Article:
Allen, JC, Thumboo, J, Lye, WK et al. (3 more authors) (2018) Novel joint selection methods can reduce sample size for RA clinical trials with ultrasound endpoints. International Journal of Rheumatic Diseases, 21 (3). pp. 693-698. ISSN 1756-1841

https://doi.org/10.1111/1756-185X.13185

Title page

Title: Novel joint selection methods can reduce sample size for RA clinical trials with ultrasound endpoints

Authors:

John C Allen Jr, Julian Thumboo, Weng Kit Lye, Philip G Conaghan, Li-Ching Chew, York Kiat Tan

Running title: Novel methods and RA trials sample size

AFFILIATION(S) OF THE AUTHOR(S):

1) J.C. Allen Jr, PhD. Office of Clinical Sciences, Centre for Quantitative Medicine, Duke-NUS Medical School Singapore.

2) J. Thumboo. MBBS (Singapore), MMED (Singapore), MRCP (UK), FAMS (Rheumatology, Singapore), FRCP (Edin). Senior Consultant, Department of Rheumatology and Immunology, Singapore General Hospital, Singapore. Professor, Duke-NUS Medical School, Singapore. Adjunct Professor, Yong Loo Lin School of Medicine, National University of Singapore, Singapore.

3) W.K. Lye, MSc. Office of Clinical Sciences, Centre for Quantitative Medicine, Duke-NUS Medical School Singapore.

4) P.G. Conaghan. MBBS, PhD, FRACP, FRCP. Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, & NIHR Leeds Biomedical Research Centre, UK.

5) L.C. Chew. B. MED. SCI (UK), B.M.B.S. (UK), M.R.C.P. (London, UK), F.A.M.S. (Rheumatology, Singapore), F.R.C.P. (London, UK). Senior Consultant, Department of Rheumatology and Immunology, Singapore General Hospital, Singapore. Adjunct Assistant Professor, Duke-NUS Medical School, Singapore. Clinical Senior Lecturer, Yong Loo Lin School of Medicine, National University of Singapore, Singapore.

6) Y.K. Tan, MBBS (Singapore), MMED (Singapore), MRCP (UK). Consultant, Department of Rheumatology and Immunology, Singapore General Hospital, Singapore. Assistant Professor, Duke-NUS Medical School, Singapore. Clinical Senior Lecturer, Yong Loo Lin School of Medicine, National University

of Singapore, Singapore.

Authors Contribution

YKT led the study and was responsible for the overall design and conduct of the study. JCA and WKL were involved in the statistical analysis. YKT performed the ultrasonography. All authors were involved in interpretation of the results as well as the drafting and preparation of the manuscript. The manuscript has been approved by all authors for publication.

CONFLICT OF INTEREST STATEMENT

The authors declare they have no conflicts of interest.

Correspondence:

York Kiat Tan. Department of Rheumatology and Immunology, Singapore General Hospital, Outram Road, Singapore 169608. Tel: +65 63214028. Email: tan.york.kiat@singhealth.com.sg

Acknowledgements

The authors thank colleagues and staff from the study site for the support and help provided in this study. This study was funded by SingHealth Duke-NUS Academic Medical Centre, through the Medicine Academic Clinical Programme Grant: Nurturing Clinician Scientist Scheme. We would also like to acknowledge the “Estate of Tan Sri Khoo Teck Puat” for the funding support (through the Khoo Clinical Scholars Pilot Award) and the National University of Singapore.

PGC is support in part by the UK National Institute for Health Research (NIHR) through the Leeds Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Main Text

ABSTRACT

Objectives

To determine whether novel methods of selecting joints through (a) ultrasonography (Individualized-Ultrasound (IUS) method), or (b) ultrasonography and clinical examination (Individualized-Composite-Ultrasound (ICUS) method) translate into smaller RA clinical trial sample sizes when compared to existing methods utilizing pre-determined joint sites for ultrasonography.

Methods

Cohen's effect size (ES) was estimated ($\widehat{ES}$) and a 95% CI ($\widehat{ES_L}$, $\widehat{ES_U}$) calculated on a mean change in 3-month total inflammatory score for each method. Corresponding 95% CIs [$n_L(\widehat{ES_U}),\ n_U(\widehat{ES_L})$] were obtained on a post-hoc sample size reflecting the uncertainty in $\widehat{ES}$. Sample size calculations were based on a 1-sample t-test as the patient numbers needed to provide 80% power at $\alpha = 0.05$ to reject a null hypothesis H_0: ES = 0 versus alternative hypotheses H_1: ES = $\widehat{ES}$, ES = $\widehat{ES_L}$ and ES = $\widehat{ES_U}$. We aimed to provide point and interval estimates on projected sample sizes for future studies reflecting the uncertainty in our study $\widehat{ES}s$.

Results

24 treated RA patients were followed up for 3 months. Utilizing the 12-joint approach and existing methods, the post-hoc sample size (95% CI) was 22 (10, 245). Corresponding sample sizes using ICUS and IUS were 11 (7, 40) and 11 (6, 38), respectively. Utilizing a 7-joint approach, the corresponding sample sizes using ICUS and IUS methods were 9 (6, 24) and 11 (6, 35), respectively.

Conclusions

Our pilot study suggests that sample size for RA clinical trials with ultrasound endpoints may be reduced using the novel methods, providing justification for larger studies to confirm these observations.

Key words: Rheumatoid Arthritis, Ultrasonography, Synovitis, Joints, Clinical trial, Sample size

INTRODUCTION

Musculoskeletal ultrasound has been increasingly used in rheumatoid arthritis (RA) trials as an outcome measurement tool [1,2]. Its use in trials has been greatly faciliated by standardized ultrasound definitions of joint pathologies [3] and the adoption of semi-quantitative grading of joint inflammation on both B-mode and power Doppler ultrasound [4]. Ultrasound monitoring using an extended number of joint sites can be time-consuming, and therefore there has been much interest in determining the optimal 'minimal' ultrasound joint count that will provide a responsive outcome [5]. Various studies have advocated the use of different reduced ultrasount joint sets based on different criteria (such as frequency of joint involvement, representativeness of joints, feasibility, and using logistic regression models) [6] and to date, there has not been a global consensus on which is the best model to adopt. It is therefore important to study the relative performance metrics of various ultrasound models [7, 8] and examine how these may impact on issues relating to RA trial design, especially the sample size requirement.

Recently, novel ultrasound outcome measures using joint selection based on patient symptoms and ultrasound findings, and enabling a reduced joint count, have been shown to be highly responsive for demonstrating improvement in joint inflammation [9] when compared to other existing methods that pre-specify joints for ultrasound scanning as well as the Disease Activity Score at 28 joints (DAS28). In this pilot study, our aim was to determine whether the use of the novel ultrasound joint selection methods might translate into smaller sample sizes for RA clinical trials with ultrasound endpoints when compared to existing methods based on pre-specified joints.

METHODS

Seropositive active RA patients with DAS28 greater than 3.2 and polyarticular joint involvement (5 or more tender and/or swollen joints) were recruited in this institutional review board (IRB) approved research study. All patients were started or escalated on systemic corticosteroids and disease modifying anti-rheumatic drugs (DMARDs). Clinical examination and ultrasound assessment was performed at 44 joint sites (which included bilateral hand and feet joints, wrists, mid-foot, ankles, knees, elbows, hips and shoulders) at baseline and 3 months as previously described [9]. This study conforms to the relevant

research ethics guidelines. All patients provided informed consent before enrolling into the research study and patient recruitment took place from March 2013 to May 2016. Details that might disclose the identity of the subjects under the study have been omitted.

Ultrasonography and clinical joint examination

Ultrasound imaging on all patients was performed in the same outpatient location using standardized scanning based on the EULAR guidelines [10]. The ultrasound machines used were either (a) General Electric Healthcare LOGIQe machine with a multi-frequency (5-13 MHz) linear array transducer or (b) Philips Medical Systems EPIQ 5G machine with a multi-frequency (5-17 MHz) linear array transducer. Ultrasound machines were pre-set for the joint sites. For each patient, the same ultrasound machine and probe was used throughout the study period. The clinical joint examination was performed by a metrologist or a study nurse. A single rheumatologist experienced in musculoskeletal ultrasound acquired and scored the ultrasound images while blinded to the findings of the clinical joint assessors.

Using grey-scale (GS) and power Doppler (PD) ultrasonography, synovial inflammation at the joints were scored using a semi-quantitative system (0-none, 1-mild, 2-moderate and 3-severe), while tenosynovitis were scored using a dichotomous system (0-absent, 1-presence) as described previously [9]. On clinical examination, joint swelling and tenderness were scored dichotomously (0-absent, 1-presence) while joint pain were scored semi-quantitatively (0-none, 1-mild, 2-moderate and 3-severe) [9].

Novel ultrasound joint selection and existing methods

The novel ultrasound joint selection and the existing methods have been previously described [9] in detail, so only a summary will be given in what follows. The novel methods selected up to 7 or 12 of the most affected joints using (a) ultrasound joint inflammatory findings only—the Individualized-Ultrasound (IUS) method, or (b) both clinical and ultrasound joint inflammatory findings—the Individualized-Composite-Ultrasound (ICUS) method. The findings were compared to existing methods which utilized pre-specified reduced ultrasound joint sets for ultrasonography. The existing methods included a pre-specified 7-joint reduced ultrasound joint set [11] and a 12-joint reduced ultrasound joint set derived from an ultrasound data

reduction method [12].

An individual joint score (IJS) at each joint site was calculated for each patient. For the IUS and existing methods, the IJS was derived by summing up the ultrasound component sub-scores and dividing it by the greatest possible score at the joint. For example, at the right elbow joint, the ultrasound sub-scores (GS and PD ultrasound scores obtained at the relevant joint recesses) are summed up and then divided by the sum of the greatest possible GS and PD ultrasound scores at the same joint recesses to obtain the IJS for the IUS and existing methods.

For the ICUS, the IJS was derived by summing up both the clinical and ultrasound component sub-scores and dividing it by the greatest possible score at the joint. This allows score weights to be equalized across the joints [9]. For example, at the right elbow joint, the ultrasound sub-scores (GS and PD ultrasound scores obtained at the relevant joint recesses) and the clinical sub-scores (joint tenderness, swelling and pain scores) are summed up and then divided by the sum of the greatest possible ultrasound scores (greatest possible GS and PD ultrasound scores at the same joint recesses) and the greatest possible clinical scores (greatest possible joint tenderness, swelling and pain scores) to obtain the IJS for the ICUS method.

In the novel ultrasound methods, the joint selection process followed an algorithm which ranked the IJSs from the 44 joints from the greatest to the least in magnitude. Joints with a greater IJS would be selected over a lesser one, and the joint selection proceeds through a pre-specified joint sequence as described previously [9]. This process was repeated with smaller magnitude IJSs until the target joint count was reached (7 or 12 joints).

Statistical analysis

Subject demographics and baseline characteristics were summarized as mean and standard deviation (SD) for continuous variables and counts and percentages for categorical variables. The IJS at the relevant joints per patient were summed to obtain the total inflammatory score (TIS). The TIS was then calculated at baseline and 3 months for the various joint selection methods via the 7- and 12-joint

approaches. The estimate of TIS effect size (ES) at 3 months ($\widehat{ES}$) was calculated as mean change in 3-month TIS divided by the standard deviation (SD). 95% confidence intervals (CIs) were calculated on the estimated ES, viz. ($\widehat{ES_L}$, $\widehat{ES_U}$), using the method described in [13]. From the 95% lower and upper confidence limits on $\widehat{ES}$, respective upper and lower limits [$n_L(\widehat{ES_U})$, $n_U(\widehat{ES_L})$] were obtained on the post-hoc sample size reflecting the uncertainty in $\widehat{ES}$. Sample size calculations were based on a 1-sample t-test as the number of subjects needed to provide 80% power at α = 0.05 to reject the null hypothesis that H_0: ES = 0 versus the respective alternative hypotheses H_1: ES = $\widehat{ES}$, ES = $\widehat{ES_L}$ and ES = $\widehat{ES_U}$. Confidence intervals on the estimated effect size were calculated using SAS software V9.4 (SAS© Cary, NC, USA). Sample size calculations were performed using PASS© 2008 (NCSS©, Kaysville, UT, USA).

RESULTS

Subjects' baseline characteristics

Twenty-four patients with RA had mean DAS28 scores at baseline and 3 months of 4.81 and 4.20, respectively. Their baseline characteristics were as follows: mean age 57.2 years; majority Chinese (75%); majority female (83.3%); baseline mean (SD) ESR, 40.7 (27.2) mm/hr; mean (SD) disease duration, 45.4 (57.3) months. All the patients were started or escalated on systemic corticosteroids and disease modifying anti-rheumatic drugs (DMARDs). Within 3 months prior to patient recruitment, 19 out of the 24 patients (79.2%) received oral prednisolone while 16 out of the 24 patients (66.7%) received oral DMARD (which included hydroxychloroquine, sulfasalazine, leflunomide, methotrexate and azathioprine). The most frequently selected joints (within the top quartile) for the novel methods using the 7-joint approach and the 12-joint approach in our study population are presented in figure 1.

Sample size analysis

Using the 12 joint approach, the baseline mean TIS (with a maximum possible score of 12) for existing, ICUS and IUS were 1.35, 2.56 and 2.43 respectively, while the 3-month mean TIS (with a maximum possible score of 12) for existing, ICUS and IUS were 1.12, 1.68 and 1.52 respectively. Based on the effect size estimates and assuming use of the existing 12-joint approach in a future study, the post-hoc sample size estimate for detecting an effect size of 0.63 was n = 22. Using the existing 12-joint approach, the wide

95% CI on sample size which ranged from 10 to 245 reflects the uncertainty in the effect size estimate (0.18, 1.06). Using the ICUS 12-joint approach, the corresponding post-hoc sample size estimate for detecting an effect size of 0.96 was n = 11, while the 95% CI on sample size ranged from 7 to 40. Using the IUS 12-joint approach, the corresponding post-hoc sample size estimate for detecting an effect size of 0.97 was n = 11, while the 95% CI on sample size ranged from 6 to 38 (Table 1). The tighter CIs using the ICUS 12-joint approach and the IUS 12-joint approach reflect a greater precision in estimating ES using the novel approaches.

Using the 7 joint approach, the baseline mean TIS (with a maximum possible score of 7) for existing, ICUS and IUS were 0.56, 1.90 and 1.89 respectively, while the 3-month mean TIS (with a maximum possible score of 7) for existing, ICUS and IUS were 0.45, 1.29 and 1.24 respectively. Based on the effect size estimates and assuming use of the existing 7-joint approach in a future study, the post-hoc sample size estimate for detecting an effect size of 0.34 was n=70. The lower bound for 95% CI on sample size using the existing 7-joint approach was 16, while the upper bound was not estimable as the confidence interval on estimated ES (-0.07, 0.75) contained the null value zero. Using the ICUS 7-joint approach, the corresponding post-hoc sample size estimate for detecting an effect size of 1.13 was n = 9, while the 95% CI on sample size ranged from 6 to 24. Using the IUS 7-joint approach, the corresponding post-hoc sample size estimate for detecting an effect size of 0.99 was n = 11, while the 95% CI on sample size ranged from 6 to 35 (Table 1).

DISCUSSION

Our study is the first to explore sample size reduction for RA clinical trials with ultrasound endpoints using novel ultrasound joint selection methods versus existing methods. From the sample estimates and 95% CI analysis, our data suggests that sample size requirement for these RA clinical trials may be reduced using the novel methods which can provide greater responsiveness.

Ultrasonography is well suited as an outcome measurement tool in RA clinical trials. For more than a decade, it’s use has been studied in numerous longitudinal observational RA cohorts, and shown to be

useful as an efficacy measure of RA drug therapies such as corticosteroids [14], conventional [15] and biological DMARDs (e.g. tumor necrotic factor inhibitors [16], tocilizumab [17] and abatacept [18]). Additionally, one RA randomized control trial (n=24) has shown that GS ultrasound and PD imaging of synovial inflammation was able to discriminate between two patient groups (one group receiving infliximab and methotrexate (MTX) versus another group receiving placebo and MTX) over an 18 weeks time period [19].

Sample size considerations are especially relevant in the modern RA trial settings. Due to ethical concerns, extended placebo treatment periods are no longer feasible. Further, longer term drug trials will need to incorporate active comparator(s) or provide avenue(s) for rescue therapy, thereby making it more challenging to discriminate between treatment arms. With smaller measurable differences between treatment arms, clinical trials may need to be substantially longer in duration as well as requiring larger sample sizes to achieve statistical differences between treatment arms [20, 21]. Conceivably, the ability to reduce sample size requirement can be invaluable in the modern trial settings, facilitating both early and late phase clinical trials and potentially speeding up the process of new drugs discovery.

The limitations of this pilot study are the small sample size, the short duration of follow-up period and the small change in mean DAS28 score in our study population. Future larger scale RA studies with longer term follow-up periods will be required to confirm our observations and to delineate any correlation between the improvements seen using clinical scores (e.g. DAS28) versus ultrasound inflammatory scores. Moreover, our study population have, on an average, long standing disease (i.e. mean disease duration of 45.4 months) and patients were escalated or initiated in both DMARDs and corticosteroid. Depending on their study objectives, other clinical trials may be testing out other medication or interventions and may have different recruitment criteria (e.g. some may recruit patients with early RA while others may recruit patients with longstanding disease) and study protocols. Hence, our novel methods will need to be further tested out in various clinical scenarios and patient profiles in future studies.

Overall, our study provides promising data that sample size may be reduced in RA trials with ultrasound endpoints using novel ultrasound outcome measures. If confirmed in larger RA cohorts, this would have important implications for optimal use of ultrasound in modern RA clinical trials.

REFERENCES

1. Tan YK, Conaghan PG. Imaging in rheumatoid arthritis. Best Pract Res Clin Rheumatol 2011;25:569-84.
2. Iagnocco A, Finucci A, Ceccarelli F, Perricone C, Iorgoveanu V, Valesini G. Power Doppler ultrasound monitoring of response to anti-tumour necrosis factor alpha treatment in patients with rheumatoid arthritis. Rheumatology (Oxford). 2015;54(10):1890-6.
3. Wakefield RJ, Balint PV, Szkudlarek M, et al. Musculoskeletal ultrasound including definitions for ultrasonographic pathology. Journal of Rheumatology 2005;32(12):2485–7.
4. Baker JF, Tan YK, Conaghan PG. Monitoring in established RA: Role of imaging and soluble biomarkers. Best Pract Res Clin Rheumatol. 2015;29(4-5):566-79.
5. Naredo E, Gamero F, Bonilla G, Uson J, Carmona L, Laffon A. Ultrasonographic assessment of inflammatory activity in rheumatoid arthritis: Comparison of extended versus reduced joint evaluation. Clinical and Experimental Rheumatology 2005;23:881-884.
6. Mandl P, Naredo E, Wakefield RJ, Conaghan PG, D’Agostino MA; OMERACT Ultrasound Task Force. A systematic literature review analysis of ultrasound joint count and scoring systems to assess synovitis in rheumatoid arthritis according to the OMERACT filter. J Rheumatol 2011;38:2055-62.
7. Dougados M, Jousse-Joulin S, Mistretta F, et al. Evaluation of several ultrasonography scoring systems for synovitis and comparison to clinical examination: results from a prospective multicentre study of rheumatoid arthritis. Ann Rheum Dis 2010;69:828-33
8. Mandl P, Balint PV, Brault Y, et al. Metrologic properties of ultrasound versus clinical evaluation of synovitis in rheumatoid arthritis: results of a multicenter, randomized study. Arthritis Rheum 2012;64:1272-82
9. Tan YK, Allen JC Jr, Lye WK, et al. Novel Ultrasound Joint Selection Methods Using a Reduced Joint Number Demonstrate Inflammatory Improvement when Compared to Existing Methods and Disease Activity Score at 28 Joints. J Rheumatol 2016;43:34-7
10. Backhaus M, Burmester GR, Gerber T, et al. Guidelines for musculoskeletal ultrasound in rheumatology. Ann Rheum Dis 2001;60:641–9
11. Backhaus M, Ohrndorf S, Kellner H, et al. Evaluation of a novel 7-joint ultrasound score in daily

rheumatologic practice: a pilot project. Arthritis Rheum 2009;61:1194-201

12. Naredo E, Rodríguez M, Campos C, et al. Validity, reproducibility, and responsiveness of a twelve-joint simplified power doppler ultrasonographic assessment of joint inflammation in rheumatoid arthritis. Arthritis Rheum 2008;59:515-22
13. Kelley, K. Confidence intervals for standardized effect sizes: theory, application, and implementation. Journal of Statistical Software 2007:20:8:2-24
14. Terslev L, Torp-Pedersen S, Qvistgaard E, Danneskiold-Samsoe B, Bliddal H. Estimation of inflammation by Doppler ultrasound: quantitative changes after intra-articular treatment in rheumatoid arthritis. Annals of the Rheumatic Diseases 2003;62:1049–53.
15. Naredo E, Collado P, Cruz A, et al. Longitudinal power Doppler ultrasonographic assessment of joint inflammatory activity in early rheumatoid arthritis: predictive value in disease activity and radiologic progression. Arthritis and Rheumatism 2007;57(1):116–24.
16. Naredo E, Möller I, Cruz A, Carmona L, Garrido J. Power Doppler ultrasonographic monitoring of response to anti-tumor necrosis factor therapy in patients with rheumatoid arthritis. Arthritis and Rheumatism 2008;58:2248–56.
17. Fukae J, Isobe M, Kitano A, et al. Radiographic prognosis of finger joint damage predicted by early alteration in synovial vascularity in patients with rheumatoid arthritis: potential utility of power Doppler sonography in clinical practice. Arthritis Care Res(Hoboken)2011;63:1247–1253.
18. D'Agostino MA, Wakefield RJ, Berner-Hammer H, et al; OMERACT-EULAR-Ultrasound Task Force. Value of ultrasonography as a marker of early response to abatacept in patients with rheumatoid arthritis and an inadequate response to methotrexate: results from the APPRAISE study. Ann Rheum Dis. 2016;75(10):1763-9.
19. Taylor PC, Steuer A, Gruber J, et al. Comparison of ultrasonographic assessment of synovitis and joint vascularity with radiographic evaluation in a randomized, placebo-controlled study of infliximab therapy in early rheumatoid arthritis. Arthritis and Rheumatism 2004;50(4):1107–16.
20. Food and Drug Administration. Guidance for Industry, Rheumatoid arthritis: Developing drug products for treatment. Draft Guidance May 2013. http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm35446

8.pdf (acessed 23012017)

21. Freeston JE, Bird P, Conaghan PG. The role of MRI in rheumatoid arthritis: research and clinical issues.Curr Opin Rheumatol. 2009;21:95-101.