Veterans Affairs (grants I01CX001388 and I01RX002604). Dr von Drygalski has received honoraria for participating in scientific advisory board panels, consulting, and speaking engagements for Baxalta/Shire, Bayer, Bioverativ/Sanofi, CSL-Behring, HemaBiologics, Novo Nordisk, and Pfizer. The JADE protocol has been copyrighted for licensure by the University of California San Diego. Dr von Drygalski is a cofounder and member of the Board of Directors of Hematherix, LLC, a biotech company that is developing superFactor V therapy for bleeding complications.

Address correspondence to Annette von Drygalski, MD, PharmD, RMSK, Hemophilia and Thrombosis Treatment Center, University of California, 9333 Genesee Ave, Suite 310, San Diego, CA 92121 USA.

E-mail: avondrygalski@ucsd.du

Abbreviations

CI, confidence interval; HJHS, Hemophilia Joint Health Score; ICC, intraclass correlation coefficient; IQR, interquartile range; JADE, Joint Tissue Activity and Damage Examination; OMERACT, Outcome Measures in Rheumatology; UCSD, University of California San Diego; US, ultrasound

doi:10.1002/jum.14846

Objectives—Musculoskeletal ultrasound (US) is used increasingly to examine hemophilic arthropathy. However, quantitative algorithms to document findings are lacking. We developed and sought to validate a protocol quantifying hemophilic joint abnormalities.

Methods—Thirty-one patients with hemophilia were examined serially for 2 years with musculoskeletal US (≈600 joint examinations and ≈6000 images). Based on the spectrum of pathologies, a quantitative algorithm, named Joint Tissue Activity and Damage Examination (JADE), was developed for soft tissue and osteochondral measurements, including power Doppler, using nominal group techniques. To study intra- and inter-rater reliability, 8 musculoskeletal US–experienced hemophilia providers performed anatomic landmark recognition and tissue measurements on 86 images with arthropathic changes, with repetition 1 month later. Twenty-three musculoskeletal US–inexperienced providers performed similar assessments. Inter-operator reliability was established by 6 musculoskeletal US–experienced hemophilia providers, each acquiring images and JADE assessments of 3 hemophilic arthropathic joints. A radiologist and musculoskeletal sonographer functioned as adjudicators. The statistical analysis was performed with the intraclass correlation coefficient (ICC), Fleiss κ, and Cohen κ where appropriate.

Results—The musculoskeletal US–experienced providers showed excellent intra- and inter-rater reliability for tissue measurements (ICCs, 0.94–0.96). Agreement was good to excellent for landmark recognition (Fleiss κ, 0.87-0.94). Inter-operator reliability was excellent for measurements and landmark recognition (ICC, 0.90; Fleiss κ, 1.0). Agreement with adjudicators was mostly good to excellent. Musculoskeletal US–inexperienced providers showed excellent inter-rater reliability for measurements (ICC, 0.96) and moderate agreement for landmark recognition (Fleiss κ, 0.58).

Conclusions—The JADE protocol appears feasible for quantifying hemophilic intra-articular abnormalities. Musculoskeletal US–trained hemophilia providers showed high intra-rater, inter-rater, and inter-operator reliability, supporting JADE as a protocol for clinical management and research.

Key Words—arthropathy; hemophilia; inter-operator reliability; inter-rater reliability; musculoskeletal ultrasound; reliability

Hemophilia is an X chromosome–linked bleeding disorder manifested by repetitive hemarthroses leading to debilitating joint arthropathy,[1,2] characterized by soft tissue hypertrophy and inflammation, as well as osteochondral deformities and destruction.[2,3] The extent to which these changes are reversible is unclear. In recent years, the use of high-resolution musculoskeletal ultrasound (US) has increased rapidly for fast and accurate detection of bleeding episodes, as well as for longitudinal assessment and management of joint health by nonradiologists.[4–10] Musculoskeletal US is a fast, accessible, and economical alternative to other imaging modalities, such as magnetic resonance imaging,[11–13] and can objectively diagnose various musculoskeletal conditions, and support treatment decisions.[4,5,7–9] Power Doppler US permits assessments of microperfusion changes and the extent of inflammation in joint tissue.[14–16] Within this realm, musculoskeletal US is evolving as a versatile tool for monitoring osteochondral health, soft tissue expansion, and vascular abnormalities over time in hemophilic joints.

However, the assessment of longitudinal changes of joint health, comprising the systematic quantification of soft tissue and osteochondral changes and widely accepted as measures of progression of hemophilic arthropathy, requires standardized, validated protocols. Toward that end, several semiquantitative musculoskeletal US scoring algorithms have been proposed,[12,16–20] especially for the detection of early changes.[18] However, these scoring algorithms lack iterative validation processes ("truth, discrimination, and feasibility"), as proposed by Outcome Measures in Rheumatology (OMERACT).[21–23] Additionally, scoring algorithms are relatively subjective, may have ceiling effects in advanced hemophilic arthropathy, and do not permit the following of individual findings in hemophilic joints, which may be central to patients' specific presentations but outside the scoring algorithms.

To fill this void, our objective was to develop a quantitative musculoskeletal US protocol evaluating soft tissue proliferation and inflammation (joint tissue activity) as well as osteochondral and cartilage alterations (joint damage), named Joint Tissue Activity and Damage Examination (JADE). The protocol was designed to capture fluctuating soft tissue changes, precisely measure osteochondral surface and cartilage alterations, and enable dynamic follow-up of individual, patient-specific intra-articular findings. Here we describe the development of JADE with the objective of reporting intra- and inter-rater as well as inter-operator reliability among a group of hemophilia providers trained and experienced in musculoskeletal US assessments of hemophilic joints, aligning our efforts with OMERACT for the validation of US imaging techniques.[21,24]

The OMERACT guidelines are considered the "reference standard" for validating imaging protocols for the assessment of arthritic changes, and following those guidelines appeared essential to provide confidence that the JADE parameters would bear relevance as outcome tools. OMERACT is a widely accepted international initiative, whereby the cornerstones of validation are composed of correct tissue annotation, the establishment of intra-rater, inter-rater, and inter-operator reliability, and clinical applicability for specific disease states and patient cohorts, emulating the OMERACT filter components truth, discrimination, and feasibility.[21,23,24] Along OMERACT recommendations, we recently established the extent of pathologic tissue discrimination with musculoskeletal US specific to hemophilic joints by comparison with conventional and ultrashort time-to-echo magnetic resonance imaging.[25]

Materials and Methods

Joint Examination With Musculoskeletal US

The JADE protocol was developed from 600 examinations of hemophilic joints by musculoskeletal US in 31 patients, yielding approximately 6000 views. Musculoskeletal US examinations were performed with high-resolution equipment (LOGIQ S8, GE Healthcare, Fairfield, CT), applying presets recommended by the manufacturer and using standardized and validated transducer positions as developed by the European League Against Rheumatism[26,27] to ascertain optimal views of joint structures sonoanatomically and in pathologic states (http.ultrasound.eular.org). Grayscale (B-mode) and power Doppler examinations of both ankles, elbows, and knees in each patient were performed with a high-frequency linear transducer (8–15 MHz). The power Doppler mode was adjusted to

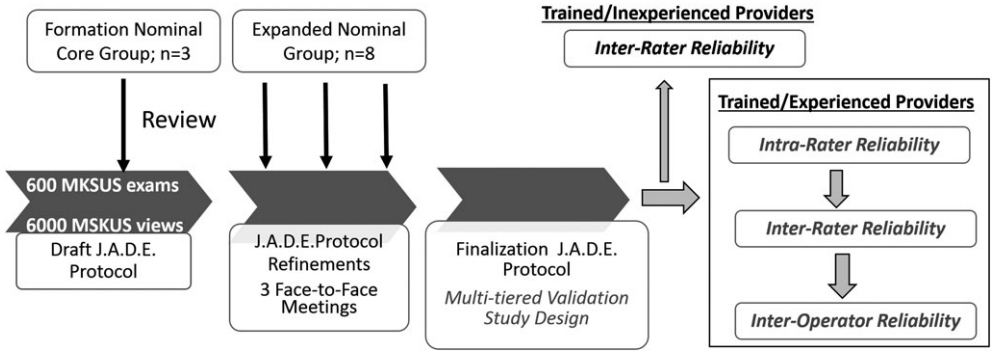
optimize image acquisition,[28] as recommended by the European League Against Rheumatism.[27,29,30] Likewise, as recommended by European League Against Rheumatism, sonopalpation and dynamic movements served to determine compressibility and displacement of intra-articular material.[27] Musculoskeletal US images were obtained by a hematologist (A.v.D.) with 5 years of experience, who was formally trained and certified in musculoskeletal US by the American Registry for Diagnostic Medical Sonography.

The 31 patients scanned to develop the JADE protocol presented with severe, moderate, or mild hemophilia A or B and at least 1 arthropathic joint. Severe, moderate, and mild hemophilia A or B were denoted as factor VIII or factor IX plasma activity levels of less than 1%, 1% to 5%, and greater than 5%, respectively. Arthropathic joints were defined either by radiographic Pettersson scores or by Hemophilia Joint Health Scores (HJHSs) in those few patients without Pettersson scores available.[28,31] Based on published correlations between the 2 joint outcome measures, Pettersson scores or HJHSs had to be at least 1 or higher or 3 or higher, respectively,[32,33] to suggest arthropathic changes. Patients were enrolled prospectively and followed with musculoskeletal US and power Doppler examinations of both ankles, elbows, and knees during pain-free intervals every 6 to 9 months over 2 years, as well as sporadic examinations of acutely painful joints. Demographic information was extracted from the electronic medical record. The study protocol was approved by the University of California San Diego (UCSD) Human Research Protection Program, and informed written consent was obtained from all patients.

Consensus Development of JADE

Following previously adopted consensus development methods for imaging techniques in other inflammatory arthritic conditions,[27,34] a nominal group method was used to determine the most appropriate transducer positions and measurements to assess osteochondral interface irregularities, cartilage changes, and soft tissue proliferation, as well as inflammatory signals within the elbow, knee, and ankle. A flowchart depicting the evolution of JADE is provided in Figure 1. The members of the nominal group were chosen for their long-standing expertise in their respective fields and consisted of a fellowship-trained musculoskeletal radiologist with 7 years of experience (E.Y.C.), a hematologist with 5 years of scanning experience (A.v.D.), and a musculoskeletal sonographer with greater than 20 years of experience (R.E.M.). Images obtained during the 600 musculoskeletal US joint evaluations (≈6000 views) were reviewed by the 3 members of the nominal group, with a special focus on results from an accompanying pathologic tissue validation study in hemophilic arthropathy, in which musculoskeletal US views were aligned with corresponding views of conventional and ultrashort time-to-echo magnetic resonance imaging,[25] obtained

Figure 1. Flowchart of JADE development. The development of JADE was a 3-year nominal process and involved core and expanded review groups. The core group members consisted of a musculoskeletal radiologist, hematologist, and sonographer, all with high expertise in their respective fields. The core group developed the JADE protocol. Five physicians and physical therapists, trained and experienced in musculoskeletal (MSK) US, were added in a second step to form the expanded group to refine and finalize the protocol in 3 face-to-face meetings. Subsequently, the JADE protocol was studied for intra- and inter-rater reliability as well as inter-operator reliability.



simultaneously. This review defined the following: (1) the extent to which musculoskeletal US discriminated different types of pathologic tissue for descriptive charting of findings; (2) the feasibility of direct measurements of soft tissue expansion, osteochondral changes, and cartilage thickness; and (3) the standardized locations of US assessments. Of note, since musculoskeletal US does not permit the distinction of soft tissue types, such as synovium, fatty tissue, and blood clots in pathologic hemophilic joints,[25] the term "soft tissue expansion" was deemed more appropriate to describe accumulation of abnormal intra-articular tissue, compared to historical terms such as "synovial proliferation."

In a second step, the nominal group was expanded by an additional 5 hemophilia care providers trained and experienced in musculoskeletal US, including 2 physicians (R.K.-J. and D.V.Q.) and 3 physical therapists (B.S., C.B., and L.M.V.). Each of the providers had 1 to 3 years of musculoskeletal US practice experience. The proposed JADE algorithm was further refined in several face-to-face meetings by consensus to establish the final JADE protocol and atlas (online Supplement; instructions provided in online modules accessible at https://cme.ucsd.edu/MSKUSWeb/webmodules.html).

Inter-rater Reliability

For evaluation of inter-rater reliability of the JADE protocol, 8 hemophilia providers (6 members of the nominal group) experienced in musculoskeletal US, including 4 physicians, 3 physical therapists, and 1 nurse, from 3 United States hemophilia treatment centers, reviewed 86 still images of the elbow (n = 20), knee (n = 36) and ankle (n = 30), acquired previously by A.v.D., in a scrambled and blinded fashion. All views of the JADE protocol were represented. An experienced musculoskeletal US provider was defined as a health care professional who was trained in musculoskeletal US at the UCSD (https://cme.ucsd.edu), with performance and interpretation of at least 3 joint US examinations weekly for the previous 6 months. The aforementioned musculoskeletal US course comprises a 20-hour theoretical and hands-on curriculum, including fundamentals of US physics, sonoanatomy, and sonopathology of articular structures in ankles, knees, and elbows, which are the main joints affected by hemophilic arthopathic changes. Participants also are taught the use of power Doppler US to detect abnormal vascularity and the appearance of complex and simple effusions. The musculoskeletal US images for review were selected by a fellowship-trained musculoskeletal radiologist (E.Y.C., with 7 years of experience) to provide a broad range of arthropathic changes equally distributed between elbow, knee, and ankle joints with midrange Pettersson scores (3–8; normal, 0; maximum, 13) or HJHSs (4–12; normal, 0; maximum, 20). The musculoskeletal US–experienced providers were asked to annotate major anatomic landmarks in all JADE positions and to provide 2-dimensional measurements of osteochondral interface irregularities and cartilage thickness in specified locations perpendicular to the bony margin, as well as soft tissue measurements in precise areas defined by the JADE protocol (online Supplement). With Sante DICOM Viewer software (Santesoft, Athens, Greece), measurements were made on images generated with a transducer frequency of 12 MHz or greater (axial resolution of ≈0.1–0.2 mm). Osteochondral surface irregularities were defined as cortical breaks or alterations of the normally smooth hyperechoic osseous line (online Supplement). Additionally, power Doppler signals were evaluated semiquantitatively as previously described,[15,35] using a score between 0 and 3 in the tissue area of interest, with a higher score indicating a stronger power Doppler signal (online Supplement).

To gauge the ability to make annotations required in the JADE protocol after a single musculoskeletal US training course, 23 hemophilia providers without previous experience in musculoskeletal US, who were freshly trained in the UCSD CME-accredited course, reviewed and annotated the same still images as the experienced providers but limited to the knee. Power Doppler scoring was excluded, since additional practice is needed to become fully familiarized with power Doppler detection. The providers performed the assessment immediately after course completion. Data collection and analysis were approved by the UCSD Human Research Protection Program.

Intra-rater Reliability

One month after the initial JADE annotation, the musculoskeletal US–experienced group of hemophilia

providers repeated the assessment of all 86 scrambled images.

Inter-operator Reliability

Six hemophilia providers (3 physicians and 3 physical therapists) from the initial nominal group individually acquired all images included in the JADE protocol for the elbow, knee, and ankle of 3 patients with severe hemophilia A and moderate joint disease, judged clinically by R.K.-J. and B.S. Image acquisition and interpretation, as specified per JADE, were completed within 1 day at Bloodworks Northwest and were approved by the Institutional Review Board of Bloodworks Northwest. Providers completed measurements of osteochondral interface irregularities (n = 3), cartilage thickness (n = 7), soft tissue expansion (n = 4), power Doppler signal scoring (n = 17), and qualitative assessments (n = 17).

Image Adjudication

All annotations and measurements provided by the musculoskeletal radiologist (E.Y.C.) and musculoskeletal sonographer (R.E.M.) were used separately for adjudication of results.

Statistical Analysis

The intra- and inter-rater reliability was established by assessment of the intraclass correlation coefficient (ICC) for continuous data and Fleiss κ (n > 2 providers) and Cohen κ (n = 2 providers) for categorical data.[36,37] The measures evaluated the level of correlation and agreement between different assessments and enabled the evaluation of the degree of reliability. Based on a level of absolute agreement, all values, including the corresponding 95% confidence intervals (CIs), were calculated by AgreeStats 2015.6 software (Advanced Analytics, LLC, Gaithersburg, MD) with a 2-way mixed-effects model for intra-rater reliability and a 2-way random-effects model for inter-rater reliability for the continuous data. A chance-corrected agreement coefficient for multiple raters (Fleiss and Cohen κ) was used to evaluate the categorical data.[37] Levels of acceptance were set at values of less than 0.5 indicating poor reliability, values between 0.5 and 0.74 corresponding to moderate reliability, values between 0.75 and 0.9 resulting in good reliability, and values above 0.9 yielding excellent reliability, as suggested by Koo and Li.[38]

Table 1. Patient Demographics and Joint Characteristics

Characteristic	Value
Hemophilia A, n	27
Mild	4
Moderate	2
Severe	21
Hemophilia B, n	4
Mild	0
Moderate	1
Severe	3
Age, y	37 (29–54)
Joint characteristics	
Arthropathic joints, n	
Total	100
Elbow	34
Knee	36
Ankle	30
HJHS	
Total	11.0 (5.3–27.8)
Elbow	4.0 (0–8.0)
Knee	2.5 (0–6.0)
Ankle	2.5 (0–10.0)
Pettersson score	
Total	26.0 (12.0–47.0)
Elbow	10.0 (1.0–16.0)
Knee	5.2 (1.5–14.5)
Ankle	4.9 (3.5–14.5)

Data are presented as median (IQR) where applicable. Maximum HJHS, 120; per joint, 20. Maximum Pettersson score, 78; per joint, 13.

Results

Patient and Joint Characteristics

Information about demographics and the joint disease status of the 31 patients whose musculoskeletal US images were used for the development of the JADE protocol as well as for the assessment of intra- and inter-rater reliability are summarized in Table 1. Briefly, a total of 109 arthropathic joints (30 elbows, 43 knees, and 36 ankles) as well as 92 non-arthropathic joints (34 elbows, 30 knees and 28 ankles) were examined serially (2 or 3 times during a 2- to 3-year period), comprising 582 evaluations (185 elbows, 213 knees, and 184 ankles), with additional evaluations during acutely painful episodes (n = 10 elbows, 8 knees, and 11 ankles), yielding approximately 6000 images, as described in "Materials and Methods." Median Pettersson scores and HJHSs of all 6 joints combined (elbows, knees, and ankles) were 26 (interquartile range [IQR], 12–47) and 11 (IQR, 5.3–27.8), respectively, suggestive of a wide

spectrum of disease severity and appearing to be suitable for developing an assessment tool.

The JADE Algorithm

After review of the 600 US examinations by the nominal group (A.v.D., E.Y.C., and R.E.M.) in association with accompanying conventional and ultrashort time-to-echo magnetic resonance imaging in 36 of the 109 arthropathic joints scanned with musculoskeletal US, a first draft of JADE was developed. The process of the JADE protocol development is outlined in Figure 1. The JADE protocol was based on the prerequisite to permit individual, direct measurements of intra-articular soft tissue expansion, cartilage thickness, and osteochondral interface irregularities in elbows, knees, and ankles in hemophilic joints. The consensus review indicated that a protocol featuring direct measurements was preferable to scoring algorithms, as previously proposed, to avoid ceiling effects in advanced arthropathic joints, enable localization and description of specific individual findings to be followed longitudinally, and create a versatile tool to assess a wide range of pathologic findings for clinical management as well as for use as an outcome measure. Additionally, it was decided also to include the description of characteristics more pertinent for clinical use, such as predominant tissue echogenicity, abnormalities of key structures (for instance, medial meniscus or Achilles tendon), and tissue compressibility by sonopalpation to differentiate between effusions and soft tissue in certain areas. Also, annotations pertaining to the complexity of effusions were enabled to take advantage of the exquisite sensitivity of musculoskeletal US to distinguish between bloody and serous effusions,[39] as well as quantification of abnormal soft tissue microcirculation and vascularity changes by power Doppler signals. Subsequently, the JADE protocol underwent several iterations and improvements based on constructive feedback from the expanded nominal group (B.S., C.B., D.V.Q., L.M.V., and R.K.-J.) during several face-to-face meetings over a 2-year period, while the providers applied JADE informally to assess the joint status during clinic visits. The JADE protocol development, including the atlas and charting tool, was a 3-year process. As a result of this process, the expanded nominal group thought that the definitive version of the JADE protocol, provided in the online Supplement, offered a versatile tool for following overall joint health longitudinally and also focusing on the progression or resolution of individual, patient-specific findings. As such, the JADE protocol can be used for all 6 joints (elbows, knees, and ankles) or as a focused assessment of individual joints (ie, target joints), for which the examination can be limited to specific areas within the joint. All JADE views are deemed optional and can be chosen on the basis of individual joint abnormalities. The complete JADE protocol includes 17 standardized transducer positions subdivided into the elbow (3 views), knee (10 views), and ankle (4 views). These standardized transducer positions are further described as follows: elbow, anterior humeroradial and humeroulnar joints in the transverse axis, anterior humeroradial joint in the longitudinal axis, and posterior olecranon recess in the longitudinal axis; knee, suprapatellar and infrapatellar recesses in the longitudinal axis, medial and lateral recesses in the transverse axis, medial meniscus in the longitudinal axis, sunrise view, and anterior/posterior medial/lateral condyle in the longitudinal axis; ankle, anterior tibiotalar joint in transverse and longitudinal axes, subtalar joint in the longitudinal axis, and Achilles tendon in the longitudinal axis. The full version of the JADE protocol, including a detailed description of all transducer positions, measurements, and qualitative assessments, is included as the online Supplement.

Number of Image Annotations and Measurements Performed by the Raters

The experienced providers were presented with the same 172 musculoskeletal US still images (2 × 86) of elbows, knees, and ankles twice, 4 weeks apart, to determine the intra-rater reliability of the JADE protocol. Providers were tasked with identifying major anatomic landmarks (categorical data) and complete direct osteochondral interface, cartilage, and soft tissue measurements (continuous data) on both occasions. The inexperienced providers were tasked with completing landmark recognition and measurements only once on 31 musculoskeletal US still images of the knee. Altogether, 3504 responses for landmark recognition were obtained from the experienced providers (1632 osseous structures, 1232 soft tissues, 400 cartilages, and 240 power Doppler signal assessments) and 1817 responses (851 osseous structures,

851 soft tissues, and 115 cartilages) from the inexperienced providers. Regarding quantitative measurements, the experienced group completed 1360 measurements (560 soft tissues, 560 cartilages, and 240 osteochondral interfaces), and the inexperienced group completed 460 measurements (230 cartilages, 115 soft tissues, 115 osteochondral interfaces).

Intra-rater Reliability: Experienced Providers

Intra-rater reliability was established for continuous and categorical parameters among musculoskeletal US–experienced providers. The continuous parameters comprised 2-dimensional measurements of soft tissue expansion, osteochondral surface irregularities, and cartilage thickness in 12 joints (3 elbow, 5 knee, and 4 ankle) in 9 key locations. Intra-rater reliability was excellent for all assessments in elbows, knees, and ankles (all ICCs, ≥ 0.96; Table 2). When divided by tissue types, intra-rater reliability was excellent for cartilage thickness and soft tissue expansion and good for measurements of osteochondral interface irregularities (Table 2).

The categorical parameters included anatomic landmark recognition and power Doppler signal assessments. Intra-rater reliability for assessments separated by joints was good to excellent (Cohen κ, 0.85–0.99) and amounted to excellent when values of all joints were combined (Cohen κ, 0.98). Isolated analyses for the detection of power Doppler signals yielded overall excellent intra-rater reliability (Cohen κ, > 0.90; Table 3).

Inter-rater Reliability: Experienced Providers

Inter-rater reliability was established for the continuous and categorical parameters among musculoskeletal US–experienced providers, assessing the same number and type of views as well as parameters used for determination of intra-rater reliability.

Table 2. Intra-rater Reliability for Continuous Parameters Among Providers Experienced in Musculoskeletal US

Parameter	ICC	95% CI
All joints and assessments	0.98	0.97–0.98
Subcategory: joints		
Elbow	0.97	0.95–0.99
Knee	0.99	0.99–0.99
Ankle	0.96	0.94–0.98
Subcategory: tissue type and assessment		
Osteochondral interface irregularities	0.85	0.73–0.93
Cartilage thickness	0.99	0.98-0.99
Soft tissue expansion	0.98	0.97–0.99

Continuous parameters comprise 2-dimensional measurements of anatomic structures.

Table 3. Intra-rater Reliability for Categorical Parameters Among Providers Experienced in Musculoskeletal US

Parameter	Cohen κ	95% CI
Anatomic landmark recognition		
All joints and assessments	0.98	0.96–1.00
Subcategory: joints		
Elbow	0.99	0.96–1.00
Knee	0.85	0.73–0.97
Ankle	0.86	0.69–1.00
Power Doppler signal		
All joints	0.93	0.88–0.97
Elbow	0.93	0.85–1.00
Knee	0.96	0.91–1.00
Ankle	0.85	0.71–0.98

Categorical parameters include anatomical landmark recognition and power Doppler signal assessment.

For the continuous parameters, excellent inter-rater reliability was shown for the entire JADE protocol, including all joint types and measurements during the initial and repeated assessments 4 weeks later (both ICC, 0.94). Separated into joint types, inter-rater reliability for the knee and elbow were excellent during the first and repeated assessments (all ICCs, ≥ 0.91) and increased for the ankle from good (ICC, 0.89) to excellent (ICC, 0.92). The ICCs for cartilage thickness (0.98–0.99) and soft tissue expansion (0.97–0.98) measurements were excellent on both occasions, whereas ICCs for osteochondral interface irregularity measurements were moderate (0.66–0.67; Table 4).

Good to excellent inter-rater reliability was shown for all categorical parameters of landmark recognition combined for all joints (Fleiss κ, 0.87–0.91) during both assessments (Table 5). Separated into joint types and assessment time points, good to excellent inter-rater reliability was demonstrated for all joints on both occasions (Fleiss κ, 0.80–0.95). Inter-rater reliability for power Doppler signal assessments was overall excellent (Fleiss κ, ≥ 0.90).

Inter-Rater Reliability: Inexperienced Providers

The group of 23 musculoskeletal US–inexperienced providers completed the evaluation of 31 musculoskeletal US images of the knee, comprising measurements of soft tissue expansion, cartilage thickness, osteochondral surface irregularities, and landmark recognition. Inter-

Table 4. Inter-rater Reliability for Continuous Parameters Among Experienced Providers in Musculoskeletal US

	Initial Assessment		Repeated Assessment	
Parameter	**ICC**	**95% CI**	**ICC**	**95% CI**
All joints and assessments	0.94	0.92–0.96	0.94	0.93–0.96
Subcategory: joints				
Elbow	0.95	0.90–0.98	0.91	0.84–0.95
Knee	0.96	0.94–0.98	0.98	0.97–0.99
Ankle	0.89	0.83–0.93	0.92	088–0.95
Subcategory: tissue type and assessment				
Osteochondral interface irregularities	0.66	0.42–0.85	0.67	0.42–0.85
Cartilage thickness	0.98	0.97–0.99	0.99	0.98–0.99
Soft tissue expansion	0.98	0.97–0.99	0.97	0.96–0.99

Continuous parameters comprise 2-dimensional measurements of anatomic structures.

rater reliability was excellent for all direct measurements combined (ICC, 0.96), with ICCs for osteochondral interface irregularities, cartilage thickness, and soft tissue expansion measurements ranging from moderate to good (0.63–0.86). The inter-rater reliability for the categorical data (landmark recognition) was moderate (Fleiss κ, 0.58).

Inter-rater Reliability Comparison to Adjudicators

Two adjudicators (E.Y.C. and R.E.M.) completed the same measurements and annotations as the experienced and inexperienced groups. The agreement between the adjudicators was found to be excellent to good (ICC, 0.84–0.99) for the continuous data and excellent (Cohen κ, 1.00) for the categorical data of the JADE protocol. Both groups of providers (experienced and inexperienced) were compared individually to the adjudicators for the continuous and categorical data.

Table 5. Inter-rater Reliability for Categorical Parameters Among Experienced Providers in Musculoskeletal US

	Initial Assessment		Repeated Assessment	
Parameter	**Fleiss κ**	**95% CI**	**Fleiss κ**	**95% CI**
Anatomic landmark recognition				
All joints and assessments	0.87	0.78–0.97	0.91	0.81–1.00
Subcategory: joints				
Elbow	0.89	0.56–1.00	0.87	0.63–1.00
Knee	0.95	0.71–1.00	0.80	0.37–1.00
Ankle	0.87	0.55–1.00	0.90	0.73–1.00
Power Doppler signal				
All joints	0.91	0.81–1.00	0.90	0.90–0.99
Elbow	0.97	0.68–1.00	0.93	0.91–1.00
Knee	0.91	0.68–1.00	0.94	0.79–1.00
Ankle	0.93	0.71–1.00	0.75	0.32–1.00

Categorical parameters include anatomic landmark recognition and power Doppler signal assessment.

The experienced group, composed of the 8 raters, measured against the 2 adjudicators on 2 separate occasions, provided 32 adjudications for continuous parameters. Inter-rater agreement was excellent (ICCs, > 0.9) and good (ICCs, 0.83–0.9) for 27 and 5 adjudications, respectively, for all joints and subcategories combined (online Supplemental Table 1). Results appeared consistent across joint types (elbows, knees, and ankles; online Supplemental Table 1). When the adjudication analysis was divided by tissue type, the inter-rater agreement was excellent for all cartilage thickness and soft tissue measurement adjudications (ICCs, ≥ 0.97). The adjudication of the measurements of osteochondral interface irregularities yielded excellent inter-rater agreement on 2 occasions, with most values indicating good (24 values; ICCs, 0.75–0.9) and 6 values indicating moderate (ICCs, 0.62–0.74) inter-rater agreement, respectively.

The adjudication for landmark recognition consisted of 32 adjudications and revealed excellent agreement for 26 adjudications (Cohen κ, > 0.9) and good agreement for 6 adjudications (Cohen κ, 0.77–0.88). Results were consistent across joint types (elbows, knees, and ankles). Power Doppler signal assessments revealed excellent (Cohen κ, > 0.9), good (Cohen κ, 0.78–0.89), and moderate (Cohen κ, 0.71) agreement for 12, 19, and 1 adjudications, respectively (online Supplemental Table 2).

The inexperienced group, composed of the 23 raters, measuring structures against the 2 adjudicators, provided 46 adjudications on agreement of continuous parameters combined within the knee. Overall agreement was excellent, with all 46 assessments showing ICCs of 0.94 or greater. Divided by tissue types, which yielded 138 individual assessments, some variability in agreement became apparent, which was neither tissue nor rater specific; however, only 28 of 138 assessments (20.2%) showed poor/moderate agreement. Landmark recognition appeared easy for most providers, as evidenced by

excellent and good agreement in 31 and 9 of 46 assessments (all ICCs, 1.0), respectively.

Inter-operator Reliability

Six musculoskeletal US–experienced providers of the nominal group and 1 adjudicator (R.E.M.) independently acquired all JADE views of 1 elbow, knee, and ankle from 3 patients with severe hemophilic arthropathy, evidenced by abnormal HJHSs (median, 24; IQR, 22.0–35.5), followed by independent measurements of cartilage thickness, soft tissue expansion, and osteochondral surface irregularities. In addition, everyone completed qualitative descriptions of findings and power Doppler signal assessments, as outlined in the JADE protocol (continuous and categorical parameters). Since patients were examined at their baseline, there were no effusions present for annotation, and inflammatory power Doppler signals were scarce. Power Doppler scores of 0 and 1 were combined into a single category for the statistical analysis, since the differentiation between these can be challenging in the presence of operational motion background scatter, and they bear little clinical relevance. A total of 102 measurements and 17 categorical assessments were completed between the operators.

Inter-operator reliability was excellent for all direct measurements in all joints combined (ICC, 0.90; 95% confidence interval [CI], 0.85–0.95). Divided into joints, inter-operator reliability was excellent for elbow (ICC, 0.93; 95% CI, 0.78–1.0) and knee (ICC, 0.93; 95% CI, 0.87–0.97) and moderate for ankle (ICC, 0.72; 95% CI, 0.44–0.93) assessments. Divided into tissue subtypes, measurements of osteochondral interface irregularities (ICC, 0.76; 95% CI, 0.38–0.99), cartilage thickness (ICC, 0.85; 95% CI, 0.71–0.94), and soft tissue expansion (ICC, 0.79; 95% CI, 0.60–0.93) showed good inter-operator reliability. Power Doppler signal assessments revealed excellent inter-rater reliability in all joints (Fleiss κ, 1.0).

Subsequently, the 6 operators were compared individually to the adjudicator (R.E.M), as displayed in the continuous data, including all measurements for the elbow, knee, and ankle. This comparison revealed excellent agreement between the adjudicator and 4 providers (ICC, 0.90–0.98), and good-to-moderate agreement between the adjudicator and 2 providers (ICC, 0.70–0.85). Whereas elbow and knee measurements showed predominantly excellent/good agreements with a small range of ICC values, the ankle analysis revealed larger discrepancies, ranging from excellent to moderate (ICC, 0.70–0.95), with 2 outliers in the poor range. When divided into tissue subtypes, cartilage and soft tissue measurements showed predominantly excellent/good ICCs, with only a small range of ICC values, whereas the assessments of the osteochondral interface irregularities revealed larger discrepancies, ranging primarily from excellent to moderate (ICC, 0.70–0.91), with 1 outlier in the poor range. Power Doppler signal assessments showed excellent inter-rater agreement between the adjudicator and the providers for all joints (Cohen κ, 1.0).

Discussion

The JADE protocol was developed to follow joint health of patients with hemophilia longitudinally in a systematic and quantitative manner, focusing on soft tissue expansion, inflammation, and osteochondral alterations, to provide precise, patient-specific assessments complementing semiquantitative scoring algorithms.[12,16–20] Two-dimensional measurements of osteochondral and soft tissue evaluation in musculoskeletal medicine have become increasingly attractive,[12,40–46] since they take advantage of the high spatial and tissue contrast resolution inherent to continuously improving US technology, enabling precise measurements in 0.1- to 0.2-mm increments with higher-frequency transducers (>12 MHz). Here we report the intra- and inter-rater reliability as well as inter-operator reliability of the JADE protocol, incorporating direct measurements by hemophilia care providers trained in musculoskeletal US, in alignment with OMERACT recommendations for the validation of imaging protocols.[21–24] Reliability is defined as the ability of a measurement to be replicated and mathematically illustrated by a value between 0 and 1, with values closer to 1 representing stronger reliabilities.[38] Most of the assessments and measurements exceeded values of 0.70, which provides confidence that the JADE protocol is suitable for implementation in hemophilia practice and can serve as an objective imaging outcome measure in clinical studies, evaluating effects of new therapies on the progression or resolution of individual pathologic findings. The reliability and reproducibility of the JADE protocol

were very reasonable, indicating that assessments can be performed by trained and experienced nonradiologists. These findings are consistent with several other studies demonstrating acceptable feasibility and reliability of various musculoskeletal US assessments mainly performed by rheumatologists.[47,48] In that setting, positive recommendations for translation of musculoskeletal US into practice and its use in clinical studies were provided if a moderate to high level of congruent readings was achieved.

Pertaining to JADE, intra-rater reliability for reading musculoskeletal US still images, including direct measurements of soft tissue expansion, cartilage thickness, and osteochondral alterations, was good to excellent, ensuring accurate longitudinal assessments of findings if read by the same health care provider. Many US machines have a "compare assist function," permitting side-by-side comparisons of sequential studies, which is expected to further enhance precision regarding landmark recognition and associated measurement placements. Similarly, there was overall good-to-excellent inter-rater reliability when reading still images between the providers, providers and adjudicators, and adjudicators. However, whereas readings detailing cartilage and soft tissue thickness were in agreement, measurements of osteochondral interface irregularities showed greater variations, even between the adjudicators. This finding suggests that measurements of osteochondral alterations were more subjective than measurements of cartilage or soft tissue thickness. Regardless, small reading discrepancies between providers may not be clinically meaningful because of the high-resolution measurements in very small increments of 0.1 to 0.2 mm. The safest recommendation is to assign longitudinal assessments of osteochondral interface irregularities to the same provider or a central reader in clinical studies.

As expected, much greater variability and discrepancies in landmark recognition and tissue measurements were present between the inexperienced providers, who had just completed a single training course. This observation emphasizes the fact that experience and hands-on practice are required beyond didactic instruction to optimize the accuracy of the readings and the interpretation of musculoskeletal US images. This and other observations support the belief that point-of-care (musculoskeletal) US imaging can be learned and correctly applied by nonradiologists but requires appropriate education, continued practice, and accumulation of experience to ensure accurate clinical assessments.[49–53]

A final step was the establishment of inter-operator reliability, which examined the extent of agreement of image interpretation between providers when acquiring images independently. The operators were experienced in musculoskeletal US and acquired the images from patients with hemophilia during a 1-day workshop. Joints with a moderate disease burden were thought to be most appropriate for this exercise to visualize a reasonable spectrum of abnormalities and measurable structures. Furthermore, the evaluation of moderate disease processes ensures a more rigorous assessment of reliability between different operators, since reading congruency is relatively easy to achieve in normal states or when assessing extremes of disease.[54] Even with this more challenging approach, inter-operator reliability and agreement with the adjudicator were very favorable, especially for elbow and knee assessments, establishing reasonable suitability of the JADE protocol for interinstitutional compilations of findings. However, greater variability was seen in the assessment of ankle abnormalities. The reasons for this variability are unclear but may be due to complicated anatomy involving multiple articulations. Similar observations were made in other studies focusing on imaging and nonimaging assessments of hemophilic arthropathy.[31,54] This factor must be taken into consideration when planning clinical studies, perhaps by centralizing readings of locally acquired images.

Finally, this study tested the reliability of power Doppler signal sampling and reading as an important component in the overall assessment of joint health. The presence of power Doppler signals reveals abnormal synovial or soft tissue microperfusion, which is a hallmark of synovial inflammation in rheumatoid arthritis.[14,15,55] Unique to hemophilic arthropathy, positive power Doppler signals also indicate vascular remodeling associated with leaky vessels and bleeding tendencies.[15,56] Therefore, it is important to capture the extent of power Doppler signal abnormalities during musculoskeletal US examinations of hemophilic joints. There is some concern that variability in equipment settings and motion artifacts can interfere with power Doppler readings, especially when performed by inexperienced operators.[30] In our experience,

rapidly evolving technical advances provide valid preset power Doppler settings across manufacturers, thereby minimizing artifacts.

The JADE protocol incorporates a widely accepted power Doppler scoring algorithm, which was devised and validated to determine the degree of inflammation in rheumatoid arthritis in mainly small joints[57] but subsequently found to be equally useful in large hemophilic joints.[15] In rheumatoid arthritis and osteoarthritis, the extent of the power Doppler signal strength and distribution correlates well with increased microvascularity on histologic examinations,[58,59] and similar correlations have been established recently in hemophilic mouse joints.[60] Here, we showed excellent reliability of power Doppler signal scoring on still images and between operators, providing confidence that it is feasible for nonradiologists to acquire and score power Doppler signals accurately.

There were several limitations to this study. First, full testing of the JADE protocol was limited to a relatively small number of hemophilia providers who also participated in protocol development and who were highly experienced in musculoskeletal US. Another limitation was the lack of joint effusion assessments, since joints were imaged at patients' baselines with a paucity of effusions present. Although even small volume joint effusions are easily detected by musculoskeletal US,[10,39,61] the accuracy of effusion assessments in hemophilia requires formal study. Also, intra-operator reliability was not assessed because we thought that the second assessment would require several weeks to pass to limit memorization of initial findings. During this time frame, sub(acute) events, including hemarthroses, could occur and possibly change previous findings. Last, study images were acquired with a single brand of high-resolution equipment, thereby removing potential confounders related to image quality.

In conclusion, findings from this study demonstrate a high degree of intra-rater, inter-rater, and inter-operator reliability of the JADE protocol, adding another critical step toward validation of JADE as a musculoskeletal US protocol for quantifying arthropathic changes in hemophilic joints. We stress that this validation remains iterative, with refinements to be added as more experience is gained with JADE, a process mirroring OMERACT efforts.[48]

References

1. Luck JV, Silva M, Rodriguez-Merchan EC, Ghalambor N, Zahiri CA, Finn RS. Hemophilic arthropathy. *J Am Acad Orthop Surg* 2004; 12:234–245.
2. Wyseure T, Mosnier LO, von Drygalski A. Advances and challenges in hemophilic arthropathy. *Semin Hematol* 2016; 53:10–19.
3. Acharya SS. Exploration of the pathogenesis of haemophilic joint arthropathy: understanding implications for optimal clinical management. *Br J Haematol* 2012; 156:13–23.
4. Ceponis A, Wong-Sefidan I, Glass CS, von Drygalski A. Rapid musculoskeletal ultrasound for painful episodes in adult haemophilia patients. *Haemophilia* 2013; 19:790–798.
5. 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.
6. Martin EJ, Cooke EJ, Ceponis A, et al. Efficacy and safety of point-of-care ultrasound-guided intra-articular corticosteroid joint injections in patients with haemophilic arthropathy. *Haemophilia* 2017; 23:135–143.
7. Aznar JA, Abad-Franch L, Perez-Alenda S, Haya S, Cid AR, Querol F. Ultrasonography in the monitoring of management of haemarthrosis. *Haemophilia* 2011; 17:826–828.
8. Aznar JA, Perez-Alenda S, Jaca M, et al. Home-delivered ultrasound monitoring for home treatment of haemarthrosis in haemophilia A. *Haemophilia* 2015; 21:e147–e150.
9. Querol F, Cortina V, Cid AR, Haya S, Aznar JA. Clinical and echographical control protocol of haemarthrosis in haemophilia patients with inhibitors: evaluation of the efficacy of recombinant factor VIIa in the evolution process (EFFISEVEN protocol). *Haemophilia* 2008; 14(suppl 6):36–44.
10. Zhou JY, Rappazzo KC, Volland L, et al. Pocket handheld ultrasound for evaluation of the bleeding haemophilic joint: a novel and reliable way to recognize joint effusions. *Haemophilia* 2018; 24:e77–e80.
11. Doria AS. State-of-the-art imaging techniques for the evaluation of haemophilic arthropathy: present and future. *Haemophilia* 2010; 16(suppl 5):107–114.
12. Doria AS, Keshava SN, Mohanta A, et al. Diagnostic accuracy of ultrasound for assessment of hemophilic arthropathy: MRI correlation. *AJR Am J Roentgenol* 2015; 204:W336–W347.
13. Ligocki CC, Abadeh A, Wang KC, Adams-Webber T, Blanchette VS, Doria AS. A systematic review of ultrasound imaging as a tool for evaluating haemophilic arthropathy in children and adults. *Haemophilia* 2017; 23:598–612.
14. Kang T, Lanni S, Nam J, Emery P, Wakefield RJ. The evolution of ultrasound in rheumatology. *Ther Adv Musculoskelet Dis* 2012; 4: 399–411.

15. Bhat V, Olmer M, Joshi S, et al. Vascular remodeling underlies rebleeding in hemophilic arthropathy. *Am J Hematol* 2015; 90: 1027–1035.
16. Melchiorre D, Linari S, Innocenti M, et al. Ultrasound detects joint damage and bleeding in haemophilic arthropathy: a proposal of a score. *Haemophilia* 2011; 17:112–117.
17. Querol F, Rodriguez-Merchan EC. The role of ultrasonography in the diagnosis of the musculo-skeletal problems of haemophilia. *Haemophilia* 2012; 18:e215–e226.
18. Martinoli C, Della Casa Alberighi O, Di Minno G, et al. Development and definition of a simplified scanning procedure and scoring method for haemophilia early arthropathy detection with ultrasound (HEAD-US). *Thromb Haemost* 2013; 109: 1170–1179.
19. Muça-Perja M, Riva S, Grochowska B, Mangiafico L, Mago D, Gringeri A. Ultrasonography of haemophilic arthropathy. *Haemophilia* 2012; 18:364–368.
20. Klukowska A, Czyrny Z, Laguna P, Brzewski M, Serafin-Krol MA, Rokicka-Milewska R. Correlation between clinical, radiological and ultrasonographical image of knee joints in children with haemophilia. *Haemophilia* 2001; 7:286–292.
21. Bruyn GA, Naredo E, Iagnocco A, et al. The OMERACT Ultrasound Working Group 10 years on: update at OMERACT 12. *J Rheumatol* 2015; 42:2172–2176.
22. Terslev L, Gutierrez M, Schmidt WA, et al. Ultrasound as an outcome measure in gout: a validation process by the OMERACT Ultrasound Working Group. *J Rheumatol* 2015; 42:2177–2181.
23. D'Agostino MA, Boers M, Kirwan J, et al. Updating the OMERACT filter: implications for imaging and soluble biomarkers. *J Rheumatol* 2014; 41:1016–1024.
24. Naredo E, Wakefield RJ, Iagnocco A, et al. The OMERACT Ultrasound Task Force: status and perspectives. *J Rheumatol* 2011; 38: 2063–2067.
25. von Drygalski A, Moore RE, Nguyen S, et al. Advanced hemophilic arthropathy: sensitivity of soft tissue discrimination with musculoskeletal ultrasound. *J Ultrasound Med* 2018; 37:1945–1956.
26. Backhaus M, Burmester GR, Gerber T, et al. Guidelines for musculoskeletal ultrasound in rheumatology. *Ann Rheum Dis* 2001; 60: 641–649.
27. Moller I, Janta I, Backhaus M, et al. The 2017 EULAR standardised procedures for ultrasound imaging in rheumatology. *Ann Rheum Dis* 2017; 76:1974–1979.
28. Pettersson H, Ahlberg A, Nilsson IM. A radiologic classification of hemophilic arthropathy. *Clin Orthop Relat Res* 1980; 149: 153–159.
29. Torp-Pedersen ST, Terslev L. Settings and artefacts relevant in colour/power Doppler ultrasound in rheumatology. *Ann Rheum Dis* 2008; 67:143–149.
30. Torp-Pedersen S, Christensen R, Szkudlarek M, et al. Power and color Doppler ultrasound settings for inflammatory flow: impact on scoring of disease activity in patients with rheumatoid arthritis. *Arthritis Rheumatol* 2015; 67:386–395.
31. Hilliard P, Funk S, Zourikian N, et al. Hemophilia joint health score reliability study. *Haemophilia* 2006; 12:518–525.
32. 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.
33. Poonnoose PM, Hilliard P, Doria AS, et al. Correlating clinical and radiological assessment of joints in haemophilia: results of a cross sectional study. *Haemophilia* 2016; 22:925–933.
34. McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. *Int J Clin Pharm* 2016; 38:655–662.
35. Backhaus M, Ohrndorf S, Kellner H, et al. Evaluation of a novel 7-joint ultrasound score in daily rheumatologic practice: a pilot project. *Arthritis Rheum* 2009; 61:1194–1201.
36. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; 86:420–428.
37. Gwet KL. *Handbook of Inter-rater Reliability*. 4th ed. Gaithersburg, MD; Advanced Analytics LLC; 2014.
38. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016; 15:155–163.
39. Nguyen S, Lu X, Ma Y, Du J, Chang EY, von Drygalski A. Musculoskeletal ultrasound for intra-articular bleed detection: a highly sensitive imaging modality compared with conventional magnetic resonance imaging. *J Thromb Haemost* 2018; 16:490–499.
40. Naredo E, Acebes C, Moller I, et al. Ultrasound validity in the measurement of knee cartilage thickness. *Ann Rheum Dis* 2009; 68: 1322–1327.
41. Moller B, Bonel H, Rotzetter M, Villiger PM, Ziswiler HR. Measuring finger joint cartilage by ultrasound as a promising alternative to conventional radiograph imaging. *Arthritis Rheum* 2009; 61: 435–441.
42. Pradsgaard DO, Fiirgaard B, Spannow AH, Heuck C, Herlin T. Cartilage thickness of the knee joint in juvenile idiopathic arthritis: comparative assessment by ultrasonography and magnetic resonance imaging. *J Rheumatol* 2015; 42:534–540.
43. Hiraga M, Ikeda K, Shigeta K, et al. Sonographic measurements of low-echoic synovial area in the dorsal aspect of metatarsophalangeal joints in healthy subjects. *Mod Rheumatol* 2015; 25:386–392.
44. Bevers K, Zweers MC, van den Ende CH, et al. Ultrasonographic analysis in knee osteoarthritis: evaluation of inter-observer reliability. *Clin Exp Rheumatol* 2012; 30:673–678.
45. Rosendahl K, Bruserud IS, Oehme N, et al. Normative ultrasound references for the paediatric wrist; dorsal soft tissues. *RMD Open* 2018; 4:e000642.
46. Couanis G, Breidahl W, Burnham S. The relationship between subacromial bursa thickness on ultrasound and shoulder pain in open water endurance swimmers over time. *J Sci Med Sport* 2015; 18:373–377.

47. Cheung PP, Dougados M, Gossec L. Reliability of ultrasonography to detect synovitis in rheumatoid arthritis: a systematic literature review of 35 studies (1415 patients). *Arthritis Care Res (Hoboken)* 2010; 62:323–334.
48. Hammer HB, Iagnocco A, Mathiessen A, et al. Global ultrasound assessment of structural lesions in osteoarthritis: a reliability study by the OMERACT ultrasonography group on scoring cartilage and osteophytes in finger joints. *Ann Rheum Dis* 2016; 75: 402–407.
49. Kissin EY, Niu J, Balint P, et al. Musculoskeletal ultrasound training and competency assessment program for rheumatology fellows. *J Ultrasound Med* 2013; 32:1735–1743.
50. Irwin RW, Smith J, Issenberg SB. Long-term retention of musculoskeletal ultrasound training during residency. *Am J Phys Med Rehabil* 2018; 97:523–530.
51. Wakefield RJ, Weerasinghe A, Tung P, et al. The development of a pragmatic, clinically driven ultrasound curriculum in a UK medical school. *Med Teach* 2018; 40:600–606.
52. Lawson W, Uy M, Strike K, et al. Point of care ultrasound in haemophilia: building a strong foundation for clinical implementation. *Haemophilia* 2017; 23:648–651.
53. Strike KL, Iorio A, Jackson S, et al. Point of care ultrasonography in haemophilia care: recommendations for training and competency evaluation. *Haemophilia* 2015; 21:828–831.
54. Stephensen D, Classey S, Harbidge H, Patel V, Taylor S, Wells A. Physiotherapist inter-rater reliability of the Haemophilia Early Arthropathy Detection with Ultrasound protocol. *Haemophilia* 2018; 24:471–476.
55. Szkudlarek M, Court-Payen M, Strandberg C, Klarlund M, Klausen T, Ostergaard M. Power Doppler ultrasonography for assessment of synovitis in the metacarpophalangeal joints of patients with rheumatoid arthritis: a comparison with dynamic magnetic resonance imaging. *Arthritis Rheum* 2001; 44:2018–2023.
56. Kidder W, Chang EY, Moran C, Rose SC, von Drygalski A. Persistent vascular remodeling and leakiness are important components of the pathobiology of re-bleeding in hemophilic joints: two informative cases. *Microcirculation* 2016; 23:373–378.
57. Szkudlarek M, Court-Payen M, Jacobsen S, Klarlund M, Thomsen HS, Ostergaard M. Interobserver agreement in ultrasonography of the finger and toe joints in rheumatoid arthritis. *Arthritis Rheum* 2003; 48:955–962.
58. Walther M, Harms H, Krenn V, Radke S, Faehndrich TP, Gohlke F. Correlation of power Doppler sonography with vascularity of the synovial tissue of the knee joint in patients with osteoarthritis and rheumatoid arthritis. *Arthritis Rheum* 2001; 44:331–338.
59. Walther M, Harms H, Krenn V, Radke S, Kirschner S, Gohlke F. Synovial tissue of the hip at power Doppler US: correlation between vascularity and power Doppler US signal. *Radiology* 2002; 225:225–231.
60. Cooke EJ, Zhou JY, Wyseure T, et al. Vascular permeability and remodelling coincide with inflammatory and reparative processes after joint bleeding in factor VIII-deficient mice. *Thromb Haemost* 2018; 118:1036–1047.
61. Zubler V, Mamisch-Saupe N, Pfirrmann CW, Jost B, Zanetti M. Detection and quantification of glenohumeral joint effusion: reliability of ultrasound. *Eur Radiol* 2011; 21:1858–1864.