

## ORIGINAL ARTICLE

## Exploring a new ultrasound score as a clinical predictive tool in patients with rheumatoid arthritis starting abatacept: results from the APPRAISE study

Maria-Antonietta D'Agostino,<sup>1</sup> Maarten Boers,<sup>2</sup> Richard J Wakefield,<sup>3</sup> Hilde Berner Hammer,<sup>4</sup> Olivier Vittecoq,<sup>5</sup> Georgios Filippou,<sup>6</sup> Peter Balint,<sup>7</sup> Ingrid Möller,<sup>8</sup> Annamaria Iagnocco,<sup>9</sup> Esperanza Naredo,<sup>10</sup> Mikkel Østergaard,<sup>11</sup> Corine Gaillez,<sup>12</sup> Manuela Le Bars,<sup>12</sup> on behalf of the OMERACT-EULAR-Ultrasound Task Force

**To cite:** D'Agostino M-A, Boers M, Wakefield RJ, *et al.* Exploring a new ultrasound score as a clinical predictive tool in patients with rheumatoid arthritis starting abatacept: results from the APPRAISE study. *RMD Open* 2016;**2**:e000237. doi:10.1136/rmdopen-2015-000237

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/rmdopen-2015-000237>).

CG: Affiliation at time of study.

Received 6 January 2016  
Revised 13 April 2016  
Accepted 16 April 2016



CrossMark

For numbered affiliations see end of article.

## Correspondence to

Professor Maria-Antonietta D'Agostino; maria-antonietta.dagostino@apr.aphp.fr

## ABSTRACT

**Objectives:** To explore whether changes in a composite (power Doppler/greyscale ultrasound (PDUS)) synovitis score, developed by the OMERACT-EULAR-Ultrasound Task Force, predict disease activity outcomes in rheumatoid arthritis (RA).

**Methods:** Patients with RA who were methotrexate inadequate responders starting abatacept were evaluated. Individual joint PDUS scores were combined in the Global OMERACT-EULAR Synovitis Score (GLOESS) for metacarpophalangeal joints (MCPs) 2–5, all joints (22 paired) and a reduced (9 paired) joint set. The predictive value of changes in GLOESS at week 1–16 evaluations for clinical status and response (Disease Activity Score (DAS) 28 (C reactive protein, CRP) <2.6; DAS28(CRP) <3.2; DAS28(CRP) ≥1.2 improvement) up to week 24, and correlations between DAS28 and GLOESS were assessed.

**Results:** Eighty-nine patients completed the 24-week treatment period. Changes in GLOESS (MCPs 2–5) from weeks 1 to 16 were unable to predict DAS28 outcomes up to week 24. However, significant improvements in GLOESS (MCPs 2–5) were observed at week 12 in patients with DAS28 ≥1.2 improvement at week 24 versus those who did not achieve that clinical response. In patients achieving DAS28 ≥1.2 improvement or DAS28 ≤3.2 at week 24, changes in GLOESS (22 and 9 paired joint sets) were greater in patients who already achieved DAS28 ≥1.2 at week 12 than in those who did not. No significant correlations were found between changes in DAS28 and GLOESS definitions at any time point.

**Conclusions:** PDUS was not correlated with clinical status or response as measured by DAS28-derived criteria, and PDUS changes were not predictive of clinical outcome. The discrepancies require further exploration.

**Trial registration number:** NCT00767325; Results.

## INTRODUCTION

Recently, the European League Against Rheumatism (EULAR) recommended that

## Key messages

## What is already known about this subject?

- Power Doppler with greyscale ultrasound (PDUS) is a non-invasive, bedside, objective and sensitive tool for visualising synovial inflammatory joint changes in rheumatoid arthritis (RA) that were not detected by conventional clinical and radiographic examinations.

## What does this study add?

- The primary analysis of the APPRAISE study of abatacept treatment in RA demonstrated the responsiveness of the composite PDUS Global OMERACT-EULAR Synovitis Score (GLOESS) when applied at a patient level, showing the rapid onset of action of abatacept at 1 week. Although improvements were also demonstrated using DAS28, a clinical improvement of ≥1.2 was reached only after 1 month. The current analyses demonstrated a lack of correlation between GLOESS outcomes and clinical status or response as measured by DAS28-derived criteria, which is an important finding to underline.

## How might this impact on clinical practice?

- The lack of correlation between PDUS and clinical measures suggests that these tools evaluate different aspects of disease activity in RA and should be considered complementary in clinical practice.

the treatment of rheumatoid arthritis (RA) should target remission or low disease activity in every patient, with adjustments in therapy if there is no improvement within 3 months or if the target is not reached within 6 months.<sup>1</sup> To follow this treat-to-target strategy, rheumatologists need to tightly monitor

patients to ensure that they reach the target using clinical indices.

The combined use of power Doppler and greyscale ultrasound (PDUS) represents an easy, non-invasive, bedside imaging modality that has been demonstrated to be an objective and sensitive tool for visualising synovial inflammatory joint changes in RA that were not detected by conventional clinical and radiographic examinations.<sup>2-6</sup> Several factors, such as machine characteristics and operator-dependent interpretation, are known to influence the sensitivity of detecting synovitis by PDUS. Therefore, the Outcome Measures in Rheumatology-Ultrasound (OMERACT-US) Task Force, with funding from EULAR, developed a standardised composite PDUS scoring system for synovitis in RA designed to be applicable to all joints and consistent between machines. To facilitate the assessment of global synovial activity, the group also developed a patient PDUS activity score, the Global OMERACT-EULAR Synovitis Score (GLOESS), calculated from the sum of composite PDUS scores for all joints examined. GLOESS has since been validated in cross-sectional and longitudinal data sets.<sup>7</sup>

APPRAISE was the first prospective, multicentre, international study to use this composite PDUS score at joint and patient levels to measure the early signs of response to treatment with abatacept in biological-naïve adult patients with active RA despite methotrexate (MTX) therapy.<sup>8</sup> This study demonstrated the responsiveness of the composite PDUS GLOESS when applied at a patient level, showing the rapid onset of action of abatacept, independently of the number of joints examined. The responsiveness of GLOESS equalled that of clinical assessment by Disease Activity Score (DAS)28 C reactive protein (CRP). Despite the clear capability of PDUS for monitoring the effects of treatment in RA demonstrated in APPRAISE and other published clinical studies,<sup>9-12</sup> discordant correlations have been found between PDUS scores and clinical outcomes measured at the same time point.<sup>13 14</sup>

The aim of this paper was to present the results of pre-defined secondary, exploratory and post hoc analyses, which investigated whether changes in GLOESS at assessments from weeks 1 to 16 were predictive of clinical response measured by DAS28 at later assessments, and also whether GLOESS could differentiate between multiple definitions of clinical response or status using DAS28 up to week 12 and at week 24.

## METHODS

The APPRAISE study methodology has been reported previously.<sup>8</sup> Briefly, APPRAISE (NCT00767325) was a 24-week, Phase IIIb, open-label, multicentre, single-arm study investigating the responsiveness of the OMERACT-EULAR-composite PDUS score in biological-naïve patients ( $\geq 18$  years) with active RA and an inadequate response to MTX therapy starting abatacept. Patients received intravenous abatacept

(approximately 10 mg/kg) at baseline (day 1), and at weeks 2, 4, 8, 12, 16, 20 and 24, in addition to stable doses of concomitant MTX ( $\geq 15$  mg/week). Oral corticosteroid use (stable dose of  $\leq 10$  mg prednisone/day) was permitted. Patients had American College of Rheumatology (ACR)-defined RA (1987 classification criteria) for at least 6 months, and were receiving MTX ( $\geq 15$  mg/week) for at least 3 months prior to baseline, with a stable MTX dose for at least 28 days before baseline (except in cases of intolerance). Patients were required to have active disease, defined by a baseline DAS28 (CRP) score of  $>3.2$  or tender and swollen joint counts (TJC and SJC) of  $\geq 6$  and a CRP level of greater than the upper normal limit.

## PDUS evaluations

Patients underwent bilateral PDUS examinations of metacarpophalangeal joints (MCPs) 2-5 at screening and baseline, and of 44 (22 paired) joints (MCPs 1-5, proximal interphalangeal joints (PIPs) 1-5, wrist, elbow, shoulder (glenohumeral), knee, ankle (tibiotalar), hind foot (talonavicular and calcaneocuboidal) and metatarsophalangeal joints (MTPs) 1-5) at baseline (day 1), and at weeks 1, 2, 4, 6, 8, 12, 16, 20 and 24. The PDUS examinations were performed at each site by an independent expert in musculoskeletal ultrasound who was blinded to the clinical evaluations. Medium-level to high-level ultrasound machines were used (Esaote Technos MPX, MyLab 70, Toshiba Aplio, GE Logic (series 5, 7, 9 and E 9) or Siemens Acuson Antares), employing high-frequency (12-18 MHz) transducers. Doppler parameters were adjusted according to the device used (range of pulse repetition frequency 400-800 Hz; Doppler frequency 7-11.1 MHz).<sup>8</sup>

The presence of hypoechoic synovial hyperplasia (SH) and joint effusion (JE), both assessed using greyscale, and of synovial vascularisation, assessed using power Doppler (PD), were scored using semiquantitative scales. The presence of synovitis (SH and PD, without JE) was scored for each joint according to the semiquantitative OMERACT-EULAR-US composite PDUS scale, giving a score of 0-3 for each joint. GLOESS was calculated for MCPs 2-5 of both hands and for the 22 paired joints, using the sum of the composite PDUS scores for all joints examined, giving a potential score of 0-24 for MCPs 2-5, and of 0-132 for the 22 paired joints. A new reduced, 9 paired joint set score (including both large and small joints: shoulder, elbow, wrist, MCP1, MCP4, PIP2, knee, MTP3 and MTP5) was also determined using principal component analysis and was found to adequately represent the comprehensive 22 paired joint GLOESS.<sup>8</sup>

## Clinical evaluations

DAS28 (CRP) was evaluated at baseline (day 1) and at weeks 1, 2, 4, 6, 8, 12, 16, 20 and 24. Mean change in DAS28 from baseline, the proportion of patients achieving DAS28 improvement  $\geq 1.2$ , and the proportion of

patients achieving DAS28 <2.6 or DAS28  $\leq$ 3.2 were assessed.

### Statistical analyses

Receiver operating characteristic (ROC) analysis was used to assess whether changes in GLOESS (MCPs 2–5) or in any of the component scores (SH, JE and PD) at any of the assessments from weeks 1 to 16 were predictive of clinical status (DAS28 <2.6; DAS28  $\leq$ 3.2) and response (improvement in DAS28  $\geq$ 1.2) at later time points. An area under the ROC curve of  $\geq$ 0.7 was considered acceptable for prediction of disease activity, as suggested by Hosmer and Lemeshow.<sup>15</sup> Descriptive statistics were used to compare changes from baseline to weeks 12 and 24 in GLOESS (MCPs 2–5) and the component scores in patients who achieved a clinical status or response at week 24 versus those who did not.

Data were analysed to see if PDUS could identify patients with a meaningful clinical status or response at week 12 versus week 24. In patients achieving DAS28  $\leq$ 3.2 or improvement in DAS28  $\geq$ 1.2 at week 24, descriptive statistics were used to assess whether changes in GLOESS at week 12 could differentiate patients who also achieved DAS28  $\geq$ 1.2 improvement at week 12 from those who did not. These analyses were completed using MCPs 2–5, the 22 paired and reduced 9 paired joint sets.

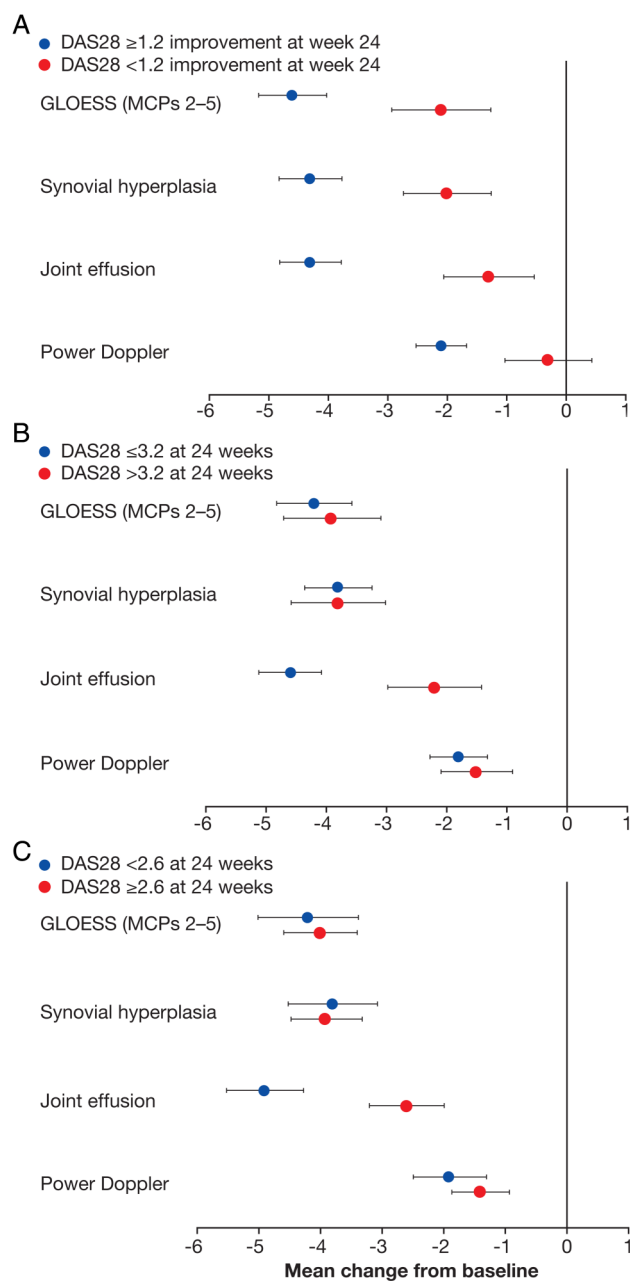
### Correlation analyses

To explore further the relationship between GLOESS and clinical response, additional post hoc correlation analyses were performed using Pearson's (parametric) or Spearman's (non-parametric) correlation coefficient with 95% CIs. Correlations between changes at different time points were assessed within and between GLOESS and DAS28 outcomes. Details of the correlation analyses performed are presented in online supplementary table S1. Effect size, expressed as standardised response means (SRM) of GLOESS and DAS28, was investigated on the basis of mean changes from baseline to weeks 1, 12 and 24. Correlation analyses using absolute values and changes from baseline were completed with the last observation carried forward to quantify the treatment effect of abatacept plus MTX over time. Analyses were performed using all available GLOESS and component scores (no imputation of missing data). Missing DAS28 values were imputed. No correction for multiple statistical tests was applied.

## RESULTS

In total, 89 (86%) of the 104 patients enrolled between December 2008 and October 2011 completed the 24-week, open-label treatment period. Demographic and baseline characteristics of the study population have been reported previously, as well as the results on the responsiveness of the score.<sup>8</sup> In patients with DAS28 measurements available at week 12 and week 24 ( $n=98$ , which included 9 patients with DAS28 measurements

but who had not completed the open-label treatment period), the number of patients who achieved DAS28  $\geq$ 1.2 improvement at week 12 and who also achieved DAS28  $\geq$ 1.2 improvement or DAS28  $\leq$ 3.2 at week 24 were 50/98 (51%) and 37/98 (38%), respectively, while the number of patients who did not achieve DAS28  $\geq$ 1.2 improvement at week 12 but achieved DAS28  $\geq$ 1.2



**Figure 1** Mean change from baseline in GLOESS and components at week 12 in patients who did, or did not, achieve (A) DAS28  $\geq$ 1.2 improvement, (B) DAS28  $\leq$ 3.2 or (C) DAS28 <2.6 at week 24. Error bars represent SEM. DAS28, Disease Activity Score 28; GLOESS, Global OMERACT-EULAR Synovitis Score; MCPs, metacarpophalangeal joints. GLOESS and the component scores were calculated for MCPs 2–5 using both hands, giving a potential score of 0–24.

**Table 1** Mean change in GLOESS from baseline to weeks 12 and 24 in patients with DAS28  $\leq 3.2$  at week 24, with or without DAS28  $\geq 1.2$  improvement at week 12

		DAS28 $\geq 1.2$ improvement at week 12 and DAS28 $\leq 3.2$ at week 24	DAS28 $< 1.2$ change at week 12 and DAS28 $\leq 3.2$ at week 24
MCPs 2–5 joint set	Week 12	-4.2 (-5.8 to -2.5)	-4.3 (-6.2 to -2.4)
	Week 24	-5.5 (-7.0 to -4.0)	-6.9 (-10.6 to -3.2)
22 paired joint set	Week 12	-12.6 (-17.2 to -8.0)	-10.6 (-16.7 to -4.5)
	Week 24	-17.1 (-22.4 to -11.8)	-18.2 (-28.6 to -7.8)
Reduced (9 paired joints) set	Week 12	-6.1 (-8.0 to -4.1)	-4.3 (-6.2 to -2.4)
	Week 24	-7.3 (-9.6 to -5.1)	-6.7 (-11.1 to -2.3)

Data are mean change (95% CI).

DAS28, Disease Activity Score 28; GLOESS, Global OMERACT-EULAR Synovitis Score; MCP, metacarpophalangeal joint.

improvement or DAS28  $\leq 3.2$  at week 24 were 10/98 (10%) and 10/98 (10%), respectively.

### Relationship between GLOESS and meaningful clinical status or response

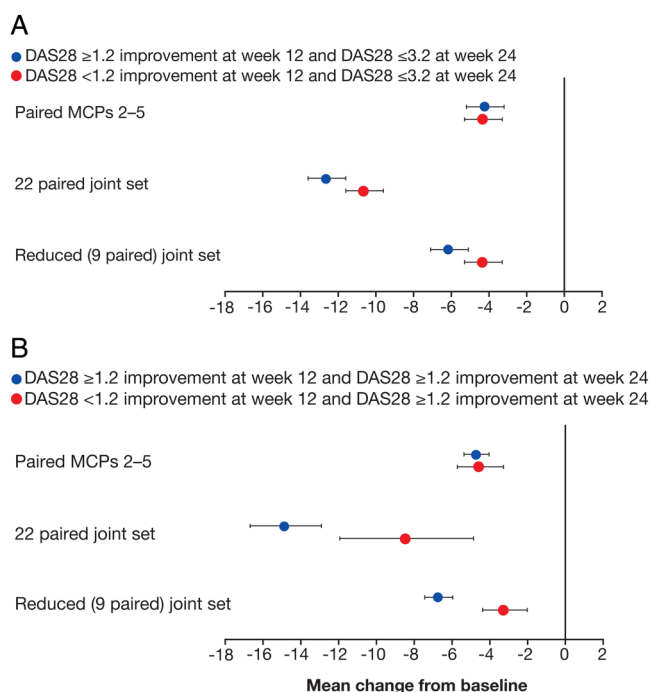
All patients showed an improvement in GLOESS (MCPs 2–5) at week 12, regardless of clinical status or response at week 24 (figure 1). In particular, significantly greater improvements were observed from baseline to week 12 in GLOESS and in all three component scores (SH, JE and PD) in patients who achieved DAS28  $\geq 1.2$  improvement at week 24 compared with those who did not (figure 1A). However, the ROC area under the curve never reached or exceeded the predefined cut-off of 0.7; therefore, changes in GLOESS and in all component scores for MCPs 2–5 at any time point up to week 16 were unable to adequately predict clinical status or response at week 24 or at any other time point, regardless of criteria used. Moreover, no difference was observed in the mean decrease of GLOESS (MCPs 2–5) from baseline to week 12 in patients who achieved or did not achieve DAS28  $< 2.6$  or DAS28  $\leq 3.2$  at week 24 (figure 1B, C). The only component score that was different in patients who reached DAS28  $< 2.6$  or DAS28  $\leq 3.2$  at week 24 versus those who did not was the JE score. Similar results were observed using the 22 paired and 9 paired joint sets (table 1).

In patients who achieved DAS28  $\leq 3.2$  or DAS28  $\geq 1.2$  improvement at week 24, changes from baseline to week 12 in GLOESS for the 22 paired and 9 paired joint sets were able to detect some differences between patients who achieved DAS28  $\geq 1.2$  improvement at week 12 versus those who did not (figure 2).

### Correlation analyses

Previous correlation analyses demonstrated the absence of correlations between changes in DAS28 from baseline with changes in GLOESS or component scores; very low correlations between GLOESS (MCPs 2–5) or component scores and swollen joint counts (SJC) and/or tender joint counts (TJC); no correlations between early changes in GLOESS or components and changes in the sum of swollen joints at week 12 or 24.<sup>8</sup>

When we looked at correlations within each assessment method (ie, between DAS28 scores at different time points or between GLOESS scores at different time points), moderate-to-high positive correlations were observed between improvements up to week 12 and improvements at week 24 in DAS28 (table 2). Moderate-to-high positive correlations were also observed between changes in GLOESS from baseline to weeks 2, 4, 8 and



**Figure 2** Mean change from baseline in GLOESS at week 12 in patients who did, or did not, also achieve DAS28  $\geq 1.2$  improvement, in (A) patients who achieved DAS28  $\leq 3.2$  at week 24 or (B) patients who achieved DAS28  $\geq 1.2$  improvement at week 24. Patients who discontinued due to lack of efficacy, adverse event or withdrawal of consent are considered as non-responders. Error bars represent SEM. DAS28, Disease Activity Score 28; GLOESS, Global OMERACT-EULAR Synovitis Score; MCPs, metacarpophalangeal joints. GLOESS was calculated for each joint set using both hands, giving a potential score of 0–24 for MCPs 2–5, 0–132 for the 22 paired joint set, and 0–54 for the reduced (9 paired) joint set.

**Table 2** Correlation between changes from baseline in DAS28 up to week 12 and change from baseline to week 24

	n	Pearson's correlation coefficient (95% CI)
Baseline to week 1	82	0.37 (0.16 to 0.54)
Baseline to week 2	84	0.44 (0.25 to 0.60)
Baseline to week 4	87	0.61 (0.45 to 0.72)
Baseline to week 8	86	0.66 (0.51 to 0.76)
Baseline to week 12	87	0.71 (0.59 to 0.80)

DAS28, Disease Activity Score 28.

12 and change in GLOESS (both 22 paired and 9 paired joint sets) from baseline to week 24 (table 3). However, changes in GLOESS during the first week of treatment were weakly correlated with overall change in GLOESS during the 24-week treatment period. Overall, SRM values for GLOESS (MCPs 2–5, 22 paired and reduced 9 paired joint sets) were smaller than SRM values for DAS28 for mean change from baseline at weeks 1, 12 and 24 (see online supplementary tables S1 and S2). A summary of the findings of the correlation analysis is presented in table 4.

## DISCUSSION

This study demonstrated an almost complete absence of association between ultrasound and clinical status or response in RA in patients starting treatment with abatacept. Whereas both modalities (PDUS and composite DAS) were responsive to treatment in this setting,<sup>8</sup> the extent of response in one modality was not reflected in the extent of improvement in the other. In addition, early changes in GLOESS were not predictive of a good clinical status or response at 24 weeks. This lack of correlation or predictive capacity was not caused by inconsistency of results over time, as correlations within modality (ie, between early and late GLOESS, and between early and late DAS28 outcomes) were moderate to strong, as were SRM values. This lack of correlation is a novel finding, although previous studies have reported a lack of complete overlap between ultrasound-assessed disease activity and composite clinical measures,<sup>16–18</sup> with one study suggesting that the simplified disease activity index is closer to PDUS assessment of disease state than DAS28.<sup>19</sup> Another study found that ultrasound had low correlations with disease activity assessment at

baseline, while after 12 months of adalimumab treatment, correlation coefficients had improved.<sup>20</sup>

Several explanations can be discussed for interpreting these results. First, the sensitivity to change as estimated by SRM calculations (which was better for DAS28 than for GLOESS, whatever the joint set) can suggest that clinical efficacy measures were more sensitive to change than ultrasound. However, these discrepancies can also be explained by the wider SD for PDUS assessments than DAS28, as the SRM calculation is dependent on the SD of the measure.<sup>21</sup> Second, we identified some patients who appeared to have a complete disconnect between GLOESS (and component scores) and DAS28, SJC or TJC (data not shown). This may be explained by the greater sensitivity of PDUS for detecting minimal and/or subclinical synovitis compared with clinical examination. A recent paper confirmed that, in patients in disease remission (clinical disease activity index  $\leq 2.8$ ), only grade 3 (severe) PDUS-assessed synovitis correlated with clinical examination and clinical activity scores, demonstrating that when clinical evaluation is taken as the gold standard, the capability of PDUS is reduced.<sup>22</sup> A number of patients with RA can score pain and TJC higher than PDUS assessments would indicate. Finally, the severity of PDUS synovitis (MCPs 2–5) was relatively low (grade 1–2) at baseline, supporting the apparent discordance between clinical and PDUS evaluations when assessing whether a joint was inflamed or not. This latter aspect requires further investigation. Overall, this study suggests that PDUS and DAS28 capture different aspects of disease activity that present different kinetics of response over time.

The findings of this study should be considered as exploratory, and the limitations of the study have been discussed previously and include the single-arm, open-label design, the sample size and the variables relating to the technology used.<sup>8</sup> The analyses presented here are intended to inform future prospective studies of PDUS in RA. The present study suggests that PDUS scores improve as a consequence of effective treatment, but DAS28 cannot be used as a comparator for ultrasound, as they do not reflect the same aspects of the arthritis inflammatory process. The use of other measures of disease activity that are more stringent than DAS28, such as Clinical Disease Activity Index (CDAI) and Simple Disease Activity Index (SDAI), would probably provide better correlation results.

**Table 3** Correlation between changes from baseline in GLOESS up to week 12 and change from baseline to week 24

	Pearson's correlation coefficient (95% CI)			
	n	22 paired joint set	n	Reduced (9 paired) joint set
Baseline to week 1	79	0.29 (0.07 to 0.48)	75	0.27 (0.05 to 0.47)
Baseline to week 2	86	0.37 (0.17 to 0.54)	82	0.35 (0.14 to 0.52)
Baseline to week 4	84	0.61 (0.46 to 0.73)	79	0.53 (0.35 to 0.67)
Baseline to week 8	82	0.62 (0.46 to 0.74)	77	0.61 (0.44 to 0.73)
Baseline to week 12	82	0.71 (0.57 to 0.80)	81	0.62 (0.46 to 0.74)

GLOESS, Global OMERACT-EULAR Synovitis Score.

**Table 4** Correlation analysis summary

Variables	Correlation
Changes in DAS28 up to week 12 vs week 24	Moderate-to-high positive correlations ( <a href="#">table 2</a> )
Changes in GLOESS (22 paired joints, reduced 9 paired joint set) from baseline up to week 12 vs baseline to week 24	Moderate-to-high positive correlations ( <a href="#">table 3</a> )
Changes in GLOESS (MCPs 2–5, 22 paired joints, reduced 9 paired joint set) from baseline up to week 12 and lower levels of synovitis* at week 24	Low-to-moderate negative correlations (see online supplementary table S2)
Changes in GLOESS at week 1 and changes in GLOESS during the entire treatment period (baseline to week 24)	Low correlation
Changes in DAS28 and changes in GLOESS (MCPs 2–5, 22 paired joints, reduced 9 paired joint set, 28 joints†) or component scores (SH, PD, JE) from baseline	No correlation at any time points
GLOESS (MCPs 2–5) or component scores (SH, PD, JE) vs SJC	Very low correlation at all time points
GLOESS (MCPs 2–5) or component scores (SH, PD, JE) vs TJC	Very low correlation at all time points
GLOESS (MCPs 2–5) or component scores (SH, PD, JE) vs SJC + TJC combined	Very low correlation at all time points
Changes in GLOESS (MCPs 2–5, 22 paired joints, reduced 9 paired joint set) or components (SH, PD, JE) up to week 4 vs change in the sum of swollen joints (MCPs 2–5, 22 paired joints, reduced 9 paired joint set) at weeks 12 and 24	No correlation

\*Defined by a GLOESS at week 24 that was less than the median GLOESS at that time point.

†The 28 joints assessed for tenderness and swelling as part of the DAS28.

DAS28, Disease Activity Score 28; GLOESS, Global OMERACT-EULAR Synovitis Score; JE, joint effusion; MCPs, metacarpophalangeal joints; PD, power Doppler-assessed synovial vascularisation; SH, hypoechoic synovial hyperplasia; SJC, swollen joint count; TJC, tender joint count.

In conclusion, PDUS was a responsive tool of disease activity in patients with RA starting abatacept plus MTX, but the extent of PDUS response was not correlated with clinical status or response, as measured by DAS28 derived criteria, and early PDUS changes were not predictive of a later good clinical outcome. This suggests that PDUS adds independent information on treatment response, and its contribution should be explored further.

#### Author affiliations

<sup>1</sup>APHP, Hôpital Ambroise Paré, Service de Rhumatologie, 92100 Boulogne-Billancourt; INSERM U1173, Laboratoire d'Excellence INFLAMEX, Université Paris Ouest-Versailles St-Quentin, 78180 Saint Quentin en Yvelines, France

<sup>2</sup>Departments of Epidemiology & Biostatistics, The Amsterdam Rheumatology & Immunology Center, VU University Medical Center, Amsterdam, The Netherlands

<sup>3</sup>Leeds Institute of Rheumatic and Musculoskeletal Medicine and NIHR Leeds Musculoskeletal Biomedical Research Unit (LMBRU), University of Leeds, Leeds, UK

<sup>4</sup>Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway

<sup>5</sup>Rheumatology Department & CIC-CRB-1404, Inserm U905, Rouen University Hospital, Rouen University, Rouen, France

<sup>6</sup>Department of Medicine, Surgery and Neurosciences, Rheumatology Section, University of Siena, Siena, Italy

<sup>7</sup>3rd Rheumatology Department, National Institute of Rheumatology and Physiotherapy, Budapest, Hungary

<sup>8</sup>Rheumatology Department, Instituto Poal, Barcelona, Spain

<sup>9</sup>Rheumatology Unit, Dipartimento di Medicina Interna e Specialità Mediche, Sapienza Università di Roma, Roma, Italy

<sup>10</sup>Department of Rheumatology, Hospital GU Gregorio Marañón, Madrid, Spain

<sup>11</sup>Copenhagen Center for Arthritis Research, Center for Rheumatology and Spine Diseases, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

<sup>12</sup>Medical Affairs, Bristol-Myers Squibb, Rueil-Malmaison, France

**Acknowledgements** The authors would like to thank the OMERACT-EULAR-US Task Force and principal investigators of the APPRAISE study. APPRAISE principal investigators: Silvano Adami, Vivi Bakkenheim, HBH, Stefano Bombardieri, M-AD'A, Paul Emery, Liana Euler-Ziegler, Gianfranco Ferraccioli, Maurizio Galeazzi, Philippe Gaudin, Walter Grassi, Annamaria Iagnocco, Herbert Kellner, Thierry Lequerré, IM, EN, MØ, Fredeswinda Romero, Istvan Szombati, Lene Terslev, Jacqueline Uson, Esther Vicente, OV and RW. The authors also thank Emilie Barré, Karina Van Holder, Wendy Kerselaers, Harry Goyvaerts, Stephane Munier and Nathalie Schmidely from Bristol-Myers Squibb, Coralie Poncet from DOCS International and Christel Perrone from the CRO ICON for their contribution to the study design, analysis and study conduct. Professional medical writing and editorial assistance was provided by Gary Burd, PhD, CMPP, and Laura McDonagh, PhD, at Caudex Medical and was funded by Bristol-Myers Squibb.

**Funding** This study was funded by Bristol-Myers Squibb.

**Competing interests** M-AD'A has received speaker's bureau fees from Bristol-Myers Squibb, AbbVie, UCB, MSD, Novartis and Roche Pharma, as well as research grants from Pfizer. MB has received consulting fees from Bristol-Myers Squibb and personal fees from Mundipharma, Roche, Pfizer, GSK and Novartis. HBH has received honoraria for scientific lectures from AbbVie, Roche, BMS, Pfizer and MSD, as well as grants for scientific studies from AbbVie, Roche and Pfizer. IM has received consulting fees from Bio Iberica pharma, AbbVie, GE and ESAOTE. EN has received consulting fees from AbbVie, Roche Pharma, Bristol-Myers Squibb, Pfizer, UCB, General Electric and Esaote, as well as research grants from MSD. MØ has received personal fees from AbbVie, BMS, Boehringer-Ingelheim, Celgene, Eli Lilly, Janssen, Genmab, GSK, Merck, Pfizer, Regeneron, Roche, Sanofi and UCB, grants from AbbVie, Merck and UCB, as well as non-financial support from AbbVie, BMS, Janssen, Merck, Pfizer, Roche and UCB. CG is a former employee of Bristol-Myers Squibb. MLB is an employee of Bristol-Myers Squibb.

**Ethics approval** Hôpital Ambroise Paré, Comité De Protection Des Personne IdF VIII Lab D'Anatomopathologie, 9 Ave Charles De Gaulle, Boulogne-Billancourt 92100, France. Comitato Etico Per La Sperimentazione Del Farmaco, Asur, Territoriale 5 Di Jesi Via Gallodoro 68, Jesi (An) 60035,

Italy. Azienda Ospedaliera Universitaria Policlinico G. Martino, Comitato Etico Scientifico Via Consolare Valeria, 1, Messina 98124, Italy. Leeds East Research Ethics Committee, Yorkshire andamp; Humber Rec Cntr Off. First Floor, Millside Mill Pond Lane, Leeds, England LS6 4RA, UK. Ceic Fundacio Unio Catalana D'Hospitals, Area De Serveis C/ Bruc 72-74 1a, Barcelona 08009, Spain. Azienda Universitaria Senese, Comitato Etico Locale La Sperimentazione Clinica Dei Medicinali Farmacia Aous—Viale Bracci, Siena 53100, Italy. De Videnskabetiske Komiteer For Region Hovedstaden, Kongens Vaenge 2, Hillerod 3400, Denmark. Com Sperimentazione Clin Dei Med Azienda Osp-Univ Pisana, Via Roma 67, Pisa 56126, Italy. Comitato Etico Dell'Universita Cattolica Del Sacro Cuore, Policlinico Universitario Agostino Gemelli Di Roma Largo A Gemelli 8, Roma 00168, Italy. Rek Sorost, Frederik Holsts Hus Ullevall Terrasse Ullevall Sykehus Kirkeveien 166, Oslo 0450, Norway. Ceic—Fundacion Jimenez Diaz- Ute, Avda. Reyes Catolicos, 2-2a, Madrid 28040, Spain. Hospital Mostoles, Ceic Area 8 C/ Rio Jucar S/N, Madrid 28935, Spain. Universita Di Roma La Sapienza, Comitato Etico Dell'Azienda Policlinico Umberto I, Roma 00161, Italy. Ceic Area 9-Hosp Severo Ochoa De Leganes Y Hosp De Fuenlabrad, Avenida Orellana S/N Leganes, Madrid 28911, Spain. Hospital Universitario La Princesa, C/ Diego De Leon, 62, Madrid 28006, Spain. Farmakologiai Etk Bizottsaga, Arany J.U. 6-8., Budapest 1051, Hungary. Comitato Etico Per La Sperimentazione Dell'Azienda, Ospedaliera Istituti Ospitalieri Di Verona Piazzale Stefani 1, Verona 37126, Italy. Ethikkommission Der Ludwig-Maximilians Universitaet, Marchioninstr. 15, Muenchen 81377, Germany.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** No additional data are available.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

## REFERENCES

- Smolen JS, Landewé R, Breedveld FC, *et al*. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2013 update. *Ann Rheum Dis* 2014;73:492–509.
- Grassi W, Salaffi F, Filippucci E. Ultrasound in rheumatology. *Best Pract Res Clin Rheumatol* 2005;19:467–85.
- Kane D, Grassi W, Sturrock R, *et al*. Musculoskeletal ultrasound—a state of the art review in rheumatology. Part 2: clinical indications for musculoskeletal ultrasound in rheumatology. *Rheumatology (Oxford)* 2004;43:829–38.
- Lopez-Ben R, Bernreuter WK, Moreland LW, *et al*. Ultrasound detection of bone erosions in rheumatoid arthritis: a comparison to routine radiographs of the hands and feet. *Skeletal Radiol* 2004;33:80–4.
- Østergaard M, Szkudlarek M. Ultrasonography: a valid method for assessing rheumatoid arthritis? *Arthritis Rheum* 2005;52:681–6.
- Wakefield RJ, Gibbon WW, Conaghan PG, *et al*. The value of sonography in the detection of bone erosions in patients with rheumatoid arthritis: a comparison with conventional radiography. *Arthritis Rheum* 2000;43:2762–70.
- Naredo E, Wakefield RJ, Iagnocco A, *et al*. The OMERACT ultrasound task force—status and perspectives. *J Rheumatol* 2011;38:2063–7.
- D'Agostino MA, Wakefield RJ, Berner-Hammer H, *et al*. 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* 2015.
- Backhaus TM, Ohrndorf S, Kellner H, *et al*. The US7 score is sensitive to change in a large cohort of patients with rheumatoid arthritis over 12 months of therapy. *Ann Rheum Dis* 2013;72:1163–9.
- Filippucci E, Iagnocco A, Salaffi F, *et al*. Power Doppler sonography monitoring of synovial perfusion at the wrist joints in patients with rheumatoid arthritis treated with adalimumab. *Ann Rheum Dis* 2006;65:1433–7.
- Dougados M, Devauchelle-Pensec V, Ferlet JF, *et al*. The ability of synovitis to predict structural damage in rheumatoid arthritis: a comparative study between clinical examination and ultrasound. *Ann Rheum Dis* 2013;72:665–71.
- Kume K, Amano K, Yamada S, *et al*. Very early improvements in the wrist and hand assessed by power Doppler sonography predicting later favorable responses in tocilizumab-treated patients with rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2011;63:1477–81.
- Jousse-Joulin S, d'Agostino MA, Marhadour T, *et al*. Reproducibility of joint swelling assessment by sonography in patients with long-lasting rheumatoid arthritis (SEA-Repro study part II). *J Rheumatol* 2010;37:938–45.
- Marhadour T, Jousse-Joulin S, Chales G, *et al*. Reproducibility of joint swelling assessments in long-lasting rheumatoid arthritis: influence on Disease Activity Score-28 values (SEA-Repro study part I). *J Rheumatol* 2010;37:932–7.
- Hosmer D, Lemeshow S. *Applied logistic regression*. 2nd edn. New York, NY: John Wiley & Sons, Inc, 2000.
- Brown AK, Conaghan PG, Karim Z, *et al*. An explanation for the apparent dissociation between clinical remission and continued structural deterioration in rheumatoid arthritis. *Arthritis Rheum* 2008;58:2958–67.
- Dejaco C, Duftner C, Wipfler-Freissmuth E, *et al*. Ultrasound-defined remission and active disease in rheumatoid arthritis: association with clinical and serologic parameters. *Semin Arthritis Rheum* 2012;41:761–7.
- Saleem B, Brown AK, Keen H, *et al*. Should imaging be a component of rheumatoid arthritis remission criteria? A comparison between traditional and modified composite remission scores and imaging assessments. *Ann Rheum Dis* 2011;70:792–8.
- Balsa A, de Miguel E, Castillo C, *et al*. Superiority of SDAI over DAS-28 in assessment of remission in rheumatoid arthritis patients using power Doppler ultrasonography as a gold standard. *Rheumatology (Oxford)* 2010;49:683–90.
- Hammer HB, Sveinsson M, Kongtorp AK, *et al*. A 78-joints ultrasonographic assessment is associated with clinical assessments and is highly responsive to improvement in a longitudinal study of patients with rheumatoid arthritis starting adalimumab treatment. *Ann Rheum Dis* 2010;69:1349–51.
- Husted JA, Cook RJ, Farewell VT, *et al*. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000;53:459–68.
- Gärtner M, Mandl P, Radner H, *et al*. Sonographic joint assessment in rheumatoid arthritis: associations with clinical joint assessment during a state of remission. *Arthritis Rheum* 2013;65:2005–14.