Development of a feasible and responsive ultrasound inflammation score for rheumatoid arthritis through a data-driven approach

Anna-Birgitte Aga,1 Hilde Berner Hammer,1 Inge Christoffer Olsen,1 Till Uhlig,1 Tore K Kvien,1 Désirée van der Heijde,1,2 Hallvard Fremstad,3 Tor Magne Madland,4 Åse Stavland Lexberg,5 Hilde Haukeland,6 Erik Rødevand,7 Christian Heili,8 Hilde Stray,9 Anne Lindtner Noraas,10 Inger Johanne Widding Hansen,11 Gunnstein Bakland,12,13 Siri Lillegraven,1 Elisabeth Lie,1 Espen A Haavardsholm1

OBJECTIVE

To develop and validate a responsive and feasible ultrasound inflammation score for rheumatoid arthritis (RA).

METHODS

We used data from cohorts of early RA (development) and established RA starting/switching biologic therapy (validation). 4 tendons and 36 joints were examined by a grey scale (GSUS) and power Doppler semiquantitative ultrasound (PDUS) scoring system (full score). Ultrasound score components were selected based on factor analyses of 3-month change in the development cohort. Responsiveness was assessed by standardised response means (SRMs). We assessed the proportion of information retained from the full score by linear regression.

RESULTS

118 patients with early and 212 patients with established RA were included. The final ultrasound score included 8 joints (metacarpophalangeal 1–2–3, proximal interphalangeal 2–3, radiocarpal, metatarsophalangeal 2–3) and 1 tendon (extensor carpi ulnaris) examined bilaterally. The 6-month SRMs for the final score were −1.24 (95% CI −1.47 to −1.02) for GSUS, and −1.09 (−1.25 to −0.92) for PDUS in early RA, with 87% of total information retained for GSUS and 90% for PDUS. The new score performed somewhat better than formerly proposed scores in the validation cohort.

CONCLUSIONS

The Ultrasound in Rheumatoid Arthritis 9 joint/tendon score (USRA9) inflammation score showed good responsiveness, retained most of the information from the original full score and overall performed better than previous scores in a validation cohort.

INTRODUCTION

Ultrasound (US) is increasingly used to detect and monitor inflammation in rheumatoid arthritis (RA), in clinical practice and research. There is no agreement on which joints and tendons should be assessed for a sensitive, responsive and feasible US measure of inflammation in RA.1

The research agenda proposed in recent European League against Rheumatism (EULAR) recommendations includes evaluation of which joints should be assessed by US.2 Several reduced joint scores for US joint inflammation have been developed, but, to the best of our knowledge, their development has to a limited degree been based on data-driven approaches.3–5

Our research group recently developed two candidate sets of reduced US joint inflammation scores in early RA through a data-driven approach in cross-sectional data.
aiming to retain as much information as possible from the full score. The candidate scores performed well in a validation cohort of established RA, and performed better than previously proposed reduced joint scores. Although these results showed that a reduced US assessment might efficiently contribute to assessment in RA, an US score has to be responsive in order to be an important tool to monitor disease activity.

In the present study, longitudinal US data from the ARCTIC trial, which became available after publication of our previous paper, were used. The objective of the current study was to develop a feasible US inflammation scoring focus on sensitivity to change, and to validate the score by comparison with formerly proposed US scores in an independent longitudinal cohort.

METHODS

Patient population

Patients (>18 years) with RA were included in one of two cohorts between January 2010 and June 2013: early disease-modifying antirheumatic drug (DMARD)-naïve patients with RA fulfilling the 2010 American College of Rheumatology (ACR)/EULAR classification criteria with indication for methotrexate treatment (the ARCTIC trial, ClinicalTrials.gov identifier NCT01205854), and patients with established RA with active disease and indication for methotrexate and 212 patients with previous disease-modifying antirheumatic drug (DMARD)-naïve indication for methotrexate treatment (the ULRABIT study, Anzctr.org.au identifier ACTRN12610000284066).

Ultrasonography assessments

An extensive US examination was performed at baseline, 3 and 6 months by experienced sonographers using a 0–3 semiquantitative scoring system for grey-scale (GSUS) and power Doppler US (PDUS) in each of the following 36 joints and 4 tendons: metacarpophalangeal (MCP) 1–5, proximal interphalangeal (PIP) 2–3, radiocarpal, intercarpal, radioulnar, elbow, knee, talocrural, metatarsophalangeal (MTP) 1–5, extensor carpi ulnaris (ECU) tendon and tibialis posterior (TP) tendon bilaterally. The range of the sum scores was 0–120 for GSUS and PDUS, and the scanning protocol was a slight modification of a previously published 32-joint protocol, with the addition of bilateral PIP2–3, ECU and TP tendons. The scoring was performed with defined probe placement and patient positioning and an US atlas as a reference.

US workshops with static and dynamic hands-on exercises to calibrate readers were arranged yearly.

We used Siemens Antares Sonoline or GE Logiq E9 machines with linear probes (11.4/13.0 MHz) and power Doppler parameters were adjusted according to the device used (pulse repetition frequency 391/600 Hz; Doppler frequency 7.3/10.0 MHz) for US assessments. There were no changes in US settings during the study, and no upgrading of software.

Finally, we performed a substudy assessing the feasibility of the final Ultrasound in Rheumatoid Arthritis 9 joint/tendon score (USRA9) in clinical practice. Five sonographers with experience in structured US assessment were chosen, and recorded the time for bilateral scanning and recording of the USRA9 score in 35 consecutive patients with polyarthritis.

Laboratory and clinical examinations

Assessments in both cohorts included erythrocyte sedimentation rate (mm/hour), C reactive protein (mg/L), anticyclic citrullinated peptide (IU/mL) and visual analogue scales for investigator’s and patients global assessments of disease activity (0–100 mm). In the early RA cohort, 44 swollen joint count and Ritchie articular index were performed, while 28 swollen and tender joint counts were performed in the established RA cohort.

Statistical analyses

We performed factor analyses of GSUS and PDUS changes from baseline to 3 months (ΔGSUS and ΔPDUS) in the early RA cohort to identify joint/tendon groups with high internal correlation. The factor analyses were based on all joints/tendons examined in the US protocol. We used the principal component methodology, with the factors rotated according to the varimax criteria. The number of factors was based on eigenvalues with a cut-off of 1. Based on these analyses, we identified groups with correlating scores to guide the selection of the final US joint set. In groups with more than one joint identified by factor analysis, the selection process was based on factor loading and previous work to identify joints that retained most information from the extensive score. The final US score was defined as the sum of the joint/tendon scores in the final joint set.

Second, standardised response means (SRMs; mean change divided by the SD of the change) with 95% CIs (by bootstrapping) were calculated at 3 and 6 months for the final score and formerly proposed scores in the early RA cohort. We also calculated 3 and 6 months SRMs for the final score and previously proposed scores in the validation cohort.

Finally, we assessed the proportion of total information retained by the final score at baseline (both GSUS and PDUS) using linear regression analyses with the total US score (0–120) as the dependent variable and each individual joint/tendon score in the final joint set as independent variables. The proportion of information retained was assessed using $R^2$.

Statistical analyses were performed using SPSS (V21 SPSS, Chicago, Illinois, USA), except for factor analyses and bootstrapping, which were calculated in the SAS software package (V9.2 SAS Institute, Cary, North Carolina, USA).

RESULTS

Patient characteristics

In total, 118 DMARD-naïve patients with early RA with indication for methotrexate and 212 patients with
established RA with indication for starting or switching biological DMARD were included. Baseline demographic and patient characteristics in the two cohorts are presented in Table 1. The 18 patients in the established RA cohort who were in the Disease Activity Score-28 (DAS28) remission had active arthritis in the feet/ankles (n=15) or received intra-articular steroids prior to inclusion (n=3).

**Factor analyses**

Based on the factor analyses for ΔGSUS and ΔPDUS, we identified 12 joint groups, representing joints/tendons with correlating scores (see online supplementary table S1). Three groups (elbow, TP tendon and MTP1) were characterised by small changes in GSUS and PDUS over 3 months and were excluded from the final candidate score. In groups with multiple joints, the radiocarpal joint, MCP3 and MTP3 were selected based on the highest factor loading. Other groups were represented by one joint or tendon which all was included in the model (MCP1, MCP2, PIP2, PIP3, MTP2 and ECU tendon). The final US inflammation score, named the USRA9 score, is presented in Figure 1.

**Sensitivity to change of USRA9 compared with previously proposed US scores**

The USRA9 score displayed the same or better sensitivity to change than our previous candidate sets and other proposed reduced US scores for GSUS and PDUS at 3 and 6 months in the early RA cohort (Figure 2A, B and online supplementary table S2A, B).

**Validation**

In the validation cohort, the USRA9 score had the same level or somewhat larger SRMs compared with our previous candidate sets and other US scores (Figure 2C, D, and online supplementary table S2C,D). The USRA9 score was defined as:

\[
\text{USRA9} = \sum (a \times \text{GSUS} + b \times \text{PDUS})
\]

where a and b are regression coefficients from the factor analyses. The total score was calculated for each joint/tendon and ranged from 0 to 3. The final score was the sum of all joint/tendon scores.

---

**Table 1** Baseline demographics and patient characteristics

<table>
<thead>
<tr>
<th></th>
<th>Early RA (n=118)</th>
<th>Established RA (n=212)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years (mean (SD))</td>
<td>50.6 (13.3)</td>
<td>51.9 (13.3)</td>
</tr>
<tr>
<td>Female sex (% (n))</td>
<td>71.2 (84)</td>
<td>80.7 (171)</td>
</tr>
<tr>
<td>Positive anti-CCP (% (n))</td>
<td>78.8 (93)</td>
<td>77.4 (164)</td>
</tr>
<tr>
<td>Disease duration, years (median (25–75 centile))</td>
<td>0.5 (0.2–0.8)†</td>
<td>8.1 (2.9–14.7)†</td>
</tr>
<tr>
<td>Disease activity level‡ (% (n))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remission</td>
<td>2.5 (3)</td>
<td>8.7 (18)</td>
</tr>
<tr>
<td>Low disease activity</td>
<td>12.7 (15)</td>
<td>7.3 (15)</td>
</tr>
<tr>
<td>Moderate disease activity</td>
<td>45.8 (54)</td>
<td>45.6 (94)</td>
</tr>
<tr>
<td>High disease activity</td>
<td>39.0 (46)</td>
<td>38.3 (79)</td>
</tr>
<tr>
<td>28-swollen joint count (median (25–75 centile))</td>
<td>6 (3–11)</td>
<td>5 (2–10)</td>
</tr>
<tr>
<td>ESR, mm/hour (median (25–75 centile))</td>
<td>18 (10–32)</td>
<td>25 (13–40)</td>
</tr>
<tr>
<td>CRP, mg/L (median (25–75 centile))</td>
<td>6 (3–15)</td>
<td>6 (2–15)</td>
</tr>
<tr>
<td>Investigator's global assessment VAS (mean (SD))</td>
<td>40.9 (20.1)</td>
<td>31.4 (16.4)</td>
</tr>
<tr>
<td>Patient's global assessment VAS (mean (SD))</td>
<td>51.9 (24.7)</td>
<td>50.3 (27.1)</td>
</tr>
<tr>
<td>Number of previously used synthetic DMARDs (mean (SD))</td>
<td>0</td>
<td>2.1 (1.3)</td>
</tr>
<tr>
<td>Number of previously used biologic DMARDs (mean (SD))</td>
<td>0</td>
<td>0.8 (1.1)</td>
</tr>
</tbody>
</table>

*From patient reported first swollen joint.
†From clinical diagnosis by rheumatologist.
‡Based on DAS level in the early RA and DAS28 level in the established RA cohort.

CCP, cyclic citrullinated peptide; CRP, C reactive protein; DAS28, Disease Activity Score-28; DMARD, disease-modifying antirheumatic drug; ESR, erythrocyte sedimentation rate; RA, rheumatoid arthritis; VAS, visual analogue scale.
score retained 87% of GSUS and 90% of PDUS information from the full score (36 joints and 4 tendons) before DMARD initiation in the early RA cohort. The corresponding numbers in the established RA cohort were 90% for GSUS and 92% for PDUS (table 2). The median (range) time to complete the USRA9 score bilaterally was 3 min 14 s (1 min 50 s to 5 min 53 s).

**DISCUSSION**

In this study, we present a feasible and responsive US inflammation score for RA, the USRA9 score, which includes examination of eight joints and one tendon. The USRA9 score was selected based on factor analysis of 3-month change in US scores in a recent longitudinal cohort, as well as previous work to identify joints that contribute the best information on inflammation in US examination. Most of the information from an extensive 36 joints and 4 tendons score is retained, and the score is more feasible than our previously proposed candidate scores.6

<table>
<thead>
<tr>
<th>Modality</th>
<th>Side</th>
<th>Early RA (n=118)</th>
<th>Established RA (n=212)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSUS</td>
<td>Unilateral‡</td>
<td>0.70</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>Bilateral</td>
<td>0.87</td>
<td>0.90</td>
</tr>
<tr>
<td>PDUS</td>
<td>Unilateral‡</td>
<td>0.75</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>Bilateral</td>
<td>0.90</td>
<td>0.92</td>
</tr>
</tbody>
</table>

*Final US inflammation score (USRA9 score): MCP1, MCP2, MCP3, PIP2, PIP3, radiocarpal joint, extensor carpi ulnaris tendon, MTP2, MTP3.
†Linear regression analysis with the total US score at baseline as dependent variable and each individual joint/tendon score of the final joint set as independent variables.
‡Unilateral is dominant side.
GSUS, grey-scale ultrasound; MCP, metacarpophalangeal; MTP, metatarsophalangeal; PDUS, power Doppler ultrasound; PIP, proximal interphalangeal; RA, rheumatoid arthritis; US, ultrasound; USRA9, Ultrasound in Rheumatoid Arthritis 9 joint/tendon score.

---

**Figure 2** SRMs at 3 and 6 months for the final ultrasound inflammation score (USRA9) bilaterally and previously proposed scores (A) GSUS in the early RA cohort, (B) PDUS in the early RA cohort, (C) GSUS in the established RA cohort and (D) PDUS in the established RA cohort. *SRMs (standardised response means); mean change divided by the SD of the change with 95% CIs. †Six-joint score Perricone et al.5 wrist, MCP2, knee bilaterally. ‡Seven-joint score Backhaus et al.4 wrist, MCP2, MCP3, PIP2, PIP3, MTP2, MTP5 and extensor carpi ulnaris tendon dominant side. §Twelve-joint score Naredo et al.3 elbow, wrist, MCP2, MCP3, knee, ankle and tibialis posterior tendon bilaterally. Final ultrasound inflammation score (USRA9): radiocarpal joint, extensor carpi ulnaris tendon, MCP1, MCP2, MCP3, PIP2, PIP3, MTP2, MTP3. Dotted line indicating SRM value of most responsive score. GSUS, grey-scale ultrasound; MCP, metacarpophalangeal; MTP, metatarsophalangeal; PDUS, power Doppler ultrasound; PIP, proximal interphalangeal; RA, rheumatoid arthritis; USRA9, Ultrasound in Rheumatoid Arthritis 9 joint/tendon score.
Somewhat surprisingly, there were no substantial differences in the SRMs between unilateral and bilateral US examinations (see online supplementary table S1A, B), and it is less time consuming to examine joints/tendons only on one side of the body. The total information retained was however higher if both sides were examined (table 2), which is also supported by previous studies indicating that RA is not necessarily a symmetrical disease. Unilateral examination of dominant side may be sufficient in some settings, but our findings indicate that a bilateral US joint/tendon examination should be preferred.

Limitations of all synovial Doppler US grading systems include the varying sensitivity of scanners and variable US machine settings. In the present study, both the machines and software were standardised between the different centres, with no upgrade of software during the follow-up. However, recent developments in the hardware and software of US scanners may allow even higher sensitivity to change, and computerised determination of the colour fraction may be more objective than subjective estimation. Only dorsal examinations of the joints were performed, thus we could not assess the implications of applying different projections. The study is strengthened by the validated semiquantitative scoring system for GSUS and PDUS with high intra-rater reliability, trained and calibrated sonographers, defined probe placements, availability of an US atlas as reference, standardised equipment and an extensive approach with two large independent cohorts of RA where a substantial number of joints were examined. In a sub-study experienced sonographers familiar with the inflammation score spent a median of just over 3 min to perform bilateral scanning and recording of the USRA9 score. Our final US inflammation score includes many of the same joints as in a previous study based on clinical judgement and observed frequency of involvement in mixed cohorts of RA. Feasibility of the final score is improved compared with our previous candidate sets by reducing the number of joints.

In conclusion, we have developed a feasible and responsive reduced US inflammation score for RA, retaining most of the information from an extensive score, and which hopefully can contribute to a consensus US inflammation score for RA. The final score can be useful in clinical trials to assess US inflammation in RA. The objective measure of joint inflammation provided by US gives unique opportunities to answer research questions related to disease activity and inflammation in RA.

Author affiliations
1Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway
2Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands
3Department of Rheumatology, Ålesund Hospital, Helse Møre och Romsdal HF, Ålesund, Norway
4Department of Rheumatology, Haukeland University Hospital, Helse Bergen HF, Bergen, Norway
5Department of Rheumatology, Drammen Hospital, Vestre Viken HF, Drammen, Norway
6Department of Rheumatology, Martina Hansens Hospital AS, Baerum, Norway
7Department of Rheumatology, St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway
8Department of Rheumatology, Hospital Østfold HF, Moss, Norway
9Haugesund Rheumatism Hospital AS, Haugesund, Norway
10The Rheumatology Clinic Dovland/Benvold, Kristiansand, Norway
11Department of Rheumatology, Sarlandet Hospital HF, Kristiansand, Norway
12Department of Rheumatology, University Hospital of North Norway, Tromsø, Norway
13Institute of Clinical Medicine, University of Tromsø, Norway

Acknowledgements The authors thank the patients for participating in this study and the local rheumatology staff for data collection.

Contributors A-BA and EAH designed the study, recruited and enrolled participants, collected, analysed, and interpreted data, and wrote the report. ICO designed the study, analysed and interpreted data, and reviewed the manuscript. SL analysed, interpreted data and wrote the report. HBI designed the study, recruited participants, interpreted data and reviewed the manuscript. TD designed the study, recruited and enrolled participants, interpreted data, and reviewed the manuscript. HF, TMM, ASL, HH, ER, CH, HS, ALN, UWH and GB recruited and enrolled patients, interpreted data, and reviewed the manuscript. DvdH and TKK designed the study, interpreted data and reviewed the manuscript. All authors have approved the final draft and vouch for the accuracy and completeness of the data and analyses.

Funding This trial was supported by the Norwegian Research Council, Norwegian South-Eastern Health Region, Norwegian Women’s Public Health Association, Norwegian Rheumatism Association, investigator-initiated research grants from AbbVie, UCB Pharma, Pfizer, MSD Norway, Roche Norway. Siemens Healthcare and GE Healthcare provided technical support regarding standardisation of ultrasound equipment. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Competing interests All authors have completed the Unified Competing Interest form at http://www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: EAH has received research funding from Pfizer, UCB, Roche, MSD and AbbVie for the submitted work, honoraria as a speaker from Pfizer, UCB, Roche, and AbbVie, honoraria for development of educational material from Pfizer, and has been on advisory board for Pfizer that might have an interest on the submitted work in the previous 3 years. A-BA has been on advisory boards for UCB, AbbVie and Pfizer, and received honoraria for development of educational material for UCB. HBI has received honoraria as a speaker from AbbVie, Bristol-Myers Squibb, Roche, UCB Pharma and Pfizer. HH has been on advisory boards for UCB and AbbVie. GB has received honoraria as a speaker from AbbVie and has been on advisory board for Pfizer. DvdH has received consultancy honoraria from AbbVie, Amgen, Astellas, AstraZeneca, Bristol-Myers Squibb, Celgene, Daiichi, Eli Lilly, Galapagos, Merck, Novartis, Pfizer, Roche, Sanofi Aventis, Janssen and UCB, and is owner of Imaging Rheumatology. TKK has received consultancy honoraria from AbbVie, Bristol-Myers Squibb, Celtrion, Epirus, Hospira, Merck-Serono, MSD, Orion Pharma and Pfizer that might have an interest on the submitted work in the previous 3 years.

Patient consent Obtained.

Ethics approval The study was approved by an Independent Ethics Committee (REC South-East; reference number 2010/744) and appropriate Institutional Review Boards.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/
REFERENCES