Flares in Rheumatoid Arthritis Patients with Low Disease Activity: Predictability and Association with Worse Clinical Outcomes

Katie Bechman, Lieke Tweehuysen, Toby Garrood, David L. Scott, Andrew P. Cope, James B. Galloway, and Margaret H.Y. Ma

**ABSTRACT.** Objective. To investigate predictors of flare in rheumatoid arthritis (RA) patients with low disease activity (LDA) and to evaluate the effect of flare on 12-month clinical outcomes.

Methods. Patients with RA who were taking disease-modifying antirheumatic drugs and had a stable 28-joint count Disease Activity Score (DAS28) < 3.2 were eligible for inclusion. At baseline and every 3 months, clinical (DAS28), functional (HAQ-DI), EQ-5D, Functional Assessment of Chronic Illness Therapy Fatigue scale (FACIT-F), Medical Outcomes Study Short Form-36 (SF-36), serum biomarkers [multibiomarker disease activity (MBDA) score, calprotectin, CXCL10], and imaging data were collected. Flare was defined as an increase in DAS28 compared with baseline of > 1.2, or > 0.6 if concurrent DAS28 ≥ 3.2. Cox regression analyses were used to identify baseline predictors of flare. Biomarkers were cross-sectionally correlated at time of flare. Linear regressions were performed to compare clinical outcomes after 1 year.

Results. Of 152 patients, 46 (30%) experienced a flare. Functional disability at baseline was associated with flare: HAQ-DI had an unadjusted HR 1.82 (95% CI 1.20–2.72) and EQ-5D had HR 0.20 (95% CI 0.07–0.57). In multivariate analyses, only HAQ-DI remained a significant independent predictor of flare (HR 1.76, 95% CI 1.05–2.93). At time of flare, DAS28 and its components significantly correlated with MBDA and calprotectin, but correlation coefficients were low at 0.52 and 0.49, respectively. Two-thirds of flares were not associated with a rise in biomarkers. Patients who flared had significantly worse outcomes at 12 months (HAQ-DI, EQ-5D, FACIT-F, SF-36, and radiographic progression).

Conclusion. Flares occur frequently in RA patients with LDA and are associated with worse disease activity, quality of life, and radiographic progression. Higher baseline HAQ-DI was modestly predictive of flare, while biomarker correlation at the time of flare suggests a noninflammatory component in a majority of events. (First Release September 1 2018; J Rheumatol 2018;45:1515–21; doi:10.3899/jrheum.171375)

Key Indexing Terms:
RHEUMATOID ARTHRITIS
LOW DISEASE ACTIVITY
QUALITY OF LIFE
SERUM BIOMARKERS
FLARE
MULTIBIOMARKER DISEASE ACTIVITY SCORE
Predicting flare is therefore of direct relevance to clinical practice. Saleem, et al demonstrated that functional disability [Health Assessment Questionnaire—Disability Index (HAQ-DI)] and power Doppler ultrasound (PDUS) positivity at baseline were independently associated with flare in patients with RA in remission. Further, a previous meta-analysis revealed an association between PDUS positivity and flare in RA patients in remission.

The finding of PD positivity despite clinical remission provides evidence that flares may be related to incomplete suppression of inflammation. Based on this hypothesis, serum biomarkers may detect subclinical disease activity and consequently predict flare. In contrast to ultrasound (US), biomarkers may have smaller measurement error and may be less operator-dependent, costly, and time-consuming. In recent years, the predictive values of the multibiomarker disease activity (MBDA) score, calprotectin (S100A8/A9), and CXCL10 for treatment response in RA have been investigated. In the DRESS study, baseline MBDA score was predictive of flare and major flare in patients with LDA who did not taper treatment (usual care group). To our knowledge, calprotectin and CXCL10 have not been investigated as predictors of flare in disease-modifying anti-rheumatic drug (DMARD)-treated patients with RA in an LDA state. Calprotectin was found to be more strongly associated with US-detected synovitis than erythrocyte sedimentation rate (ESR) or C-reactive protein (CRP), and baseline calprotectin appeared to be predictive of clinical response to methotrexate. However, its predictive role as a marker of response to biologic DMARD is conflicting. CXCL10 was correlated with multiple disease activity measures in early RA, while elevated baseline levels of CXCL10 were associated with favorable response to tumor necrosis factor (TNF) inhibitor therapy in RA.

The aims of our study were 3-fold. First, we aimed to describe the frequency of flares in a cohort of prospective patients with RA in stable LDA states (including remission) over 1 year. Second, we aimed to examine the predictive value of a wide range of biomarkers (including clinical, functional, serum, and imaging variables) for flare. And third, we aimed to evaluate the effect of flare in RA patients with LDA states.

**MATERIALS AND METHODS**

**Study design and patients.** The REMIRA study is a prospective cohort study investigating RA patients with stable LDA states (including remission) over 1 year. Second, we aimed to examine the predictive value of a wide range of biomarkers (including clinical, functional, serum, and imaging variables) for flare. And third, we aimed to evaluate the effect of flare in RA patients with LDA states.

**Clinical assessments.** At baseline, demographic, disease and treatment characteristics were collected. Clinical assessments were carried out every 3 months for 1 year and included pain and fatigue (both on visual analog scale 0–100), DAS28, CRP, and ESR. Questionnaires were used to assess function and quality of life: HAQ-DI, EQ-5D-3L, Medical Outcomes Study Short Form-36 [SF-36; including physical component score (PCS) and mental component score (MCS)] and Functional Assessment of Chronic Illness Therapy Fatigue scale (FACIT-F). Flare was defined according to previously validated criteria: a DAS28 increase of > 1.2 compared with baseline or a DAS28 increase of > 0.6 compared with baseline and concurrent DAS28 ≥ 3.2.

For patients with multiple flares, only the first flare was considered in the analyses.

**Serum biomarker measurements.** Serum samples were obtained at each timepoint and stored at –80°C until being shipped frozen to the Crescendo Bioscience Clinical Laboratory (South San Francisco, California, USA) for MBDA score, calprotectin, and CXCL10 measurement. The MBDA test (Vectra DA, Crescendo Bioscience) combines the serum concentrations of 12 protein biomarkers [interleukin 6, TNF receptor type I, vascular cell adhesion molecule 1, epidermal growth factor, vascular endothelial growth factor A, YKL-40, matrix metalloproteinase 1, matrix metalloproteinase 3, CRP, serum amyloid A (SAA), leptin, and resistin] in an algorithm to provide a score that quantifies RA disease activity. The scores are on a scale of 1 to 10 with 100 categories for low (≤ 30), moderate (30–44), and high disease activity (> 44). Calprotectin and CXCL10 were measured by ELISA (Buhlmann MRP 8/14 ELISA Product Code EK-MRP8/14m; R&D Systems Human CXCL10/IFN-10 Quantikine ELISA Product Code DIP100).

**Imaging assessments.** Ultrasonography of hands and wrists, and conventional radiographs of hands and feet were carried out at baseline and 12 months. Erosive progression was defined as new or larger erosions over 1 year on radiographs. All sonographic assessments were performed using high-sensitivity US equipment (GE Logiq 9) with a 2-dimensional MI2L transducer. A single experienced sonographer (TG), blinded to clinical or laboratory data, scanned 10 metacarpophalangeal joints and 2 wrists from a dorsal aspect for greyscale US (GSUS) synovial hypervascularity and intra-articular PDUS signals. GSUS and PDUS were graded on a scale of 0 to 3 using a validated semiquantitative scoring system. The composite GSUS and PDUS scores were the sum scores of the 12 individual joints.

**Statistical analysis.** Descriptive statistics were provided with mean (± SD), median (interquartile range; IQR), or frequencies, depending on data distribution. Cross-sectional correlations between all measurements (biomarkers and DAS28 components) at time of flare were assessed by Spearman’s correlation coefficient (rs) and interpreted according to commonly used classification: very weak (rs < 0.20), weak (rs = 0.20–0.39), moderate (rs = 0.40–0.59), strong (rs = 0.60–0.79), and very strong (rs > 0.80) correlation.

To identify predictors of time to flare, we performed univariate Cox regression, in which time to flare was the dependent variable, and clinical, functional, serum, and imaging measurements the independent variables. Multivariate analyses were performed to identify factors that were independently associated with flare, adjusting for age, sex, DAS28, visual analog scale (VAS) pain, CRP, ESR, and US scores (for HAQ model only), and MBDA score (for EQ-5D model only).

Linear regression was used to determine the effect of flare on 12-month clinical outcomes (i.e., disease activity and functional status). A multivariate linear regression model was applied adjusting for baseline age, sex, disease duration < 10 years, had stable DMARD treatment for > 6 months, and 28-joint count Disease Activity Score (DAS28) < 3.2 for at least 1 month apart. Three centers across South London participated: Guy’s and St. Thomas’ Hospital, King’s College Hospital, and University Hospital Lewisham National Health Service Foundation Trusts. Patients were managed as part of routine care. The study was approved by the local ethics committee and conducted according to the guidelines of the Declaration of Helsinki (REC/09/H0803/154). Written informed consent was obtained from all patients.
duration, erosive status, baseline DAS28, HAQ, and baseline variable of interest. A p value ≤ 0.05 was regarded as significant. Because this was an exploratory study, no correction for multiple hypothesis testing was performed. Missing data were addressed using a multiple imputation module (Supplementary Data 1, available from the authors on request). All analyses were performed with STATA 14.1 statistical software.

RESULTS

Patient characteristics. In total, 152 patients were enrolled in the REMIRA study. Baseline characteristics are depicted in Table 1. The majority of patients were receiving DMARD monotherapy (n = 69, 45%) and the median disease duration was 3 (IQR 2–6) years. Ninety-seven patients (66%) fulfilled DAS28 remission criteria (DAS28 < 2.6). All patients had synovial hypertrophy (GSUS > 1) and 90% had detectable PDUS activity at baseline.

Characteristics of flare. Forty-six patients (30%) experienced at least 1 flare. Twelve patients had first flare by 3 months, 10 by 6 months, 11 by 9 months, and 13 by 12 months. Seventeen patients experienced multiple flares; 11 patients flared at 2 visits, 5 patients at 3 visits, and 1 patient at all 4 visits after baseline. When limiting the cohort to patients who were in remission defined by DAS28 < 2.6 at baseline, 24 patients of a total 97 (25%) experienced at least 1 flare.

Serum biomarkers at time of flare. There were 70 individual flare events. Seventeen percent (n = 12) of flares were driven solely by increases in patient’s global assessment (PtGA) and tender joint count (TJC), without any increase in swollen joint count (SJC) or ESR.

In total, 33% of flares (n = 23) had a concurrent high MBDA score (> 44), while 13% (n = 44) of visits without flare had a high MBDA score. The levels of ESR, CRP, MBDA score, and calprotectin were significantly higher at flare visits than at nonflare visits [median ESR 14 mm/h (IQR 5–23) vs 6 mm/h (3–12); CRP 5 mg/l (5–9) vs 5 mg/l (5–5); MBDA 38 (25–50) vs 28 (18–38); and calprotectin 2916 ng/ml (2002–4186) vs 2377 ng/ml (1504–3358)].

DAS28 significantly correlated with MBDA score (rₚ = 0.5, p = 0.0002) at time of flare. The rₚ of 0.5 suggests that the MBDA values explain only 25% of the variation in DAS28. The correlation of MBDA was stronger with the components ESR and SJC, and were nonsignificant for TJC and PtGA. Similar findings were seen for calprotectin (rₛ = 0.49, p = 0.0007). CXCL10 did not correlate with DAS28 or its components at time of flare (Supplementary Table 1, available from the authors on request).

Prediction of flare. Univariate Cox regression showed that several baseline characteristics were associated with flare (DAS28, ESR, CRP, PtGA, VAS pain, HAQ-DI, and EQ-5D; Figure 1 and Supplementary Tables 2 and 3, available from the authors on request). The strongest magnitude of association was seen with HAQ-DI and EQ-5D. Baseline US synovitis (GSUS or PDUS) and mental health (using the SF-36 MCS) were not associated with flare. Baseline MBDA scores were also not predictive of flare, although a sensitivity analysis limited to flares with a rise in MBDA score to > 44 (high disease activity) did show a relationship between baseline MBDA value and flare risk, with each unit rise in baseline MBDA score associated with a 7% increase in flare risk (1.07, 95% CI 1.02–1.11; p = 0.005; Supplementary Tables 4 and 5). Analyzing each component of the MBDA score identified SAA, leptin, and high-sensitivity CRP as the strongest predictors of flare. The remaining 9 components of the MBDA score did not individually predict flare.

The imputation model confirmed the association between flare and baseline HAQ-DI and EQ-5D but did not demonstrate any other associations. In multivariate analyses, only baseline HAQ-DI remained a significant independent
Outcomes in flare versus sustained remission group. Adjusting for baseline values, patients who had a flare experienced significantly worse clinical outcomes at 12 months than patients in sustained remission, reflected by higher disease activity, worse functional outcomes, and higher radiographic progression scores (Table 2). Having a flare was associated with a larger than minimal clinically important difference increase in HAQ-DI ($\beta = 0.32$, 95% CI 0.29–0.36; $p < 0.01$) and decrease in EQ-5D ($\beta = -0.11$, 95% CI –0.12 to –0.09; $p < 0.01$). Both the physical and mental performance measures from SF-36 were significantly worse in patients who flared in the unadjusted model. This was more marked with the PCS and did not remain significant with the MCS in the adjusted model. Patients who flared were 3.6 times (95% CI 2.77–4.67; $p < 0.01$) more likely to have erosive progression, defined as new or larger erosions over 1 year on radiographs.

DISCUSSION
In this prospective study, one-third of RA patients with LDA states experienced a flare during 12 months of follow-up. This is similar to flare rates reported in cohort studies, although these included only patients in remission$^5,9$ and in drug tapering studies in patients who remain on stable therapy. In both the DRESS$^{23}$ and the POET$^{24}$ studies, the rate of short-lived flare was significantly higher in patients who tapered or stopped their anti-TNF therapy compared to those who continued treatment, although in the DRESS study, the rate of major flares was similar between the 2 groups.

In our study, we have shown that the occurrence of a flare is hard to predict, but undeniably associated with worse clinical outcomes at 12 months. Our study highlights that identification of predictors of flare in patients with LDA states is challenging. In accordance with a previous remission cohort study$^9$, we found that HAQ-DI, a measure of functional activity, reflected by difficulties in activities of daily living, was predictive for flare. It is plausible that patients with LDA and high functional disability are more likely to flare. Functional impairment can herald a flare with the onset of morning stiffness and fatigue. A high HAQ may reflect severe rheumatoid arthritis with disease-related damage and the likelihood of grumbling disease (persistent low-grade disease activity).

Serum biomarkers were only modestly correlated with DAS28 at the time of flare. This might be because a flare is defined by worsening of the DAS28 composite score, and an increase in TJC and PtGA alone may increase the DAS28...
score to a sufficient level to define a flare. It is possible that a flare event is not solely the result of direct synovial inflammation but may be driven by other pathways, for example chronic low-grade inflammation and abnormal regulatory mechanisms. This heterogeneity may partly explain why identifying predictors of flare is challenging. The OMERACT RA flare group recognizes the limitation of DAS28 in defining flare events. They are developing a consensus-based core domain set to identify and measure flare in RA. It is likely that improving the definition of flare and establishing a scoring system may help interpret predictors of flare in the future.

We found that a higher baseline CRP and ESR were predictive of flare in the univariate analyses, while baseline MBDA score, calprotectin, and CXCL10 were not. In the sensitivity analysis limited to flare events with an associated high MBDA score at the time of flare, a relationship between baseline MBDA value and flare risk was established. This may suggest that baseline MBDA score is only predictive of flares that are driven directly by inflammation. Interestingly, when each component of the MBDA score was analyzed individually, only 3 of the 12 components (SAA, leptin, and high-sensitivity CRP) predicted flare. Studies suggest a close correlation between leptin levels and RA disease duration, activity, and severity. The rapid production of SAA and its exceptionally wide dynamic range has proved advantageous as a biomarker of disease activity, with superiority over CRP in early RA studies.

US variables, including PD signal, had no predictive value in our study. This is likely a reflection of the high proportion of patients in our cohort who had US activity at baseline. In the POET study, only 63% of patients had US signs of arthritides with positive PD signal. This is partly explained by our cohort, which included a greater proportion of patients with LDA states above the DAS28 remission cutoff. A large number of patients were taking DMARD monotherapy, and only 3 were prescribed oral corticosteroids, which may explain the difference in PD compared to other cohorts that have achieved LDA states with combination DMARD and corticosteroid therapy. Scoring of PD was also more stringent in our cohort compared to others, leading to a much higher proportion of patients with PD signal being reported. The major limitation of US is that it remains a user-dependent technique. It is increasingly sensitive at demonstrating evidence of incomplete suppression of inflammation. The joints of healthy volunteers have been shown to display PD signal, and treatment escalation studies have argued against very stringent US targets. Others have also shown that low-grade PD signal and synovial hypertrophy may not necessarily reflect the presence of active synovitis in RA joints. In our cohort, a high proportion had PD activity at baseline and did not go on to flare. It may be postulated that a binary PD cutoff might be insensitive in discriminating patients who are likely to flare.

Our study also found that patients who flare were more likely to have erosive progression, worse quality of life, and higher disease activity over 1 year. These findings are consistent with previous studies and emphasize the

---

**Table 2. Outcomes at 1 year in patients who flare compared to patients who do not flare.**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Outcome</th>
<th>β constant</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ-5D</td>
<td>Unadjusted</td>
<td>0.59</td>
<td>0.37–0.80</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Adjusted</td>
<td>0.19</td>
<td>0.04–0.32</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Imputed (adjusted)</td>
<td>0.32</td>
<td>0.29–0.36</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>SF-36 PCS</td>
<td>Unadjusted</td>
<td>−0.19</td>
<td>−0.26–−0.13</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Adjusted</td>
<td>−0.11</td>
<td>−0.18–−0.05</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Imputed (adjusted)</td>
<td>−0.11</td>
<td>−0.12–−0.09</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>SF-36 MCS</td>
<td>Unadjusted</td>
<td>−3.92</td>
<td>−7.04–−0.8</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Adjusted</td>
<td>−5.17</td>
<td>−8.51–−4.53</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Imputed (adjusted)</td>
<td>−5.42</td>
<td>−9.41–−1.42</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Adjusted</td>
<td>−2.86</td>
<td>−6.83 to 1.12</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>Imputed (adjusted)</td>
<td>−2.94</td>
<td>−3.7 to −2.18</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>DAS28</td>
<td>Unadjusted</td>
<td>−7.83</td>
<td>−11.6 to −4.06</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Adjusted</td>
<td>−4.07</td>
<td>−7.91 to −0.24</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>Imputed (adjusted)</td>
<td>−5.09</td>
<td>−5.77 to −4.42</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Erosive progression</td>
<td>Unadjusted</td>
<td>2.33*</td>
<td>0.87–6.27</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>Adjusted</td>
<td>3.51*</td>
<td>1.06–11.7</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>Imputed (adjusted)</td>
<td>3.60*</td>
<td>2.77–4.67</td>
<td>&lt; 0.01</td>
</tr>
</tbody>
</table>

* OR values. HAQ-DI: Health Assessment Questionnaire–Disability Index; SF-36 PCS: Medical Outcomes Study Short Form-36 physical component summary; MCS: mental component summary; FACT-F: Functional Assessment of Chronic Illness Therapy Fatigue scale; DAS28: 28-joint count Disease Activity Score.
importance of flare and its relationship with patient outcomes. What remains unclear is whether flares are causally implicated in clinical outcome or if they are merely a biomarker of persistent low-grade disease. A flare may imply persistent uncontrolled inflammation contributing to disease progression or a transient episode of inflammation (e.g., a 6-week flare within a stable 6-month period) that is sufficient to affect longterm outcome, or signify negative patient experience, and a lack of self-control and unpredictability of the disease, which undoubtedly have psychological health implications.

There were several strengths of our study. The cohort was selected from routine care, which is far more representative than a highly selective clinical trial population. Using patients in LDA states rather than remission enables access to a broader range of patients and is more in keeping with routine clinical care. Further, this was a deeply phenotyped cohort with extensive clinical and laboratory data at multiple timepoints across the study period.

There are potential limitations to our study. We must acknowledge the limitation of the REMIRA study sample size, and the limited number of predictors identified could reflect a type 2 error. We also acknowledge issues with missing data, particularly with incomplete available US reports. However, the pattern of missing data met the assumptions of missing at random and we were able to successfully construct an imputation model to address this. We only registered flares during a visit to a rheumatologist and the actual flare rate might be higher. Potential flares between visits could have been detected by a flare questionnaire or alternative tools that permit remote monitoring. However, we would have missed only short-lived flares (< 3 mos), and those are of less clinical importance because they are less likely to lead to worse clinical outcomes (e.g., no radiographic progression) REMIRA was an observational study and any modifications in medications were carried out according to the physicians’ and patients’ choices. Because treatment was not protocolized, this may have affected the rate of flares. A single failure model was used to identify predictors of flare, and thus changes in therapy after a flare event should not influence the analysis. It is, however, possible that treatment modifications, for example, glucocorticoids during a flare, may improve disease outcome at 12 months.

We have demonstrated that flares are common in RA patients with LDA states and are strongly associated with poor clinical outcomes. Therefore, preventing flares is clinically relevant yet relatively challenging. HAQ-DI, a measure of functional activity, was an important predictor of flare. However, flares are complex events and not simply a reflection of inflammatory disease activity. It is possible that 2 distinct subtypes of flare might exist: an “inflammatory” flare predominately driven by an increase in SJC and ESR, and a “noninflammatory” flare with a disproportionately elevated TJC and a high PtGA score. Differentiating these 2 flare types may identify potential predictors. Further research is needed to determine whether distinct flares exist and to categorize the potential predictors of each.

ACKNOWLEDGMENT
We acknowledge the Crescendo Bioscience team, in particular Eric Sasso and Nadine Defranoux, for processing the REMIRA blood samples and assisting in the completion of this manuscript. We also acknowledge Dr. Stephen Kelly for his advice on setting up the US protocols.

REFERENCES