Concise report

Ultrasound Doppler measurements predict success of treatment with anti-TNF-α drug in patients with rheumatoid arthritis: a prospective cohort study

Karen Ellegaard1, Robin Christensen1,2, Søren Torp-Pedersen1, Lene Terslev1, Christian C. Holm1, Merete J. Kønig1, Peter S. Jensen1, Bente Danneskiold-Samsøe1,3 and Henning Bliddal1,3

Abstract

Objective. To investigate the predictive ability of core outcomes applied in RA trials, including ultrasound (US) Doppler (USD) measurements differentiating patients who remain on anti-TNF-α therapy following 1 year.

Methods. Patients with RA in anti-TNF-α therapy were followed 1 year after therapy initiation. All patients had wrist involvement. At baseline, 2 weeks, 26 weeks and 1 year a USD examination, clinical examination including tender and swollen joint count, visual analogue scale (VAS) global and HAQ, biochemical measures and 28-joint DAS (DAS28) were collected for all patients. The amount of USD signal in the synovium was quantified by measuring the percentage of colour pixels—the colour fraction (CF). Predictive validity for patients who remain on anti-TNF-α therapy after 1 year was assessed for both USD measurements and other disease measures. Baseline values of disease measures of patients who remained on treatment after 1 year was compared with those who stopped therapy.

Results. The study cohort consisted of 109 patients. In this study, the baseline CF was the only measure predicting which patients would stay on the initial anti-TNF-α therapy for 1 year, evaluated using the square-root of CF (P = 0.024). The other disease markers could not significantly differentiate between the two groups of patients, with P-values of 0.86 and 0.98 for tender and swollen joint count, respectively, 0.86 for CRP, 0.24 for VAS, 0.10 for HAQ and 0.38 for DAS28.

Conclusion. There is now evidence to support that baseline USD, in contrast to clinical measures, can predict which patients will remain on anti-TNF-α 1 year after initiating therapy.

Key words: Rheumatoid arthritis, Ultrasound, Wrist, Inflammation, Synovium.

Introduction

Anti-TNF-α agents are effective in many patients with RA [1, 2], while not effective in all [3], with predictors of response being necessary [4]. Ultrasound (US) Doppler (USD) has been suggested as a further predictor of disease activity [5]. Cross-sectional studies have shown concurrent validity between Doppler and other validated measures of disease activity [6–13]. Furthermore, USD has been applied in the assessment of patients with RA in treatment with anti-TNF-α, showing the ability of USD measurements to aid in the monitoring of treatment [14–21].

USD assessment of a single wrist joint has good concurrent validity when compared with validated measures of disease activity in RA [13]. Approximately 70% of patients with RA have affection of a wrist joint [22] and the wrist joint is therefore well-suited for estimation of arthritis activity by imaging [23].

Predictive validity refers to a measurement ability to forecast future events [24, 25]. Part of the assessment
Predictive value of Doppler US in patients with RA

The aim of this study was to investigate the ability of clinical measures and USD measurements in a wrist joint to predict which patients with RA will benefit from treatment with anti-TNF-α in terms of staying on therapy for at least 1 year. A second objective was to investigate the ability of USD measurements to monitor changes over time.

Methods

Patients

Patients with RA treated with an anti-TNF-α drug (adalimumab, etanercept or infliximab) were studied, all fulfilling the ACR criteria for RA [26]. The patients were enrolled consecutively in an outpatient clinic in 2003–07 as they started treatment. The patients were investigated at baseline and 1 year onwards with clinical examinations, blood tests and a USD examination of an individually defined target joint. The target joint was the joint with most pronounced synovial hyperaemia on USD; if possible, the most affected wrist. Only patients with wrist involvement were included in our data analysis. Both the radio-carpal and inter-carpal compartments of the wrist joint were included in the evaluation. Only the dorsal aspect of the wrist was evaluated to reduce time spent on the US examination. In our experience, the dorsal aspect of the joint is more frequently involved than the volar aspect. The local ethics committee (the committees on Biomedical Research Ethics for the Capital region of Denmark) approved the study (KF01-045/03) and informed consent was obtained.

Clinical and para-clinical examination

At each visit, the patients underwent assessment of the number of tender and swollen joints by a rheumatologist unaware of the results of the US examination. Blood was tested for ESR and CRP; the patients filled in an HAQ and a general health score on a visual analogue scale (VAS) global and a 28-joint DAS (DAS28)–CRP was calculated [27].

US examination

Scanning was performed with a US machine (Siemens, Mountainview, CA) using a linear array transducer with 14 MHz centre frequency. The Doppler pre-set was made according to the recommendations made by Torp-Pedersen and Terslev [28]. No adjustments of Doppler parameters were performed. The dorsal wrist was scanned longitudinally from side to side and images were obtained in the radial, central and ulnar positions. In the three positions, the area with the most pronounced Doppler activity was identified, the transducer was held in this position for a couple of heart cycles whereupon the image was frozen. The image with most Doppler activity was then selected and stored. This image acquisition technique has shown a test-retest reliability intraclass correlation coefficient (ICC) of 0.77 [29].

Four persons performed the US examinations (K.E., S.T.-P., L.T. and M.J.K.). One investigator (S.T.-P.) is the head of the US unit, with 20 years US experience; the other investigators had several years of US training and 1 month of specific training on the wrist scans.

Image analysis

USD activity was quantified by the colour fraction (CF) [6, 30]. The CF is the number of colour pixels divided by the total number of pixels in a region of interest (ROI). We defined ROI as the synovial tissue seen as a predominantly hypo-echoic mass located dorsally to the ulnar head and the radio-carpal and inter-carpal joints. In all three positions, the synovial tissue was traced and the CF was calculated (supplementary fig. 1, available as supplementary data at Rheumatology Online). Subsequently, the average of the three CF values was computed. In a previous study, this evaluation technique showed an excellent ICC [30].

Two investigators (K.E. and P.S.J.) with long-standing experience in image evaluation performed all image analyses blinded to the patients’ clinical characteristics. The CF calculation was made in Datapro (DataPro; Noesis Courtabœuf, France).

End points

It is an underlying assumption that if patients continue on a certain therapy for a period of time it is because there is both an effect of the drug and it is well tolerated. Thus, the continuation of a therapy can be used as a proxy for overall effectiveness and safety [31]. This assumption is supported by the fact that the main reason for patients with RA to discontinue a treatment is lack of efficacy [31]. Furthermore, correlation between withdrawal and ACR-20 response has been demonstrated [32].

Statistical analyses

In order to evaluate which, if any, of the variables measured at baseline could predict patients’ survival on therapy for 1 year, we used Still-on-therapy as a dichotomization factor with two levels (yes/no). To evaluate the impact (i.e., clinical significance) of each of these possible predictive variables, we calculated the so-called effect size (ES) being the Cohen’s index or the standardized mean difference—calculated as the mean difference between the groups (Still-on-therapy: yes vs no) divided by the s.d. in the total data set [33]; i.e., the higher the absolute value of the ES—the more predictive. Subsequently, based on the patients still on therapy after 1 year, we wanted to evaluate the discriminant capacity for each of the variables included [34]. This was evaluated using the standardized response mean (SRM).

Results

Patient characteristics

One hundred and sixty-two outpatients started on anti-TNF-α therapy; of these, 109 had wrist involvement.
At 1 year follow-up, 78 patients were still on therapy with the initial anti-TNF-α. These 78 were categorized as completers. The 31 patients who discontinued treatment were considered drop-outs. The reasons for drop-out were lack of efficacy (23 patients) or side effects (8 patients).

Baseline data of the 109 patients are shown in Table 1. Predominantly females (71%) participated. The mean age was 58 years (range 26–84 years). The mean disease duration was 10.4 years (range 1–34.6 years), mean baseline CF was 0.24 (range 0.000–0.690) and mean DAS28 5.07 (range 1.40–7.80). When applying the EULAR DAS28 criteria for assessment of disease activity at baseline [34, 42] 4 had a DAS28 < 2.6 (2 completers and 2 drop-outs); 7 had < 3.2 (6 completers and 1 drop-out) and 33 had < 5.1 (19 completers and 14 drop-outs); the remaining 64 patients had a DAS28 > 5.1 (47 completers and 17 drop-outs). No statistically significant difference in concomitant therapies was seen between the completer and the drop-out groups (Table 1).

Compliers vs drop-outs in measures of disease activity

The CF had a low efficiency and poor consistency due to a large statistic variation. Following assessment of multiple transformations, it was evident that a square-root transformation resulted in an unbiased estimate with improved efficiency and better consistency (data not shown). Thus, we use the square-root CF (√CF) as the primary US variable. As presented in Table 1, the baseline √CF could predict those remaining on therapy (P = 0.024). Also, the CF was statistically significant (P = 0.020). No other statistically significant differences were found between the two groups. Furthermore, the square-root transformed CF had a moderate-to-large clinical predictive value, higher than any other measurements, in terms of differentiating patients who would still be on therapy after 1 year (ES = 0.57)—indicating that patients with low (or no) CF probably would withdraw from therapy.

In order to explore thresholds, an iterative post hoc analysis showed that the best cut-off for predicting those completing 1-year therapy were those with a √CF > 0.23 (i.e. CF > 0.55). As indicated in Table 1, the majority of the included patients had a CF value of at least 5% at baseline before initiating biologics. When using this threshold, most patients would remain on therapy (71.6%); of course, this threshold is to be considered preliminary, and need confirmation in a prospective trial.

At 1-year follow-up, USD data were only available for 69 patients, allowing us to only calculate a change score for these. Table 2 presents the discriminant capacity for the various outcome variables varied. It was evident that DAS28 was the most discriminative outcome measure when assessing changes following therapy—with an SRM of −1.45 (95% CI −1.83, −1.07)—having the best signal-to-noise ratio. In comparison, the SRM for √CF was −0.62 (95% CI −0.91, −0.32) and CF was −0.52 (95% CI −0.81, −0.23).

Discussion

The main finding in this prospective cohort study was that USD measurement using CF obtained at baseline was the only outcome measure that could significantly predict which patients would remain on anti-TNF-α therapy. The ES for the CF was substantially higher than for any of the other disease activity measures [35]. CF gives the clinician a measure of hyperaemia, an integral part of inflammation, and the values are readily perceptible as high values of CF equal high inflammatory activity. It seems natural that an individual with high baseline inflammatory activity would benefit most from anti-inflammatory medication. We anticipated that DAS28 also could predict which patients would remain on therapy, but our results could not confirm this. This finding is in accordance with another study showing lack of predictive value of DAS28 [36]. Since lack of efficacy is the main reason for dropping out of therapy [31], this suggests that CF can be used as predictor of anti-TNF-α treatment success in patients with RA [31, 32].

Some overlap between the responders and non-responders was seen. This overlap naturally weakens the value of the CF as a predictive marker in a clinical setting; however, compared with the other markers of disease activity, the ES of CF was considerably higher. It can be argued that estimation of CF is too time consuming to be used in daily clinical praxis. However, software that allows estimation of the CF on the US machine is under preparation. This application will make it possible to estimate the CF at the US examination.

Some limitations of USD measurements exist when they are assessed according to the OMERACT filter [34]. There is very sparse information of US findings in healthy persons [37, 38]; differences in Doppler findings due to differences in machines or settings remain to be addressed [28] and furthermore, there is no consensus in the use of scoring systems. The perfusion in the synovial tissue is estimated from the part of the synovial area covered by colour. When given as a fraction, a theoretical bias is that a decrease in perfusion may be accompanied by a shrinking in synovial volume; this would result in an increase in the relative amount of colour, and a benefit of treatment might be overlooked, no matter which scoring system is applied. The only way to overcome the error caused by shrinking synovial tissue is to use an ROI defined by anatomical structures surrounding the synovial tissues [29].

The OMERACT filter emphasizes that the outcome may discriminate between groups. Thus, a good outcome measure needs to have a high signal-to-noise ratio. We found DAS28 to be a better discriminator than CF with an SRM of −1.45. The differentiation between best predictor (CF) and best discriminator (DAS28) is intriguing. The higher SRM of DAS28 as compared with CF may reflect that DAS28 overestimates the effect of therapy, and that the smaller decrease in CF indicates that the patients did only to some extent benefit from treatment. This interpretation is supported by a number of studies showing that patients with RA assessed clinically as being in remission...
<table>
<thead>
<tr>
<th>Variable</th>
<th>Total sample (n = 109)</th>
<th>Completers (n = 78)</th>
<th>All drop-outs (n = 31)</th>
<th>Drop-outs (lack of efficacy) (n = 23)</th>
<th>P-value completers vs all drop-outs ES</th>
<th>P-value completers vs drop-outs due to lack of efficacy ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td>78 (71.1)</td>
<td>56 (78.8)</td>
<td>22 (28.2)</td>
<td>17 (74)</td>
<td>0.930(^c)</td>
<td>0.28</td>
</tr>
<tr>
<td>Age, years</td>
<td>57.9 (13.9) (25.5–84.3)</td>
<td>58.4 (13.3)</td>
<td>56.5 (15.5)</td>
<td>54.4 (16.3)</td>
<td>0.545(^d)</td>
<td>0.14</td>
</tr>
<tr>
<td>Duration of RA</td>
<td>10.4 (8.0) (1.0–34.6)</td>
<td>9.7 (8.7)</td>
<td>12.0 (9.8)</td>
<td>11.8 (9.7)</td>
<td>0.355(^a)</td>
<td>–0.25</td>
</tr>
<tr>
<td>Height</td>
<td>170 (8) (148–191)</td>
<td>170 (9)</td>
<td>169 (7)</td>
<td>169 (7)</td>
<td>0.818(^d)</td>
<td>0.05</td>
</tr>
<tr>
<td>Weight</td>
<td>72.8 (15.6) (44.0–130)</td>
<td>73.9 (16.6)</td>
<td>70.2 (12.5)</td>
<td>69.8 (11)</td>
<td>0.215(^d)</td>
<td>0.24</td>
</tr>
<tr>
<td>BMI</td>
<td>25.3 (4.8) (16.9–45.0)</td>
<td>25.5 (4.9)</td>
<td>24.6 (4.6)</td>
<td>24.6 (4)</td>
<td>0.334(^d)</td>
<td>0.20</td>
</tr>
<tr>
<td>DAS28</td>
<td>5.07 (1.36) (1.40–7.80)</td>
<td>5.14 (1.34)</td>
<td>4.88 (1.42)</td>
<td>5.1 (1.2)</td>
<td>0.375(^d)</td>
<td>0.20</td>
</tr>
<tr>
<td>CRP</td>
<td>22 (28) (1–152)</td>
<td>23 (29)</td>
<td>20 (27)</td>
<td>24 (29)</td>
<td>0.453(^d)</td>
<td>0.11</td>
</tr>
<tr>
<td>TJC</td>
<td>11 (8) (0–28)</td>
<td>11 (8)</td>
<td>10 (10)</td>
<td>11 (9.5)</td>
<td>0.321(^d)</td>
<td>0.14</td>
</tr>
<tr>
<td>SJC</td>
<td>8 (6) (0–28)</td>
<td>8 (6)</td>
<td>8 (8)</td>
<td>8 (7.4)</td>
<td>0.486(^d)</td>
<td>0.01</td>
</tr>
<tr>
<td>Patient global VAS</td>
<td>63 (26) (66–100)</td>
<td>61 (23)</td>
<td>66 (20)</td>
<td>70 (19)</td>
<td>0.240(^d)</td>
<td>–0.24</td>
</tr>
<tr>
<td>HAQ-20</td>
<td>1.19 (0.63) (0.00–2.63)</td>
<td>1.25 (0.59)</td>
<td>1.03 (0.71)</td>
<td>1.1 (0.8)</td>
<td>0.098(^a)</td>
<td>0.35</td>
</tr>
<tr>
<td>CF</td>
<td>0.243 (0.176) (0.000–0.690)</td>
<td>0.270 (0.180)</td>
<td>0.176 (0.151)</td>
<td>0.179 (0.151)</td>
<td>0.013(^b)</td>
<td>0.53</td>
</tr>
<tr>
<td>√CF</td>
<td>0.450 (0.202) (0.000–0.831)</td>
<td>0.484 (0.188)</td>
<td>0.363 (0.215)</td>
<td>0.371 (0.206)</td>
<td>0.008(^d)</td>
<td>0.60</td>
</tr>
<tr>
<td>MTX</td>
<td>49.5%</td>
<td>38 (50.6)</td>
<td>16 (59.2)</td>
<td>–</td>
<td>0.602</td>
<td>–</td>
</tr>
<tr>
<td>Prednisolone</td>
<td>37.6%</td>
<td>28 (37.3)</td>
<td>13 (48.1)</td>
<td>–</td>
<td>0.779</td>
<td>–</td>
</tr>
<tr>
<td>SSZ</td>
<td>10.1%</td>
<td>9 (12)</td>
<td>2 (7.4)</td>
<td>–</td>
<td>0.475</td>
<td>–</td>
</tr>
<tr>
<td>Etanercept</td>
<td>30 (27.5)</td>
<td>25 (83.3)</td>
<td>5 (16.7)</td>
<td>4 (17.4)</td>
<td>0.110(^d)</td>
<td>–</td>
</tr>
<tr>
<td>Adalimumab</td>
<td>61 (56.0)</td>
<td>43 (70.5)</td>
<td>18 (29.5)</td>
<td>14 (60.8)</td>
<td>0.110(^d)</td>
<td>–</td>
</tr>
<tr>
<td>Infliximab</td>
<td>18 (16.5)</td>
<td>10 (55.6)</td>
<td>8 (44.4)</td>
<td>5 (21.7)</td>
<td>0.110(^d)</td>
<td>–</td>
</tr>
</tbody>
</table>

Continuous data are presented as mean (S.D.) and (range); counts are presented as the observed n (%). \(^a\)Patients treated with the same anti-TNF-\(\alpha\) drug for year. \(^b\)Patients treated with the same anti-TNF-\(\alpha\) drug for \(<\)1 year. \(^c\)Analysed using a chi-square test with a factor for sex and a factor for completer. \(^d\)Analysed using PROC TTEST, with unequal variances (i.e. the Satterthwaite method). \(^e\)Analysed using PROC NPAR1WAY (Wilcoxon) (i.e. the Kruskal–Wallis test). \(^f\)Analysed using a chi-square test with a factor for the label of anti-TNF applied and a factor for completer. ES calculated as the standardized mean difference; TJC: tender joint count; SJC: swollen joint count.
All change scores are calculated as post- and pre-treatment with anti-TNF-\(\alpha\), from the mean change from baseline (i.e. one-sample \(t\)-test); \(\rho\): the correlation between the pre- and post-scores. SRM: SRM from baseline; \(\Delta \sqrt{CF}\): \(\sqrt{CF_{\text{post}}} - \sqrt{CF_{\text{pre}}}\); TJ: tender joint count; SJC: swollen joint count.

### Rheumatology key messages

- Colour US predicts treatment success with anti-TNF-\(\alpha\) in patients with RA.
- DAS28 do not predict treatment success with anti-TNF-\(\alpha\) in patients with RA.

### Acknowledgements

**Funding:** This study was supported by the Oak Foundation, the Danish Physiotherapy Foundation, Aase and Ejnar Danielsen Foundation, Beckett-Foundation, Danish Rheumatism Association, Direktor Ib Henriksens Foundation, C.C. Klesstrup Foundation, Civilingenier Frode V. Nyegaard Foundation and Fonden til Lægevidenskabens Fremme.

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

### Supplementary data

Supplementary data are available at *Rheumatology* Online.

### References


<table>
<thead>
<tr>
<th>Variable</th>
<th>(n)</th>
<th>Mean (s.d.)</th>
<th>(t)-value</th>
<th>(P)-value</th>
<th>(\rho)</th>
<th>SRM (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Delta)CRP</td>
<td>75</td>
<td>-14.67 (23.90)</td>
<td>-5.31</td>
<td>&lt;0.0001</td>
<td>0.532</td>
<td>-0.61 (-0.85, -0.37)</td>
</tr>
<tr>
<td>(\Delta)DAS28</td>
<td>75</td>
<td>-2.16 (1.49)</td>
<td>-12.56</td>
<td>&lt;0.0001</td>
<td>0.298</td>
<td>-1.45 (-1.83, -1.07)</td>
</tr>
<tr>
<td>(\Delta)HAQ20</td>
<td>75</td>
<td>-0.43 (0.47)</td>
<td>-7.98</td>
<td>&lt;0.0001</td>
<td>0.687</td>
<td>-0.92 (-1.14, -0.71)</td>
</tr>
<tr>
<td>(\Delta)TJC</td>
<td>75</td>
<td>-8.21 (8.46)</td>
<td>-8.41</td>
<td>&lt;0.0001</td>
<td>0.170</td>
<td>-0.97 (-1.32, -0.62)</td>
</tr>
<tr>
<td>(\Delta)SJC</td>
<td>75</td>
<td>-5.97 (5.82)</td>
<td>-8.89</td>
<td>&lt;0.0001</td>
<td>0.300</td>
<td>-1.03 (-1.36, -0.70)</td>
</tr>
<tr>
<td>(\Delta)CF</td>
<td>69</td>
<td>-0.10 (0.19)</td>
<td>-4.31</td>
<td>&lt;0.0001</td>
<td>0.318</td>
<td>-0.52 (-0.81, -0.23)</td>
</tr>
<tr>
<td>(\Delta\sqrt{CF})</td>
<td>69</td>
<td>-0.14 (0.23)</td>
<td>-5.13</td>
<td>&lt;0.0001</td>
<td>0.339</td>
<td>-0.62 (-0.91, -0.32)</td>
</tr>
<tr>
<td>(\Delta)VAS</td>
<td>75</td>
<td>-29.12 (30.11)</td>
<td>-8.38</td>
<td>&lt;0.0001</td>
<td>0.138</td>
<td>-0.97 (-1.33, -0.61)</td>
</tr>
</tbody>
</table>


