Effect of using the Simple Erosion Narrowing Score or Sharp/van der Heijde score on the power of rheumatoid arthritis clinical trials to detect differences in radiographic progression

Broeder, N. den; Bolhuis, T.E.; Herwaarden, N. van; Bergstra, S.A.; Bouman, C.A.M.; Hoogen, F.H.J. van den; ... ; Maas, A. van der

Rheumatology, 2025, **64**, 1464–1468
https://doi.org/10.1093/rheumatology/keae328
Advance access publication 12 June 2024
Concise Report

British Society for Rheumatology RHEUMATOLOGY OXFORD

Clinical science

Effect of using the Simple Erosion Narrowing Score or Sharp/van der Heijde score on the power of rheumatoid arthritis clinical trials to detect differences in radiographic progression

Nathan den Broeder [1,2,*], Thomas E. Bolhuis [1,2], Noortje van Herwaarden [1], Sytske Anne Bergstra [3], Chantal A. M. Bouman [4], Frank H. J. van den Hoogen [1], Lise M. Verhoef [1], Alfons A. den Broeder [1], Aatke van der Maas [1]

[1]Department of Rheumatology, Sint Maartenskliniek, Nijmegen, The Netherlands
[2]Radboud Institute for Health Sciences, Radboud University Medical Centre, Nijmegen, The Netherlands
[3]Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands
[4]Department of Rheumatology, Radboudumc, Nijmegen, The Netherlands

*Correspondence to: Nathan den Broeder, Department of Rheumatology, Sint Maartenskliniek, Postbus 9011, 6500 GM Nijmegen, The Netherlands.
E-mail: N.denbroeder@maartenskliniek.nl

Abstract

Objectives: The Simple Erosion Narrowing Score (SENS) is a simplification of the Sharp/van der Heijde score (SHS). Previous studies found SENS and SHS to have very similar measurement properties, but suggest that SENS has a lower discriminative ability that may result in reduced power. Therefore, we aimed to quantify the effect of using SENS rather than SHS on the power to show between-group differences in radiographic progression.

Methods: Using data from two clinical trials in RA (DRESS and BeSt), SENS was derived from the SHS. Criterion validity of the SENS in relation to the SHS was assessed by calculating the Spearman correlation. The power of both scores to show a difference between groups was compared using bootstrapping to generate 10 000 replications of each study. Then, the number of replications with a significant difference in progression (using analysis of covariance adjusted for baseline scores) were compared.

Results: Correlations between SENS and SHS were all >0.9, indicating high criterion validity of SENS compared with SHS as a reference standard. There was one exception—the DRESS study showed a somewhat lower correlation for the change score at 18 months (0.787). The loss in power of SENS over SHS was limited to at most 19% (BeSt year 5). In addition, the difference in power between SENS and SHS is smaller at higher levels of power.

Conclusion: SENS appears to be a reasonable alternative to SHS, with only a limited loss of power to show between-group differences in radiographic progression.

Keywords: rheumatoid arthritis, radiographic progression, Simple Erosion Narrowing Score, statistical power.

Rheumatology key messages

- The Simple Erosion Narrowing Score (SENS) correlates strongly with the Sharp/van der Heijde score (SHS).
- The loss in power from using SENS instead of SHS is limited.

Introduction

Joint damage and progression thereof is one of the main outcomes used in the assessment of RA treatment and is most often quantified using the Sharp/van der Heijde score (SHS) [1]. The Simple Erosion Narrowing Score (SENS) is a simplification of the SHS and has been recommended for use in clinical practice to assess radiographic progression of RA because it combines a simpler, less time-consuming scoring method with measurement properties comparable to the SHS [2]. The SHS grades the erosions from 0 to 5 per joint in the hands and from 0 to 10 per joint in the feet, and joint space narrowing (JSN) from 0 to 4 in each joint. In comparison, the SENS assesses the presence of both erosions and JSN in the same joints as the SHS, but without grading them.

Previous studies comparing the SENS to the SHS have found very similar measurement properties [2–6], but have suggested that the SENS may lack some discriminative ability compared with the SHS, which may result in a lower power

Received: 6 November 2023. **Accepted:** 8 May 2024

to detect between-group differences in progression. This claim is based on the idea that the SENS score cannot measure any progression in a joint that already had erosions or JSN, resulting in a ceiling effect on the joint level, as demonstrated in data from the COBRA trial [4]. In a reanalysis of radiographic data from this study, authors showed that progression of erosions from previously eroded joints contributed relevantly to total progression of joint damage. In addition, when only focusing on joints not previously eroded, radiographic progression could not be shown past the initial phase of the study. An analysis of data from the BeSt study strengthens this finding by showing that between 11 and 27% of patients with SHS progression of ≥1 point did not have SENS progression [6]. As a result, the SENS has thus far not been recommended for use as an outcome in RA research.

While perhaps directionally convincing, the claim that SENS lacks between-group discriminative ability has not been quantified. Furthermore, there are several arguments that this effect may be quite limited. Firstly, a study using data from the TEMPO trial failed to show indications of a ceiling effect affecting the total SENS score, as detection of progression using SENS was equally sensitive in all quartiles of baseline SHS [5]. Secondly, the studies that do suggest a lack of discriminative ability have some limitations: the analyses of the COBRA trial focused only on the erosion subscore and ignored the effects of JSN, which may somewhat compensate for a lack of discriminative ability based on erosion scores alone [4]. The analyses of the BeSt study focused on classifying individual patients as progressors or not, which is not as relevant at a group level and scores for research purposes need only to show between-group differences in progression [6].

In summary, the SENS is a simplification of the SHS score with very similar measurement properties, but the SHS requires a factor of 3.5 more time to perform [2]. There are arguments that the SENS may lack discriminative ability, but the effect of this on the power of a study to show differences in radiographic progression has not been quantified and may not be of relevant magnitude. A quantification of the effect is relevant, because it allows researchers to judge whether the advantages of SENS (reduction in time and associated costs) weigh up to the disadvantage of potentially lower power (which may offset the reduction in time and costs if a much larger sample size is then required). The objective of this study is therefore to quantify the effect of using SENS rather than SHS on the power of RA trials to show differences in radiographic progression.

Methods

Data sources

This study uses data from two well-known clinical trials in RA: the DRESS and BeSt studies. These studies were selected because they detected relatively minor differences in SHS progression between treatment arms, meaning any lack of discriminative ability of SENS should become clear more easily than in studies with a large difference in progression. Both studies are on a treat-to-target background, aiming for (maintaining) low disease activity.

The DRESS study (Dutch trial register, NTR 3216, CMO region Arnhem-Nijmegen, NL37704.091.11) is an open label non-inferiority randomized controlled trial in which RA patients with low disease activity on a stable TNFi dose (adalimumab or etanercept) were randomized 2:1 to disease activity–guided tapering or full dose continuation [7]. Radiographs of hands and feet were made at baseline and 18 months and independently scored by two trained readers using the SHS blinded for allocation and in chronological order. The DRESS study was approved by Commissie Mensgebonden Onderzoek region Arnhem-Nijmegen (number NL37704.091.11).

The BeSt study is a four-arm randomized controlled trial in early RA patients of sequential monotherapy *vs* step-up combination therapy *vs* initial combination therapy with tapered high-dose prednisone *vs* initial combination therapy with infliximab (Dutch trial register, NTR262 and NTR265) [8]. Radiographs of hands and feet were made every 12 months and independently scored by two trained readers using SHS blinded for order and treatment allocation. The Medical Ethical Committee Leids Universitair Medisch Centrum approved the study (number P258/99), and all patients gave written informed consent.

For each study, the SENS was calculated based on the SHS and not scored separately. SENS scores were derived from the individual joint scores of the SHS per reader individually, then summed to a total score. Finally, for both scores, the mean of the two readers was used to minimize measurement error.

Analyses

First, we considered the criterion validity of SENS in relation to the SHS as a reference standard. This was assessed through correlations both cross-sectionally by determining the Spearman correlation between the absolute SENS and SHS scores, and longitudinally by determining the Spearman correlation between the change in SENS and SHS scores over the course of each study. We used the criteria set forth by the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) initiative, regarding good measurement properties, to determine whether these correlations were sufficient (≥0.70) [9].

Secondly, we compared the power of both methods to show a difference in progression between treatment arms. For this purpose, we used bootstrapping to generate 10 000 replications of each study and assessed for both SENS and SHS in how many replications a significant difference in progression could be shown using ANCOVA adjusted for baseline score. In addition, because the effect on power might be dependent on the sample size, this process was repeated using random subsamples of each study from 90% of the sample size down to 10% in increments of 10 percentage points. For the BeSt study, this analysis was repeated for progression at years 1 through 5 as the between-group differences first increase over time and then gradually reduce again. Years 6 through 10 of the BeSt study were not considered as little between-arm contrast remained at this point. This allowed us to assess different between-group contrasts and the resulting relative power of the SENS and SHS. As a sensitivity analysis, we also assessed power for the non-parametric Wilcoxon rank-sum test (for DRESS) or Kruskal–Wallis test (for BeSt).

Results

Baseline characteristics for the DRESS and BeSt are shown in Supplementary Tables S1 and S2 (available at *Rheumatology* online). DRESS included patients with long-standing RA and higher baseline radiographic damage [median SHS (interquartile

range, IQR) 23 (6.0–50.0) in the tapering arm and 18 (9.0–47.0) for the usual care arm]. BeSt included early RA patients with almost no radiographic damage at baseline. The median SHS (IQR) was 1.5 (0.0–4.5) in the sequential monotherapy arm, 2.0 (0.0–6.0) in the initial combination with infliximab arm, 1.5 (0.0–3.0) in the initial combination with prednisone arm and 1.5 (0.0–6.5) in the step-up combination therapy arm. Information on the reader performance is shown in Supplementary Table S3 (available at *Rheumatology* online).

Correlation between SENS and SHS

The correlations between both absolute SENS and SHS scores and changes in both scores over time are shown in Table 1. Further descriptives for both scores are shown in Supplementary Table S4 (available at *Rheumatology* online). For all time points, absolute correlations were high or very high, indicating high criterion validity of the SENS compared with the SHS as a reference standard. For change scores, only the DRESS study showed a somewhat lower correlation, at 0.787 for progression at 18 months. This could be explained by the higher baseline radiographic damage in this group, which means there is no possibility for progression with SENS in those joints with baseline damage.

Power of SENS and SHS

Fig. 1 shows the bootstrapped estimated power for the DRESS and the BeSt study at different times of follow-up. It can be seen that the loss in power by using SENS over SHS is limited to at most 19% (BeSt study year 5). In addition, the difference in power between SENS and SHS is smaller at higher levels of power, and in some cases SENS even outperformed SHS. Interestingly, despite the lower correlation between SENS and SHS progression in the DRESS study, this is one of the analyses in which SENS actually had higher power than SHS. The sensitivity analyses using non-parametric approaches showed overall lower power and slightly bigger but still limited differences between SHS and SENS power (Supplementary Fig. S1, available at *Rheumatology* online).

Discussion

This study shows that, at the trial level, the loss in power from using SENS over SHS is fairly limited. Interestingly, the results seem to suggest that the difference between the SHS and SENS in terms of power is smaller when SHS had higher power, which is where a difference in power is most relevant if a trial is powered to detect radiographic progression.

These results are consistent with earlier studies that have argued that the SENS may have lower discriminative ability due to its inability to score progressive erosions in joints with a previous erosion [3–6]. This is also reflected by the fact that criterion validity of SENS progression is lowest in the DRESS study which had a higher level of baseline radiographic damage. The added value of this study compared with the existing literature is a quantification of the limited degree with which this limitation impacts study power. This may be explained by several factors. For example, it is possible that patients with progression of previous erosions also develop new erosions in other joints, though there are indications that inflammation tends to recur in the same joints resulting in progression [10, 11]. Another possibility is that the grading of damage does add some information, but also increases the amount of measurement error, as the subtle differences between two grades of erosions of JSN in SHS may be harder to score than simply the presence of erosions or JSN in SENS.

Results from this study can aid researchers in selecting a scoring system for radiographic damage in RA studies. The ideal scoring system for a specific study depends on the relative importance of the radiographic progression as an outcome, and on other considerations that may influence the sample size. For example, in a study that has radiographic progression as a secondary outcome, which already requires a high sample size for its primary outcome, SENS may be considered. On the other hand, a study with a limited sample size in established RA patients with high baseline radiographic damage, with radiographic progression as a primary outcome, may require the use of the more laborious SHS. The latter scenario may become increasingly uncommon as the diagnosis and treatment of RA improves and results in lower levels of radiographic damage. Another consideration may be power for secondary and sub-analyses that may be lower with the less detailed information provided by SENS. Of note, the substantial savings in scoring time and cost when using SENS could also be reinvested in optimizing power for radiological outcome by other means, such as scoring with an additional reader.

Strengths of this study are the inclusion of data from two landmark trials that reflect a broad spectrum of RA patients. Furthermore, long-term follow-up was available for the BeSt study. A limitation of this study is the use of SENS derived from SHS instead of directly scoring radiographs using SENS, but previous research has shown the correlation between

Table 1. Spearman correlation between SENS and SHS in terms of absolute scores and as change from baseline

	Absolute scores Median (IQR)		Spearman correlation	Change from baseline Median (IQR)		Spearman correlation
	SHS	SENS		SHS	SENS	
DRESS baseline	18.5 (6.5–49.0)	9.5 (4.0–21.5)	0.988			
DRESS 18 months	19.5 (7.0–49.0)	10 (4.0–21.5)	0.987	0.0 (0.0–1.0)	0.0 (0.0–0.5)	0.787
BeSt baseline	1.5 (0.0–5.0)	1.0 (0.0–2.5)	0.976			
BeSt 1 year	3.0 (0.0–7.5)	1.5 (0.0–4.0)	0.977	0.0 (0.0–1.5)	0.0 (0.0–1.0)	0.931
BeSt 2 years	2.5 (0.5–9.0)	1.5 (0.5–5.0)	0.986	0.0 (0.0–2.5)	0.0 (0.0–1.5)	0.961
BeSt 3 years	3.0 (0.5–10.0)	1.5 (0.5–5.0)	0.986	0.5 (0.0–3.0)	0.0 (0.0–1.5)	0.965
BeSt 4 years	3.5 (1.0–11.0)	2.0 (0.5–6.0)	0.982	1.0 (0.0–4.0)	0.5 (0.0–2.0)	0.957
BeSt 5 years	4.0 (1.0–11.5)	2.0 (0.5–6.0)	0.975	1.0 (0.0–5.5)	0.5 (0.0–3.0)	0.971

SENS: Simple Erosion Narrowing Score; SHS: Sharp/van der Heijde score; IQR: interquartile range.

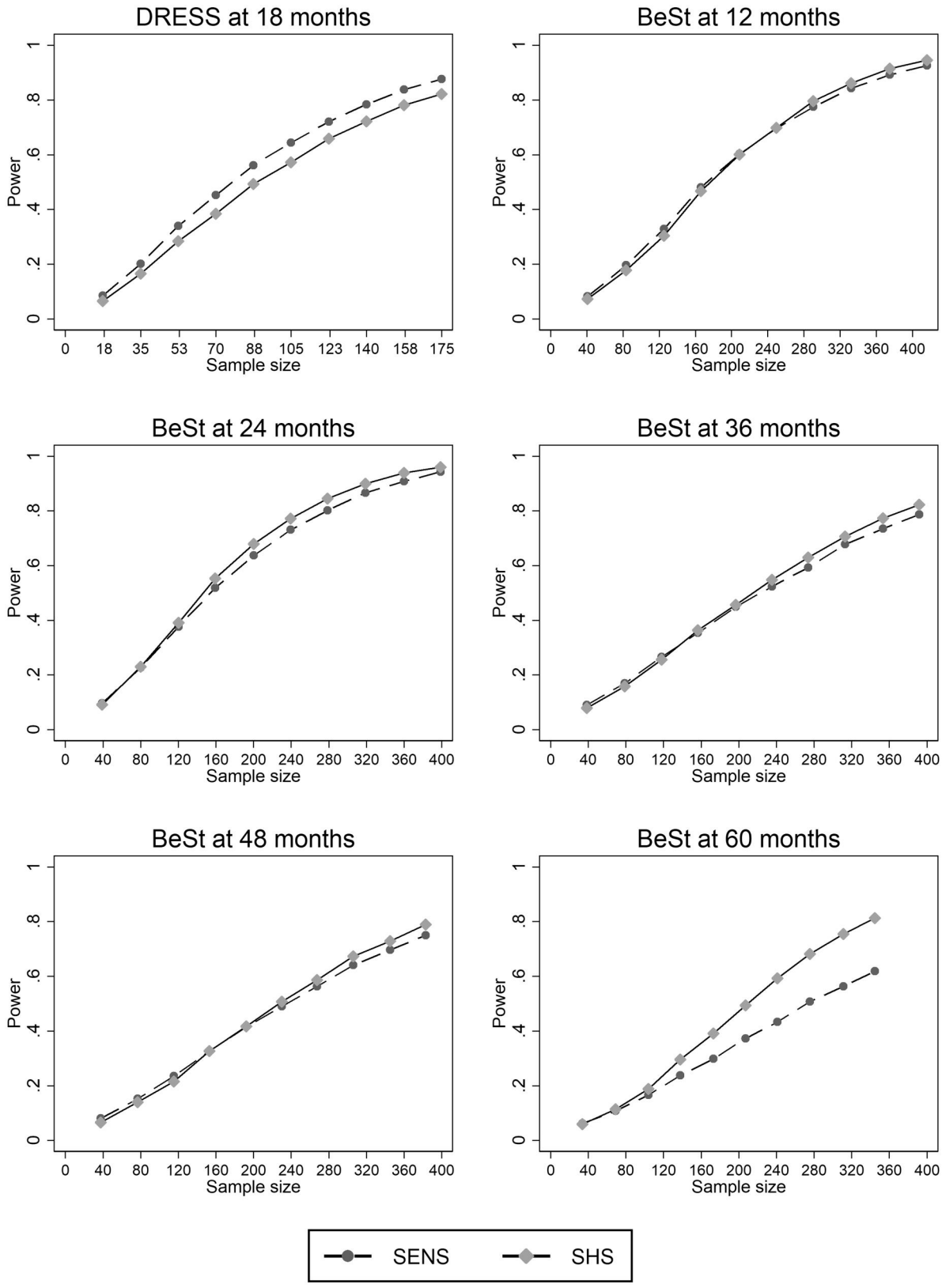

Figure 1. Power of SENS and SHS scores in bootstrapped replications of the DRESS and BeSt studies at various sample sizes and time points. Power is calculated as the proportion of bootstrapped replications with $P < 0.05$ for ANCOVA adjusted for baseline score for both scoring systems. SENS: Simple Erosion Narrowing Score; SHS: Sharp/van der Heijde score; ANCOVA: analysis of covariance

calculated and derived SENS to be very high [2]. Moreover, as this study was focused on power of the SHS and SENS at the group level, it does not provide information on the use of SENS to assess progression in individual patients. An apparent limitation is the low intra-class correlation coefficient for the change scores in the DRESS study, but this can be explained by the low overall progression as the smallest detectable change shows that measurement error was in fact slightly lower than in the BeSt data.

In conclusion, SENS appears to be a reasonable alternative to SHS, with only a limited loss of power to show between-group differences in progression in clinical trials.

Supplementary material

Supplementary material is available at *Rheumatology* online.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

Funding

No specific funding was received from any bodies in the public, commercial or not-for-profit sectors to carry out the work described in this article.

Disclosure statement: S.A.B. reports grants from Pfizer and speaker's fees from Benecke, and A.v.d.M. reports grants from Dutch Arthritis foundation and The Netherlands Organisation for Health Research and Development (ZonMw). All other authors report no conflicts of interests.

References

1. Committee for Medicinal Products for Human Use (CHMP): European Medicines Agency. Guideline on clinical investigation of medicinal products for the treatment of rheumatoid arthritis: European Medicines Agency. 2017. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-investigation-medicinal-products-treatment-rheumatoid-arthritis_en.pdf (25 August 2023, date last accessed).
2. van der Heijde D, Dankert T, Nieman F, Rau R, Boers M. Reliability and sensitivity to change of a simplification of the Sharp/van der Heijde radiological assessment in rheumatoid arthritis. Rheumatology (Oxford) 1999;38:941–7.
3. Barnabe C, Hazlewood G, Barr S, Martin L. Comparison of radiographic scoring methods in a cohort of RA patients treated with anti-TNF therapy. Rheumatology (Oxford) 2012;51:878–81.
4. Bruynesteyn K, Van Der Heijde D, Boers M *et al.*; COBRA Trial Group. Contribution of progression of erosive damage in previously eroded joints in early rheumatoid arthritis trials: COBRA trial as an example. Arthritis Rheum 2002;47:532–6.
5. Dias EM, Lukas C, Landewé R, Fatenejad S, van der Heijde D. Reliability and sensitivity to change of the Simple Erosion Narrowing Score compared with the Sharp-van der Heijde method for scoring radiographs in rheumatoid arthritis. Ann Rheum Dis 2008;67:375–9.
6. Klarenbeek NB, Güler-Yüksel M, van der Heijde DMFM *et al.* A comparison between the simplified erosion and narrowing score and the Sharp-van der Heijde score: post hoc analysis from the Best study. Ann Rheum Dis 2011;70:714–6.
7. van Herwaarden N, van der Maas A, Minten MJM *et al.* Disease activity guided dose reduction and withdrawal of adalimumab or etanercept compared with usual care in rheumatoid arthritis: open label, randomised controlled, non-inferiority trial. BMJ 2015;350:h1389.
8. Goekoop-Ruiterman YPM, de Vries-Bouwstra JK, Allaart CF *et al.* Clinical and radiographic outcomes of four different treatment strategies in patients with early rheumatoid arthritis (the BeSt study): a randomized, controlled trial. Arthritis Rheum 2005;52:3381–90.
9. Prinsen CAC, Mokkink LB, Bouter LM *et al.* COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res 2018;27:1147–57.
10. Heckert SL, Bergstra SA, Matthijssen XME *et al.* Joint inflammation tends to recur in the same joints during the rheumatoid arthritis disease course. Ann Rheum Dis 2022;81:169–74.
11. Heckert SL, Bergstra SA, Goekoop-Ruiterman YPM *et al.* Frequency of joint inflammation is associated with local joint damage progression in rheumatoid arthritis despite long-term targeted treatment. RMD Open 2023;9:e002552.

Rheumatology, 2025, 64, 1464–1468
https://doi.org/10.1093/rheumatology/keae328
Concise Report