TITLE PAGE

Title: Comparing psoriatic arthritis low-field magnetic resonance imaging, ultrasound and clinical outcomes: data from the TICOPA trial

Authors: Philip S Helliwell (0000-0002-4155-9105)[1], Laura C Coates (0000-0002-4756-663X)[2], Ne Siang Chew[3], Giovanni Lettieri[1,3], Anna R Moverley[4], Jane E Freeston (0000-0003-0358-338X)[1], Jackie Nam (0000-0003-4944-7922)[1], Robin Waxman[1], Paul Emery (0000-0002-7429-8482)[1], Philip G Conaghan (0000-0002-3478-5665)[1]

Key Indexing Terms: Psoriatic arthritis, Magnetic Resonance Imaging, Ultrasound, Joint inflammation, Joint erosions

Funding

The TICOPA study was supported by Arthritis Research UK (grant no 18825) and Pfizer. This study presents independent research supported by the National Institute for Health Research (NIHR) Leeds Biomedical Research Centre (BRC). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Departments and institutions:

[1] Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, and NIHR Leeds Biomedical Research Centre, Leeds Teaching Hospitals NHS Trust

[2] Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford UK

[3] Bradford Teaching Hospitals Foundation Trust, Bradford, UK

[4] York Teaching Hospitals NHS Foundation Trust, Scarborough General Hospital, Department of Rheumatology, Woodlands Drive, Scarborough, YO12 6QL

Academic degrees

PS Helliwell, MA, PhD, DM, FRCP, Professor of Clinical Rheumatology

Laura C Coates, MBChB, MRCP, NIHR Clinician Scientist

Ne Siang Chew, MRCP, FRCR, Consultant Musculoskeletal Radiologist

Giovanni Lettieri, MD, Hon Consultant Radiologist, Clinical Research Fellow

Anna R Moverley, MBBS, BSc, FRCP, Consultant Rheumatologist

Jane E Freeston,MA, MD, MRCP, Consultant Rheumatologist and Honorary Clinical Associate Professor

Jackie Nam, MBBCh FCP, PhD, Consultant in Rheumatology and Honorary Lecturer

Robin Waxman, MPH, Research coordinator

Paul Emery, MA, MD, FRCP, FMedSci, Arthritis Research UK Professor of Rheumatology

Philip G Conaghan, MBBS, PhD, FRCP, FRACP, Professor of Musculoskeletal Medicine

Corresponding author:

Dr Philip Helliwell

LIRMM

Chapel Allerton Hospital

Chapeltown Road

Leeds LS7 4SA

UK

Tel: +44 (0)113 392 3064 Fax: +44 (0)113 392 4991

Email: p.helliwell@leeds.ac.uk

Running title: Multi-modality imaging in TICOPA

Word Count: 2769

Number of Tables: 3

Number of supplementary tables: 0

Number of Figures: 0

Number of appendices: 0

Abstract

Objective

The Tight Control of Psoriatic arthritis (TICOPA, ISCRCTN30147736) trial, compared standard care (StdC) and tight control (TC) in early PsA, demonstrating better outcomes for TC. This sub-study evaluated the performance metrics of modern imaging outcomes and compared them to the clinical data.

Methods

Non-contrast 0.2TMRI (single hand) was assessed using the OMERACT PsAMRIS with an additional global inflammation score. Ultrasound (US, same hand) was scored for grey scale, power Doppler and erosions at the metacarpophalangeal (MCP) and proximal interphalangeal (PIP) joints and scores summated.

Results

78 patients had paired (baseline and 48 weeks) US data and 63 paired MRI data; 50 had matched clinical, MR and US data. Significant within-group changes were seen for the inflammatory PsAMRIS components at MCP level: MRI global inflammation (Median difference (range), Standardised Response Mean, SRM): 3.25 (- 5.0 – 12.0) 0.68, 1.0 (-4.5 – 17.5), 0.45 for TC and StdC respectively. Similar within group differences were obtained for US: 1.0 (-13.0 – 23.0), 0.45, 3.0 (-6.0 – 21.0), 0.77 for TC and StdC respectively. No differences were seen between treatment groups. Significant correlations were found between baseline and change MRI and US scores. A significant correlation was found between baseline PsA disease activity scores and MRI global inflammation scores (Spearman's rho for MCP, PIP 0.46, 0.63 respectively). No differences in erosion progression were observed.

Conclusion

The PsAMRIS and US inflammation scores demonstrated good responsiveness. No between group differences were demonstrated but this sub-study was likely under-powered to determine differences between the two treatment strategies.

Introduction

The emphasis placed on treating inflammatory arthritis as early as possible to minimise damage and functional disability has been shown to be effective in rheumatoid arthritis (RA)[1] and the concept has been extended to other inflammatory arthritides such as psoriatic arthritis (PsA). The Tight Control of inflammation in Psoriatic arthritis (TICOPA) trial targeted early, treatment naïve patients and demonstrated improved clinical outcomes above usual care, but was unable to demonstrate an advantage in terms of radiographic progression in hands and feet [2].

Modern imaging modalities such as magnetic resonance imaging (MRI) or ultrasound (US) provide sensitive tools to explore both objective inflammation and damage responses, though there are extremely few PsA studies using these modalities [3]. It is also unclear, given the known patterns of PsA joint involvement, whether imaging a single hand (as is typically done in RA trials using MRI) will provide a responsive tool.

The aim of this study was therefore to describe and compare the performance metrics of commonly-used MRI and US scores in an imaging sub-study of the TICOPA study, and to compare these imaging outcomes with the clinical data obtained in this randomised trial.

Patients and Methods

The full trial protocol and clinical results of the TICOPA study have been previously reported (ISCRCTN30147736) [2, 4]. In brief this randomised, controlled, parallel group, open label, multi-centre clinical trial recruited people with early (less than 2 years), treatment naive PsA. The trial had ethical approval from North East York Ethics Committee (14/NE/1090) and all participants gave written informed consent.

The primary objective of the main trial was to compare tight control (TC) with standard care (StdC), using minimal disease activity (MDA[5]) as the treatment target. Participants received either TC or StdC for a period of 48 weeks. Participants randomised to TC were seen every 4 weeks by the study physician and treated according to a predefined treatment protocol. Participants randomised to the StdC arm were treated in a general rheumatology outpatient clinic supervised by a consultant rheumatologist. These patients were generally reviewed every 12 weeks but were seen more often if clinically indicated, with no formal measures of disease activity used in clinical decision making. A blinded assessor collected clinical assessments and patient reported outcomes every 12 weeks. Disease activity was measured using the Psoriatic arthritis Disease Activity Score (PASDAS) which assesses patient and physician global assessment of disease, tender and swollen joint counts, dactylitis and enthesitis, CRP and the physical summary subscale of the short form 36 health related quality of life [6].

Magnetic Resonance Imaging

At the Leeds site, consenting patients were invited to participate in both MRI and US sub-studies, imaging the hand (the most affected hand, or the dominant hand if both were asymptomatic). Imaging was performed with both a non-contrast low field extremity MRI (0.2T C-scan, Esaote, Genova, Italy) and an US scan at baseline and 48 weeks.

MRI imaging

For the MRI scan the imaging sequences and details of scoring are as follow:

Scout. Whole hand FOV 140*140 TR 140ms. Matrix 192*128

STIR coronal. TR 2620ms. 160*160 matrix 192*144. 3 slices. 24 echoes.

STIR sagittal. TR 2840ms. 190*190. 192*144. 4 slices. 25 echoes.

T1 3D coronal. T3D T1. TR 35ms. 140*140 80 matrix 192*160 72. 2 slices 88 echoes.

Images were scored for the second to fifth fingers at each level in the hand (metacarpophalangeal joint, proximal interphalangeal joint, distal interphalangeal joint) for the following features based on the OMERACT PsAMRIS score [7].:

Synovitis: Grading scale: 0 is normal, while 1–3 is mild, moderate, severe, by thirds of the maximum potential volume of tissue. Score range at each level for eachfinger, 0 – 36.

Tenosynovitis: Grading scale: the maximal thickness of signal as follows: 0: none; 1: < 1/2 tendon thickness; 2: ≥ 1/2 and < 1 tendon thickness; 3: ≥ 1 tendon thickness. Score range at each level for each finger, 0 – 36.

Periarticular inflammation (distal and proximal): Grading scale, 0 absent, 1 present on both dorsal and volar aspects. Score range at each level for each finger, 0 – 24.

Bone edema (distal and proximal): Grading scale: the scale is based on the proportion of bone with edema, compared to the "assessed bone volume," judged on all available images: 0: no edema; 1: 1–33% of bone edematous; 2: 34–66%; 3: 67–100% scored on either side of the joint. Score range at each level for each finger, 0 – 72.

Scores for synovitis, tenosynovitis, peri-articular inflammation and bone oedema were summed to give a 'global inflammation' score at each level, for each finger, score range 0 – 168. The aggregate scores at each level were retained in order to examine the responsiveness of 'global inflammation' in very small joints such as the DIPJ, and to determine which joints demonstrated most change with respect to this feature.

Bone erosion (distal and proximal): Grading scale: the scale is 0–10, based on the proportion of eroded bone compared to the "assessed bone volume," judged on all available images: 0:no erosion; 1: 1–10% of bone eroded; 2: 11–20%, and so on. Scored at either side of the joint. Score range at each level for all 4 fingers 0 – 240.

Bone proliferation: Grading scale: 0 absent, 1 present. Score range at each level for all 4 fingers 0 – 12.

The images were read by two independent readers (NC, GL), anonymized to patient demographics, treatment group and time order. Inter-rater reliability for domain scores at each joint level was calculated by intra-class correlation coefficients.

Ultrasound imaging

One of two ultrasonographers (JF and JN) scanned the same hand as the MRI using a Philips HDI 5000 (Best, The Netherlands) machine employing 12-5 and 15-7 MHz linear transducers and were unaware of the clinical examination findings. The inter-rater agreement between these assessors for this group of patients has been previously reported [8]. PD was assessed using a pulse repetition frequency of 750 Hz and medium wall filter and gain was adjusted until background signal was removed. Each joint was scanned in both longitudinal and transverse planes from the dorsal aspect. For the small joints of the hand the second to fifth metacarpophalangeal (MCP) joints, and the second to fifth proximal inter-phalangeal (PIP) joints were examined. GS and PD were scored separately on a 0-3 semi-quantitative scale for each joint imaged. A GS score of ≥2 and/or a PD score ≥1 were used to identify US active joints. The GS and PD scores were summated to give an overall score for 'inflammation' (total possible score of 48) [9]. Erosions were

defined as a definite cortical break seen in two planes and scored as present or absent at the joint level, so the maximum score for erosions per hand was 8.

Statistical analysis

The original TICOPA study was appropriately powered for its clinical outcome, but no formal power calculation was made for this sub-study. Only matched (baseline and follow up) MRI and US data, and combined MRI, US and clinical data were used. There was no data imputation. The clinical composite outcome (the Psoriatic arthritis Disease Activity Index, PASDAS), was derived as previously described [6]. Significance was assumed at a level of 5%; no correction was made for multiple comparisons. Inter-rater reliability for aggregate MR scores was assessed using the intra-class correlation coefficient. The magnitude of MR parameter response was compared using the standardised response mean (SRM), calculated as the mean difference between time points divided by the standard deviation of the difference [10]. Statistical testing was carried out using SPSS v21.

Results

In the TICOPA study 206 patients were recruited and of these 85 entered imaging sub-studies. Clinical characteristics of the patients in this study were: Male/Female 40/45; mean age 45.1 years; mean tender joint count 11.7; mean swollen joint count 7.3; mean skin score (Psoriasis Area and Severity Score: PASI) 2.7; mean CRP (mg/dl) 23.9. The majority of patients (n=59, 69%) presented with polyarticular disease (≥5 joints involved). Baseline disease activity was high (mean PASDAS score 5.1 and significant within group changes in clinical outcomes were seen (Tight control group, mean change in PASDAS score 2.2, $p < 0.0001$; Standard care: 1.1, $p = 0.03$) but between group differences were not significant ($F = 3.6$, $p = 0.06$). In the imaging sub-studies paired observations (baseline and 48 weeks) were available for

61 participants for the MRI and 78 participants for the US groups, with complete paired MRI, US and clinical data for 50 participants. The demographics of each of these groups (MRI, US and matched) were very similar (see online Supplementary table).

MRI results

Inter-observer intra-class correlation for paired observations varied by feature: ICC scores (95% confidence intervals) for synovitis 0.85 (0.74 – 0.91), flexor tenosynovitis 0.73 (0.54 – 0.85), periarticular inflammation 0.82 (0.69 – .89), bone oedema 0.76 (0.59 – 0.86), bone erosion 0.86 (0.76 – 0.92) and bone proliferation 0.25 (0.30 – 0.57). The data for both readers was combined and expressed as the mean. The results for the MRI scores, for each joint level, at each time point, and each treatment group are given in Table 1. At the MCP joint a significant difference between baseline and 48 weeks was seen in the TC arm for synovitis, flexor tenosynovitis, periarticular inflammation, bone oedema and global inflammation but not for bone erosion and bone proliferation Comparable changes were seen in the StdC arm of the study. At the PIP joint the changes were similar with the exception of bone oedema. At the distal inter-phalangeal joint, the differences were less pronounced, with only flexor tenosynovitis and global inflammation for both arms of the study significantly different between baseline and follow up. SRMs varied from 0.70 (periarticular inflammation at the MCP joint in the TC arm), to – 0.39 (erosions at the DIP joint in the TC arm) and were generally larger for the TC arm. Analysis of covariance for individual components of the score (synovitis, tenosynovitis, peri-articular inflammation, bone oedema, global inflammation, bone erosion and bone proliferation) at each joint level, did not show any difference between the two treatment groups at 48 weeks for any of the comparisons (statistics not shown).

Ultrasound results

The results for the US examination at each time point, and each treatment group, for McP and PiP joints, are given in Table 2. For approximately two thirds of cases inflammation (synovitis) was represented by a GS score of ≥ 2. A significant difference was seen for the inflammation score between baseline and 48 weeks for both treatment groups. However, there was no difference in scores between treatment groups at 48 weeks ($F = 0.38$, $p = 0.75$). For erosions, scores were low (median of 0 for both groups at baseline and 48 weeks) and no significant differences within or between groups were seen at joint or aggregate level.

Relationship between MRI and US data and clinical outcomes

MRI and US scores at baseline, and their change scores, were highly significantly correlated (Table 3). A significant correlation was found between baseline PASDAS scores and MRI global inflammation scores from the MCP and PIP joint regions (Spearman's rho for MCP, PIP and DIP joint inflammation and PASDAS were 0.46, 0.63 and 0.35 respectively). However, a non-significant positive correlation was found between baseline US inflammation and baseline PASDAS score. Non-significant positive correlations were found between the change in PASDAS score from baseline to week 48 and the change in global inflammation MRI score over the same time period. A significant positive correlation was found between the change in PASDAS score from baseline to week 48 and the change in US 'inflammation' score (rho = 0.37, p = 0.02).

Discussion

In this sub-study of the TICOPA trial, the individual low-field MRI inflammation scores reflected a modest degree of inflammation but consistent with another report using the PsAMRIS scoring method in PsA [11]. Although a within-group

improvement in the inflammation components of the PsAMRIS score was demonstrated for the TC group, the improvements were modest overall, as reflected by the standardised response mean, but larger than those seen in the StdC group. However, the 'whole body' clinical improvements were reflected in the single hand MR improvement scores, thus indicating construct validity of the change scores. It is also worth noting that the MRI scans in this analysis were low-field scans where there are limitations to the images, such as low resolution and difficulty visualising the distal inter-phalangeal joints, and there was lack of contrast agent to help define inflammation. The relatively oligoarticular nature of PsA, where individual joints may be affected in an asymmetrical distribution, compared to RA which is more symmetrical and polyarticular, should also be recognised [12]. In this situation imaging may show large changes in individual joints but, collectively, over the whole hand, the magnitude of change may be smaller when compared to polyarticular disease.

US inflammation scores improved in both treatment groups, and there was a significant association between baseline and change in US score and the equivalent clinical scores. In this study, therefore, both US and MRI were responsive, aligned with baseline clinical scores, and in the case of US, aligned with change in clinical scores. It should be noted that MRI and US assessed slightly different joint sets.

In all the imaging/clinical comparisons made in this study, it must be remembered that the imaging focussed on the peripheral joints of a single hand, whereas the clinical score is more comprehensive, with both patient reported measures, joint counts, measures of dactylitis and enthesitis and an acute phase reactant. Although the PASDAS response has been shown to correlate with radiographic progression scores [13], in this study the use of treatments without proven disease modifying

abilities, such as methotrexate, could lower the effect size and interfere with attempts to demonstrate relationships between clinical course and imaging changes, and the TICOPA study was not powered to demonstrate this. It is also worth noting that the design of the TICOPA study does not allow direct comparison of drug efficacy between conventional synthetic DMARDs and biologic drugs.

The TICOPA study demonstrated improved clinical outcomes using a treat-to-target approach in early PsA but there were no differences in X-ray progression between groups. In the current analysis, a sub-study of TICOPA, there were similar within-group improvements in clinical outcomes but a significant change in most of the inflammatory components of an extremity MRI score in the tight control group over the 48-week study, and a significant improvement in US inflammation scores in both groups. A significant difference between the treatment groups for the change in clinical scores was not found in this sub-study, and the imaging modalities also did not demonstrate a between-group difference. It must be remembered that both groups received active treatment for 48 weeks, there being no placebo group in this study. Good correlation between baseline and change scores for MRI and US was found, and good correlation between baseline MRI imaging and clinical scores. Overall, few erosions were seen and there was little progression over 48 weeks in either group.

MRI assesses a greater range of pathologies compared to US yet more recent US machines can now give much better detail compared to those used in this paper. Future studies of this kind using US could include an assessment of enthesitis and tenosynovitis, which may improve responsiveness of a more ‘global’ inflammation construct. In this context, dactylitis reflects many of the pathologies seen in PsA,

including synovitis, enthesitis and tenosynovitis, but reliable US assessments of dactylitis have not yet been developed [14].

The limitations of this study relate to both the modalities and the clinical context. This substudy was not powered to show a significant difference in imaging outcomes between two active therapies. Secondly, as noted above, the clinical composite used relates to total disease burden yet the imaging was confined to a single hand. It may be that more extensive joint assessment, such as obtained with total body MR, are more closely related to clinical scores such as the PASDAS. Thirdly, the MR technique, being a peripheral scanner without the use of contrast, will have limited ability to demonstrate improvement in inflammation in any tissue.

In conclusion, the imaging substudy of TICOPA reported in this paper provides further validation for the use of both imaging modalities as outcome measures in this disease. The somewhat sporadic joint involvement of PsA, where only a few individual joints may be affected, makes aggregate imaging scores less responsive to change and future imaging studies should perhaps focus on polyarticular disease inclusion, or one manifestation, such as dactylitis, to demonstrate within and between group changes in response to treatment.

Acknowledgement

We thank Mr Brian Witham and Dr Jill Halstead for technical assistance with the MR scans.

References

1. Nell VP, Machold KP, Eberl G, Stamm TA, Uffmann M, Smolen JS. Benefit of very early referral and very early therapy with disease-modifying anti-rheumatic drugs in patients with early rheumatoid arthritis. Rheumatology 2004;43:906-914.
2. Coates LC, Moverley AR, McParland L, Brown S, Navarro-Coy N, O'Dwyer J, et al. Effect of tight control of inflammation in early psoriatic arthritis (TICOPA): a UK multicentre, open-label, randomised controlled trial. Lancet 2015;386:2489-2498.
3. Felbo S, Terslev L, Ostergaard M. Imaging in peripheral and axial psoriatic arthritis: contributions to diagnosis, follow-up, prognosis and knowledge of pathogenesis. Clin Exp Rheumatol 2018;36S:24-34.
4. Coates LC, Navarro-Coy N, Brown SR, Brown S, McParland L, Collier H, et al. The TICOPA protocol (TIght COntrol of Psoriatic Arthritis): a randomised controlled trial to compare intensive management versus standard care in early psoriatic arthritis. BMC Musculoskelet Disord 2013;14:101.
5. Coates LC, Fransen J, Helliwell PS. Defining minimal disease activity in psoriatic arthritis: a proposed objective target for treatment. Ann Rheum Dis 2010;69:48-53.
6. Coates LC, Mahmood F, Emery P, Conaghan PG, Helliwell P. The dynamics of response as measured by multiple composite outcome tools in the TIght COntrol of inflammation in early Psoriatic Arthritis (TICOPA) trial. Ann Rheum Dis 2017;76:1688-1692.
7. Østergaard M, McQueen F, Wiell C, Bird P, Boyesen P, Ejbjerg B, et al. The OMERACT Psoriatic Arthritis Magnetic Resonance Imaging Scoring System (PsAMRIS): Definitions of Key Pathologies, Suggested MRI Sequences, and Preliminary Scoring System for PsA Hands. J Rheumatol 2009;36:1816-1824.
8. Freeston JE, Coates LC, Nam JL, Moverley AR, Hensor EMA, Wakefield RJ, et al. Is there subclinical synovitis in early psoriatic arthritis? A clinical comparison with grey scale and power Doppler ultrasound. Arthrit Care Res 2014;66:432-439.
9. Bruyn GA, Iagnocco A, Naredo E, Balint PV, Gutierrez M, Hammer HB, et al. OMERACT Definitions for Ultrasonographic Pathology and Elementary Lesions Of Rheumatic Disorders Fifteen Years On. J Rheumatol 2019;46S:1-6.
10. Liang MH, Fossel AH, Larson MG. Comparison of five health status instruments for orthopaedic evaluation. Med Care 1990;28:632-642.
11. Strube H, Becker-Gaab C, Saam T, Reiser M, Schewe S, Schulze-Koops H, et al. Feasibility and reproducibility of the PsAMRIS-H score for psoriatic arthritis in low-field-strength dedicated extremity magnetic resonance imaging. Scan J Rheumatol 2013;42:379–382.
12. Helliwell PS, Helthen J, Sokoll K, Green MJ, Marchesoni A, Lubrano E, et al. Joint symmetry in early and late rheumatoid and psoriatic arthritis: comparison with a mathematical model. Arthritis Rheum 2000;43:865-871.
13. Helliwell PS, Kavanaugh A. Radiographic progression is less in psoriatic arthritis achieving a good response to treatment: data using newer composite indices of disease activity. Arthritis Care Res 2018;70:797-800.
14. Bakewell C, Olivieri I, Aydin S, Dejaco C, Ikeda K, Gutierrez M, et al. Ultrasound and Magnetic Resonance Imaging in the Evaluation of Psoriatic Dactylitis: Status and Perspectives. J Rheumatol 2013;40:1951-1957.

Table 1. MRI scores for each PsAMRIS feature at each joint level, for each treatment group at each time point.

(a) Metacarpophalangeal joint

Score	Tight control n=31		SRM	z	p	Standard Care n=30		SRM	z	p
	Baseline	48 weeks				Baseline	48 weeks			
Synovitis score	1.5 (0 – 7.5)	0.75 (0 – 2.5)	0.55	2.9	0.003	1.5 (0 – 6.0)	1.0 (0 – 4.5)	0.44	2.1	0.037
Flexor tenosynovitis score	3.0 (0 – 6.0)	1.5 (0 – 4.5)	0.39	2.3	0.020	3.0 (0 – 4.5)	2.25 (0 – 5.0)	0.29	1.3	ns
Periarticular inflammation score	0.5 (0 – 5.5)	0 (0 – 3.5)	0.70	3.2	0.001	0.5 (0 – 5.0)	0 (0 – 2.0)	0.48	2.4	0.016
Bone oedema score	0 (0 – 8.5)	0 (0 – 3.5)	0.35	2.4	0.016	0 (0 – 9.5)	0 (0 – 0)	0.20	1.3	ns
Global inflammation score	6.0 (1.0 – 22.5)	2.5 (0 – 10.5)	0.68	3.3	0.001	5.5 (0 – 20.0)	3.5 (0 – 8.5)	0.45	2.1	0.04
Bone erosion score	0 (0 – 12.0)	0 (0 – 11.5)	0.02	0.7	ns	0 (0 – 4.5)	0 (0 – 6.5)	0.30	0.7	ns
Bone proliferation score	0 (0 – 2.5)	0 (0 – 0.5)	0.19	1.4	ns	0 (0 – 0.5)	0 (0 – 0.5)	-0.23	1.0	ns

Figures are median (range). Z Wilcoxon paired ranks test statistic. SRM: standardised response mean.

(b) Proximal inter-phalangeal joint

Score	Tight control n=31		SRM	z	p	Standard Care n=30		SRM	z	p
	Baseline	48 weeks				Baseline	48 weeks			
Synovitis score	1.5 (0 – 8.5)	0.5 (0 – 8.0)	0.61	2.8	0.006	1.0 (0 – 5.0)	0.5 (0 – 4.0)	0.29	0.9	ns
Flexor tenosynovitis score	3.0 (0 – 5.5)	2.0 (0 – 6.0)	0.53	2.5	0.014	3.0 (0 – 5.0)	2.25 (0 – 4.5)	0.29	1.3	ns
Periarticular inflammation score	1.0 (0 – 7.5)	0 (0 – 4.0)	0.68	3.4	0.001	0.5 (0 – 5.0)	0 (0 – 1.5)	0.67	2.9	0.004
Bone oedema score	0 (0 – 10.0)	0 (0 – 10.0)	-0.05	0.2	ns	0 (0 – 6.5)	0 (0 – 4.0)	0.04	0.4	ns
Global inflammation score	4.75 (1.5 – 24.0)	3.5 (0 – 27.0)	0.55	2.5	0.011	4.0 (0 – 16.0)	3.0 (0 – 10.0)	0.32	1.7	ns
Bone erosion score	0 (0 – 6.0)	0 (0 – 15.0)	0.04	0.9	ns	0 (0 – 3.0)	0 (0 – 1.5)	-0.19	0.40	ns
Bone proliferation score	0 (0 – 2.5)	0 (0 – 3.0)	0.26	0.7	ns	0 (0 – 2.0)	0 (0 – 2.5)	0.05	0.8	ns

Figures are median (range). Z Wilcoxon paired ranks test statistic. SRM: standardised response mean.

(c) Distal inter-phalangeal joint

Score	Tight control n=31		SRM	z	P	Standard Care n=30		SRM	z	p
	Baseline	48 weeks				Baseline	48 weeks			
Synovitis score	0.25 (0 – 3.0)	0 (0 – 1.5)	0.42	1.8	Ns	0 (0 – 3.0)	0 (0 – 3.0)	0.09	0.6	ns
Flexor tenosynovitis score	1.5 (0 – 3.5)	0 (0 – 4.0)	0.70	2.4	0.015	1.0 (0 – 3.5)	0 (0 – 2.5)	0.58	2.0	0.04
Periarticular inflammation score	0 (0 – 7.5)	0 (0 – 2.0)	0.27	1.2	Ns	0 (0 – 2.0)	0 (0 – 0)	0.40	1.6	ns
Bone oedema score	0 (0 – 5.5)	0 (0 – 2.0)	0.14	0.5	Ns	0 (0 – 1.5)	0 (0 – 0)	0.21	1.0	ns
Global inflammation score	2.25 (0 – 19.0)	0 (0 – 6.5)	0.46	2.0	0.05	2.0 (0 – 7.0)	0 (0 – 5.0)	0.57	2.0	0.042
Bone erosion score	0 (0 – 1.0)	0 (0 – 2.5)	-0.39	1.6	Ns	0 (0 – 0.5)	0 (0 – 0.5)	-0.29	0	ns
Bone proliferation score	0 (0 – 2.5)	0 (0 – 1.5)	0.27	0	Ns	0 (0 – 2.0)	0 (0 – 2.0)	0.04	0.5	ns

Figures are median (range). Z Wilcoxon paired ranks test statistic. SRM: standardised response mean.

Table 2. Ultrasound scores for each group at baseline and 48 weeks. Figures are median (range). For TC n = 39, for StdC n = 39. Figures are median (range). Z Wilcoxon paired ranks test statistic. SRM: standardised response mean.
*GS score of ≥2 and/or a PD score ≥1 aggregated for both metacarpophalangeal (McP) and proximal interphalangeal (PiP) joints
+Erosion score combined for McP and PiP joints

Score	Tight control		SRM	z	p	Standard Care		SRM	z	p
	Baseline	48 weeks				Baseline	48 weeks			
Inflammation										
McP G/S ≥ 2	2 (0 – 11)	0 (0 – 11)	0.26	1.6	ns	2 (0 – 12)	2 (0 – 6)	0.71	3.7	0.0001
McP PD ≥ 1	0 (0 – 4)	0 (0 – 2)	0.41	2.3	0.02	0 (0 – 5)	0 (0 – 3)	0.53	3.0	0.003
PiP G/S ≥ 2	0 (0 – 12)	0 (0 – 9)	0.36	2.2	0.03	2 (0 – 11)	0 (0 – 9)	0.57	3.0	0.002
PiP PD ≥ 1	0 (0 – 6)	0 (0 – 2)	0.38	2.2	0.03	0 (0 – 7)	0 (0 – 1)	0.42	2.7	0.007
Inflammation*	4.5 (0 – 28)	2 (0 – 16)	0.64	2.5	0.01	5 (0 – 20)	2 (0 – 16)	0.95	4.2	0.0001
Erosions										
McP	0 (0 – 1)	0 (0 – 1)	0.07	-0.5	ns	0 (0 – 2)	0 (0 – 1)	0.05	-0.3	ns
PiP	0 (0 – 3)	0 (0 – 4)	0.26	-1.3	ns	0 (0 – 1)	0 (0 – 3)	0.22	-1.1	ns
Erosion score+	0 (0 – 3)	0 (0 – 4)	0.41	-1.2	ns	0 (0 – 3)	0 (0 – 4)	0.33	-0.5	ns

Table 3. Relationship between MRI and US scores at baseline and difference between scores at 48 weeks.

US inflammation at baseline			US difference between baseline and 48 weeks		
MRI global inflammation at baseline	rho	p	MRI global inflammation difference between baseline & 48 weeks	rho	p
McP	0.54	0.002	McP	0.62	0.001
PiP	0.53	0.003	PiP	0.64	0.001
Combined McP/PiP	0.62	0.001	Combined McP/PiP	0.67	0.001

Spearman rho correlation coefficients.