



# Effect of Training on Patient Self-Assessment of Joint Counts in Rheumatoid Arthritis: A Systematic Review

Keith Tam,<sup>1</sup> Glen S. Hazlewood,<sup>2</sup>  and Claire E. H. Barber<sup>2</sup> 

**Objective.** Patient self-assessed joint counts, if accurate and reliable, could potentially serve as a useful clinical assessment tool in rheumatoid arthritis (RA). This systematic review examines the effect of patient training on the inter-rater reliability of joint counts between patients and clinicians.

**Methods.** The review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. A search was performed in PubMed, Embase, Cochrane Library, and CINAHL for articles that incorporated patient training and measured the reliability of patient self-assessed joint counts in RA. Articles were included if they reported on the inter-rater reliability between patient and clinician joint counts in both trained and untrained patients with RA. Data were extracted on characteristics of patients, structure and components of the training interventions, joint count reliability of patients with and without training, and patient feedback on training interventions. The relevant data were summarized and described.

**Results.** Multiple training methods have been studied ( $n = 5$ ), including in-person sessions run by rheumatologists and instructional videos on the joint examination. Overall, training improved the reliability of patient self–joint counts, with more marked improvement in reliability of swollen joint counts than tender joint counts. Patients had positive feedback when surveyed on their experiences with training.

**Conclusion.** Various training modalities (in-person and video-based) may be effective at improving reliability of patient self–joint counts. More research is needed on this topic, with potential areas for future research including 1) comparison between the efficacy of different modalities of training, and 2) impact of patient factors (education level and disease severity) on the efficacy of training.

## INTRODUCTION

Joint count assessments, including swollen joint counts (SJC) and tender joint counts (TJC), form an integral part of the clinical assessment in rheumatoid arthritis (RA). The commonly used 28-SJC, 28-TJC, and Disease Activity Score for RA (DAS28) composite score have been widely validated for use in assessing disease severity and guiding treatment decisions in RA (1,2). Traditionally, joint counts are performed by clinicians, who are trained in the clinical detection of synovitis. However, patient self–joint counts have been explored for their potential benefits in clinical practice, including their incorporation as a clinical tool for disease monitoring and increasing patient self-engagement with their disease (3,4). The utility of

patient self–joint counts has become an increasingly important area to explore, with remote disease monitoring and telehealth taking on a larger role across rheumatology practices (5,6).

Prior research has suggested that patient self–joint count assessments have a high degree of inter-rater reliability when compared with joint counts performed by trained clinicians (7). However, limitations to using patient self–joint counts have also been identified, including a relatively low degree of reliability for patient SJC in comparison with the high reliability seen with patient TJC (8,9). Training programs for performing joint counts designed for clinicians have been shown to improve the reliability of joint count assessments among both experienced and inexperienced clinicians, raising the possibility that similar training programs could be designed for patients (10,11).

This was an unfunded study. Dr. Barber is funded through the Arthritis Stars Career Development Award from the Canadian Institutes of Health Research-Institute of Musculoskeletal Health and Arthritis (STAR-19-0611/CIHR SI2-169745). Dr. Hazlewood is supported by a Canadian Institute for Health Research New Investigator Salary Award

<sup>1</sup>Keith Tam, MD: Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada; <sup>2</sup>Glen S. Hazlewood, MD, PhD, FRCPC, Claire E. H. Barber, MD, PhD, FRCPC: Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada, and Arthritis Research Canada, Vancouver, British Columbia, Canada.

Dr. Hazlewood is the current Canadian Rheumatology Association (CRA) Guidelines Committee Chair and Dr. Barber is the current co-chair of the CRA Human Resources Committee. No other disclosures relevant to this article were reported.

Address correspondence to Claire Barber, MD, PhD, FRCPC, Division of Rheumatology, University of Calgary, HRC3AA20-3280, Hospital Drive, NW Calgary, Alberta T2N 4Z6, Canada. Email: cehbarbe@ucalgary.ca.

Submitted for publication June 22, 2021; accepted in revised form August 16, 2021.

**SIGNIFICANCE & INNOVATIONS**

- Training methods, including in-person teaching sessions and instructional videos, have been used to teach patients to perform their own joint counts in rheumatoid arthritis.
- Structured training may be effective at improving the inter-rater reliability between patient and clinician joint counts.
- More research is needed to quantify the effect of training and determine whether any differences exist between different modalities of training.

Although prior reviews have examined the general reliability of patient self-joint counts, to our knowledge, no reviews to date have explored the role of training on improving the reliability of patient self-joint counts (7,12). Patient education is known to play

a large role in the management of RA, from improving disease self-monitoring to facilitating shared decision-making for disease management (11,13). Structured programs for patient education have been shown to have positive effects on reducing disability, improving pain and global assessment scores, and improving psychological status for patients (14). Educating patients with RA to conduct their own joint counts may provide additional benefits for facilitating remote disease monitoring. Determining whether any differences exist between different training methods also has important implications for recommending optimal strategies for training.

We conducted this systematic review to examine the topic of training patients to perform self-joint counts. Specifically, we were interested in identifying different training methods and exploring whether the use of training results in any measurable impact on the inter-rater reliability of patient self-joint counts compared with clinician joint counts.

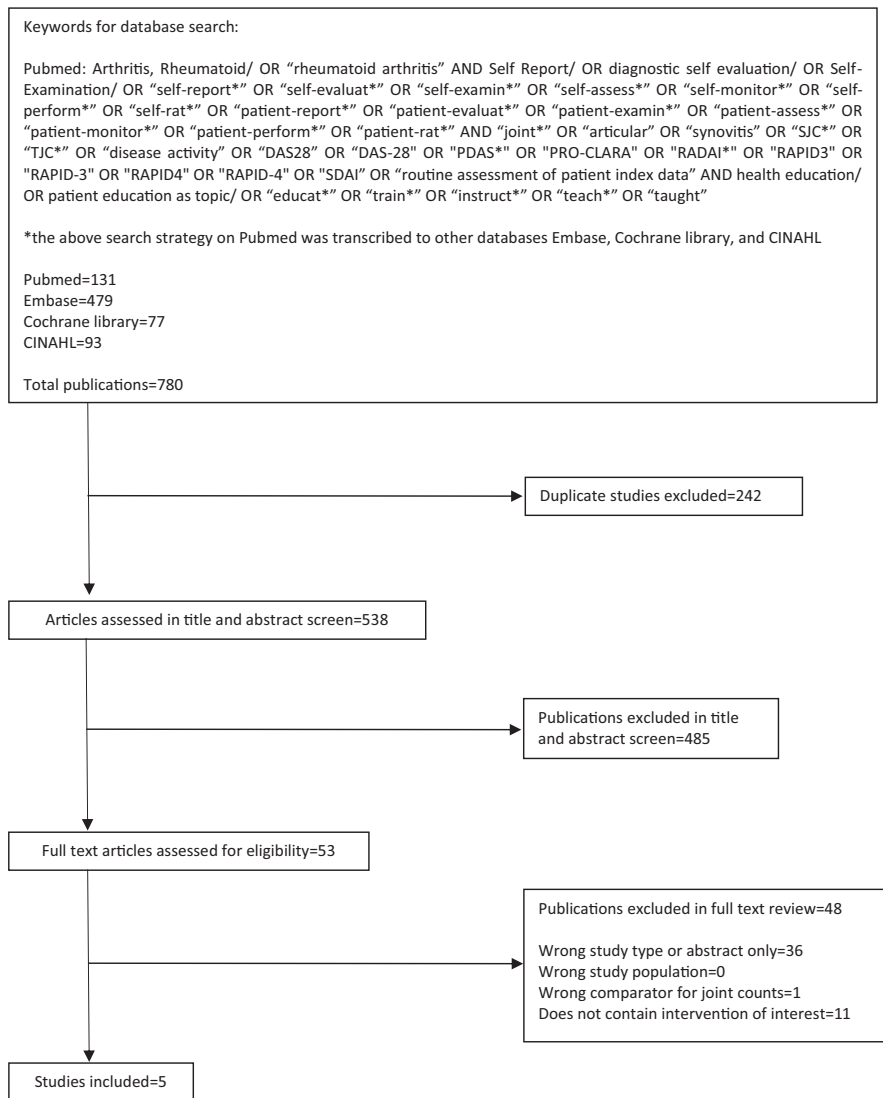


Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram and search strategy.

## PATIENTS AND METHODS

Methodology for this review was conducted in accordance with PRISMA guidelines (15). A protocol was submitted to and published in the International Prospective Register of Systematic Reviews (PROSPERO), prior to conducting the search (reference ID: CRD42021226557) (16).

**Search strategy.** A search was performed in PubMed, Embase, Cochrane Library, and CINAHL databases for articles published up to January 12, 2021. The search strategy was developed in consultation with two university librarians and was conducted by combining the following search concepts using the Boolean operator “AND”: 1) rheumatoid arthritis, 2) joint or joint assessments, 3) self-report/evaluation, and 4) training/education. Our search strategy can be found in Figure 1. The references of all included studies were also searched for other potential articles. The search results were limited to articles published in English language.

**Article selection criteria.** Articles were managed using the online tool Covidence after the removal of duplicates. Two authors (KT and CB) independently reviewed the titles and abstracts of all articles identified in the search to screen articles for relevance. Disagreements in screening were resolved through consensus. The same two authors independently reviewed the full texts of included articles to assess eligibility for inclusion. The full-text review was performed according to the following inclusion criteria: 1) appears in English language peer-reviewed publications, 2) includes patients with RA, and 3) evaluates the effect of patient training on the inter-rater reliability of joint counts (patient vs clinician) either between groups of patients or before/after training intervention in the same group of patients. Case reports, case series, review articles, and letters were excluded. Disagreements in the full-text review were resolved through consensus between the two reviewers.

**Quality assessment.** Randomized control trials (RCTs) and pseudorandomized studies were independently assessed for risk of bias using the Cochrane RoB 2 tool by two reviewers (17). Studies evaluating the inter-rater reliability between patient and clinician joint counts before and after training were evaluated using a recommended modification of the Cochrane ROBINS-I tool for uncontrolled before-and-after studies, as outlined in the Cochrane Handbook for Systematic Reviews (18).

**Data extraction.** Data extraction was performed independently by two authors (KT and CB) using a pilot-tested data abstraction tool. General information was extracted from each study, including author, year, country, study design, study setting, and study duration.

We extracted the following data on the participants of each study: method of recruitment, eligibility criteria, number of participants (including number in intervention and nonintervention groups), age, sex, education level, ethnicity, disease (RA) duration, seropositivity (rheumatoid factor, RF and anti-cyclic citrullinated peptide, anti-CCP), baseline disease activity, baseline functional status, and baseline RA medications. We also extracted data on whether statistical analysis was performed to assess for baseline differences between participant groups.

Our primary intervention of interest was training provided to patients with the aim of teaching them how to perform self-joint counts (SJC or TJC). We extracted the following data on the intervention: modality of training (in-person, video, etc), intervention provider, timing of intervention, difficulties in compliance/execution of training, and co-interventions. We also extracted any data that were provided on the content of the training sessions.

Our primary outcome was the comparison between trained and untrained patients in the inter-rater reliability of their self-administered joint counts as compared with clinician joint counts. We extracted the following data on the outcome of the studies: comparator for patient self-joint counts, timing of outcome measurement, statistical measure used to assess reliability, reliability of

**Table 1.** Characteristics of study participants

Author, Year	Study Type	Number of Participants	Age, Mean, Y	Female Sex, %	Disease Duration, Mean, Y	Baseline Measures on Disease Activity, mean (SD)
Cheung, 2015 (19)	Randomized controlled trial	101	54.1	81	6.8	SJC = 1.5 (2.3) TJC = 2.1 (3.4)
Radner, 2012 (20)	Cohort study with pseudorandomization	144a	56.3	80	11.4	SJC = 2.4 (1.9) TJC = 3.0 (4.1)
Grainger, 2020 (21)	Before–after study	10	49.5b	90	14.8	SJC = 5.2c TJC = 8.3c
Levy, 2007 (22)	Before–after study	60a	54.1	77	7.4	SJC = 2.0 (2.0) TJC = 3.9 (3.7)
Skougaard, 2020 (23)	Before–after study	60a	55.1	81	11.9	DAS28 = 3.0 (1.0)

Abbreviations: DAS28, Disease Activity Score for Rheumatoid Arthritis; SJC, swollen joint count; TJC, tender joint count.

<sup>a</sup> Subgroup of total study participants with complete data eligible for analysis for the primary outcome of interest.

<sup>b</sup> Median age; mean age not reported.

<sup>c</sup> Mean scores were not reported in the study; values were calculated for this review using the individual scores provided.

patient self-joint counts with and without training, and any relevant secondary outcomes. We also extracted data on patients' qualitative feedback toward the training interventions they received in studies in which this was available, and major themes were identified.

**Data analysis.** The different training interventions used and the primary results from each of the studies were analyzed and summarized descriptively. Because of the heterogeneity in the training interventions used and in the statistical measures of reliability, no meta-analysis was performed. Secondary outcomes that

were relevant to the topic of patient training for self-joint counts were also included for descriptive purposes.

**RESULTS**

A total of 780 articles were identified across the four databases, and 242 duplicate articles were excluded, resulting in 538 articles included in the title and abstract screen. A total of 485 articles were excluded after screening, and 48 articles were excluded after full-text review, resulting in five studies being included for final analysis (Figure 1).

**Table 2.** Characteristics of training interventions & clinician assessment of patient self-joint counts

Study	Characteristics of Patient Training for Self-Reported Joint Counts		Characteristics of the Clinician Assessors		
	Training Components	Timing of Training	Assessors	Blinded (Yes/No)	Standardization of Assessors
Cheung (19)	1. In-person training by rheumatologist lasting 5-10 minutes 2. Doppler ultrasound of 28 joints by trained ultrasonographer to provide live feedback on joints with and without active synovitis	Immediately after randomization and again at 3-month follow-up	One physician per patient; two physicians total in study	Yes	Not reported between assessors
Radner (20)	1. In-person training session by physician with special emphasis on synovial versus bony swelling	At baseline visit	One physician and one biometrician per patient; total number of assessors not clearly specified	Yes	Not reported between assessors, only ICC between patients and different assessor types reported
Grainger (21)	1. Instructional videob demonstrating how to perform joint counts 2. In-person discussion between patients and rheumatologist, facilitated by the principal investigator, focusing on patient opinions on their knowledge and training needs	At baseline visit	One physician pair (two rheumatologists) per patient; two physician pairs (four rheumatologists) total in study	No	Reliability between rheumatologist pairs was excellent for the 28TJC (ICC = 0.95) and moderate for the 28SJC (ICC = 0.53)
Levy (22)	1. In-person training session by clinician lasting 5 minutes with focus on distinguishing actively swollen from chronically enlarged joint	At follow-up visit (average of 50 days after initial visit)	One rheumatologist per patient; total number of assessors in study not specified	Yes	N/A
Skougaard (23)	1. Instructional video with general information about joint assessment, including focus on assessment of wrist, finger, elbow, shoulder, and knee joint groups 2. In-person training session by an HCP	Both training components were administered at baseline to all patients; a subgroup of patients had repeated training with videos only at follow-up (40-68 days)	One rheumatologist and one medical studentc per patient; total number of assessors in study not specified	No	ICCs reported between assessors at all visits (≥0.86)

Abbreviations: HCP, health care provider; ICC, intraclass correlation coefficient; N/A, not applicable; SJC, swollen joint count; TJC, tender joint count.

<sup>a</sup> We have only included the physician's joint count and not the biometrician's joint counts for analysis for comparability between other studies.  
<sup>b</sup> Patients in this study attended a workshop and initially were presented with a 20-minute presentation describing the joint structure, cause of synovitis, treat-to-target principles and an introduction to the joint counts. They then performed their self-examination without further instruction (baseline). Following a physician joint examination, patients watched an instructional video (originally developed for HCPs) on how to perform a joint count. This video, in combination with the discussion between the HCPs, was counted as the "training," after which the second self-joint count was measured.  
<sup>c</sup> We have only included the rheumatologist's and not the medical student's joint counts for analysis.

**Characteristics of the studies and patient populations.** Two of the included studies allocated patients to training and nontraining groups for comparison of joint count inter-rater reliability, one using a randomized method and the other using a pseudorandomized method of allocation (Table 1) (19,20). The three remaining studies examined the reliability of patient self-joint counts before and after a training intervention within the same group of patients (21–23). Of these studies, Levy et al calculated joint count reliability for all 60 patients at baseline (untrained), which was then compared with joint count reliability in a subset of 30 patients who received training at the follow-up visit (22). Although the study by Skougaard et al also allocated patients to an intervention group as part of their study design, the intervention in question was a repeat training session at the follow-up visit, and all patients in the study received an initial round of training at the baseline visit (23). Therefore, for the purposes of data extraction of our primary review question, it was best characterized as a before-and-after study.

Six countries were represented across the five studies (Austria, the Czech Republic, Denmark, New Zealand, Singapore, and the United States). All five studies were conducted within a clinic setting (19–23). The majority of studies (n = 3) included patients with established RA (19,20,23), whereas two of the studies did not distinguish between early versus established RA (21,22). Not all studies reported on their method of recruitment, but, of the studies that reported on this (n = 3), two recruited consecutive patients attending a clinic (20,23) and one used purposive sampling to recruit patients (21). Apart from the study by Grainger et al, which was conducted as a workshop over the course of 1 day (21), the remaining studies had follow-up between 2 and 6 months.

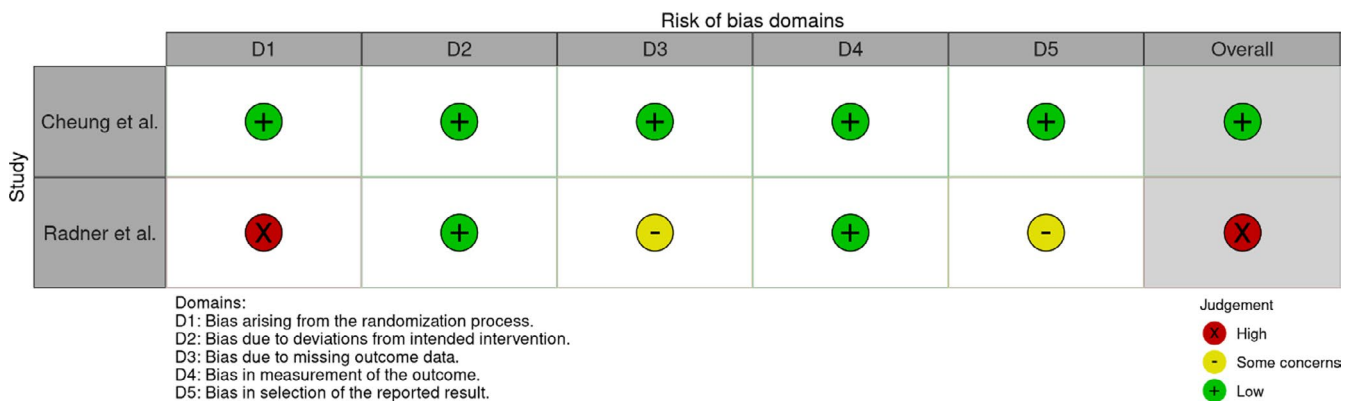
For the primary outcome (comparison of joint count inter-rater reliability with and without training), the number of participants included ranged from 10 to 144 (Table 1) (19–23). This number did not necessarily reflect the total number of participants in the studies, as three studies conducted an analysis for our outcome of interest only within a subgroup of participants (typically those with complete data sets including follow-up data). The patient

populations of all studies were similar in age (mean 50–55 years) and sex (77–90% female). There was a mean disease duration of 10.5 years. In three studies that reported baseline DAS28-C-reactive protein (DAS28-CRP) scores by a clinician, there was an average score of 3.27 (19,20,23). Four studies reported baseline RA medications for patients, and an average of 88% of patients were receiving disease-modifying antirheumatic drugs, whereas the proportion of patients receiving biologic therapy ranged greatly depending on country, from 1% to 40% (19–21,23).

**Methodology of patient training.** Training protocols used to teach patients to perform their own joint counts were heterogeneous across the studies (Table 2). Three studies included more than one training modality as part of their protocol. There was some overlap in training modalities used, with all five studies including an in-person component that was typically administered by a physician. Two studies included instructional videos on how to perform joint counts (21,23). Both the in-person training sessions and instructional videos typically lasted from 5 minutes to 20 minutes (19–23). One study by Cheung et al also used Doppler ultrasound performed by a trained technician to provide patients with live feedback on joints with and without active synovitis (19). Lastly, the study by Grainger et al used a two-way discussion between the patient and rheumatologist, allowing for a more open format to the training (21).

The content covered in the training sessions generally consisted of an overview and description of how to perform joint count assessments (SJs and TJs) (19,21–23). Of note, three studies highlighted that a focus of the training was learning to differentiate between bony enlargement and synovial (soft tissue) swelling (20–22). Teaching on how to perform joint counts for specific regions of the body (eg, fingers and wrists) was also listed as content covered during training (23).

**Quality of studies.** Using the RoB 2 tool for the RCT by Cheung et al, we identified a low risk of bias within this study (Figure 2) (17,19). Although not an RCT, the study by Radner et al was similarly assessed using the RoB 2 tool, given that there was



**Figure 2.** Risk of bias assessment using the Cochrane RoB 2 tool.



allocation of patients to training and nontraining groups in a pseudorandomized manner (20). Assessment using this tool found high risk of bias because of concerns about the randomization process and statistically significant differences in baseline characteristics of patients (baseline evaluator global status and clinician SJC).

The other studies were assessed for potential sources of bias in domains of the Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-1) tool applicable to before-and-after studies (Table 3) (18). The study by Levy et al allocated the first 30 patients returning for follow-up to the training intervention group (22). It is possible this could act as a potential source of participant selection bias, as patients seen earlier in follow-up may generally have more severe disease, and severity of disease potentially affects the inter-rater reliability between patient and clinician joint counts (24). A potential source of bias by confounding identified in all three studies was potential changes to the patient self-joint counts that resulted from undergoing clinicians' assessment, independent of any training intervention received. For instance, it is plausible that patients, after undergoing a clinician joint count in which swollen joints were identified, may subsequently remember which joints were swollen and demonstrate improved inter-rater reliability in their postintervention joint counts through a separate mechanism unrelated to their training. The short time interval between the clinicians' joint assessment and patients' subsequent joint counts after intervention (performed within 1 day of training for all three studies) may increase the effects of this confounder (21–23).

**Reliability of patient self-joint counts.** Reliability of patient self-joint counts was assessed compared with joint counts by one or more clinician and/or biometricians in all studies, although standardization between assessors was not consistently measured (Table 2). To minimize the potential for the clinicians' assessment affecting patients' self-joint counts, four studies had study designs specifying that patients conducted self-joint

counts prior to the clinician's assessment (19–22), whereas the timing of these assessments was not clear in one study (23). Furthermore, to limit this effect in the opposite direction, three studies blinded clinicians to the results of patient self-joint counts or to the allocation of patient groups (19,20,22).

All the studies measured reliability of patient self-joint counts at a baseline visit. The timing of reassessment for reliability of patient self-joint counts after training intervention differed across the studies (Table 4). One study reassessed this outcome at 3 and 6 months after training (19), two studies reassessed immediately after training (during the same clinic visit) (21,22), and two studies reassessed both immediately after training and at follow-up 2 to 3 months after the initial visit (20,23).

Four studies primarily assessed the reliability of SJCs and TJCs (19–22), and one study primarily assessed the reliability of DAS28-CRP scores (23). Although not directly equivalent to a joint count, the reporting of DAS28-CRP by Skougaard et al was felt to be appropriate for inclusion, given that joint counts are central to its calculation. The intraclass correlation coefficient (ICC) was the most common statistical method of reporting reliability of patient self-joint counts as compared with clinician joint counts and was used by three studies (20,21,23). One study used Pearson and Spearman correlation coefficients (22), and one study used prevalence-adjusted bias-adjusted  $\kappa$  (PABAK) as the primary statistical measure (19). Overall, pretraining reliability of patient self-joint counts was considered high across all the studies (19–23). In general, there was also higher reliability for patient TJCs than for SJCs. Given the heterogeneity in the training methodology used and in the reporting methods for joint count reliability, no meta-analysis was performed.

**Effect of training on patient self-joint counts.** Four of five studies found improvement in the inter-rater reliability between patient and clinician joint counts following a training session (19,21–23), whereas one did not find significant differences in reliability of

**Table 3.** Risk of bias assessment in before–after studies

Domain of Bias	Grainger (21)	Levy (22)	Skougaard (23)
Bias due to confounding	Some concerns (post-training measurement made immediately after intervention)	Some concerns (post-training measurement made immediately after intervention)	Some concerns (post-training measurement made immediately after intervention)
Bias in selection of participants into the study	High risk (patients were selectively recruited for participation)	High risk (patients returning earlier for follow-up were allocated to intervention group)	Low risk
Bias in classification of interventions	Low risk	Low risk	Low risk
Bias due to deviations from intended interventions	Low risk	Low risk	Low risk
Bias due to missing data	Low risk	Low risk	Some concerns (subgroup analysis was conducted on participants with complete data)
Bias in measurement of the outcome	Low risk	Low risk	Low risk
Bias in selection of the reported result	Low risk	Low risk	Low risk

**Table 4.** Reliability of patient self-joint counts with and without training

Studies	Timing of Assessment After Training	Statistical Measure of Reliability	Type of Joint Count	Result Without Training	Result With Training
Cheung (19)	6 months	PABA $\kappa$ a	SJCs	0.84±0.23	0.87±0.14
Radner (20)	3 months	ICCb	SJCs	0.41 (95% CI: 0.22-0.57)	0.48 (95% CI: 0.25-0.66)
			TJCs	0.81 (95% CI: 0.72-0.87)	0.84 (95% CI: 0.74-0.90)
Grainger (21)	Immediately after	ICCc	SJCs	0.21 and 0.35	0.65 and 0.71
			TJCs	0.76 and 0.86	0.87 and 0.96
Levy (22)	Immediately after	pp and psd	SJCs	pp: 0.41; ps: 0.64	pp: 0.93; ps: 0.52
			TJCs	pp: 0.79; ps: 0.83	pp: 0.94; ps: 0.89
Skougaard (23)	Immediately after	ICCe	DAS28-CRP	0.69 (95% CI: 0.51-0.84)	0.75 (95% CI: 0.61-0.84)

Abbreviations: CI, confidence interval; DAS28-CRP, Disease Activity Score for Rheumatoid Arthritis C-Reactive Protein; ICC, intraclass correlation coefficient; PABA $\kappa$ , prevalence-adjusted bias-adjusted  $\kappa$ ; pp, Pearson correlation; ps, Spearman correlation; SJC, swollen joint count; TJC, tender joint count.

<sup>a</sup> PABA $\kappa$  >0.6 was considered to be high agreement. Although there was minimal change in the PABA $\kappa$  seen in the trained group before and after training, the proportion of patients reaching high agreement increased from 84% at baseline to 98% at 6 months for the training intervention group, compared with a decrease from 90% at baseline to 85% at 6 months (mean PABA $\kappa$  = 0.89 ± 0.19 to 0.83 ± 0.37) in the no training control group.

<sup>b</sup> Values provided indicate ICC scores at 3-month follow-up. The ICC was also measured at the baseline visit, which was also similar between groups: 0.33 (95% CI: 0.12-0.51) for trained and 0.39 (95% CI: 0.23-0.52) for untrained groups for SJCs, and 0.66 (95% CI: 0.51-0.77) for trained and 0.85 (95% CI: 0.80-0.90) for untrained groups for TJCs.

<sup>c</sup> Patient self-joint counts were compared with two pairs of rheumatologists for reliability, with separate calculations for ICC.

<sup>d</sup> pp and ps calculations for the untrained group include all 60 patients assessed at the baseline visit, whereas calculations for the trained group include only the first 30 patients who returned for follow-up and received training.

<sup>e</sup> Values provided indicate ICC scores at initial visit. ICC scores at 40- to 68-day follow-up were similar between groups with and without repeated training: 0.87 (95% CI 0.71-0.94) and 0.81 (95% CI 0.47-0.93) respectively.

joint counts in patients with or without training (20). The RCT by Cheung et al used PABA $\kappa$  as the statistical measure of agreement (19). Although this is not technically a measure of reliability, it also provides information about the amount of error in measuring self-joint counts and was included in our review. Although the mean PABA $\kappa$  values were fairly comparable between trained and untrained groups (0.87 and 0.84, respectively), using the authors' predefined measure of high agreement (PABA $\kappa$  > 0.6) (19), a statistically significant higher proportion of trained patients achieved high agreement (98%) at 6 months after the intervention compared with patients in the untrained group (85%). No significant differences were noted between the groups for TJCs in this study. On the other hand, the study by Radner et al found no significant difference in the reliability of patient self-joint counts between trained and untrained groups at 3 months, although both groups showed improved reliability at 3 months compared with baseline (20).

Three remaining studies showed improved reliability of patient self-joint counts following training, but it is unclear whether the observed differences in correlation were statistically significant (21–23). Although the study by Skougaard et al reported a change in the reliability of patient DAS28-CRP from an ICC of 0.69 before training to an ICC of 0.75 after training, the authors commented that this may not represent a significant difference (23). Meanwhile, the studies by Levy and Grainger reported more sizable improvements in reliability of both patient SJCs and TJCs following training, which were interpreted as clinically significant changes (21,22). This improvement was more dramatic in SJCs compared with TJCs, likely because of the pre-training reliability of SJCs being much lower than that of TJCs.

Given the limited number of studies that met the inclusion criteria for this review and the significant overlap between different

types of training present in the included studies, we were unable to elucidate whether significant differences between the effectiveness of different training modalities existed (ie, in-person training vs videos).

**Feedback on training sessions.** Several studies in this review also collected data on qualitative feedback provided by patients on their respective training sessions. In general, patients responded positively to the training they received, ranging from the in-person discussions to interactive ultrasound sessions (19,21). The study by Grainger et al, which used an open discussion format for patient feedback, found that clearly defining terminology (eg, “tender” used to mean “sore”) was important to patients, whereas having detailed information about the anatomy of joints was less important (21).

## DISCUSSION

The results of this systematic review suggest that structured training directed for patients to perform self-joint counts may improve the reliability of these counts as compared with joint counts by clinicians, although more research is likely needed to better quantify and confirm these findings. Four of five included studies showed quantitative differences in the reliability of patient self-joint counts between those with and without training, but the significance of these differences is difficult to interpret given the heterogeneity of statistical measures used for reporting this outcome. Several studies noted improvements with patient training in both SJCs and TJCs, whereas the RCT by Cheung et al reported improved reliability for SJCs but not TJCs.

Given that prior studies have found a higher degree of patient–clinician inter-rater reliability for TJs than for SJs, the accurate detection of swollen joints should be an important objective to consider when developing training interventions for patient self–joint counts (8,9). Apart from providing general instruction on how to perform joint count assessments, the training sessions described in the studies also shared other thematic similarities, including the differentiation between bony enlargement and soft-tissue swelling (20–22). Providing information on what constitutes active synovitis may be one method in targeting improvement for the reliability of patient SJs.

There was some evidence of the longitudinal effects of training within our included studies. Three studies reassessed reliability of patient self–joint counts at follow-up visits ranging from 2 to 6 months after the initial visit when the training intervention was administered (19,20,23). Overall, patients who were reassessed at follow-up months after the initial training still demonstrated improvements in the reliability of their joint counts, with similar reliability to when they were measured immediately after training. The study by Skougaard et al looked at the effect of repeated training and did not find a significant impact of repeated training on the reliability of patient self–joint counts (23). Further research would be needed to fully examine the effectiveness of training over time, including the potential for decay of training.

Disease severity has been cited as a potential factor in determining the reliability of patient self–joint counts, with some studies suggesting higher reliability for patients with low disease activity or in remission (24). Overall, patients included in this review exhibited low to moderate disease activity (mean DAS28 = 3.27), and the study by Radner et al specifically looked at patients with RA at or near remission as part of their inclusion criteria (20). The role of training may hold particular importance in this population, as remote monitoring could be a useful strategy in managing patients with RA in remission who typically have less frequent follow-up visits (25,26). Training for self-performed joint counts could also be of value for patients requiring close self-monitoring at home, including those with high disease activity or in whom medications are being adjusted. Prior research has shown that patient self-reported joint counts could be useful in the early detection of disease flares, highlighting the potential for joint count training in a population of patients with active disease (27).

There were several limitations with the available data from studies included in this review. Firstly, most of the studies were conducted in clinic settings at academic centers, potentially affecting the generalizability of the findings to nonacademic centers, which may have comparatively limited access to resources needed to administer the training interventions. Secondly, not all studies reported data on baseline patient characteristics (including education level, socioeconomic status, or ethnicity), which could impact the reliability of patient self–joint counts. Factors such as socioeconomic status have been shown to affect other outcomes in RA, including disease activity and access to health

care resources (28,29). Health literacy, which may be linked to these other demographic factors, has also been shown to be independently associated with functional status in patients with RA (30). Thirdly, inter-rater variability of joint counts between clinicians has been well documented previously (31), and this effect could have potentially affected the accuracy of results in studies in which joint counts by more than one clinician were analyzed (19–21).

An additional limitation was the variability between training interventions used across the studies. Although there were several common themes to the training as noted previously (eg, in-person sessions directed by a rheumatologist, instructional videos, etc), there was lack of a generalized protocol for any of these training modalities, including standardization of factors such as the length, content, and structure of the training. Although this potentially reduces the generalizability of the findings, as they may be subject to differences in training execution, it should also be viewed as an area for future research. It could be beneficial to explore a wide range of different types of training in future research, including direct comparisons between training protocols, to determine which ones show greater efficacy in improving patient self–joint counts. Furthermore, given the heterogeneity of the RA population, it may be possible that certain training methodologies are more effective for certain groups of patients than for others, which would signal the importance of developing training programs specific to individual patient needs. For instance, live feedback using Doppler ultrasound in the training intervention by Cheung et al provided a visual component, which could be beneficial for patients with lower health literacy (19).

The data reported on the primary outcome of patient self–joint count reliability were also heterogeneous across the studies. Although all the studies included physician joint counts as a comparator, several different statistical measures were used, including ICC, Pearson and Spearman correlations, and PABAK. Pearson correlations in particular measure relationship but not reliability, making the ICC a more appropriate statistical measure to use in this setting (32). The heterogeneity in methods of reporting poses a challenge in interpreting the significance of the observed differences between trained and untrained groups. Ultimately, the differences in the study designs precluded the possibility of performing meta-analysis on these results.

A potential source of confounding in several of the studies was the possibility that undergoing joint count assessment by a clinician could independently affect the reliability of patient self–joint counts even without training interventions (21–23). If patients were aware of and remembered which joints were swollen during the clinician’s assessment, an artificial increase in their postintervention joint count reliability would be observed. This effect would likely be more significant for studies without a significant time interval for longitudinal follow-up (ie, a shorter time interval for patients to remember which joints were swollen). To explore the potential for this effect, we examined a previous study by Wong et al that measured patient self-assessed joint counts before and



1 day after a clinician joint count assessment, without other specific training in between (24). Their study did not find a significant difference in inter-rater reliability between patient and clinician joint counts before and after the clinician assessment, which would suggest a relatively minor impact from this confounder on the results of our review. In addition, most of the studies had patients perform their self-joint counts before the clinician-performed joint counts, further minimizing the potential for confounding.

Overall, this systematic review identified five studies with data on the impact of patient training on reliability of patient self-joint counts in RA. The results of these studies suggest that training patients to perform their own joint counts may improve the inter-rater reliability of their joint counts with those of clinicians, both immediately after training and at longitudinal follow-up up to 6 months after training. However, more research is needed to quantify the size of this effect, provide direct comparison between different modalities of training (eg, in-person sessions, instructional videos, and Doppler ultrasound assessment), and determine the effect of patient factors (eg, socioeconomic status and disease severity) on the efficacy of training. Developing structured training programs teaching patients to perform their own joint counts may have significant utility for the future of RA management as telehealth and remote disease monitoring develop increasingly important roles in rheumatology.

## AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Drs. Barber and Tam had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

**Study conception and design.** Tam, Hazlewood, Barber.

**Acquisition of data.** Tam, Barber.

**Analysis and interpretation of data.** Tam, Hazlewood, Barber.

## REFERENCES

- Smolen JS, Breedveld FC, Eberl G, Jones I, Leeming M, Wylie GL, et al. Validity and reliability of the twenty-eight-joint count for the assessment of rheumatoid arthritis activity. *Arthritis Rheum* 1995;38:38–43.
- Wells G, Becker JC, Teng J, Dougados M, Schiff M, Smolen J, et al. Validation of the 28-joint Disease Activity Score (DAS28) and European League Against Rheumatism response criteria based on C-reactive protein against disease progression in patients with rheumatoid arthritis, and comparison with the DAS28 based on erythrocyte sedimentation rate. *Ann Rheum Dis* 2009;68:954–60.
- Osborne RH, Wilson T, Lorig KR, McColl GJ. Does self-management lead to sustainable health benefits in people with arthritis? A 2-year transition study of 452 Australians. *J Rheumatol* 2007;34:1112–7.
- Escalante A. What do self-administered joint counts tell us about patients with rheumatoid arthritis? *Arthritis Care Res* 1998;11(4):280–290. doi: 10.1002/art.1790110409
- Chew LC, Xin X, Yang H, Thumboo J. An evaluation of the Virtual Monitoring Clinic, a novel nurse-led service for monitoring patients with stable rheumatoid arthritis. *Int J Rheum Dis* 2019;22:619–25.
- Ferucci ED, Holck P, Day GM, Choromanski TL, Freeman SL. Factors associated with use of telemedicine for follow-up of rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2020;72:1404–9.
- Cheung PP, Gossec L, Mak A, March L. Reliability of joint count assessment in rheumatoid arthritis: a systematic literature review. *Semin Arthritis Rheum* 2014;43:721–9.
- Gordon JK, Girish G, Berrocal VJ, Zhang M, Hatzis C, Assassi S, et al. Reliability and validity of the tender and swollen joint counts and the modified rodnan skin score in early diffuse cutaneous systemic sclerosis: analysis from the Prospective Registry of Early Systemic Sclerosis Cohort. *J Rheumatol* 2017;44:791–4.
- 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 (Hoboken)* 2010;62:1112–9.
- Cheung PP, Dougados M, Andre V, Balandraud N, Chalès G, Chary-Valckenaere I, et al. The learning curve of nurses for the assessment of swollen and tender joints in rheumatoid arthritis. *Joint Bone Spine* 2014;81:154–9.
- Grunke M, Witt MN, Ronneberger M, Schnez A, Laubender RP, Engelbrecht M, et al. Use of the 28-joint count yields significantly higher concordance between different examiners than the 66/68-joint count. *J Rheumatol* 2012;39:1334–40.
- 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.
- Mathews AL, Coleska A, Burns PB, Chung KC. Evolution of patient decision-making regarding medical treatment of rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2016;68:318–24.
- Riemsma RP, Kirwan JR, Taal E, Rasker JJ. Patient education for adults with rheumatoid arthritis. *Cochrane Database Syst Rev* 2003:CD003688.
- Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009;339:b2700.
- Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, et al. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. *Syst Rev* 2012;1:2.
- Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:14898.
- Sterne J, Hernán M, McAleenan A, Reeves B, Higgins J. Chapter 25: assessing risk of bias in a non-randomized study. In: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al, editors. *Cochrane handbook for systematic reviews of interventions*. Version 6.2.2021. 2021.
- Cheung PP, Lahiri M, Teng GG, Lim AY, Lau TC, Lateef A, et al. A randomized controlled trial for improving patient self-assessment of synovitis in rheumatoid arthritis with education by ultrasonography: the RAEUS Study. *Rheumatology (Oxford)* 2015;54:1161–9.
- 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.
- Grainger R, Townsley HR, Stebbings S, Harrison AA, Taylor WJ, Stamp LK. Codevelopment of patient self-examination methods and joint count reporting for rheumatoid arthritis. *ACR Open Rheumatol* 2020;2:705–9.
- Levy G, Cheatham 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.
- Skougaard M, Bliddal H, Christensen R, Ellegaard K, Nielsen SM, Zavada J, et al. Patients with rheumatoid arthritis acquire sustainable skills for home monitoring: a prospective dual-country cohort study (ELECTOR Clinical Trial I). *J Rheumatol* 2020;47:658–67.

24. Wong AL, Wong WK, Harker J, Sterz M, Bulpitt K, Park G, et al. Patient self-report tender and swollen joint counts in early rheumatoid arthritis. *J Rheumatol* 1999;26:2551–61.
25. Van Riel P, Alten R, Combe B, Abdulganieva D, Bousquet P, Courtenay M, et al. Improving inflammatory arthritis management through tighter monitoring of patients and the use of innovative electronic tools. *RMD Open* 2016;2:e000302.
26. National Guideline Centre (UK). Rheumatoid arthritis in adults: diagnosis and management. London: National Institute for Health and Care Excellence (UK); 2018.
27. Kuettel D, Terslev L, Weber U, Østergaard M, Primdahl J, Petersen R, et al. Flares in rheumatoid arthritis: do patient-reported swollen and tender joints match clinical and ultrasonography findings? *Rheumatology* 2020;59(1):129–136. doi:<http://dx.doi.org/10.1093/rheumatology/kez231>
28. Jacobi CE, Mol GD, Boshuizen HC, Rupp I, Dinant HJ, van den Bos GA. Impact of socioeconomic status on the course of rheumatoid arthritis and on related use of health care services. *Arthritis Rheum* 2003;49:567–73.
29. Hernández-García C, Vargas E, Abásolo L, Lajas C, Bellajdell B, Morado IC, et al. Lag time between onset of symptoms and access to rheumatology care and DMARD therapy in a cohort of patients with rheumatoid arthritis. *J Rheumatol* 2000;27:2323–8.
30. Caplan L, Wolfe F, Michaud K, Quinzanos I, Hirsh JM. Strong association of health literacy with functional status among rheumatoid arthritis patients: a cross-sectional study. *Arthritis Care Res (Hoboken)* 2014;66:508–14.
31. Scott DL, Choy EH, Greeves A, Isenberg D, Kassiror D, Rankin E, et al. Standardising joint assessment in rheumatoid arthritis. *Clin Rheumatol* 1996;15:579–82.
32. Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–10.