

# Patients with Rheumatoid Arthritis Acquire Sustainable Skills for Home Monitoring: A Prospective Dual-country Cohort Study (ELECTOR Clinical Trial I)

Marie Skougaard, Henning Bliddal, Robin Christensen, Karen Ellegaard, Sabrina M. Nielsen, Jakub Zavada, Sabina Oreska, Niels S. Krogh, Christian C. Holm, Merete L. Hetland, Jiri Vencovsky, Henrik Røgind, Peter C. Taylor, and Henrik Gudbergson

**ABSTRACT. Objective.** In an eHealth setting, to investigate intra- and interrater reliability and agreement of joint assessments and Disease Activity Score using C-reactive protein (DAS28-CRP) in patients with rheumatoid arthritis (RA) and test the effect of repeated joint assessment training.

**Methods.** Patients with DAS28-CRP  $\leq 5.1$  were included in a prospective cohort study (clinicaltrials.gov: NCT02317939). Intrarater reliability and agreement of patient-performed joint counts were assessed through completion of 5 joint assessments over a 2-month period. All patients received training on joint assessment at baseline; only half of the patients received repeated training. A subset of patients was included in an appraisal of interrater reliability and agreement comparing joint assessments completed by patients, healthcare professionals (HCP), and ultrasonography. Cohen's  $\kappa$  coefficients and intraclass correlation coefficients (ICC) were used for quantifying of reliability of joint assessments and DAS28-CRP. Agreement was assessed using Bland-Altman plots.

**Results.** Intrarater reliability was excellent with ICC of 0.87 (95% CI 0.83–0.90) and minimal detectable change of 1.13. ICC for interrater reliability ranged between 0.69 and 0.90 (good to excellent). Patients tended to rate DAS28-CRP slightly higher than HCP. In patients receiving repeated training, a mean difference in DAS28-CRP of  $-0.08$  was observed (limits of agreements of  $-1.06$  and  $0.90$ ). After 2 months, reliability between patients and HCP was similar between groups receiving single or repeated training.

**Conclusion.** Patient-performed assessments of joints and DAS28-CRP in an eHealth home-monitoring solution were reliable and comparable with HCP. Patients can acquire the necessary skills to conduct a correct joint assessment after initial and thorough training. [clinicaltrials.gov (NCT02317939)] (First Release February 1 2020; J Rheumatol 2020;47:658–67; doi:10.3899/jrheum.181362)

## Key Indexing Terms:

RHEUMATOID ARTHRITIS  
HOME MONITORING

DISEASE ACTIVITY SCORE

SELF-MANAGEMENT  
RELIABILITY

From The Parker Institute, Copenhagen University Hospital Bispebjerg and Frederiksberg, Copenhagen; Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Odense; Zitellab ApS, Copenhagen, Denmark; Copenhagen Center for Arthritis Research (COPECARE), Center for Rheumatology and Spine Diseases, Centre of Head and Orthopaedics, Rigshospitalet, Glostrup; Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; Institute of Rheumatology, Prague; Department of Rheumatology, First Faculty of Medicine, Charles University, Prague, Czech Republic; Botnar Research Centre, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK.

This study was funded by The European Commission's Research and Innovation Framework Programme as a part of the ELECTOR project (E-Health in Rheumatology, a HORIZON 2020 project). The Parker Institute, Bispebjerg and Frederiksberg Hospital is supported by a core grant from the Oak Foundation (OCAY-13-309).

N.S. Krogh is CEO of Zitellab ApS and developer of the ELECTOR online platform.

M. Skougaard, MD, The Parker Institute, Copenhagen University Hospital Bispebjerg and Frederiksberg; H. Bliddal, MD, DMSc, Professor, The

Parker Institute, Copenhagen University Hospital Bispebjerg and Frederiksberg; R. Christensen, BSc, MSc, PhD, Professor, The Parker Institute, Copenhagen University Hospital Bispebjerg and Frederiksberg, and Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital; K. Ellegaard, PT, PhD, The Parker Institute, Copenhagen University Hospital Bispebjerg and Frederiksberg; S.M. Nielsen, BSc, MSc, The Parker Institute, Copenhagen University Hospital Bispebjerg and Frederiksberg, and Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital; J. Zavada, MD, PhD, Institute of Rheumatology, and Department of Rheumatology, First Faculty of Medicine, Charles University; S. Oreska, MD, Institute of Rheumatology, and Department of Rheumatology, First Faculty of Medicine, Charles University; N.S. Krogh, CEO, Zitellab ApS; C.C. Holm, MSc, The Parker Institute, Copenhagen University Hospital Bispebjerg and Frederiksberg; M.L. Hetland, MD, DMSc, Professor, COPECARE, Center for Rheumatology and Spine Diseases, Centre of Head and Orthopaedics, Rigshospitalet, and Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen; J. Vencovsky, Professor, Institute of Rheumatology, and Department of Rheumatology, First Faculty of Medicine, Charles University; H. Røgind, MD, PhD, COPECARE, Center for Rheumatology and Spine Diseases, Centre of Head and Orthopaedics,

Telemonitoring and eHealth solutions for assessing patients with chronic illnesses such as diabetes, asthma, and hypertension have previously shown great advantages in better control of illnesses and improvement of symptoms<sup>1</sup>. A similar eHealth solution for patients with rheumatoid arthritis (RA) may be advantageous both for patients and healthcare systems.

The eHealth in Rheumatology (ELECTOR, [www.elector.eu](http://www.elector.eu)) project is part of the Horizon 2020 Programme, set up to develop and implement an eHealth platform for home-based monitoring of patients with RA. For monitoring of RA, the European League Against Rheumatism (EULAR) recommends tight control of disease activity to ensure optimal treatment<sup>2,3</sup>. Currently, tight control is managed in outpatient clinics by healthcare professionals (HCP), and is expensive and time-consuming for both parties<sup>4,5,6</sup>. The possibility for self-management of stable patients as part of an eHealth solution would leave more time in the clinics for patients in specific need of care and provide greater involvement of patients. Previously, therapeutic adjustments based on monitoring of disease activity by patient-reported outcomes for patients in stable remission or with low disease activity has maintained disease control similar to routine care<sup>7</sup>.

To assess disease activity by 28-joint count Disease Activity Score using C-reactive protein (DAS28-CRP), one needs data concerning CRP, the visual analog scale (VAS) on patient assessment of global health, and swollen and tender joint counts. All items of this score may be made available at the patient's home in a future home-monitoring solution.

Previous studies of the subject have shown poor to moderate reliability and agreement as well as intra- and inter-observer variations in joint assessments, especially of joint swelling<sup>4,6,8–16</sup>. Consequently, before home monitoring can become a reality there is a need to investigate whether patient-performed joint counts reported through a home-monitoring platform are reliable and whether the agreement between patient and HCP joint counts is sufficient. Joint assessments by use of ultrasonography (US) in RA have been shown to be more sensitive than clinical examination in evaluating both swelling and tenderness<sup>17</sup>; consequently, US was included to explore the validity of patient and HCP joint assessments.

The aim of our study was to examine the intra- and inter-rater reliability and agreement of joint assessments and the corresponding DAS28-CRP, and to assess the effect of repeated training in joint assessment.

---

Rigshospitalet; P.C. Taylor, Professor, Botnar Research Centre, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford; H. Gudbergson, MD, PhD, The Parker Institute, Copenhagen University Hospital Bispebjerg and Frederiksberg.

Address correspondence to Dr. H. Gudbergson, The Parker Institute, Copenhagen University Hospital Bispebjerg and Frederiksberg, Nordre Fasanvej 57, Road 8, Entrance 19, 2000 Frederiksberg, Denmark.  
E-mail: [rindelgudbergson@gmail.com](mailto:rindelgudbergson@gmail.com)

Accepted for publication August 2, 2019.

## MATERIALS AND METHODS

**Design.** The ELECTOR clinical trial I was initially designed as a randomized controlled trial with a 2:2:1:1 allocation ratio. The randomization algorithm appeared impossible to implement because of organizational challenges. It was decided to continuously include patients for a prospective cohort study with consecutive recruitment of participants in the investigation of intrarater reliability and agreement (group 1), or the investigation of intra- and inter-rater reliability and agreement (group 2). The trial was registered at [clinicaltrials.gov](http://clinicaltrials.gov) (NCT02317939). All patients received joint assessment training at baseline, whereas only half of the patients were given repeated training (group A); the other half of patients did not receive repeated training (group B). This resulted in a total of 4 possible groups: 1A, 1B, 2A, and 2B (Figure 1). When allocated to one of these 4 groups, the platform kept track of the study protocol, making sure that the content related to the recurrent training at followup was only shown to patients allocated to this group.

**Setting.** During the period from December 2014 to April 2017, patients routinely visiting the outpatient clinics of rheumatology in Frederiksberg, Denmark (DK), and in Prague, Czech Republic (CZ), were screened for eligibility and invited to participate. As part of the ELECTOR project, an online Web-based platform<sup>18</sup> was developed for filling in questionnaires and joint assessments. The study was performed according to the Helsinki Declaration, and with approval from the Ethics Committee of the Capital Region (J.no.: H-3-2014-108), DK, and the institutional Ethics Committee (J.no.: 9220/2015), CZ. Further, approval was obtained from the Danish Data Protection Agency.

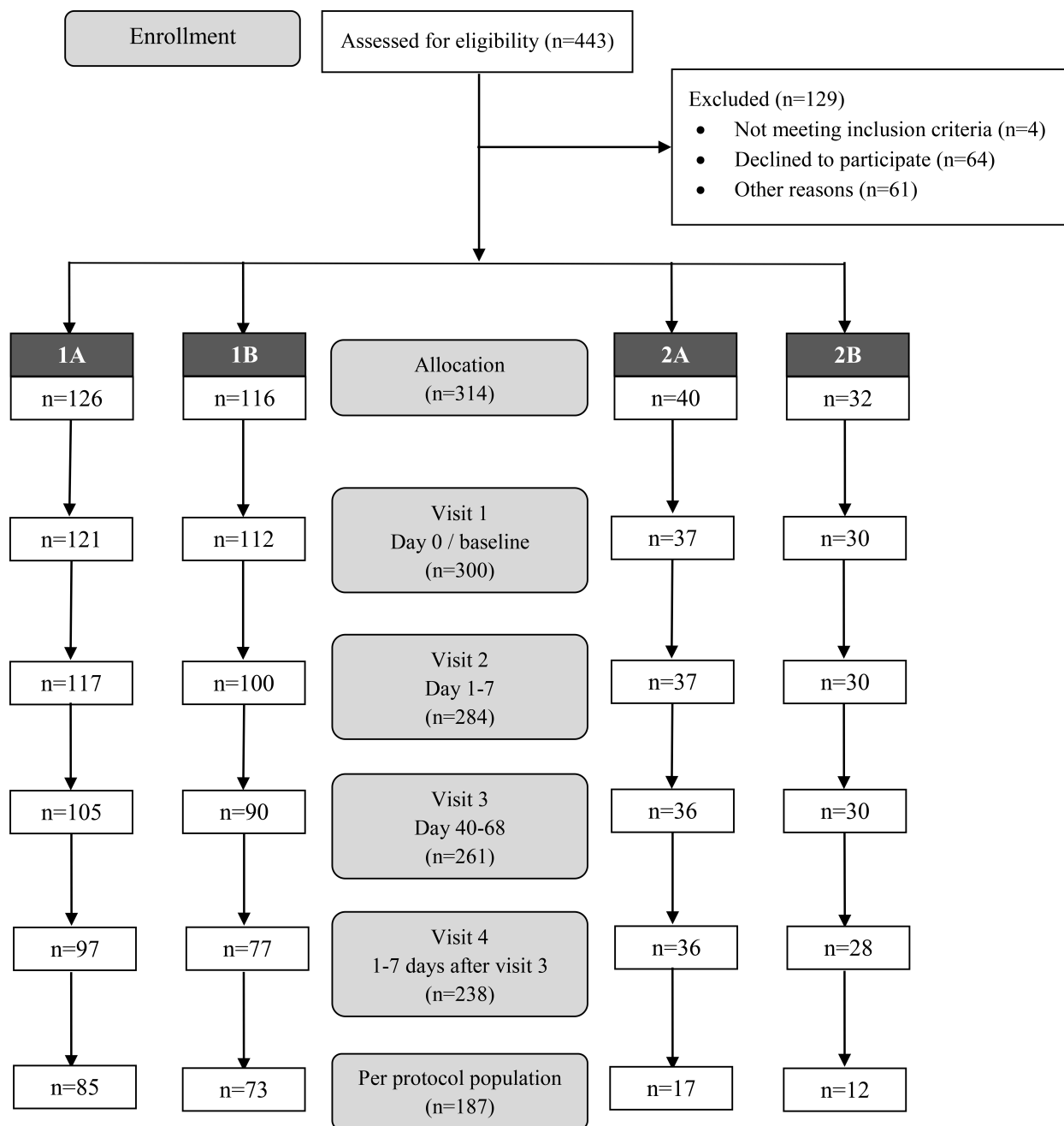
**Patients.** Patients were considered eligible for inclusion if they had been diagnosed with RA for  $\geq 12$  months, were between 18 and 85 years of age, and had a DAS28-CRP  $\leq 5.1$  assessed by an HCP at a screening visit. Exclusion criteria included vision impairment that prevented use of computers. All patients included in the study provided written informed consent.

**Patient instructions.** At baseline, all patients received video instruction and individual face-to-face guidance on how to fill in results on the online platform. Baseline training on how to perform joint assessments was delivered through a Web-based service consisting of 3 instruction videos with focus on (1) general information about joint assessment; (2) assessment of wrist and finger joints; and (3) assessment of elbow, shoulder, and knee joints, followed by a short training session delivered by the HCP. The followup training session, given to half of the patients, included only the videos.

**Clinical investigations.** Assessment of 28 tender and/or swollen joints [proximal interphalangeal joints (PIP) 1–5, metacarpophalangeal joints (MCP) 1–5, wrist, elbow, shoulder, and knee on both sides] was performed by patients 5 times on 4 different days over a period of 2 months (visit schedule is in Supplementary Table 1, available with the online version of this article). For investigation of interrater reliability and agreement, 2 joint assessments were completed within a week at 2 occasions with an interval of 2 months, resulting in a total of 4 joint assessments. Joint assessments were conducted by an experienced rheumatologist (HCP1) and a medical student trained in joint assessments (HCP2). Timing of HCP assessments and CRP measurements were aligned with the patient-performed joint assessments. Completion of the Health Assessment Questionnaire and VAS was done at all visits (Supplementary Table 1).

**Examination by US.** An US specialist performed an examination of 12 joints (MCP 2–5, wrist, and elbow on both sides) that contained an assessment of synovial hypertrophy (swollen joint) and Doppler activity (tender joint). Synovial hypertrophy was assessed on greyscale images. Doppler mode was used for the visualization of increased blood flow due to inflammation. Only 12 joints were included in the US examination, comprising wrists, elbows, and MCP joints, all of which are considered the most frequently involved joints in RA. MCP-1 was excluded because this joint may also be affected by osteoarthritis.

**Statistical methods.** Prior to conducting the analyses, a statistical analysis



**Figure 1.** Patient flow chart demonstrating the total number of patients filling in questionnaires through the online platform at visits 1 to 4. The per-protocol population is defined as patients with complete data for the calculation of DAS28-CRP at all visits. 1A: patients receiving joint count training at baseline and followup. Patients' joints were assessed only by patients themselves. 1B: patients receiving joint count training only at baseline. Patients' joints were assessed only by patients themselves. 2A: patients receiving joint count training at baseline and followup. Patients' joints were assessed by patients themselves, HCP, and US. 2B: patients receiving joint count training only at baseline. Patients' joints were assessed by patients themselves, HCP, and US. Group 1A and 1B included patients from the Czech Republic and Denmark. Group 2A and 2B included patients from Denmark. DAS28-CRP: 28-joint count Disease Activity Score using C-reactive protein; HCP: healthcare professional; US: ultrasound.

plan was published online ([www.parkerinst.dk](http://www.parkerinst.dk)). Statistical analysis was conducted on the per-protocol (PP) population for assessment of intrarater reliability and agreement. The PP population was defined as patients with complete data to calculate DAS28-CRP at all visits within the predefined time slots, with a short time followup within 1–3 days (situation 1, Supplementary Table 1, available with the online version of this article). The same analysis was carried out for the PP population with all data to calculate

DAS28-CRP at an extended time slot, with a short time followup within 1–7 days (situation 2, Supplementary Table 1). Statistical analysis for interrater assessment was conducted on the complete case (CC) population for interrater assessments because this analysis was independent of time compared to the intrarater analysis. CC was defined as patients having complete data to calculate DAS28-CRP for all 4 raters.

*Comparison of DAS28-CRP ratings.* Intraclass correlation coefficients (ICC)

were calculated for quantification of intra- and interrater reliability for DAS28-CRP. ICC were calculated for quantification of intrarater reliability (i.e., patient scores from visit 1 were compared with visit 2, and visit 3 was compared with visit 4) and interrater reliability (i.e., patient scores were compared to HCP scores at visit 1 and visit 3). Values <0.40 were interpreted as poor reliability, 0.40–0.59 as fair reliability, 0.60–0.74 as good reliability, and 0.75–1.00 as excellent reliability<sup>19</sup>.

**Comparison of swollen/tender joint ratings.** Cohen's  $\kappa$  coefficients were used for quantification of reliability for classification of tender and swollen joints on single-joint level. A  $\kappa$  coefficient <0 was considered poor, 0.0–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial, and 0.81–1.00 almost perfect level of reliability<sup>20</sup>. Intra- and interrater agreement were assessed using Bland-Altman plots and limits of agreement were calculated. A difference in DAS28-CRP > 0.6 was considered minimal clinically important difference (MCID) because that is the cutoff in EULAR's response criteria<sup>21,22</sup> where changes in medical management might be considered<sup>23</sup>. Differences in DAS28-CRP > 0.6 were considered as proxies for when additional contact with the patient might be considered during home monitoring.

**US comparison.** Observed agreement was reported (i.e., number of patients who had identical ratings between visits). Interrater agreement was reported as absolute agreement. Because not all 28 joints were included in every US examination, the correlation between US and joint count was made for each single joint.

**Sensitivity analyses.** To account for missing data, a series of sensitivity analyses were carried out to compare the predefined PP and CC populations, and moreover for an intention-to-treat population in which missing data on DAS28-CRP were imputed using grand mean and an as-observed population defined by patients who had any data on the outcome of interest.

Calculations were carried out using the statistical software R (version 3.3.3; R Foundation for Statistical Computing) with the package "psych"<sup>24</sup>.

## RESULTS

**Patient flow.** Of 443 patients screened for eligibility, a total of 314 patients were included in the trial. For the primary analysis, data from 187 patients were included for assessment of intrarater reliability and agreement (PP population, Figure 1). For analysis of interrater reliability and agreement, data were included from 60 patients (CC population).

**Demographics.** At inclusion, the PP population had a mean age of 55.1 (SD 13.5) years, 81% were women, mean disease duration was 11.9 (SD 8.7) years, and 86%, 24%, and 22% received conventional synthetic disease-modifying antirheumatic drugs (csDMARD), biologic DMARD (bDMARD), and prednisolone, respectively (Table 1). They had a mean DAS28-CRP (assessed by an HCP2) of 3.0 (SD 1.0), and median VAS patient global, VAS pain, and VAS fatigue of 21.5, 18, and 28, respectively. When comparing patients from DK and CZ (Supplementary Table 2, available with the online version of this article), the patients from DK were generally older, included fewer females, fewer were treated with csDMARD and prednisolone, and more were in treatment with bDMARD. There was no statistically significant difference between the patients regarding DAS28-CRP, VAS global, or VAS pain, but patients from DK had a statistically significant higher VAS fatigue score.

**Comparing repeated patient-performed joint assessments by DAS28-CRP (intrarater reliability and agreement).** Patients'

Table 1. Patient characteristics.

Characteristics	Patients PP, n = 187
Age, yrs	55.1 (13.5) <sup>†</sup>
Female, n (%)	151 (81)
Disease duration, yrs, mean (SD)	11.9 (8.7) <sup>‡</sup>
csDMARD use, n (%)	154 (86) <sup>§</sup>
bDMARD use, n (%)	43 (24) <sup>††</sup>
Prednisolone use, n (%)	39 (22) <sup>††</sup>
DAS28-CRP*	3.0 (1.0) <sup>  </sup>
VAS global (0–100), median (IQR)	21.5 (7.3–44.8)
VAS pain (0–100), median (IQR)	18 (7.0–38)
VAS fatigue (0–100), median (IQR)	28 (11.0–51.0)

Values are mean  $\pm$  SD unless otherwise specified. \* Baseline data for DAS28-CRP are those measured by HCP2 at visit 1. <sup>†</sup> Data missing for 1 patient. <sup>‡</sup> Data missing for 81 patients. <sup>§</sup> Data missing for 8 patients. <sup>††</sup> Data missing for 9 patients. <sup>||</sup> Data missing for 44 patients. PP: per protocol; csDMARD: conventional disease-modifying antirheumatic drug; bDMARD: biologic DMARD; DAS28-CRP: 28-joint count Disease Activity Score using C-reactive protein; VAS: visual analog scale; IQR: interquartile range; HCP2: healthcare professional (medical student trained in joint assessments).

intrarater reliability for DAS28-CRP at visit 1 and visit 2 was excellent, with an ICC of 0.87 (0.83–0.90) and minimal detectable change (MDC) of 1.13 (Table 2).

For patients' intrarater agreement on DAS28-CRP at visit 1 and 2, illustrated with Bland-Altman plots, the mean difference between the 2 visits was 0.11 (i.e., patients tended to rate DAS28-CRP higher at visit 2; Figure 2). Limits of agreement were –1.01 and 1.23 (i.e., the 2 sets of ratings may not be considered interchangeably, because the differences within the limits of agreement were > 0.6). Forty-four patients (23%) had a DAS28-CRP difference of > 0.6 between visit 1 and 2, 69 (37%) between visit 2 and 3, and 27 (14%) between visit 3 and 4.

Patients' intrarater reliability and agreement for visit 3 and 4 were overall slightly improved compared to visit 1 and 2, with ICC of 0.92 (0.90–0.94), MDC of 1.00, mean difference of –0.02, and limits of agreement of –1.03 and 0.98 (Table 2 and Figure 2).

Observed agreement for patients' assessment of swollen and tender joints at visit 1 and 2 ranged between 85.6–98.9% and 81.3–94.7%, respectively (data not shown). The  $\kappa$  coefficient estimates for swollen and tender joints ranged from 0.20 to 0.79 (slight to substantial agreement), and 0.31 to 0.66 (fair to substantial agreement), respectively. Joints resulting in slight or fair reliability were right fifth PIP and left MCP (both tender and swollen), left second PIP (tender), and left fourth MCP (swollen).

Separate analyses according to country (Supplementary Table 3, available with the online version of this article) showed excellent patients' intrarater reliability for DAS28-CRP at visit 1 and visit 2 for both Danish and Czech patients, with an ICC of 0.87 (0.81–0.91) and ICC of 0.88 (0.80 to 0.92), respectively. For the Czech patients, ICC was



Table 2. DAS28-CRP intra- and interrater reliability.

Variables	Intrarater Reliability (PP)						Interrater Reliability (CC)			
	n	Situation 1 † ICC (95% CI)	MDC	n	Situation 2 †† ICC (95% CI)	MDC	n	Pt vs HCP1, ICC (95% CI)	Pt vs HCP2, ICC (95% CI)	HCP1 vs HCP2, ICC (95% CI)
Visit 1 <sup>§</sup> , all (primary)	121	0.87 (0.80–0.91)	1.12	187	0.87 (0.83–0.90)	1.13	60	0.75 (0.61–0.84)	0.79 (0.68–0.87)	0.86 (0.78–0.91)
Visit 3 <sup>§</sup> , all	121	0.93 (0.90–0.95)	0.92	187	0.92 (0.90–0.94)	1.00	46	0.86 (0.71–0.93)	0.87 (0.75–0.93)	0.90 (0.82–0.94)
Visit 3 <sup>§</sup> , subgroup A	67	0.94 (0.90–0.96)	0.84	102	0.93 (0.90–0.96)	0.92	25	0.87 (0.71–0.94)	0.87 (0.72–0.94)	0.90 (0.78–0.95)
Visit 3 <sup>§</sup> , subgroup B	54	0.92 (0.87–0.95)	1.00	85	0.91 (0.86–0.94)	1.08	21	0.81 (0.47–0.93)	0.84 (0.63–0.93)	0.87 (0.71–0.95)
Visit 1 pretraining <sup>‡</sup> , all	121	0.88 (0.84–0.92)	1.04	187	0.90 (0.87–0.92)	0.99	60	0.69 (0.51–0.84)	0.74 (0.59–0.84)	0.86 (0.78–0.91)

† Situation 1: assessment of intrarater reliability in the PP population with short time followup performed within 1–3 days. †† Situation 2: assessment of intrarater reliability in the PP population with short time followup performed within 1–7 days. § For intrarater reliability, visit 1 is compared with visit 2, and visit 3 is compared with visit 4. ‡ For intrarater reliability, visit 1 pretraining is compared to visit 1 post-training. DAS28-CRP: 28-joint count Disease Activity Score using C-reactive protein; ICC: intraclass correlation coefficient; MDC: minimal detectable change; PP: per-protocol population; CC: complete-case population; Pt: patient; HCP1: healthcare professional (experienced rheumatologist); HCP2: medical student trained in joint assessments.

improved at visit 3 versus 4 compared to visit 1 versus 2 with an ICC of 0.95 (0.93–0.97).

*Comparing differences in joint assessment by DAS28-CRP and on single-joint level performed by patients, HCP, and US (intrarater reliability and agreement).* For DAS28-CRP assessments at visit 1, interrater reliability between patients and HCP1 was good, with an ICC of 0.75 (0.61–0.84) with CI ranging from good to excellent agreement (Table 2). Slightly better reliability was seen with patients versus HCP2, and HCP1 versus HCP2. For the interrater agreement at visit 1, the mean differences and limits of agreement were –0.21 (–1.66; 1.25), –0.15 (–1.55; 1.26), and 0.06 (–0.94; 1.07) for patients versus HCP1, patients versus HCP2, and HCP1 versus HCP2, respectively (Figure 3). Hence, patients tended to rate DAS28-CRP higher than HCP1 and HCP2, while assessments performed by HCP2 and HCP1 were very similar.

Agreement between patients, HCP, and US was rated on single-joint level. Agreement on swollen joints comparing patients and HCP joint assessments ranged from 73.3% to 100% (Supplementary Figure 1, available with the online version of this article). Hence, in 73.3–100% of the assessments, patients and HCP were agreeing on whether the given joint was swollen. Comparing patients and HCP with US showed an observed agreement for swollen joints between 43.3–78.3% and 41.7–83.3% (12 joints assessed on single-joint level), respectively (Supplementary Figure 1). Lowest agreement comparing US with patients and HCP was seen for both wrists ranging from 43.4% to 45% and 41.7% to 45%, respectively.

For the ratings of tender joints, the observed agreement ranged from 78.3% to 100% for joints assessed by patients and HCP, and between 75% to 93.3% when comparing patients and HCP individually with US (Supplementary Figure 1, available with the online version of this article). Lowest agreement for tender joints was seen for right second, fourth, and fifth MCP, and for left wrist when comparing with US.

*Effect of joint count training on intra- and interrater reliability and agreement.* At visit 1, patients' DAS28-CRP assessments before and after joint score training had excellent intrarater reliability with ICC of 0.90 (0.87–0.92) and MDC of 0.99, which was similar to the reliability between visit 1 and 2 (Table 2). The mean difference was –0.08 with limits of agreements of –1.06 and 0.90. Hence, assessments before and after training were very similar (Figure 4). Reliability between patients and HCP was generally better after initial training (Table 2).

Comparing groups receiving repeated training with groups receiving no training at visit 3 showed similar intrarater reliability between visit 3 and 4; however, the MDC was slightly higher in group B (no repeated training) compared to group A (repeated training; Table 2). Intrarater agreement tended to be slightly better in group A than in group B, with mean difference and limits of agreement of –0.07 (–0.98; 0.84) and 0.03 (–1.06; 1.12), respectively (Figure 4).

*Sensitivity analyses.* For intrarater reliability, the sensitivity analyses showed results similar to the PP analyses when comparing with the intention-to-treat population and the as-observed population. However, the interrater reliability was generally lower.

## DISCUSSION

This study showed patient-performed joint assessments to be reliable and in overall agreement with joint counts performed by HCP in this group of patients with low to moderate disease activity. Further, the findings were consistent in people with RA across 2 European countries with different language and routines.

Patients' intrarater reliability on DAS28-CRP on all visits was excellent with small variations in patient assessments from day to day. These results were independent of the need for repeated instruction (i.e., the group receiving only baseline training did as well in the test as the group that was given repeated training). Based on the observed, but not significant, effect of training, patients with chronic RA

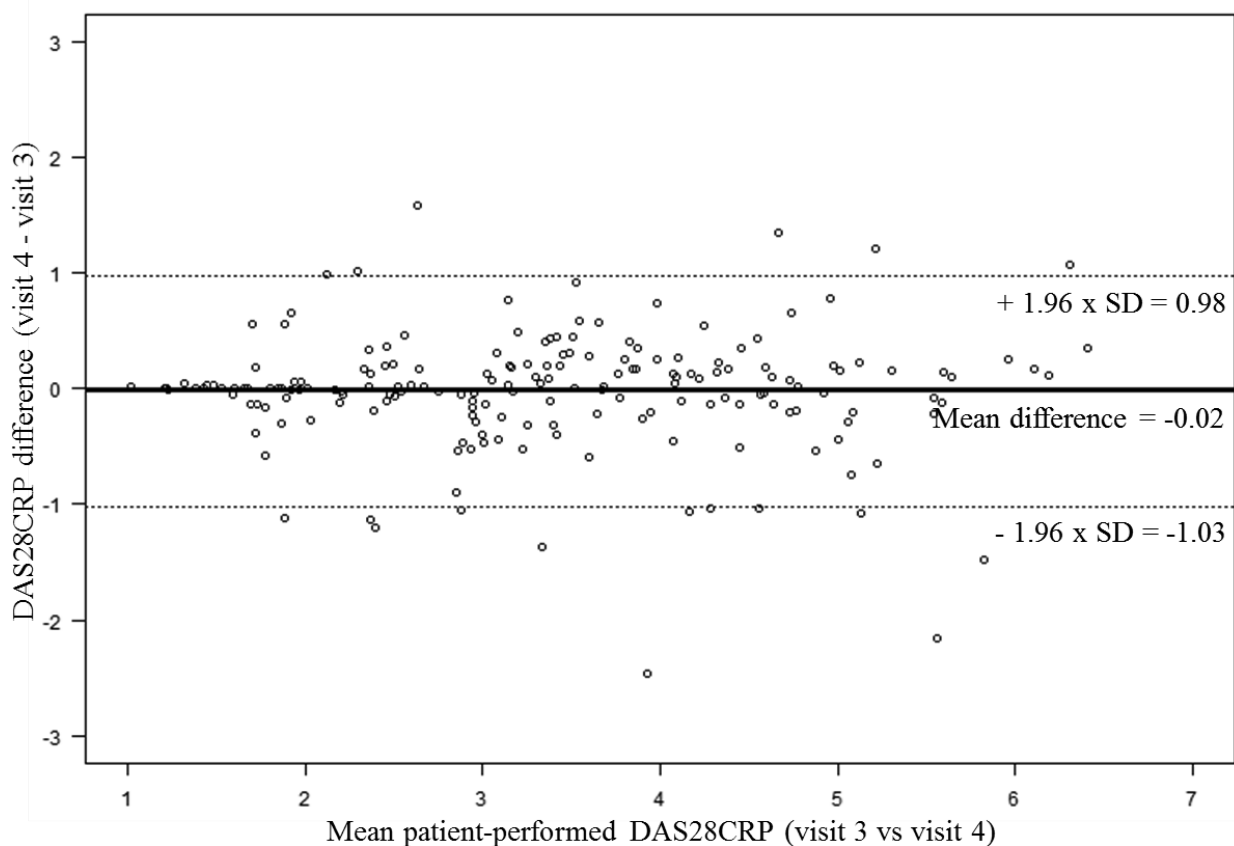
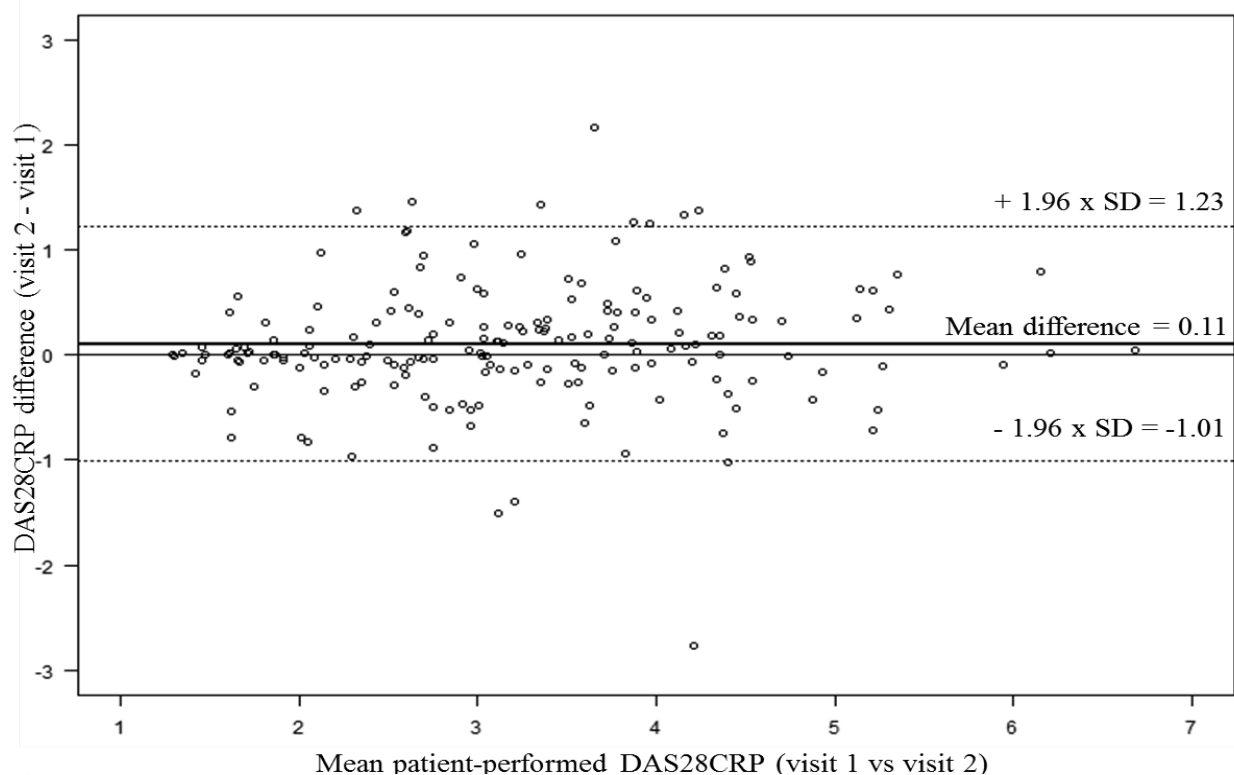


Figure 2. Patients' DAS28-CRP intrarater agreement at visit 1 versus 2 and visit 3 versus 4. DAS28-CRP: 28-joint count Disease Activity Score using C-reactive protein.

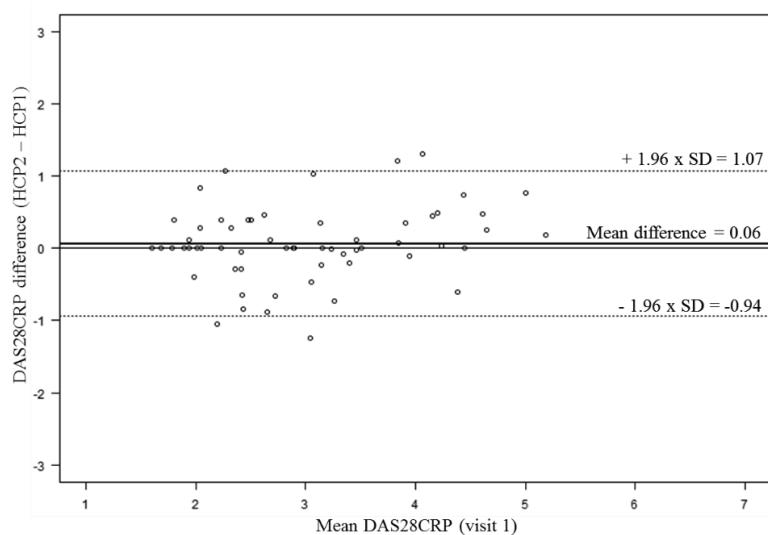
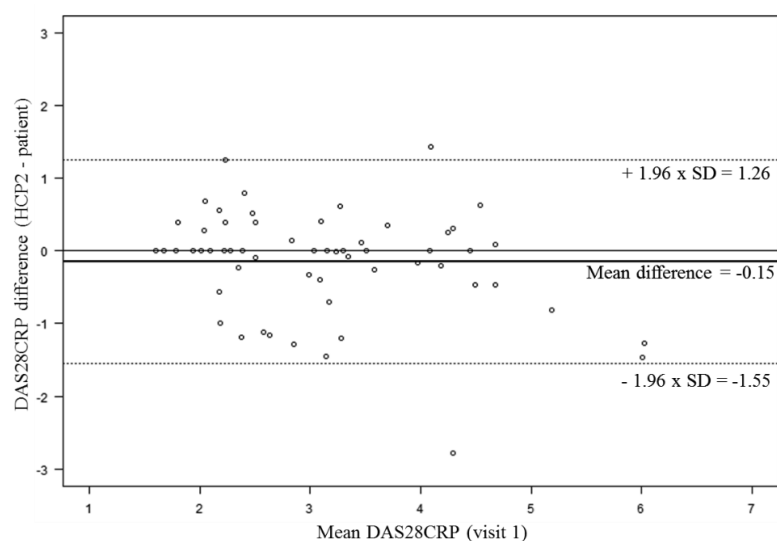
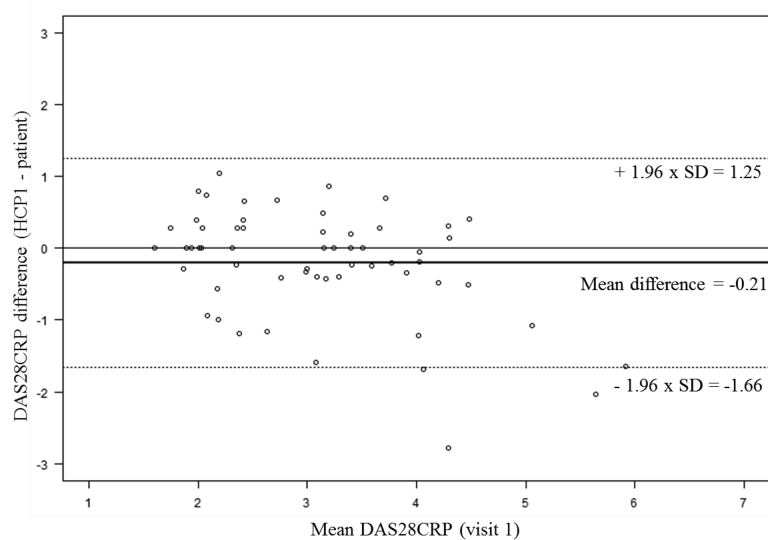


Figure 3. DAS28-CRP interrater agreement comparing patients and HCP at visit 1. DAS28-CRP: 28-joint count Disease Activity Score using C-reactive protein; HCP: healthcare professional.

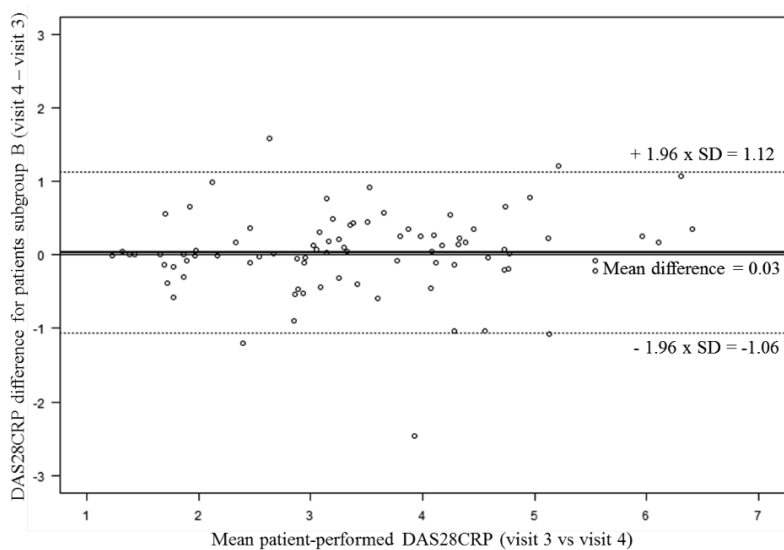
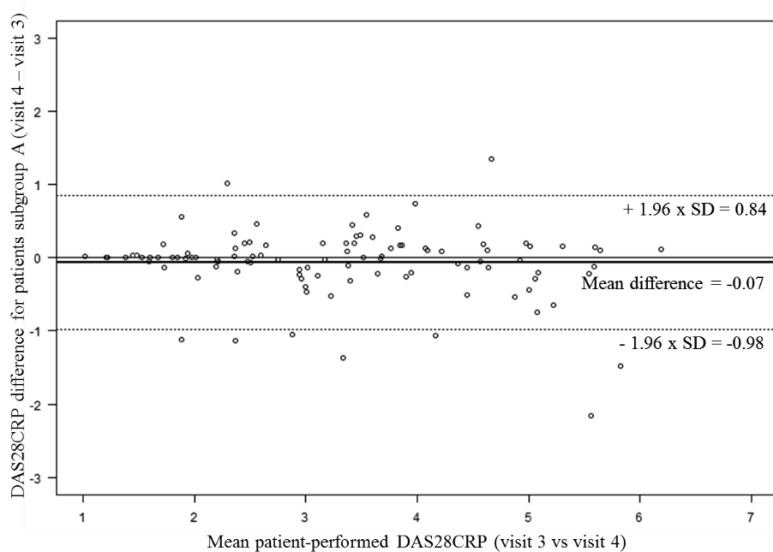
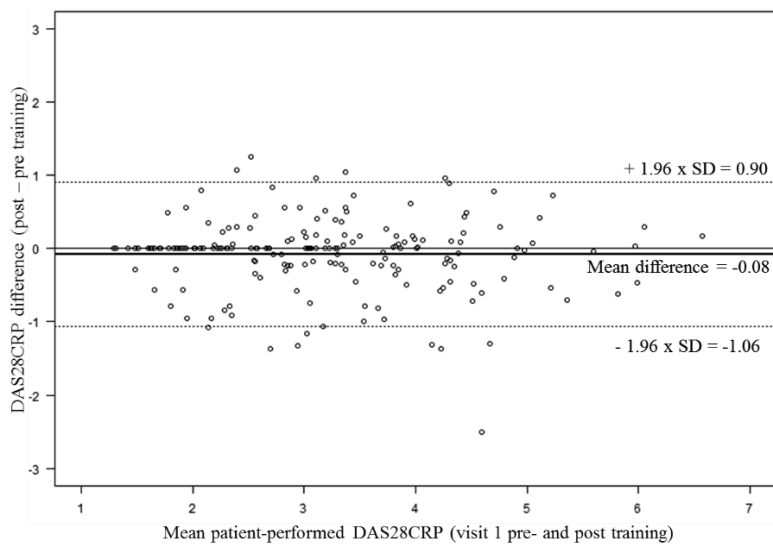


Figure 4. Effect of repeated training for DAS28-CRP evaluation by patients. DAS28-CRP: 28-joint count Disease Activity Score using C-reactive protein.



(disease duration: mean 11.9 yrs, SD 8.7) have a good idea about the state of their joints, which might be an advantage during home monitoring. Although this may have a limitation regarding generalizability for all patients, it might explain why this study found excellent reliability.

Reliability was excellent, though with MDC of 1.13 [i.e., above the predefined limit for non-important difference (MCID DAS28-CRP 0.6)]. The implications of this difference between MDC and MCID in the context of home monitoring are that a real change of 0.6 in DAS28-CRP would not be reliably detected. A potential consequence of this in a home-monitoring setting would be that HCP might adjust therapy based on spurious DAS28-CRP changes caused by variations in joint assessments.

Assessing agreement revealed the number of patients showing a change in DAS28-CRP exceeding the predefined cutoff limit of 0.6. Fourteen to thirty-seven percent of the patients had a DAS28-CRP difference > 0.6 between visits, which was above the limit where treatment intervention might be considered in a home-monitoring solution. During home monitoring, this would necessitate HCP contacting patients either by phone or requesting an outpatient review. However, the majority of patients' evaluations at home would not cause any concern or elicit a visit to the clinic. Alternatively, it may cause some concern if patients at home do not rate themselves as worsening, when a treatment change in fact is needed. Reassuringly, however, the results of interrater analyses suggest that this would not generally be an issue.

Overall, reliability of DAS28-CRP was considered good to excellent when comparing patients with HCP. Only slightly better reliability was seen at followup. Results suggested that patients' joint assessments were as reliable as assessments performed by HCP, because we identified < 10 cases in which a difference above the limit for clinically important change in DAS28-CRP<sup>21</sup> occurred during the comparison of patients and both HCP. As seen in previous studies, patients tended to rate DAS28-CRP slightly higher than HCP<sup>4,13</sup>, and individual HCP assessments were very similar, though with minor variations<sup>13</sup>.

When comparing patients and HCP at the single-joint level, results showed slightly higher agreement for swollen joints than for tender joints. However, comparing joint assessments performed by patients and HCP with US, higher agreement was seen on tender joints compared to swollen joints for all assessors. The discordance comparing US with HCP and patients, respectively, has no clinical significance because disease activity is based on joint assessments — now performed by HCP and in the future by patients during home monitoring. The discordance, especially in wrists, may reflect that wrists are more frequently affected by inflammation<sup>25,26</sup>, and may reflect overall difficulties and variations in assessments when assessing swollen joints as described in previous studies<sup>4,6,12,13</sup>.

Results of reliability testing suggested that patient-performed joint assessments are just as reliable as HCP-performed joint assessments. This finding is supported by results from the analysis of observed agreement, which showed high agreement comparing patients, HCP1, and HCP2, respectively. Combined with the finding that assessors all showed lack of agreement with US, results indicated that the quality of patient-performed joint assessments was in line with HCP-performed joint assessments.

Our results are in accordance with the suggestion that patient-performed joint counts may be used in clinical research and management<sup>12</sup>, because patients' DAS28-CRP assessments were in line with HCP-performed assessments, which indicates that patients experiencing a worsening of symptoms do in fact react appropriately. These results support the notion that patients' self-assessments can be used in home monitoring of disease activity in patients in remission or with low to moderate disease activity as a supplement to assessments performed by HCP at outpatient clinics<sup>4,6</sup>.

A strength of our study is that it included patients with DAS28-CRP over a broad range from remission to moderate disease activity. A previous study<sup>6</sup> discusses the possibility that it might be easier to obtain agreement in patients with low disease activity because of the low numerical discrepancy and less room for numerical error. Our results showed that moderate to good reliability and agreement can be achieved even when the setup includes patients with moderate disease activity.

One potential limitation may be the extension of the period between assessments (visits 1 and 2, and visits 3 and 4, respectively) to within 7 days instead of 3 days. The latter turned out to be too strict for followup, considering patients' everyday lives. This adaptation did not cause a decrease in quality of the results regarding the interrater reliability and agreement because data collection was time independent (patient and HCP assessments were performed on the same day). The analysis of intrarater reliability and agreement revealed similar results regardless of whether the analyses were performed on data within a 3- or 7-day period.

Patient-performed joint assessments are reliable and comparable with joint assessments performed by HCP or by US, making them useful in the integration of home monitoring in outpatient clinics. Moreover, it is feasible for patients in remission or with low to moderate disease activity to perform joint assessments in a contextual setup, and to examine the possibility of replacing or supplementing hospital-based joint assessments to inform a therapeutic decision. Our data suggest that patient-performed joint assessments may also be applicable for monitoring RA patients with moderate disease activity.

## ACKNOWLEDGMENT

We acknowledge all patients participating in the study and their contribution to the ELECTOR project.

## ONLINE SUPPLEMENT

Supplementary material accompanies the online version of this article.

## REFERENCES

1. Paré G, Moqadem K, Pineau G, St-Hilaire C. Clinical effects of home telemonitoring in the context of diabetes, asthma, heart failure and hypertension: a systematic review. *J Med Internet Res* 2010;12:e21.
2. Knevel R, Schoels M, Huizinga TW, Aletaha D, Burmester G, Combe B, et al. Current evidence for a strategic approach to the management of rheumatoid arthritis with disease-modifying antirheumatic drugs: a systematic literature review informing the EULAR recommendations for the management of rheumatoid arthritis. *Ann Rheum Dis* 2010;69:987-94.
3. Smolen JS, Landewé R, Bijlsma J, Burmester G, Chatzidionysiou K, Dougados M, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. *Ann Rheum Dis* 2017;76:960-77.
4. Cheung PP, Ruysen-Witrand A, Gossec L, Paternotte S, Le Boulout C, Mazieres M, et al. Reliability of patient self-evaluation of swollen and tender joints in rheumatoid arthritis: a comparison study with ultrasonography, physician, and nurse assessments. *Arthritis Care Res* 2010;62:1112-9.
5. Kirwan JR, Snow SM. Which patients see a rheumatologist? SWRACR (South Western Regional Advisory Committee on Rheumatology). *Rheumatology* 1991;30:285-7.
6. Radner H, Grisar J, Smolen JS, Stamm T, Aletaha D. Value of self-performed joint counts in rheumatoid arthritis patients near remission. *Arthritis Res Ther* 2012;14:R61.
7. de Thurah A, Stengaard-Pedersen K, Axelsen M, Fredberg U, Schougaard LM, Hjollund NH, et al. Tele-health followup strategy for tight control of disease activity in rheumatoid arthritis: results of a randomized controlled trial. *Arthritis Care Res* 2018;70:353-60.
8. Barton JL, Criswell LA, Kaiser R, Chen YH, Schillinger D. Systematic review and metaanalysis of patient self-report versus trained assessor joint counts in rheumatoid arthritis. *J Rheumatol* 2009;36:2635-41.
9. Figueroa F, Braun-Moscovici Y, Khanna D, Voon E, Gallardo L, Luinstra D, et al. Patient self-administered joint tenderness counts in rheumatoid arthritis are reliable and responsive to changes in disease activity. *J Rheumatol* 2007;34:54-6.
10. Hanly JG, Mosher D, Sutton E, Weerasinghe S, Theriault D. Self-assessment of disease activity by patients with rheumatoid arthritis. *J Rheumatol* 1996;23:1531-8.
11. Heegaard C, Dreyer L, Egsmose C, Madsen OR. Test-retest reliability of the Disease Activity Score 28 CRP (DAS28-CRP), the Simplified Disease Activity Index (SDAI) and the Clinical Disease Activity Index (CDAI) in rheumatoid arthritis when based on patient self-assessment of tender and swollen joints. *Clin Rheumatol* 2013;32:1493-500.
12. Houssien DA, Stucki G, Scott DL. A patient-derived disease activity score can substitute for a physician-derived disease activity score in clinical research. *Rheumatology* 1999;38:48-52.
13. Kavanaugh A, Lee SJ, Weng HH, Chon Y, Huang XY, Lin SL. Patient-derived joint counts are a potential alternative for determining disease activity score. *J Rheumatol* 2010;37:1035-41.
14. Levy G, Cheetham C, Cheatwood A, Burchette R. Validation of patient-reported joint counts in rheumatoid arthritis and the role of training. *J Rheumatol* 2007;34:1261-5.
15. Salaffi F, Filippucci E, Carotti M, Naredo E, Meenagh G, Ciapetti A, et al. Inter-observer agreement of standard joint counts in early rheumatoid arthritis: a comparison with grey scale ultrasonography—a preliminary study. *Rheumatology* 2007;47:54-8.
16. Uhlig T, Kvien TK, Pincus T. Test-retest reliability of disease activity core set measures and indices in rheumatoid arthritis. *Ann Rheum Dis* 2009;68:972-5.
17. Naredo E, Bonilla G, Gamero F, Usón J, Carmona L, Laffon A. Assessment of inflammatory activity in rheumatoid arthritis: A comparative study of clinical evaluation with grey scale and power Doppler ultrasonography. *Ann Rheum Dis* 2005;64:375-81.
18. Lee A, Sandvei M, Asmussen HC, Skougaard M, Macdonald J, Zavada J, et al. The development of complex digital health solutions: formative evaluation combining different methodologies. *JMIR Res Protoc* 2018;7:e165.
19. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess* 1994;6:284-90.
20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
21. Fransen J, Van Riel P. The Disease Activity Score and the EULAR response criteria. *Clin Exp Rheumatol* 2005;5 Suppl 39:S93-9.
22. Fransen J, Van Riel PL. The Disease Activity Score and the EULAR response criteria. *Rheum Dis Clin North Am* 2009;35:745-57.
23. Wells G, Beaton D, Shea B, Boers M, Simon L, Strand V, et al. Minimal clinically important differences: review of methods. *J Rheumatol* 2001;28:406-12.
24. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [Internet. Accessed December 16, 2019.] Available from: [www.R-project.org](http://www.R-project.org)
25. Dejaco C, Duftner C, Wipfler-Freissmuth E, Weiss H, Schneider T, Schirmer M. Elderly- versus younger-onset rheumatoid arthritis: higher levels of ultrasound-detected inflammation despite comparable clinical disease activity. *Arthritis Care Res* 2013;65:304-8.
26. Terao C, Hashimoto M, Yamamoto K, Murakami K, Ohmura K, Nakashima R, et al. Three groups in the 28 joints for rheumatoid arthritis synovitis—analysis using more than 17,000 assessments in the KURAMA database. *PLoS One* 2014;8: e59341.