Contributions of ultrasound beyond clinical data in assessing inflammatory disease activity in rheumatoid arthritis: current insights and future prospects

Peter Mandl¹, Reka Kurucz², Dora Niedermayer², Peter V. Balint² and Josef S. Smolen¹,3

Abstract

Appropriate measures of disease activity need to be valid, reliable and sensitive to change for use in clinical studies while remaining at the same time feasible and practicable for utilization in daily clinical practice. Ultrasonography was shown to be a valid, sensitive and reliable imaging modality for the detection of synovitis in RA, however, it has so far failed to demonstrate superior sensitivity to change as compared with clinical examination. This review examines the current evidence for the use of established measures and/or US, either as an alternative or as a supplementary measure to clinical examination, as tools for monitoring synovitis in RA. It also includes a summary of results of recent studies evaluating clinical examination–based as well as clinical- and US-based multimodal disease activity indices. We review the rationale and limitations of incorporating US into composite disease activity indices and suggest a research roadmap for further studies in this field.

Key words: ultrasound, disease activity indices, rheumatoid arthritis, clinical evaluation.

Introduction

The monitoring of disease activity in rheumatic diseases in general and RA in particular remains challenging in both daily practice and clinical studies. It requires reliable and valid instruments that are sensitive to change, with the ability to capture and accurately measure the multifaceted status of the disease in a standardized way for the purpose of performing clinical studies, while remaining at the same time feasible for utilization in daily clinical practice [1]. An instrument/variable should also not be redundant, i.e. not reflect an aspect of disease activity that another instrument/variable covers inherently. Not all clinical variables characteristic of RA are useful for disease activity assessment and many of the more suitable ones vary in their predominance between and even longitudinally within patients. This inadequacy of applying individual measures of disease activity for assessing outcomes, which made clinical trials of earlier decades difficult to interpret, was overcome by the introduction and successful application of composite measures of disease activity that have become the standard assessment tools of therapeutic efficacy and overall disease activity in RA [2–8].

Clinical examination has traditionally been complemented by radiography [9–11]. X-rays, however, can only depict bony changes relatively accurately, while they allow assessment of cartilage damage only indirectly and are usually unable to determine soft tissue abnormalities. In this sense radiography mirrors the past with respect to inflammation, while MRI and US reflect the present. Its potential to detect and monitor both inflammatory and structural changes and the fact that it is free of radiation, is applicable in an outpatient setting and is not overly time-consuming make US an ideal imaging modality for the evaluation of pathological changes in arthritis at any point in time as well as longitudinally. This review...
examines the current evidence, utility and limitations of incorporating US into composite disease activity indices.

US as a tool for monitoring synovitis in RA

A number of studies using arthroscopy or MRI as the gold standard have demonstrated that both grey-scale US (GSUS) and Doppler, particularly power Doppler US (PDUS), have greater sensitivity for detecting synovitis as compared with physical examination [12–15] and are also more reliable [16–18] (Fig. 1). On the other hand, US is considered to be one of the most operator-dependent imaging techniques, and concerns over the reliability of musculoskeletal US have been raised and noted since its initial introduction to the field of rheumatology [19]. Standardized scanning techniques, definitions of pathological findings and increased availability of training have led to improvements in its reproducibility [20–22]. A recent systematic literature review on the reliability of US in assessing synovitis in RA reported high intra- and interobserver reliability, especially for PDUS [23]. However, when interpreting these findings we should consider that most studies report intra- and interreader reliability by using still images, while it may be more informative (and important) to assess intra- and interacquirer reliability assessing variabilities in acquisition and immediate interpretation of dynamic images. Further, most published data have evaluated highly trained observers as well as high-end US equipment, while intermachine variability and the question of experienced vs less experienced users have not been sufficiently addressed and remain important issues in multicentre monitoring studies. Finally, each of these aspects has been addressed in only a small number of studies [24–27]. Our recent systematic evaluation of global synovitis scoring systems concluded that it is currently difficult to suggest a minimum number of joints to be included in a generally accepted global US synovitis score. Having such a score available might allow better interpretation of data from available studies [28]. Nonetheless, scores based on a reduced number of joints have shown similar correlations with clinical and laboratory variables as scores including an extensive number of joints [29–32]. A recent editorial even goes so far as to suggest the scanning of a single joint as a quick and simple US measure of disease activity [33]. A quantitative joint count, however, regardless of the modality through which it is obtained, may not be sufficient to serve as a sole measure to assess and monitor all individual patients with RA, since patients may suffer from diverse symptoms and yet will ultimately be treated with identical therapies [2, 34].

US has also been shown to be a capable tool for the monitoring of synovitis, in both early and established RA, and GSUS/PDUS signs of synovitis decrease upon

---

**Fig. 1 Clinical and ultrasonographic examination**

(A and B) Palpation and corresponding US image of a clinically swollen MCP joint (GSUS/PDUS grade 3 = marked synovial thickening/markd vessels signals in more than 50% of the synovium). (C and D) Palpation and corresponding US image of a clinically non-swollen and non-tender MCP joint (GSUS/PDUS grade 2 = moderate synovial thickening/moderate vessel signals in < 50% of the synovium). GSUS: grey-scale US; PDUS: power Doppler US.
treatment [35–39]. Several studies have demonstrated GSUS/PDUS signals in RA patients in clinical remission, implying that residual active synovitis may exist in clinical remission [40–45]. PD signals have been suggested to confer a 12-fold risk of relapse in patients with persistent clinical remission [46]. However, interpretation of these findings is limited by the lack of long-term outcome data for US and the issue of feasibility in clinical practice. Importantly, remission portends excellent clinical, functional and structural outcomes over time and these beneficial effects of reaching remission have been established using purely clinical markers (with or without inclusion of acute phase reactant levels), and indeed, long-standing clinical remission is usually associated with the halt of radiographic damage progression [47–49], a reduction of cardiovascular risk [49] and optimal physical function [50, 51]. Comparative data using US are currently not available and US-based criteria of remission have to be validated further.

**Established measures of disease activity in RA**

The primary rationale behind the development of composite indices was to enhance validity and sensitivity to change, particularly for clinical trials where the distinction between an effective and a less effective or ineffective therapy is absolutely key; indeed, composite indices, pooled from core set measures, have been shown to be more valid distinguishers of RA-related disease activity than individual measures [4–8]. Among core set measures, some patient-reported outcomes (PROs), despite their apparent subjective nature, have shown remarkable sensitivity to change, and some, like physical function, are also predictive for long-term outcomes such as mortality [52]. Although well validated, studies on the interobserver variance, sensitivity and specificity of clinical assessments for the detection of joint inflammation have produced conflicting results [53–58]. Nonetheless, clinical measures should ideally reflect the complex reality of RA disease activity, which is best captured by composite indices [59]. In the absence of measures with superior discriminative capacity, presently clinical composite indices remain the best predictors of overall long-term outcomes, including joint damage, and have been the measures utilized in almost all available randomized controlled clinical trials in the past two decades.

On the other hand, potential future improvements in predictive capacity of disease activity assessment may either allow better insights into treatment response or enable a reduction in the numbers of patients needed in clinical trials. In this respect the high reliability, sensitivity and specificity of US in detecting inflammatory joint changes needs to be considered. Thus, the question arises as to whether US would add information or even replace clinical assessment during follow-up (Fig. 2).

**US as a potential tool for monitoring disease activity in RA**

While several studies have advocated the use and utility of GSUS and PDUS evaluation for the monitoring of synovitis as an individual measure, only a few studies have actually attempted to incorporate US within a composite index. A study evaluating a sonographic composite disease activity index (US DAS) in which the clinical joint counts were replaced by GSUS and PDUS counts for 28 and 22 joints, respectively, reported strong to moderate correlation with the clinical 28-joint DAS (DAS28) score, inflammatory markers, as well as PROs, including the HAQ [56].

---

**Fig. 2** The evolution of clinical indices and ultrasonographic indices and their synthesis into a clinical- and US-based multimodal index

APR: acute phase reactants; GSUS: grey-scale US; PDUS: power Doppler US; VAS: visual analogue scale.
This approach deliberately discounts for most, if not all, clinical information on joint involvement, replacing the clinical joint counts of the index with US joint counts. On the other hand, one can choose to merge the two worlds by incorporating US into a composite index: instead of replacing clinical data with US-derived data, the two may be combined into so-called clinical- and US-based indices [59]. In such indices the 28-joint swollen joint count was either supplemented or replaced by clinically non-swollen joints with GSUS and/or PDUS signs of synovitis, and this approach was shown to improve intra-observer reliability when compared with its purely clinical counterparts. However, no major advantages were observed regarding validity, and sensitivity to change was even slightly lower for the clinical- and US-based composite indices compared with the clinical DAS28 or SDAI [59]. These findings are in accordance with several studies that, while demonstrating the superiority of GSUS and PDUS over clinical examination with regard to reliability, did not show enhanced sensitivity to change [17, 18]. To our knowledge there are no further studies that have evaluated clinical- and US-based composite disease activity indices [56, 59]. Looking at the present study arena, it is currently difficult to conclude whether GSUS or PDUS is superior regarding inclusion into such indices, although certain studies have suggested that PDUS reflects clinical activity better than GSUS [46, 60].

The use of US is seen as a complementary procedure and not as an alternative to clinical evaluation in the recent ACR recommendation on the reasonable use of US in rheumatology clinical practice [61]. US, along with MRI, is now widely accepted as being superior to clinical examination for the detection of joint inflammation and is recommended over clinical criteria alone to improve diagnostic accuracy in RA [62]. Indeed, the new ACR–European League Against Rheumatism (EULAR) RA classification criteria [63] agree with the use of US to determine joint involvement and thus how many points are given to the joint-related component of the criteria, provided that at least one joint is unequivocally clinically swollen. US is also increasingly used as a tool for monitoring the disease process in daily rheumatologic practice, but its exact role in this respect needs further clarification. There is no agreement on the type and number of joints to monitor [28] and disagreement on the relevance of US findings of synovitis both in early disease and in remission [60, 64, 65]. Replacing information derived from clinical examination entirely by information derived from US raises questions regarding the currently unconfirmed added value of sono graphic assessment beyond clinical scoring and of its feasibility in routine clinical practice (availability of equipment and trained staff, time constraints, costs, etc.). In addition, although the utility of US to predict subsequent radiographic damage has been suggested by several studies [43, 66–69], a study that compared its predictive validity with that of clinical examination failed to demonstrate its superiority [70], casting doubt on its utility as a replacement for clinical evaluation of synovitis within a composite index.

Since both GSUS and PDUS have been shown to be more sensitive than clinical examination [9–12], one would expect to see improvement in sensitivity to change upon adding US-derived information to a clinical composite index. An explanation for the apparent, if minor decrease in the sensitivity to change of clinical- and US-based multimodal indices compared with clinical indices as demonstrated in our study may derive from the fact that both US and clinical examination measure the same component of RA, namely synovitis. A combination of individual measures, which are highly related, may actually result in poorer overall sensitivity to change than the combination of more diverse core set measures [50]. Nevertheless, the responsiveness to change of a health status measurement instrument is closely related to its reliability [71]. The majority of validated global US scoring systems have been shown to be more reliable than clinical examination for the monitoring of synovitis [14, 15, 26–29]. However, we currently lack an answer as to whether US as compared with clinical examination is truly superior regarding long-term implications, and whether multimodal indices containing also US-derived data might potentially serve as more responsive measures than clinical indices. The essential nature of this question is highlighted by the fact that US is already used to monitor synovitis in RA in routine clinical practice, despite the limited available data in this regard. The first results of a clinical trial evaluating US- and DAS28-guided DMARD therapy showed overall similar clinical outcomes for the primary endpoint (change in DAS44) at 18 months, although in the US group DAS44 remission rates were significantly (though only slightly) higher at 18 months, but not earlier [72].

**Conclusion**

In order to gain definitive insight into this matter, sono graphic assessment of synovitis needs to be tested as thoroughly as clinical measures have been tested, both in clinical trials and in daily practice. For this purpose, US as a follow-up tool need not only demonstrate sensitivity to change, but should add different and additional prognostic information beyond that obtained through pure clinical measures. Thus at some point a decision will have to be made if US should be generally utilized to monitor disease activity in patients or be confined to only occasional use in follow-up and primarily exploited for its very high diagnostic and differential diagnostic value.

This is a very important research agenda for the near future and the first studies in this regard are currently under way [72, 73]. Nonetheless, a roadmap of studies are needed, including (i) assessment of the metric properties of US scoring systems of synovitis in large, preferably multicentre clinical studies; (ii) assessment of the metric properties of US scoring systems of synovitis in routine practice setting featuring less experienced operators and mid-range US equipment; (iii) validation and further evaluation of multimodal indices combining clinical and US-derived data in both multicentre clinical studies and routine practice; (iv) evaluation of adding US to standard clinical assessment for the monitoring of RA disease.
activity on clinical and radiological outcomes in clinical studies.

In line with the recent EULAR recommendations for the use of imaging of the joints in the clinical management of RA [62], the assessment of algorithms using clinical examination and imaging modalities, particularly ultrasonography, needs to be evaluated for cost-effectiveness and feasibility in clinical practice diagnosis, disease prognosis and outcome measurement in RA.

**Rheumatology key messages**

- Musculoskeletal US is increasingly used for the monitoring of synovitis in patients with RA.
- Multimodal indices need to demonstrate improved sensitivity to change and correlation with long-term outcomes in RA as compared with clinical measures.

**Acknowledgements**

The authors would like to thank Adrienn Kovacs for her assistance.

**Disclosure statement**: The authors have declared no conflicts of interest.

**References**

23. Cheung PP, Dougados M, Gossec L. Reliability of ultrasonography to detect synovitis in rheumatoid arthritis: a...


38 Kawashir SY, Kawakami A, Iwamoto N et al. The power Doppler ultrasonography score from 24 synovial sites or 6 simplified synovial sites, including the metacarpophalan-geal joints, reflects the clinical disease activity and level of serum biomarkers in patients with rheumatoid arthritis. Rheumatology 2011;50:962–5.


50 Klarenbeek NB, Koevoets R, van der Heijde DM et al. Association with joint damage and physical functioning of nine composite indices and the 2011 ACR/EULAR