

1 **Patients with rheumatoid arthritis acquire sustainable skills for home**  
2 **monitoring: a prospective dual-country cohort study (ELECTOR clinical trial I)**

3 *Marie Skougaard<sup>1</sup>, Henning Bliddal<sup>1</sup>, Robin Christensen<sup>1,2</sup>, Karen Ellegaard<sup>1</sup>, Sabrina M. Nielsen*  
4 *<sup>1,2</sup>, Jakub Zavada<sup>3</sup>, Sabina Oreska<sup>3</sup>, Niels S. Krogh<sup>4</sup>, Christian C. Holm<sup>1</sup>, Merete L. Hetland<sup>5,6</sup>,*  
5 *Jiri Vencovsky<sup>3</sup>, Henrik Røgind<sup>5</sup>, Peter C. Taylor<sup>7</sup>, Henrik Gudbergensen<sup>1</sup>*

6

7 **KEY INDEXING TERMS:** rheumatoid arthritis / self-management / home-monitoring / disease  
8 activity score / reliability

9

10 **AFFILIATIONS**

11 1: The Parker Institute, Copenhagen University Hospital Bispebjerg and Frederiksberg, Copenhagen,  
12 Denmark; 2: Research Unit of Rheumatology, Department of Clinical Research, University of  
13 Southern Denmark, Odense University Hospital, Denmark; 3: Institute of Rheumatology, Prague, and  
14 Department of Rheumatology, First Faculty of Medicine, Charles University, Prague, Czech  
15 Republic; 4: Zitelab ApS, Copenhagen, Denmark; 5: Copenhagen Center for Arthritis Research  
16 (COPECARE), Center for Rheumatology and Spine Diseases, Centre of Head and Orthopaedics,  
17 Rigshospitalet, Glostrup, Denmark; 6: Department of Clinical Medicine, Faculty of Health and  
18 Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 7: Botnar Research Centre,  
19 Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of  
20 Oxford, Oxford, United Kingdom.

21

22 *ELECTOR: E-Health in Rheumatology, a HORIZON 2020 project.*

23 **FUNDING**

1 The project was funded by The European Commission's Research and Innovation Framework  
2 Programme as a part of the ELECTOR project. The Parker Institute, Bispebjerg and Frederiksberg  
3 Hospital is supported by a core grant from the Oak Foundation (OCAY-13-309).

#### 4 **CONFLICT OF INTEREST**

5 M. Skougaard: none, H. Bliddal: none, R. Christensen: none, K. Ellegaard: none, S.M. Nielsen: none,  
6 J. Zavada: none, S. Oreska: none, N.S. Krogh: CEO of Zitelab ApS and developer of the ELECTOR  
7 online platform, C.C. Holm: none, M. L. Hetland: none, J. Vencovsky: none, H. Røgind: none, P. C.  
8 Taylor: none, H. Gudbergson: none

9

#### 10 **AUTHOR INFORMATION**

11 M. Skougaard, MD

12 H. Bliddal, MD, DMSc, Professor

13 R. Christensen, BSc, MSc, PhD, Professor

14 K. Ellegaard, PT, PhD

15 S.M. Nielsen, BSc, MSc

16 J. Zavada, MD, PhD

17 S. Oreska, MD

18 N.S. Krogh, CEO

19 C.C. Holm, MSc

20 M. L. Hetland, MD, DMSc, Professor

21 J. Vencovsky, Professor

22 H. Røgind, MD, PhD

23 PC. Taylor, Professor

24 H. Gudbergson, MD, PhD

1   **CORRESPONDING AUTHOR**

2   Henrik Gudbergson, MD, PhD,

3   The Parker Institute, Copenhagen University Hospital Bispebjerg and Frederiksberg,

4   Nordre Fasanvej 57, Road 8, Entrance 19, 2000 Frederiksberg, Denmark

5   Mail: [rindelgudbergson@gmail.com](mailto:rindelgudbergson@gmail.com)

6

7   **RUNNING HEAD:** ELECTOR: e-Health in Rheumatology

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

## 1    **ABSTRACT**

2    **Objective:** In an eHealth setting, to investigate intra- and inter-rater reliability and agreement of joint  
3    assessments and disease activity score (DAS28CRP) in patients with rheumatoid arthritis (RA) and  
4    test the effect of repeated joint assessment training.

5    **Methods:** Patients with DAS28CRP  $\leq 5.1$  were included in a prospective cohort study  
6    (clinicaltrials.gov: NCT02317939). Intra-rater reliability and agreement of patient-performed joint  
7    counts were assessed through completion of five joint assessments over a two-month period. All  
8    patients received training on joint assessment at baseline, only half of the patients received repeated  
9    training. A subset of patients was included in an appraisal of inter-rater reliability and agreement  
10    comparing joint assessments completed by patients, healthcare professionals (HCPs), and  
11    ultrasonography. Cohen's Kappa coefficients and intraclass correlation coefficients (ICC) were used  
12    for quantifying of reliability of joint assessments and DAS28CRP. Agreement was assessed using  
13    Bland-Altman plots.

14    **Results:** Intra-rater reliability was excellent with ICC of 0.87 (95%CI: 0.83 to 0.90) and minimal  
15    detectable change of 1.13. ICCs for inter-rater reliability ranged between 0.69 and 0.90 (good to  
16    excellent). Patients tended to rate DAS28CRP slightly higher than HCPs. In patient receiving repeated  
17    training, a mean difference in DAS28CRP of -0.08 was observed (limits of agreements of -1.06 and  
18    0.90). After 2 months, reliability between patients and HCPs was similar between groups receiving  
19    single or repeated training.

20    **Conclusion:** Patient-performed assessments of joints and DAS28CRP in an eHealth home-  
21    monitoring solution were reliable and comparable with HCPs. Patients can acquire the necessary  
22    skills to conduct a correct joint assessment after initial and thorough training.

23

24

## 1 INTRODUCTION

2 Telemonitoring and eHealth solutions for assessing patients with chronic illnesses such as diabetes,  
3 asthma, and hypertension have previously shown great advantages in better control of illnesses and  
4 improvement of symptoms (1). A similar eHealth solution for patients with rheumatoid arthritis (RA)  
5 may be advantageous both for patients and health care systems.

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

16 To assess disease activity by Disease Activity Score in 28 joints and C-reactive protein (DAS28CRP)  
17 one needs data concerning CRP, the visual analogue scale (VAS) on patient assessment of global  
18 health, as well as swollen and tender joint counts. All items of this score may be made available at  
19 the patient's home in a future home-monitoring solution.

20 Previous studies on the subject have shown poor to moderate reliability and agreement as well as  
21 intra- and inter-observer variations in joint assessments, especially of joint swelling (4, 6, 8-16).  
22 Consequently, before home monitoring can become a reality there is a need to investigate whether  
23 patient-performed joint counts reported via a home-monitoring platform are reliable and whether the  
24 agreement between patient and HCP joint counts is sufficient. Joint assessments by use of

1 ultrasonography (US) in patients with RA has been shown to be more sensitive than clinical  
2 examination in evaluating both swelling and tenderness (17), consequently US was included to  
3 explore the validity of patient and HCP joint assessments.

4 The aim of this study was to explore the intra- and interrater reliability and agreement of joint  
5 assessments and the corresponding DAS28CRP, and assess the effect of repeated training in joint  
6 assessment.

## 8 **MATERIALS AND METHODS**

### 9 **Design**

10 The ELECTOR clinical trial #1 was initially designed as a randomised controlled trial with a 2:2:1:1  
11 allocation ratio. The randomisation algorithm appeared impossible to implement due to organisational  
12 challenges. It was decided to continuously include patients for a prospective cohort study with  
13 consecutive recruitment of participants in the investigation of intra-rater reliability and agreement  
14 (group 1), or the investigation of intra- and inter-rater reliability and agreement (group 2). The trial  
15 was registered at clinicaltrials.gov (NCT02317939). All patients received joint assessment training at  
16 baseline whereas only half of the patients were subjected to a repeated training (group A). Patients  
17 not receiving repeated training (group B). This resulted in a total of 4 possible groups; 1A, 1B, 2A,  
18 and 2B (figure 1). When allocated, the platform kept track of the study protocol making sure that the  
19 content related to the recurrent training at follow up was only shown to patients allocated to this  
20 group.

### 21 **Setting**

22 During the period from December 2014 to April 2017 patients routinely visiting the outpatient clinics  
23 of rheumatology in Frederiksberg, Denmark (DK), and in Prague, The Czech Republic (CZ), were  
24 screened for eligibility and invited to participate. As part of the ELECTOR project an online web-

1 based platform (18) was developed for filling in questionnaires and joint assessments. The study was  
2 performed according to the Helsinki Declaration, and with approval from the Ethics Committee of  
3 the Capital Region (J.no.: H-3-2014-108), DK, and the institutional Ethics Committee (J.no.:  
4 9220/2015), CZ. Furthermore, approval was obtained from the Danish Data Protection Agency.

## 5 **Patients**

6 Patients were considered eligible for inclusion if they had been diagnosed with RA for  $\geq 12$  months,  
7 were between 18 and 85 years of age, and had a DAS28CRP  $\leq 5.1$  assessed by a HCP at a screening  
8 visit. All patients included in the study provided written informed consent.

## 9 **Patient instructions**

10 At baseline, all patients received video instruction and individual face-to-face guidance on how to fill  
11 in results on the online platform. Baseline training on how to perform joint assessments was delivered  
12 via a web-based service consisting of three instruction videos with focus on; 1) general information  
13 about joint assessment, 2) assessment of wrist and finger joints, and 3) assessment of elbow, shoulder,  
14 and knee joints, followed by a short training session delivered by the HCP. The follow up training  
15 session, given to half of the patients, included only the videos.

## 16 **Clinical investigations**

17 Assessment of 28 tender and/or swollen joints (proximal interphalangeal joints (PIP) 1-5,  
18 metacarpophalangeal joints (MCP) 1-5, wrist, elbow, shoulder and knee on both sides) were  
19 performed by patients five times on four different days over a period of two months (visit schedule;  
20 suppl. 1). For investigation of inter-rater reliability and agreement, two joint assessments were  
21 completed within a week at two occasions with an interval of two month, resulting in a total of four  
22 joint assessments. Joint assessments were conducted by an experienced rheumatologist (HCP1) and  
23 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 (HAQ) and Visual Analogue Scale (VAS) were done at all visits (suppl. 1).

### **Ultrasound examination**

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

### **Statistical methods**

Prior to conducting the analyses, a statistical analysis plan was published online ([www.parkerinst.dk](http://www.parkerinst.dk)). Statistical analysis was conducted on the per-protocol (PP) population for assessment of intra-rater reliability and agreement. The PP population was defined as patients with complete data to calculate DAS28CRP at all visits within the predefined time slots; short time follow up within 1-3 days (situation 1 – suppl. 1). The same analysis was carried out for the PP population with all data to calculate DAS28CRP at an extended time slot; short time follow up within 1-7 days (situation 2 - suppl. 1). Statistical analysis for inter-rater assessment was conducted on the complete case (CC) population for inter-rater assessments as this analysis was independent of time compared to the intra-rater analysis. CC was defined as patients having complete data to calculate DAS28CRP for all four raters.



1 *Comparison of DAS28CRP ratings:* Intraclass correlation coefficients (ICC) were calculated for  
2 quantification of intra- and interrater reliability for DAS28CRP. ICC were calculated for  
3 quantification of intra-rater reliability, i.e. patient scores from visit 1 was compared with visit 2, and  
4 visit 3 was compared with visit 4, interrater reliability, i.e. patient scores were compared to HCP  
5 scores at visit 1 and visit 3. Values  $<0.40$  were interpreted as poor reliability, 0.40 to 0.59 as fair  
6 reliability, 0.60 to 0.74 as good reliability, and 0.75-1.00 as excellent reliability (19).

7 *Comparison of swollen/tender joint ratings:* Cohen's Kappa coefficients were used for quantification  
8 of reliability for classification of tender and swollen joints on single joint level. A kappa coefficient  
9  $<0$  was considered poor, 0.0–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial,  
10 and 0.81–1.00 almost perfect level of reliability (20). Intra- and interrater agreement were assessed  
11 using Bland-Altman plots and limits of agreement were calculated. A difference in DAS28CRP  $>0.6$   
12 was considered minimal clinically important difference (MCID) as this is the cut off in EULAR's  
13 response criteria (21, 22) where changes in medical management might be considered (23).  
14 Differences in DAS28CRP  $>0.6$  were considered as a proxy for when additional contact to the patient  
15 might be considered during home-monitoring.

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

20 *Sensitivity analyses:* To account for missing data a series of sensitivity analyses were carried out to  
21 compare the pre-defined PP and CC populations, and moreover for an intention-to-treat (ITT)  
22 population where missing data on DAS28CRP were imputed using grand mean and an as-observed  
23 (AsO) population defined by patients, who had any data on the outcome of interest.

1 Calculations were carried out using the statistical software R (version 3.3.3) with the package “psych”  
2 (24).

### 3 **RESULTS**

#### 4 **Patient flow**

5 Of 443 patients screened for eligibility, a total of 314 patients were included in the trial. For the  
6 primary analysis, data from 187 patients were included for assessment of intra-rater reliability and  
7 agreement (PP population – figure 1). For analysis of inter-rater reliability and agreement data were  
8 included from 60 patients (CC population).

#### 9 **Demographics**

10 At inclusion, the PP population had a mean age of 55.1 (standard deviation [SD] 13.5) years, 81%  
11 were women, mean disease duration was 11.9 (SD 8.7) years, and 86%, 24%, and 22% received  
12 conventional synthetic disease modifying anti-rheumatic drugs (csDMARD), biologic DMARD  
13 (bDMARD), and prednisolone, respectively (Table 1). They had a mean DAS28CRP (assessed by a  
14 HCP2) of 3.0 (SD 1.0), and median VAS patient global, VAS pain, and VAS fatigue of 21.5, 18, and  
15 28, respectively. When comparing patients from DK and CZ (suppl. 2), the patients from DK were  
16 generally older, included fewer females, fewer were treated with csDMARD and prednisolone, and  
17 more were in treatment with bDMARDs. There was no statistically significant difference between  
18 the patients regarding DAS28CRP, VAS global or VAS pain, but patients from DK had a statistically  
19 significant higher VAS fatigue score.

20

21 **Comparing repeated patient-performed joint assessments by DAS28CRP (Intra-rater**  
22 **reliability and agreement)**

1 Patients' intra-rater reliability for DAS28CRP at visit 1 and visit 2 was excellent with an ICC of 0.87  
2 (0.83 to 0.90) and minimal detectable change (MDC) of 1.13 (Table 2).

3 For patients' intra-rater agreement on DAS28CRP at visit 1 and 2, illustrated with Bland-Altman  
4 plots, the mean difference between the two visits was 0.11 (Figure 2), i.e. patients tended to rate  
5 DAS28CRP higher at visit 2. Limits of agreement were -1.01 and 1.23, i.e. the two sets of ratings  
6 may not be considered interchangeably, since the differences within the limits of agreement were  
7 >0.6. 44 patients (23%) of the patients had a DAS28CRP difference of >0.6 between visit 1 and 2, 69  
8 (37%) between visit 2 and 3, and 27 (14%) between visit 3 and 4.

9 Patients' intra-rater reliability and agreement for visit 3 and 4 were overall slightly improved  
10 compared to visit 1 and 2, with ICC of 0.92 (0.90 to 0.94), MDC of 1.00, mean difference of -0.02,  
11 and limits of agreement of -1.03 and 0.98 (Table 2 and figure 2).

12 Observed agreement for patients' assessment of swollen and tender joints at visit 1 and 2 ranged  
13 between 85.6-98.9% and 81.3-94.7%, respectively (data not shown). The kappa coefficient estimates  
14 for swollen and tender joints ranged from 0.20 to 0.79 (slight to substantial agreement), and 0.31 to  
15 0.66 (fair to substantial agreement), respectively. Joints resulting in slight or fair reliability were right  
16 fifth PIP and left MCP (both tender and swollen), left second PIP (tender), and left fourth MCP  
17 (swollen).

18 Separate analyses according to country (suppl. 3) showed excellent patients' intra-rater reliability for  
19 DAS28CRP at visit 1 and visit 2 for both Danish and Czech patients, with an ICC of 0.87 (0.81 to  
20 0.91) and ICC of 0.88 (0.80 to 0.92), respectively. For the Czech patients, ICC was improved at visit  
21 3 vs 4 compared to visit 1 vs 2 with an ICC of 0.95 (0.93 to 0.97).

22

1    **Comparing differences in joint assessment by DAS28CRP and on single joint level performed**  
2    **by patients, HCPs, and US (Interrater reliability and agreement)**

3    For DAS28CRP assessments at visit 1, inter-rater reliability between patients and HCP1 was good;  
4    ICC of 0.75 (0.61 to 0.84) with confidence interval ranging from good to excellent agreement (Table  
5    2). Slightly better reliability was seen with patients vs HCP2, and HCP1 vs HCP2. For the inter-rater  
6    agreement at visit 1, the mean differences and limits of agreement were -0.21 (-1.66; 1.25), -0.15 (-  
7    1.55; 1.26), and 0.06 (-0.94; 1.07) for patients vs. HCP1, patients vs. HCP2, and HCP1 vs HCP2,  
8    respectively (figure 3). Hence, patients tended to rate DAS28CRP higher than HCP1 and HCP2, while  
9    assessments performed by HCP2 and HCP1 were very similar.

10   Agreement between patients, HCPs, and US was rated on single joint level. Agreement on swollen  
11   joints comparing patients and HCPs joint assessments ranged from 73.3-100% (suppl. 4). Hence, in  
12   73.3-100 % of the assessments patients and HCPs were agreeing whether the given joint was swollen  
13   or not. Comparing patients and HCPs with US showed an observed agreement for swollen joints  
14   between 43.3-78.3% and 41.7-83.3% (12 joints assessed on single joint level), respectively (suppl.  
15   4). Lowest agreement comparing US with patients and HCPs was seen for both wrists ranging from  
16   43.4-45% and 41.7-45%, respectively.

17   For the ratings of tender joints (suppl. 4), the observed agreement ranged from 78.3-100% for joints  
18   assessed by patients and HCPs, and between 75-93.3% when comparing patients and HCPs  
19   individually with US (suppl. 4). Lowest agreement for tender joints was seen for right second, fourth  
20   and fifth MCP and for left wrist when comparing with US.

21

22   **Effect of joint count training on intra- and inter-rater reliability and agreement**

23   At visit 1, patients' DAS28CRP assessments before and after joint score training had excellent intra-  
24   rater reliability with ICC of 0.90 (0.87 to 0.92) and MDC of 0.99, which was similar to the reliability

1 between visit 1 and 2 (Table 2). The mean difference was -0.08 with limits of agreements of -1.06  
2 and 0.90. Hence, assessments before and after training were very similar (Figure 4). Reliability  
3 between patients and HCPs was generally better after initial training (Table 2).

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

### 9 **Sensitivity analyses**

10 For intrarater reliability, the sensitivity analyses showed similar results as the per-protocol analyses  
11 when comparing with the intention-to-treat population and as-observed population, however, the  
12 interrater reliability was generally lower.

## 13 **DISCUSSION**

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

### 18 **Comparing repeated patient-performed joint assessments by DAS28CRP**

19 Patients' intra-rater reliability on DAS28CRP on all visits was excellent with small variations in  
20 patient assessments from day to day. These results were independent on the need for repeated  
21 instruction, i.e. the group only receiving baseline training did as well in the test as the group, who  
22 were subjected to repeated training. Based on the observed, but not significant, effect of training, it  
23 is believed that chronic RA patients (disease duration: mean 11.9 years, SD 8.7) included have a good

1 idea about the state of their joints which might be an advantage during home-monitoring. Though,  
2 this may have a limitation in terms of generalizability for all patients, it might explain why this study  
3 found excellent reliability.

4 Reliability was excellent, though with MDC of 1.13, i.e. above the pre-defined limit for non-important  
5 difference (MCID DAS28CRP 0.6). The implications of this difference between MDC and MCID in  
6 the context of home monitoring are that a real change of 0.6 in DAS28CRP would not reliably be  
7 detected. A potential consequence of this in a home-monitoring setting would be that HCPs might  
8 adjust therapy based on spurious DAS28CRP changes caused by variations in joint assessments.

9 Assessing agreement revealed the number of patients showing a change in DAS28CRP exceeding the  
10 pre-defined cut-off limit of 0.6. 14-37% of the patients had a DAS28CRP difference >0.6 between  
11 visits which was above the limit where treatment intervention might be considered in a home-  
12 monitoring solution. During home-monitoring, this would necessitate HCPs contacting patients either  
13 by phone or requesting an outpatient review. However, the majority of patients' evaluations at home  
14 would not cause any concern or elicit a visit to the clinic. Alternatively, it may cause some concern  
15 if patients at home do not rate themselves as worsening, when a treatment change in fact is needed.

16 Reassuringly, however, the results of interrater analyses suggest that this would not generally be an  
17 issue.

#### 18 **Comparing differences in joint assessment by DAS28CRP and on single joint level performed** 19 **by patients, HCPs, and US**

20 Overall, reliability of DAS28CRP was considered good to excellent when comparing patients with  
21 HCPs. Only, slightly better reliability was seen at follow-up. Results suggested that patients' joint  
22 assessments were as reliable as assessments performed by HCPs as we did only identify <10 cases  
23 where a difference above the limit for clinically important change in DAS28CRP (21) during the  
24 comparison of patients both HCPs. As seen in previous studies, patients tended to rate DAS28CRP

1 slightly higher than HCPs (4, 13), and individual HCP assessments were very similar, though with  
2 minor variations (13).

3 When comparing patients and HCPs at single joint level, results showed slightly higher agreement  
4 for swollen joints than for tender joints. However, comparing joint assessments performed by patients  
5 and HCPs with US, higher agreement was seen on tender joints compared to swollen joints for all  
6 assessors. The discordance comparing US with HCPs and patients, respectively, have no clinical  
7 significance as disease activity is based on joint assessments – now performed by HCPs and for the  
8 future by patients during home-monitoring. The discordance, especially in wrists, may reflect that  
9 wrists are more frequent affected by inflammation (25, 26), and overall difficulties and variations in  
10 assessments when assessing swollen joints as described in previous studies (4, 6, 12, 13).

11 Results of reliability testing suggested patient-performed joint assessments to be just as reliable as  
12 HCP-performed joint assessments, which was supported by results from the analysis of observed  
13 agreement showing high agreement comparing patients, HCP1, and HCP2, respectively. Together  
14 with the fact that assessors all showed lack of agreement with US, results indicated that the quality  
15 of patient-performed joint assessments were in line with HCP-performed joint assessments.

16 Our results are in accordance with the suggestion that patient-performed joint counts may be used in  
17 clinical research and management (12) as patients' DAS28CRP assessments were in line with HCP-  
18 performed assessments which indicate that patients experiencing worsening in symptoms do in fact  
19 react to these. These results support that patients' self-assessments can be used in home-monitoring  
20 of disease activity in patients in remission or with low-moderate disease activity as a supplement to  
21 assessments performed by HCPs at outpatient clinics (4, 6).

## 22 **Limitations and strengths**

23 A strength of this study is that it included patients with DAS28CRP over a broad range from remission  
24 to moderate disease activity. A previous study (6) discusses the possibility that it might be easier to

1 obtain agreement in patients with low disease activity because of the low numerical discrepancy and  
2 with less room for numerical error. Our results showed that moderate to good reliability and  
3 agreement can be achieved even when the setup includes patients with moderate disease activity.  
4 One potential limitation may be the extension of the period between assessments (visit 1 and 2 & visit  
5 3 and 4, respectively) to be performed within seven days instead of three days, as the latter turned out  
6 to be too strict for follow up in terms of adapting to patients' everyday life. This adaptation did not  
7 cause a decrease in quality of the results with respect to the interrater reliability and agreement as  
8 data collection was time independent (patient and HCP assessments were performed on the same  
9 day). The analysis of intra-rater reliability and agreement revealed similar results regardless of  
10 whether or not the analyses were performed on data within a three- or a seven-day time period.

11

## 12 **CONCLUSION**

13 Patient-performed joint assessments are reliable and comparable with joint assessments performed  
14 by HCPs or via US, allowing them to be useful in the integration of a home-monitoring solution in  
15 outpatient clinics. Moreover, it is feasible for patients in remission or with low to moderate disease  
16 activity to perform patient-performed joint assessments in a contextual setup examining the  
17 possibility to replace or supplement hospital-based joint assessments in order to inform a therapeutic  
18 decision. In conclusion, our data suggests that patient-performed joint assessments may also be  
19 applicable for monitoring RA patients with moderate disease activity.

20

## 21 **ACKNOWLEDMENT**

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

23

24



## 1 REFERENCES

- 2 1. Paré G, Moqadem K, Pineau G, St-Hilaire C. Clinical effects of home telemonitoring in the context  
3 of diabetes, asthma, heart failure and hypertension: A systematic review. *J Med Internet Res*. 2010;12
- 4 2. Knevel R, Schoels M, Huizinga TW, Aletaha D, Burmester G, Combe B, et al. Current evidence  
5 for a strategic approach to the management of rheumatoid arthritis with disease-modifying  
6 antirheumatic drugs: A systematic literature review informing the eular recommendations for the  
7 management of rheumatoid arthritis. *Ann Rheum Dis* 2010;69:987-94.
- 8 3. Smolen JS, Landewé R, Bijlsma J, Burmester G, Chatzidionysiou K, Dougados M, et al. Eular  
9 recommendations for the management of rheumatoid arthritis with synthetic and biological disease-  
10 modifying antirheumatic drugs: 2016 update. *Ann Rheum Dis* 2017;76:960-77
- 11 4. Cheung PP, Ruysen-Witrand A, Gossec L, Paternotte S, Le Boulout C, Mazieres M, et al.  
12 Reliability of patient self-evaluation of swollen and tender joints in rheumatoid arthritis: A  
13 comparison study with ultrasonography, physician, and nurse assessments. *Arthritis Care Res*.  
14 2010;62:1112-9.
- 15 5. Kirwan J, Snow S. Which patients see a rheumatologist? Swracr (south western regional advisory  
16 committee on rheumatology). *Rheumatology* 1991;30:285-87.
- 17 6. Radner H, Grisar J, Smolen JS, Stamm T, Aletaha D. Value of self-performed joint counts in  
18 rheumatoid arthritis patients near remission. *Arthritis Res Ther* 2012;14:R61.
- 19 7. De Thurah A, Stengaard-Pedersen K, Axelsen M, Fredberg U, Schougaard LM, Hjollund NH, et  
20 al. Tele-health followup strategy for tight control of disease activity in rheumatoid arthritis: Results  
21 of a randomized controlled trial. *Arthritis Care Res*. 2018;70:353-60.
- 22 8. Barton JL, Criswell LA, Kaiser R, Chen Y-H, Schillinger D. Systematic review and metaanalysis  
23 of patient self-report versus trained assessor joint counts in rheumatoid arthritis. *J Rheumatol*  
24 2009;36:2635-41.

- 1 9. Figueroa F, Braun-Moscovici Y, Khanna D, Voon E, Gallardo L, Luinstra D, et al. Patient self-  
2 administered joint tenderness counts in rheumatoid arthritis are reliable and responsive to changes in  
3 disease activity. *J Rheumatol* 2007;34:54-6.
- 4 10. Hanly J, Mosher D, Sutton E, Weerasinghe S, Theriault D. Self-assessment of disease activity by  
5 patients with rheumatoid arthritis. *J Rheumatol* 1996;23:1531-8.
- 6 11. Heegaard C, Dreyer L, Egsmose C, Madsen OR. Test–retest reliability of the disease activity  
7 score 28 crp (das28-crp), the simplified disease activity index (sdai) and the clinical disease activity  
8 index (cdai) in rheumatoid arthritis when based on patient self-assessment of tender and swollen  
9 joints. *Clin Rheumatol* 2013;32:1493-500.
- 10 12. Houssien D, Stucki G, Scott D. A patient-derived disease activity score can substitute for a  
11 physician-derived disease activity score in clinical research. *Rheumatology* 1999;38:48-52.
- 12 13. Kavanaugh A, Lee SJ, Weng HH, Chon Y, Huang X-Y, Lin S-L. Patient-derived joint counts are  
13 a potential alternative for determining disease activity score. *J Rheumatol* 2010;37:1035-41.
- 14 14. Levy G, Cheetham C, Cheatwood A, Burchette R. Validation of patient-reported joint counts in  
15 rheumatoid arthritis and the role of training. *J Rheumtol* 2007;34:1261-5.
- 16 15. Salaffi F, Filippucci E, Carotti M, Naredo E, Meenagh G, Ciapetti A, et al. Inter-observer  
17 agreement of standard joint counts in early rheumatoid arthritis: A comparison with grey scale  
18 ultrasonography—a preliminary study. *Rheumatology* 2007;47:54-8.
- 19 16. Uhlig T, Kvien TK, Pincus T. Test–retest reliability of disease activity core set measures and  
20 indices in rheumatoid arthritis. *Ann Rheum Dis* 2009;68:972-5.
- 21 17. Naredo E, Bonilla G, Gamero F, Uson J, Carmona L, Laffon A. Assessment of inflammatory  
22 activity in rheumatoid arthritis: A comparative study of clinical evaluation with grey scale and power  
23 doppler ultrasonography. *Ann Rheum Dis* 2005;64:375-81.

- 1 18. Lee A, Sandvei M, Asmussen HC, Skougaard M, Macdonald J, Zavada J, et al. The Development  
2 of Complex Digital Health Solutions: Formative Evaluation Combining Different Methodologies.  
3 JMIR Res Protoc 2018;7:7
- 4 19. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized  
5 assessment instruments in psychology. Psychological assessment 1994;6:284-90.
- 6 20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics  
7 1977;33:159-74.
- 8 21. Fransen J, Van Riel P. The disease activity score and the eular response criteria. Clin Exp  
9 Rheumatol 2005;23:93-99.
- 10 22. Fransen J, Van Riel PL. The disease activity score and the eular response criteria. Rheum Dis  
11 Clin North Am 2009;35:745-57.
- 12 23. Wells G, Beaton D, Shea B, Boers M, Simon L, Strand V, et al. Minimal clinically important  
13 differences: Review of methods. J Rheumatol 2001;28:406-12.
- 14 24. Team RC. R: A language and environment for statistical computing. 2013:1-16
- 15 25. DeJaco C, Duftner C, Wipfler-Freissmuth E, Weiss H, Schneider T, Schirmer M. Elderly- Versus  
16 Younger-Onset Rheumatoid Arthritis: Higher Levels of Ultrasound-Detected Inflammation Despite  
17 Comparable Clinical Disease Activity. Arthritis Care Res (Hoboken) 2013;65:304-8
- 18 26. Terao C, Hashimoto M, Yamamoto K, Murakami K, Ohmura K, Nakashima R et al. Three groups  
19 in the 28 joints for rheumatoid arthritis synovitis – Analysis using more than 17,000 assessments in  
20 the KURAMA database. PLoS One 2014:8

1 **FIGURE LEGENDS**

2

3 **FIGURE 1: PATIENT FLOW**

4 Patient flow demonstrating the total number of patient filling in questionnaires through the online  
5 platform at the giving visit. The per-protocol population is defined as patients with complete data for  
6 the calculation of DAS28CRP at all visits. 1A: Patients receiving joint count training at baseline and  
7 follow up. Patients' joints were only assessed by patients themselves. 1B: Patients receiving joint  
8 count training only at baseline. Patients' joints were only assessed by patients themselves. 2A:  
9 Patients receiving joint count training at baseline and follow up. Patients' joints were assessed by  
10 patients themselves, HCPs, and US. 2B: Patients receiving joint count training only at baseline.  
11 Patients' joints were assessed by patients themselves, HCPs, and US. Group 1A and 1B included  
12 patients from the Czech Republic and Denmark. Group 2A and 2B included patients from Denmark.

13

14 **FIGURE 2: PATIENTS DAS28CRP INTRA-RATER AGREEMENT AT VISIT 1 VS. 2 &**  
15 **VISIT 3 VS. 4**

16 DAS28CRP; disease activity score calculated based on assessment of 28 joints and C-reactive  
17 protein, SD; standard deviation

18

19 **FIGURE 3: DAS28CRP INTER-RATER AGREEMENT COMPARING PATIENTS AND**  
20 **HCPs AT VISIT 1**

21 DAS28CRP; disease activity score calculated based on assessment of 28 joints and C-reactive  
22 protein, SD; standard deviation

23

24 **FIGURE 4: EFFECT OF REPEATED TRAINING FOR DAS28CRP EVALUATION BY**  
25 **PATIENTS**

- 1 DAS28CRP; disease activity score calculated based on assessment of 28 joints and C-reactive
- 2 protein, SD; standard deviation