BMJ Open

Clinical predictors of flare and drug-free remission in rheumatoid arthritis: preliminary results from the prospective BIO-FLARE experimental medicine study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2024-092478
Article Type:	Original research
Date Submitted by the Author:	15-Aug-2024
Complete List of Authors:	Rayner, Fiona; Newcastle University Translational and Clinical Research Institute; Newcastle Upon Tyne Hospitals NHS Trust, Rheumatology Hiu, Shaun; Newcastle University Population Health Sciences Institute, Newcastle University Melville, Andrew; University of Glasgow Institute of Infection Immunity and Inflammation Bigirumurame, Theophile; Newcastle University Population Health Sciences Institute Anderson, Amy; Newcastle University Translational and Clinical Research Institute Dyke, Bernard; NIHR Birmingham Biomedical Research Centre Kerrigan, Sean; University of Glasgow Institute of Infection Immunity and Inflammation McGucken, Andrew; University of Glasgow Institute of Infection Immunity and Inflammation Prichard, Jonathan; Newcastle University Population Health Sciences Institute Shahrokhabadi, Mohadeseh Shojaei; Newcastle University Population Health Sciences Institute Hilkens, Catharien M U; Newcastle University Translational and Clinical Research Institute Buckley, Christopher D; NIHR Birmingham Biomedical Research Centre; University of Oxford Kennedy Institute of Rheumatology McInnes, Iain; University of Glasgow Institute of Infection Immunity and Inflammation Ng, Wan-Fai; Newcastle University Translational and Clinical Research Institute; Newcastle Upon Tyne Hospitals NHS Foundation Trust, Musculoskeletal Unit Goodyear, Carl; University of Glasgow Institute of Infection Immunity and Inflammation Teare, Dawn; Newcastle University Population Health Sciences Institute Filer, Andrew; NIHR Birmingham Biomedical Research Centre, Institute of Inflammation and Ageing Siebert, Stefan; University of Glasgow, Institute of Infection, Immunity and Inflammation; NHS Greater Glasgow and Clyde, Raza, Karim; NIHR Birmingham Biomedical Research Centre, School of Immunity and Infection; Sandwell and West Birmingham Hospitals NHS Trust Pratt, Arthur; Newcastle University, Translational and Clinical Research

	Institute; Newcastle Upon Tyne Hospitals NHS Foundation Trust, Musculoskeletal Unit Baker, Kenneth F; Newcastle University Translational and Clinical Research Institute; Newcastle Upon Tyne Hospitals NHS Foundation Trust, Rheumatology Department Isaacs, John; Newcastle University Translational and Clinical Research Institute; Newcastle Upon Tyne Hospitals NHS Foundation Trust, Rheumatology Department
Keywords:	Rheumatology < INTERNAL MEDICINE, Drug Therapy, IMMUNOLOGY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Clinical predictors of flare and drug-free remission in rheumatoid arthritis: preliminary results from the prospective BIO-FLARE experimental medicine study

Fiona Rayner[1,2], Shaun Hiu[3]*, Andrew Melville[4]*, Theophile Bigirumurame[3], Amy Anderson[1], Bernard Dyke[5], Sean Kerrigan[4], Andrew McGucken[4], Jonathan Prichard[3], Mohadeseh Shojaei Shahrokhabadi[3], Catharien M U Hilkens[1], Christopher D Buckley[5,6], Iain McInnes[4], Wan-Fai Ng[1,2], Carl Goodyear[4], M Dawn Teare[3], Andrew Filer[5], Stefan Siebert[4], Karim Raza[5,7], Arthur G Pratt[1,2], Kenneth F Baker[1,2], John D Isaacs[1,2].

*Contributed equally as second authors

Author Information

1. Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, UK
2. Rheumatology Department, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK
3. Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK
4. School of Infection and Immunity, University of Glasgow, Glasgow, UK
5. NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and Institute for Inflammation and Ageing, University of Birmingham, Birmingham, UK
6. Kennedy Institute of Rheumatology, University of Oxford, Oxford, UK
7. Department of Rheumatology, Sandwell and West Birmingham NHS Trust, Birmingham, UK

Corresponding author:

Professor John Isaacs

Translational and Clinical Research Institute, 4th Floor, William Leech Building, Framlington Place, Newcastle upon Tyne, NE2 4HH

john.isaacs@ncl.ac.uk

ORCiD: https://orcid.org/0000-0002-6103-7056

Keywords:

Rheumatoid arthritis, Flare, Remission, Drug-free, cessation, withdrawal, disease-modifying anti-rheumatic drug (DMARD)

Abstract

Objectives

Huge advances in rheumatoid arthritis (RA) treatment mean an increasing number of patients now achieve disease remission. However, long term treatments can carry side effects and associated financial costs. In addition, some patients still experience painful and debilitating disease flares, the mechanisms of which are poorly understood. High rates of flare and a lack of effective prediction tools can limit attempts at treatment withdrawal. The BIO-FLARE experimental medicine study was designed to study flare and remission immunobiology. Here we present the clinical outcomes and predictors of drug-free remission and flare.

Design, setting and participants

BIO-FLARE was a multicentre, prospective, single-arm, open-label experimental medicine study conducted across seven NHS Trusts in the UK. Participants had established RA in clinical remission (DAS28-CRP<2.4) and were receiving methotrexate, sulfasalazine or hydroxychloroquine (monotherapy or combination).

Interventions

The intervention was DMARD cessation, followed by observation for 24-weeks or until flare, with clinical and immune monitoring.

Outcome measures

The primary outcome measure was proportion of participants experiencing a confirmed flare, defined as DAS28-CRP ≥3.2, or DAS28-CRP ≥2.4 twice within two weeks, and time to flare.

Exploratory predictive modelling was also performed using multivariable Cox regression, to understand risk factors for flare.

Results

121 participants were recruited between September 2018 and December 2020. Flare rate by week 24 was 52.3% (95% CI 43.0 to 61.7) with a median (IQR) time to flare of 63 (41-96) days. Female sex, baseline methotrexate use, ACPA level and RF level were associated with flare. An exploratory prediction model incorporating these variables allowed estimation of flare risk, with acceptable classification (C index 0.709) and good calibration performance.

Conclusion

The rate of flare was approximately 50%. Several baseline clinical parameters were associated with flare. The BIO-FLARE study design provides a robust experimental medicine model for studying flare and remission immunobiology.

Strengths and Limitations

- Rheumatoid arthritis flare immunobiology is poorly understood. The BIO-FLARE study represents a robust experimental medicine model for the investigation of flare and remission immunobiology in RA.
- We have used routine baseline clinical parameters to develop an exploratory model for the prediction of flare following immunomodulatory drug cessation.
- Limitations include the open-label approach, which could allow for disease flares caused by the nocebo effect
- A short follow-up time of 6 months means flares after this time were not recorded

1.0 Introduction

Rheumatoid arthritis (RA) is a chronic disease characterised by relapsing-remitting episodes of immune-mediated inflammation known as flares, which pose far-ranging negative consequences for patients (1). RA flares have been associated with impaired physical function, increased fatigue, and reduced quality of life (2), as well as serious long-term sequelae including incremental joint damage (3) and increased risk of cardiovascular events (4). Despite their importance, RA flares remain poorly understood at a mechanistic level, and are challenging to investigate scientifically because of their sporadic and unpredictable nature.

Historically most patients with RA suffered from frequent flares, though early diagnosis and rapid initiation of modern regimens of disease-modifying anti-rheumatic drugs (DMARDs) now mean that sustained remission is increasingly an achievable goal in around half of patients. Nevertheless, DMARDs carry risks of drug toxicity, are expensive to prescribe and monitor, and require regular blood testing. International guidelines now advocate consideration of DMARD dose reduction for patients in sustained remission (5), albeit with a risk of arthritis flare in around half of patients who attempt this (6, 7, 8). DMARD cessation thus provides an experimental human model, acceptable to patients and endorsed by international treatment guidelines, by which to study the immunobiology of RA flare. In turn this could identify hitherto elusive biomarkers to guide individualised therapeutic decisions.

The BIOlogical Factors that Limit sustAined Remission in rhEumatoid arthritis (BIO-FLARE) study is an experimental medicine study in which patients with established RA in remission underwent complete DMARD cessation, with the over-arching aim of advancing understanding of the biological factors underpinning RA remission and flare through multi-parameter immune monitoring (9). In this preliminary report we describe the clinical characteristics and outcomes of the BIO-FLARE cohort, and generate an exploratory clinical model to predict risk of flare at the individual patient level. This model, based on clinical predictors alone, provides a baseline which we will subsequently strengthen by the addition of immune biomarkers, informed by our laboratory studies.

2.0 Methods and Materials

BIO-FLARE was a multi-centre, prospective, single-arm, open-label, experimental medicine study of complete DMARD cessation in RA patients who had achieved remission on conventional synthetic DMARDs (csDMARDs: methotrexate, sulfasalazine and hydroxychloroquine; either as monotherapy or in combination) (9). All participants who fulfilled eligibility criteria stopped all DMARDs at enrolment without tapering. There was no randomisation or control arm, the comparator groups being those who flared versus those who remained in remission. Participants were followed up for 24 weeks, or until confirmed flare, whichever occurred earlier. The primary clinical outcomes were time to flare (in days) following DMARD cessation, and occurrence of flare (binary) during the 24-week study period.

2.1 Recruitment Criteria

Inclusion criteria included the following: 1) RA fulfilling the 1987 ACR or 2010 ACR/EULAR classification criteria, 2) stable dose csDMARDs, with no dose increase in the previous six months, and 3) clinical remission according to disease activity score in 28 joints (DAS28) with C-reactive protein (DAS28-CRP) <2.4 (10). Exclusion criteria included current use of csDMARDs other than methotrexate, sulfasalazine or hydroxychloroquine; use of leflunomide within the previous 12 months (owing to its extended half-life due to enterohepatic recirculation); use of any biologic or targeted synthetic DMARDs in the previous 6 months; use of glucocorticoids in the previous 3 months (other than inhaled or topical forms); any previous ever use of cell-depleting therapies (e.g. rituximab). Potential participants were identified by their usual rheumatology teams across seven participating National Health Service (NHS) trusts in the United Kingdom (UK), between September 2018 and December 2020.

2.2 Procedures and Definitions

As shown in Figure 1, participants underwent a screening visit to confirm eligibility. Eligible and consenting participants stopped all DMARDs, with no dose tapering. An optional baseline ultrasound guided synovial biopsy was performed in consenting participants prior to DMARD cessation. Ultrasound findings did not influence study eligibility. Subsequent study visits took place at weeks 2, 5, 8, 12 and 24 following DMARD cessation. Participant-initiated ad-hoc study visits could also be arranged at any time in response to suspected flare. At all study visits, participants underwent clinical assessment, including DAS28-CRP, adverse event (AE) and serious adverse event (SAE) recording, and blood and urine sampling.

Flare was defined as occurrence of any of the following: 1) DAS28-CRP ≥3.2 at any study visit, 2) DAS28-CRP ≥2.4 on two occasions within a 14-day period: if DAS28-CRP was ≥2.4 but <3.2 at any study visit, then another visit was arranged within 2 weeks, with flare confirmed if DAS28-CRP was ≥2.4 at second review, or 3) clinical indication for glucocorticoid rescue therapy and/or DMARD restart despite DAS28-CRP <2.4, e.g. for disease activity not captured by DAS28-CRP such as ankle or foot joint synovitis. Clinician discretion was permitted where DAS28-CRP ≥3.2 was felt to be driven by identifiable non-RA factors, e.g. concurrent infection. In such cases, an ad-hoc visit was arranged within 2 weeks and participants were considered to have remained in remission if subsequent DAS28-CRP was <2.4.

In the event of confirmed flare, an ultrasound guided synovial biopsy was performed within 7 days (if there was a joint deemed suitable for biopsy). Systemic or intra-articular glucocorticoid therapy could be administered immediately after biopsy, where indicated. Participants were then referred back to their usual rheumatology team for re-initiation of DMARDs.

2.3 Baseline Data Collection

Baseline data collected at the screening visit included participant demographics, RA history, current and previous treatments, medical history including significant co-morbidities (Charlson Comorbidity

Index), and patient reported outcome measures including functional status (HAQ-DI) (Table 1). A full schedule of events is included in the *supplementary material Table 1.*

2.4 Statistical analysis

The primary outcome for the current study was time to disease flare (in days). The Kaplan-Meier estimate of the survivor curve was computed along with numbers at risk at the scheduled visit dates (week 2, 5, 8, 12 and 24). Participants who were lost to follow-up or withdrawn from the study were censored at the last available visit.

Sixteen candidate baseline variables were considered for exploratory prediction model inclusion: age, sex, disease duration, time from symptom onset to first DMARD, baseline methotrexate use, glucocorticoids within 3-12 months of baseline visit, baseline rheumatoid factor (RF) level, baseline anti-citrullinated peptide antibody (ACPA) level, DAS28-CRP, ACR/EULAR Boolean remission status (11), education level, employment status, body mass index, smoking status, alcohol intake, and Charlson comorbidity index. These were chosen based on prior knowledge and before reviewing study data. Owing to the presence of some missing data points, analyses were performed with 10 imputed datasets using multiple imputation by chained equations (MICE) (12).

To provide clinical context, univariate analyses were performed to assess the strength of association between each candidate variable and time to flare, with hazard ratios (HR) and 95% confidence intervals (CI) determined.

A predictive model for flare containing baseline clinical variables was built using a Cox proportional hazards (PH) model following a sequential process of variable selection, estimation of shrinkage, and internal validation, described in detail in *supplementary materials*. Predictive performance was internally validated using bootstrapping and evaluated with optimism-corrected indices of discrimination (*C* index) and calibration (13, 14, 15). We report our predictive model as an equation

for calculating the prognostic index (PI), representing an individual's "propensity" to flare, and a baseline survival function, which together allow calculation of estimated risk of flare by a given time following DMARD cessation.

2.5 COVID-19 mitigation and sensitivity analysis

The latter stages of the BIO-FLARE study overlapped with the onset of the COVID-19 pandemic, meaning some follow-up visits were disrupted. A mitigation strategy was adopted whereby affected participants received telephone consultations when their study visits were due, with assessments of flare/remission status based on participant-reported symptoms rather than DAS28-CRP, and face-to-face visits reserved for those with suspected flare. Seven participants were lost to follow-up during this period, while 4 had telephone consultations up to week 24. For our primary analyses, participants with telephone consultations up to week 24, and no symptoms of flare, were classified as having remained in remission. A sensitivity analysis of our predictive modelling process was conducted using last face-to-face study visits only (i.e. last available DAS28-CRP).

2.6 Study subpopulations

Overall baseline characteristics and adverse event data are described for all participants who stopped DMARDs (n=121, the total study population). Time-to-event analyses, including predictive modelling, were performed for participants with ≥1 follow-up visit (n=120, the analysis population which excludes 1 participant who withdrew soon after baseline because of the COVID-19 pandemic). Flare rate was calculated for participants with confirmed flare or remission status (n=111, which we term the 'modified per-protocol' population following the COVID-19 mitigation strategy), i.e. excluding 10 participants who did not experience flare but withdrew (n=3) or were lost to follow-up (n=7) before week 24.

2.7 Patient and Public Involvement

The Newcastle Patient and public Involvement in Musculoskeletal reSearch (PIMS) group were consulted at the planning stage of the project. The importance of the research topic and design of the study protocol was informed by their views and discussions. Clinical results from the study have been presented at national Versus Arthritis meetings with patient partners present. We will present more results of the study to local, regional and national PPIE groups as they become available.

3.0 Results

3.1 Baseline Characteristics and Adverse Events

One hundred and twenty-one participants met the inclusion criteria including DAS28-CRP <2.4 and stopped DMARD therapy (Figure 2). The overall baseline characteristics are presented in Table 1, along with the baseline characteristics for participants who flared and those who remained in remission at 24 weeks (n=111, the modified per-protocol population). For the total study population, mean (SD) age was 64.1 (11.9) years, 60.3% were female, and median (IQR) disease duration was 6.3 (4.5–12.3) years. 67/119 (56.3%) were RF positive and 76/114 (66.7%) were ACPA positive, with 64/113 (56.6%) double positive. Only 1 participant had previous biologic therapy (etanercept, stopped 7.5 years before study entry). 101/121 participants (83.5%) were treated with methotrexate at baseline (monotherapy or combination use) with a median (IQR) dose of 15 (12.5-20) mg weekly. Of 20 participants not on methotrexate at baseline, 7/20 had previously received methotrexate treatment. Mean (SD) baseline DAS28-CRP was 1.61 (0.32); 61.2%, 78.5% and 84.9% fulfilled ACR/EULAR Boolean remission criteria, Boolean 2.0 remission criteria (16), and simplified disease activity index (SDAI) remission criteria at baseline, respectively. There were 155 AEs (*Supplementary materials Table 2*), 4 SAEs relating to hospitalisations (*Supplementary materials Table 3*), and no deaths. The four SAEs were all considered to be unrelated to study participation or procedures.

3.2 Flare Characteristics

The flare rate at 24 weeks (168 days) was 52.3% (58/111, 95% CI 43.0 to 61.7). Flare-free probability is presented in a Kaplan-Meier plot in Figure 3. For the 58 participants who experienced flare, median time to flare was 63 days (IQR 41–96 days, range 13–155 days).

Mean (SD) DAS28-CRP at time of flare was 3.81 (0.78). DAS28-CRP components at time of flare were as follows: median (IQR) tender joint count (TJC) 4 (1–5), swollen joint count (SJC) 2 (1–3), CRP 8.0 (4.5–14.4) mg/dL; mean (SD) patient global health VAS 48.4/100 (22.7).

Confirmation of flare was based on a single DAS28-CRP result ≥3.2 in 39/58 cases (of which 18 were scheduled study visits and 21 were ad-hoc visits), two DAS28-CRP results ≥2.4 within a 14-day period in 16/58 cases (of which 8 had DAS28-CRP ≥3.2 at the 2nd visit), and clinician discretion in 3/58 cases (described in *Supplementary material Table 4*).

3.3 Univariate Cox proportional hazards models

Of the 16 variables considered in univariate analyses: female sex, methotrexate use at baseline, RF level, ACPA level, and longer time from symptom onset to first DMARD were statistically significantly associated with time to flare (Table 2).

3.4 Exploratory prediction model

Our prediction modelling procedure, including variable selection, resulted in the inclusion of sex, methotrexate use at baseline, RF level, and ACPA level into the prediction model (*see Supplementary materials Table 5*). A square root transformation of RF, and two non-linear expressions of ACPA (inverse of ACPA and inverse square root of ACPA) were chosen as the best-fitting transformations. Thus, our prediction model consisted of five terms: sex, methotrexate use, $(RF + 0.1)^{0.5}$, $(ACPA + 0.1)^{-1}$, and $(ACPA + 0.1)^{-0.5}$.

The predicted probability of flare within *t* days after DMARD cessation can be computed as:

$$Predicted\ risk\ of\ flare\ by\ t\ days\ after\ DMARD\ cessation = 1 - \hat{S}_0(t)^{\wedge}\exp(PI)$$

Where $\hat{S}_0(t)$ is the estimated baseline survival function at time *t*, PI is the prognostic index, and *exp*(.) is the exponential function. The value of $\hat{S}_0(t)$ at t=168 days after DMARD cessation is 0.672. Additional values at 30, 60, 90 and 120 days are available in the *supplementary material*. The PI is computed as:

$$PI = (-0.55814869 \times Sex) + (1.05775338 \times Methotrexate\ use) + (0.03734463 \times \sqrt{RF + 0.1}) + f(ACPA)$$

Where $f(ACPA) = \left(\frac{0.55920681}{ACPA+0.1}\right) - \left(\frac{1.86737912}{\sqrt{ACPA+0.1}}\right)$, sex coded as female=0, male=1, methotrexate use coded as no=0, yes=1.

Thus, as an example, for a female patient, who was not taking methotrexate at baseline, has a RF measurement of 60 IU/ml, and an ACPA measurement of 150 IU/ml, the PI would be 0.141, and the predicted risk of flare by 168 days after DMARD cessation would be 36.7%.

The model had an optimism-corrected C index of 0.709 and calibration slope of 1.00, indicating acceptable classification performance and good agreement between estimates of flare risk and observed risk (see Statistical analysis section of *Supplementary material and Supplementary Figure 2).* The sensitivity analysis of the prediction model, using last face-to-face study visits, demonstrated comparable predictive properties (C-index 0.707, calibration slope 0.996).

4.0 Discussion

BIO-FLARE is an experimental medicine study designed to provide insights into the biological processes that trigger episodes of flare in patients with RA. The ability to compare patients who remain in remission upon DMARD cessation with those who flare provides a well-controlled biological model. In this current work we describe the clinical characteristics of the BIO-FLARE

cohort, report the main clinical outcomes, and explore predictors of flare among routine baseline clinical parameters.

Approximately 50% of participants experienced a flare over the six-month study period, which is similar to the results of previous csDMARD withdrawal studies in RA (6, 17, 18, 19).

Among baseline parameters we identified methotrexate use, female sex, RF level, and ACPA level as significant predictors of flare following DMARD cessation. Higher RF and ACPA levels have been associated with adverse outcomes in RA, including radiographic progression, and may indicate a more aggressive disease phenotype (20), and seropositivity is associated with progression from pre-clinical to clinically apparent RA (21, 22), which might be analogous mechanistically to flare. Similarly, female sex has been associated with progression to RA from early undifferentiated arthritis (23).The increased risk of flare following methotrexate cessation might reflect more severe underlying disease, confounding by indication for other reasons (i.e. reasons for avoiding or previous discontinuation of methotrexate might be protective), and/or a particular pharmacodynamic mechanism of action that leads to highly effective suppression of disease activity but not true biological remission. Longer time from symptom onset to DMARD initiation had a borderline association with flare and was not selected for inclusion in the final model, but does hint at early and effective treatment modifying the probability of achieving drug-free remission, in line with the "window of opportunity" concept (24, 25).

The association between female sex, RF and ACPA positivity and flare has been noted in previous DMARD withdrawal studies and lends face validity to our results (17, 26). In the BioRRA study, a precursor to BIO-FLARE, RF positivity and longer time from diagnosis to first DMARD were also associated with flare, while a borderline association was seen for baseline methotrexate use (6).

Unlike previous DMARD withdrawal studies, we adopted a predictive modelling approach towards our baseline clinical parameters and developed an exploratory prediction model that allows

estimation of risk of flare for an individual patient by a given time following csDMARD cessation. To our knowledge, no comparable clinical model has been described previously in this context. Our final model had acceptable performance in classifying flare versus remission, with good agreement overall between observed and predicted risks. An easy-to-use online version of the formula can be found at https://research.ncl.ac.uk/bioflare/outputs/. Using this tool, sex, methotrexate use, baseline RF and ACPA values can be entered and a predicted risk of flare at 90 or 168 days obtained. Given the lack of external validation, we do not recommend that this tool is used to guide clinical decisions. Nevertheless, in producing a predictive model using only routinely collected data, we present a benchmark against which future molecular or multimodal models can be compared.

Strengths of our study include the number of participants, which compares favourably with previous DMARD withdrawal studies, the prospective study design, and the minimal missing data among baseline parameters. Our predictive modelling followed a robust statistical approach, thereby reducing risk of bias from sensitivity to sampling variability through bootstrapping and overfitting through shrinkage. Nevertheless, our study does have some limitations. BIO-FLARE included participants on csDMARDs only, meaning the relevance of our findings to patients treated with biologic or targeted synthetic DMARDs is uncertain. However, recent studies suggest that up to 40 – 50% of real-world RA patients are treated with csDMARDs alone (27, 28), and it is possible that the immunobiological mechanisms underlying flare may be intrinsic to RA disease processes and thus independent of DMARD treatment. The DAS28-CRP score, used in our study to define remission and flare, has been criticised in the past for being overly permissive of active inflammation (29). However, we used a stringent cut-off of < 2.4, and found similar percentages of participants achieving Boolean and SDAI remission at baseline between subsequent flare and remission groups, suggesting flare was not simply driven by discrepancies in uncaptured initial disease activity. The open-label treatment withdrawal creates a risk of flares driven by the nocebo effect, but this was a pragmatic study design that reflects clinical practice. Musculoskeletal imaging was not performed at baseline, meaning the predictive potential of radiographic erosions, or ultrasonographic

synovitis/tenosynovitis could not be assessed. The COVID-19 pandemic limited face-to-face assessments for a relatively small proportion of study participants, but the close similarity between our primary and sensitivity analyses suggests our mitigation strategy was valid, without an obvious impact on the performance of the prediction model. Finally, the six-month follow-up period means that longer term outcomes, such as occurrence of flares beyond 24 weeks, response to csDMARD re-initiation, and long term sequelae that might be associated with flares, were not captured by the current study.

In conclusion, approximately half of RA patients in remission on csDMARDs experienced a flare within 6 months of stopping therapy, with a median time-to-flare of 9 weeks. Among baseline clinical parameters, RF and ACPA levels, female sex and methotrexate use were found to be predictive of flare. Our predictive model allows estimation of risk of flare at the individual level based on clinical parameters alone. We will subsequently strengthen this by the addition of immune biomarkers emerging from our BIO-FLARE laboratory analyses.

Contributors

JDI, AA, CMUH, CDB, IM, W-FN, CG, AF, SS, KR, AGP and KFB were involved in study conception, funding acquisition, study design and supervision. FR, BD, SK and AMc performed data collection and FR, BD and SK were involved in study design. FR, SH and AM directly accessed the full dataset and performed data analysis, interpretation and visualisation as well as performing a literature search, drafting the original text and reviewing and editing the text for final submission. SH, TB, MSS and MDT were involved in data analysis and validation. JP was responsible for data curation and software. All authors reviewed the final manuscript. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

The lead author (the manuscript's guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Funding

The study was funded by a Medical Research Council experimental medicine grant (reference number MR/N026977/1) and sponsored by the Newcastle upon Tyne Hospitals NHS Foundation Trust. The funders and sponsor had no role in study design, data collection, data analysis, data interpretation or report writing.

Declarations of interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/disclosure-of-interest/ and declare: JDI discloses research grants from Pfizer, Janssen and GSK; conference support from Eli Lilly; and has consulted for: Abbvie, Anaptys Bio, AstraZeneca, BMS, Eli Lilly, Galapagos NV, Gilead Sciences Ltd, GSK, Istesso Ltd, Kira Biotech, Ono Pharma, Pfizer, Revelo Biotherapeutics, Roche, Sanofi. SS reports institutional research grants from Amgen (previously Celgene), Boehringer Ingelheim, Bristol Myers Squibb, Eli Lilly, GSK, Janssen and UCB; speaker/consulting fees from AbbVie, Amgen, AstraZeneca, Eli Lilly, GSK, Janssen, UCB. IM declares honoraria or research support from Abbvie, Janssen, Novartis, Eli Lilly, Astra Zeneca, GSK, BMS, Moonlake, Evelo, Causeway THerapeutics, Cabaletta, Roche, Pfizer and Compugen. CMUH declares research funding from GSK. KFB declares research support from Genentech, clinical improvement funding from Pfizer, and consulting fees from Modern Biosciences. AGP, KFB and JDI are named as inventors on a patent application by Newcastle University ("Prediction of Drug-Free Remission in Rheumatoid Arthritis"; International Patent Application Number PCT/GB2019/050902). KR has received research grant support from Bristol Myers Squibb and personal fees for lectures / consultancy from Abbvie and Sanofi. AF reports institutional research grants from BMS, Roche, UCB, Nascient, Mestag, GSK, Janssen and speaker/consulting fees from Roche, Janssen, Sonoma.

Ethics statements

Patient consent for publication

Not applicable

Ethics approval

The study was reviewed and approved by the North East – Newcastle and North Tyneside 1 Research Ethics Committee (REC) (reference 17/NE/0386). Written informed consent was provided by all participants before study enrolment.

Data availability statement

Data are available from the corresponding author upon reasonable request. All data relevant to the study are included in the article or uploaded as supplementary information.

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

Acknowledgements

This work was funded by the Medical Research Council and supported by the National Institute for Health and Care Research (NIHR) Newcastle and Birmingham Biomedical Research Centres and NIHR Newcastle and Birmingham Clinical Research Facilities, the Research into Inflammatory Arthritis Centre Versus Arthritis, and Rheuma Tolerance for Cure (European Union Innovative Medicines Initiative 2). JDI is a NIHR Senior Investigator. KFB was funded by a National Institute for Health and Care Research (NIHR) Clinical Lectureship (CL-2017-01-004 to KFB), a Newcastle Health Innovation Partners Senior Clinical Fellowship, and the Newcastle Hospitals Charity (8033). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

References

1. McInnes IB, Schett G. The Pathogenesis of Rheumatoid Arthritis. New England Journal of Medicine. 2011;365(23):2205-19.
2. van Mulligen E, Weel A, Kuijper TM, Hazes JMW, van der Helm-van Mil AHM, de Jong PHP. The impact of a disease flare during tapering of DMARDs on the lives of rheumatoid arthritis patients. Seminars in arthritis and rheumatism. 2020;50(3):423-31.
3. Smolen JS, Pedersen R, Jones H, Mahgoub E, Marshall L. Impact of flare on radiographic progression after etanercept continuation, tapering or withdrawal in patients with rheumatoid arthritis. Rheumatology (Oxford). 2020;59(1):153-64.
4. Myasoedova E, Chandran A, Ilhan B, Major BT, Michet CJ, Matteson EL, et al. The role of rheumatoid arthritis (RA) flare and cumulative burden of RA severity in the risk of cardiovascular disease. Ann Rheum Dis. 2016;75(3):560-5.
5. Smolen JS, Landewé RBM, Bergstra SA, Kerschbaumer A, Sepriano A, Aletaha D, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2022 update. Annals of the rheumatic diseases. 2023;82(1):3-18.
6. Baker KF, Skelton AJ, Lendrem DW, Scadeng A, Thompson B, Pratt AG, et al. Predicting drug-free remission in rheumatoid arthritis: A prospective interventional cohort study. J Autoimmun. 2019;105:102298.
7. Tanaka Y, Hirata S, Kubo S, Fukuyo S, Hanami K, Sawamukai N, et al. Discontinuation of adalimumab after achieving remission in patients with established rheumatoid arthritis: 1-year outcome of the HONOR study. Ann Rheum Dis. 2015;74(2):389-95.
8. Henaux S, Ruyssen-Witrand A, Cantagrel A, Barnetche T, Fautrel B, Filippi N, et al. Risk of losing remission, low disease activity or radiographic progression in case of bDMARD discontinuation or tapering in rheumatoid arthritis: systematic analysis of the literature and meta-analysis. Annals of the rheumatic diseases. 2018;77(4):515-22.
9. Rayner F, Anderson AE, Baker KF, Buckley CD, Dyke B, Fenton S, et al. BIOlogical Factors that Limit sustAined Remission in rhEumatoid arthritis (the BIO-FLARE study): protocol for a non-randomised longitudinal cohort study. BMC rheumatology. 2021;5(1).
10. Fleischmann R, van der Heijde D, Koenig AS, Pedersen R, Szumski A, Marshall L, et al. How much does Disease Activity Score in 28 joints ESR and CRP calculations underestimate disease activity compared with the Simplified Disease Activity Index? Ann Rheum Dis. 2015;74(6):1132-7.
11. Bykerk VP, Massarotti EM. The new ACR/EULAR remission criteria: rationale for developing new criteria for remission. Rheumatology. 2012;51(suppl 6):vi16-vi20
12. van Buuren SG-O, K. mice: Multivariate Imputation by Chained Equations in R. HJournal of Statistical Software. 2011;45(3):1-67
13. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in medicine. 1996;15(4):361-87.
14. Stevens RJ, Poppe KK. Validation of clinical prediction models: what does the "calibration slope" really measure? Journal of Clinical Epidemiology. 2020;118:93-9.
15. Crowson CS, Atkinson EJ, Therneau TM. Assessing calibration of prognostic risk scores. Statistical Methods in Medical Research. 2016;25(4):1692-706.
16. Studenic P, Aletaha D, De Wit M, Stamm TA, Alasti F, Lacaille D, et al. American College of Rheumatology/EULAR Remission Criteria for Rheumatoid Arthritis: 2022 Revision. Arthritis & Rheumatology. 2023;75(1):15-22.
17. Verstappen M, Van Mulligen E, De Jong PHP, Van Der Helm-Van Mil AHM. DMARD-free remission as novel treatment target in rheumatoid arthritis: A systematic literature review of achievability and sustainability. RMD open. 2020;6(1):e001220.

18. Lee JS, Oh JS, Hong S, Kim YG, Lee CK, Yoo B. Six-month flare risk after discontinuing long-term methotrexate treatment in patients having rheumatoid arthritis with low disease activity. International journal of rheumatic diseases. 2020;23(8):1076-81.
19. Kuijper TM, Lamers-Karnebeek FBG, Jacobs JWG, Hazes JMW, Luime JJ. Flare Rate in Patients with Rheumatoid Arthritis in Low Disease Activity or Remission When Tapering or Stopping Synthetic or Biologic DMARD: A Systematic Review. The Journal of rheumatology. 2015;42(11):2012-22.
20. Syversen SW, Gaarder PI, Goll GL, Ødegård S, Haavardsholm EA, Mowinckel P, et al. High anti-cyclic citrullinated peptide levels and an algorithm of four variables predict radiographic progression in patients with rheumatoid arthritis: results from a 10-year longitudinal study. Annals of the rheumatic diseases. 2008;67(2):212-7.
21. van Steenbergen HW, Mangnus L, Reijnierse M, Huizinga TWJ, van der Helm-van Mil AHM. Clinical factors, anticitrullinated peptide antibodies and MRI-detected subclinical inflammation in relation to progression from clinically suspect arthralgia to arthritis. Annals of the rheumatic diseases. 2016;75(10):1824-30.
22. ten Brinck RM, van Steenbergen HW, van Delft MAM, Verheul MK, Toes REM, Trouw LA, et al. The risk of individual autoantibodies, autoantibody combinations and levels for arthritis development in clinically suspect arthralgia. Rheumatology. 2017;56(12):2145-53.
23. Van Der Helm-Vanmil AHM, Le Cessie S, Van Dongen H, Breedveld FC, Toes REM, Huizinga TWJ. A prediction rule for disease outcome in patients with Recent-onset undifferentiated arthritis: How to guide individual treatment decisions. Arthritis & Rheumatism. 2007;56(2):433-40.
24. Van Nies JAB, Tsonaka R, Gaujoux-Viala C, Fautrel B, Van Der Helm-Van Mil AHM. Evaluating relationships between symptom duration and persistence of rheumatoid arthritis: does a window of opportunity exist? Results on the Leiden Early Arthritis Clinic and ESPOIR cohorts. Annals of the rheumatic diseases. 2015;74(5):806-12.
25. van der Linden MPM, le Cessie S, Raza K, van der Woude D, Knevel R, Huizinga TWJ, et al. Long-term impact of delay in assessment of patients with early arthritis. Arthritis & rheumatism. 2010;62(12):3537-46.
26. Vittecoq O, Desouches S, Kozyreff M, Nicolau J, Pouplin S, Rottenberg P, et al. Relapse in rheumatoid arthritis patients undergoing dose reduction and withdrawal of biologics: are predictable factors more relevant than predictive parameters? An observational prospective real-life study. BMJ open. 2019;9(12):e031467.
27. Yates M, Ledingham JM, Hatcher PA, Adas M, Hewitt S, Bartlett-Pestell S, et al. Disease activity and its predictors in early inflammatory arthritis: findings from a national cohort. Rheumatology. 2021;60(10):4811-20.
28. Mease PJ, Stryker S, Liu M, Salim B, Rebello S, Gharaibeh M, et al. Treatment patterns in rheumatoid arthritis patients newly initiated on biologic and conventional synthetic disease-modifying antirheumatic drug therapy and enrolled in a North American clinical registry. Arthritis research & therapy. 2021;23(1).
29. Aletaha D, Smolen JS. Joint damage in rheumatoid arthritis progresses in remission according to the Disease Activity Score in 28 joints and is driven by residual swollen joints. Arthritis Rheum. 2011 Dec;63(12):3702-11.

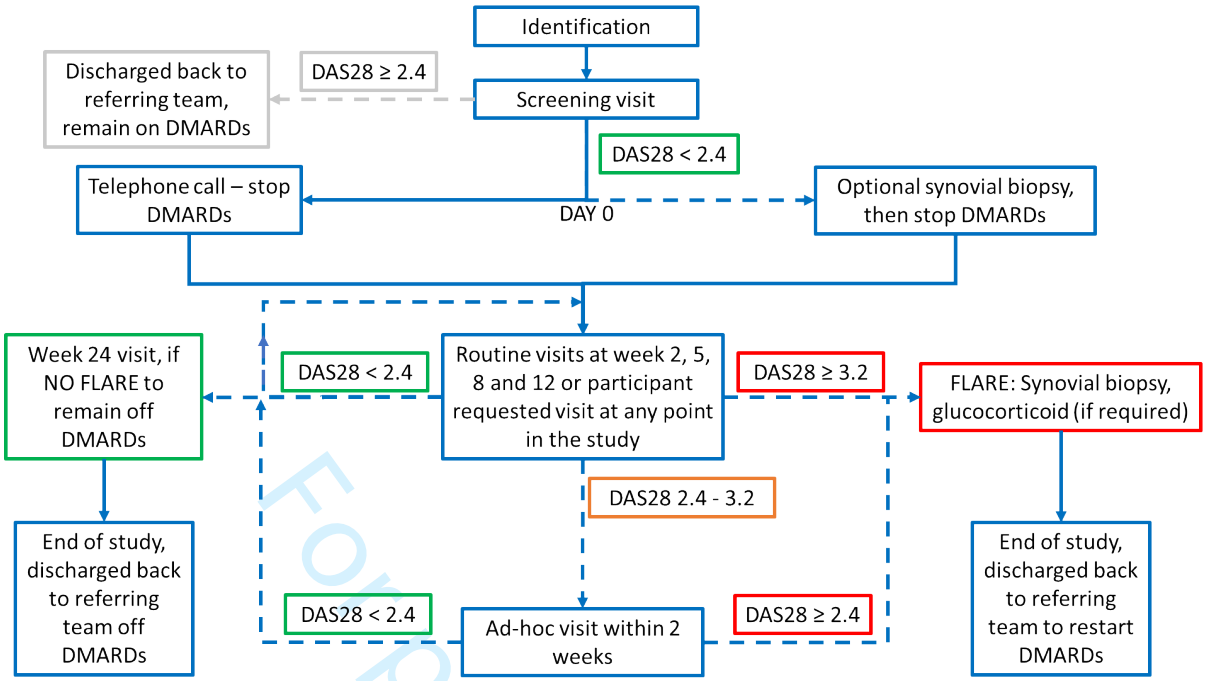


Figure 1. Participant pathway through the study.

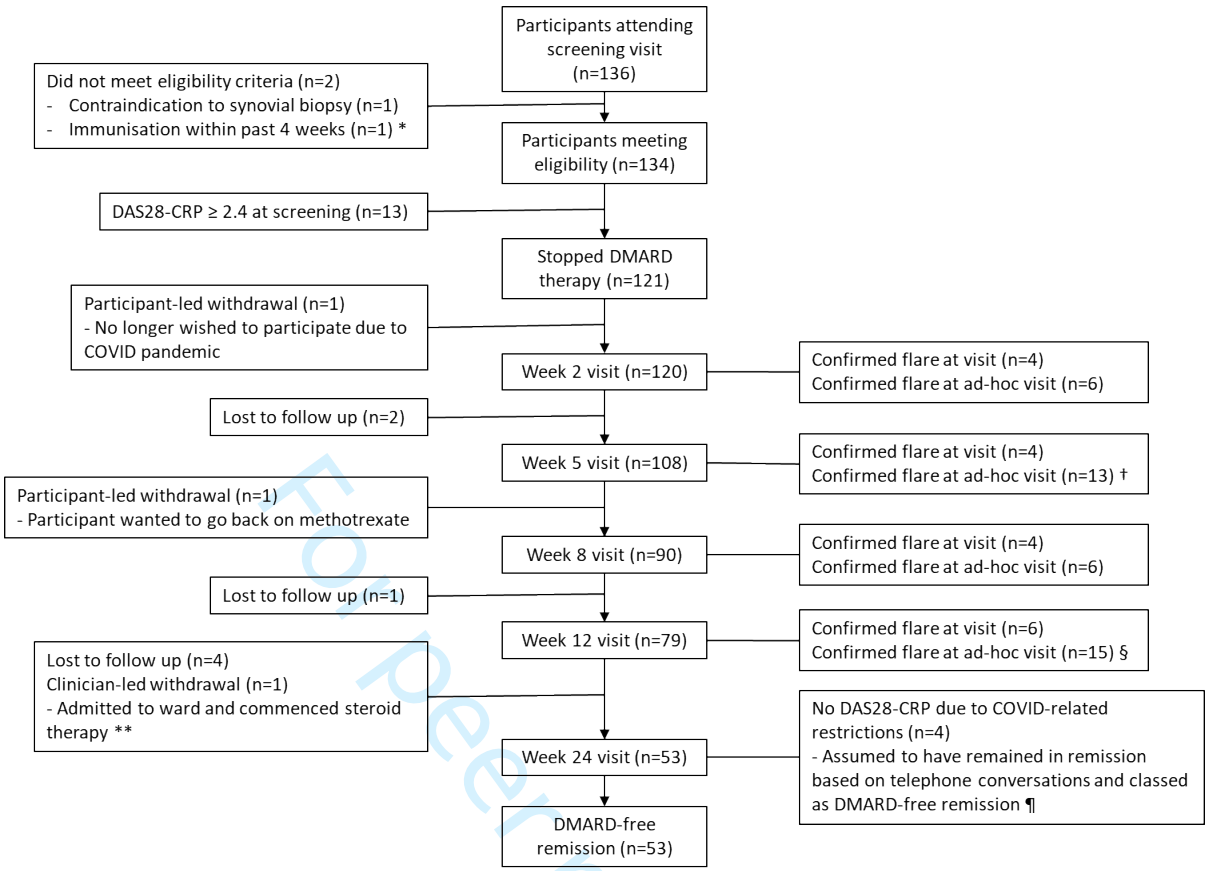


Figure 2: Participant flow diagram. * Participant discovered to have had an immunisation prior to screening at their week 2 visit; † n=2 participants flared based on clinical discretion at face-to-face visit; § n=1 flare based on clinical discretion at face-to-face visit; ** participant was censored at day 84 visit as discovered to have an intercurrent illness at week 24 visit and was withdrawn from the study; ¶ participants had last face-to-face visits at week 2 (n=1), week 5 (n=1), ad hoc visit after week 5 (n=1), and week 12 (n=1) visits.

		Total study population (n=121)	Modified per-protocol population (n=111)*		Missing data (n=121), N %
			Flare (n=58)	Remission up to week 24 (n=53)	
Age, years	Mean (SD)	64.07 (11.9)	64.76 (11.6)	64.68 (11.3)	0 (0.0)
Sex, Female	N (%)	73 (60.3)	41 (70.7)	27 (50.9)	0 (0.0)
Body mass index, kg/m^2	Mean (SD)	28.20 (5.7)	27.24 (5.3)	29.36 (5.9)	5 (4.1)
Charlson comorbidity index	Median (IQR)	2.00 (1.0, 3.0)	2.00 (1.0, 4.0)	3.00 (1.0, 3.0)	0 (0.0)
Tobacco smoking status	N (%)				0 (0.0)
Current		8 (6.6)	5 (8.6)	3 (5.7)	
Ex-smoker		64 (52.9)	32 (55.2)	26 (49.1)	
Never smoked		49 (40.5)	21 (36.2)	24 (45.3)	
Current alcohol use	N (%)	73 (60.8)	32 (55.2)	36 (67.9)	1 (0.8)
Ethnicity	N (%)				0 (0.0)
White British/Other White		113 (93.4)	55 (94.8)	51 (96.2)	
Asian/Asian British		6 (5.0)	3 (5.2)	1 (1.9)	
Black/Black British - Caribbean		2 (1.7)	0 (0.0)	1 (1.9)	
Highest educational qualification	N (%)				2 (1.7)
GCSEs or equivalent		33 (27.3)	12 (21.0)	18 (34.0)	
A-Level or equivalent		13 (10.7)	7 (12.1)	6 (11.3)	
Undergraduate		20 (16.5)	12 (20.7)	7 (13.2)	
Postgraduate		14 (11.6)	6 (10.3)	5 (9.4)	
NVQ or equivalent		14 (11.6)	5 (8.6)	6 (11.3)	
None of the above		24 (19.8)	15 (25.9)	9 (17.0)	
Not stated or missing		3 (2.5)	1 (1.7)	2 (3.8)	
Employment status	N (%)				0 (0.0)
Full-time		30 (24.8)	13 (22.4)	15 (28.3)	
Part-time		13 (10.7)	7 (12.1)	4 (7.6)	
Unemployed		3 (2.5)	0 (0.0)	0 (0.0)	
Self employed		2 (1.7)	1 (1.7)	1 (1.9)	
Retired		71 (58.7)	37 (63.8)	31 (58.5)	
Other		2 (1.7)	0 (0.0)	2 (3.8)	
Time from symptom onset to first DMARD, years	Median (IQR)	0.51 (0.3, 1.3)	0.54 (0.3, 2.1)	0.51 (0.3, 1.0)	4 (3.3)
Time from symptom onset to baseline, years	Median (IQR)	6.33 (4.5, 12.4)	6.34 (5.0, 13.7)	6.17 (3.9, 10.8)	4 (3.3)
Time from RA diagnosis to baseline, years	Median (IQR)	5.48 (3.7, 10.5)	5.48 (4.2, 10.7)	5.36 (3.3, 9.7)	2 (1.7)
MTX use at baseline	N (%)	101 (83.5)	53 (91.4)	39 (73.6)	0 (0.0)
MTX dose, mg/week	Median (IQR)	15 (12.5, 20)	15 (12.5, 20)	15 (12.5, 20)	0 (0.0)
MTX monotherapy	N (%)	72 (59.5)	37 (63.8)	28 (52.8)	0 (0.0)
MTZ + SZN	N (%)	5 (4.1)	3 (5.2)	2 (3.8)	0 (0.0)
MTX + HCQ	N (%)	22 (18.2)	11 (19.0)	9 (17.0)	0 (0.0)
MTX + SZN + HCQ	N (%)	2 (1.7)	2 (3.5)	0 (0.0)	0 (0.0)
SZN monotherapy	N (%)	10 (8.3)	3 (5.2)	7 (13.2)	0 (0.0)

		Total study population (n=121)	Modified per-protocol population (n=111)*		Missing data (n=121), N %
			Flare (n=58)	Remission up to week 24 (n=53)	
HCQ monotherapy	N (%)	8 (6.6)	2 (3.5)	5 (9.4)	0 (0.0)
SZN + HCQ	N (%)	2 (1.7)	0 (0.0)	2 (3.8)	0 (0.0)
Previous biologic therapy	N (%)	1 (0.8)	1 (1.7)	0 (0.0)	0 (0.0)
Corticosteroid use in past 12 months	N (%)	7 (5.8)	3 (5.2)	4 (7.6)	
Any		7 (5.8)	3 (5.2)	4 (7.6)	0 (0.0)
Oral		3 (2.5)	1 (1.7)	2 (3.8)	0 (0.0)
Intramuscular		1 (0.8)	0 (0.0)	1 (1.9)	0 (0.0)
Intra-articular		2 (1.7)	2 (3.5)	1 (1.9)	0 (0.0)
RF positive	N (%)	67 (56.3)	42 (72.4)	19 (36.5)	2 (1.7)
ACPA positive	N (%)	76 (66.7)	45 (77.6)	25 (49.0)	7 (5.8)
RF, IU/ml	Median (IQR)	32.00 (0.0, 94.1)	53.15 (14.0, 130.0)	12.65 (0.0, 40.1)	2 (1.7)
ACPA, U/ml	Median (IQR)	96.50 (1.1, 300.0)	207.00 (31.0, 306.5)	1.70 (0.8, 196.0)	7 (5.8)
DAS28-CRP	Mean (SD)	1.61 (0.3)	1.62 (0.3)	1.60 (0.3)	0 (0.0)
ACR/EULAR 2011 Boolean remission	N (%)	74 (61.2)	38 (65.5)	31 (58.5)	0 (0.0)
ACR/EULAR Boolean 2.0 remission	N (%)	95 (78.5)	48 (82.8)	41 (77.4)	0 (0.0)
SDAI remission	N (%)	101 (84.9)	49 (86.0)	45 (84.9)	2 (1.7)
HAQ-DI	Median (IQR)	0.00 (0.0, 0.6)	0.13 (0.0, 0.7)	0.00 (0.0, 0.4)	0 (0.0)
Follow-up time, days	Median (IQR)	115.5 (55.5, 167.5)	63.0 (41.0, 96.0)	168.0 (167.0, 174.0)	0 (0.0)

Table 1: Baseline characteristics. *The modified per-protocol population includes all participants with known outcome status and excludes those lost to follow-up (n=7) or withdrawn (n=3) before week 24 visit. ACR=American College of Rheumatology, ACPA=Anti-Citrullinated Peptide Antibody, DMARD=Disease Modifying Anti-Rheumatic Drug, EULAR=European Alliance of Associations for Rheumatology, HAQ-DI=Health Assessment Questionnaire Disability Index, HCQ=Hydroxychloroquine, IQR=Interquartile Range, MTX=Methotrexate, N=Number, NVQ=National Vocational Qualification, RF=Rheumatoid Factor, SD=Standard Deviation, SDAI=Simplified Disease Activity Index, SZN=Sulfasalazine.

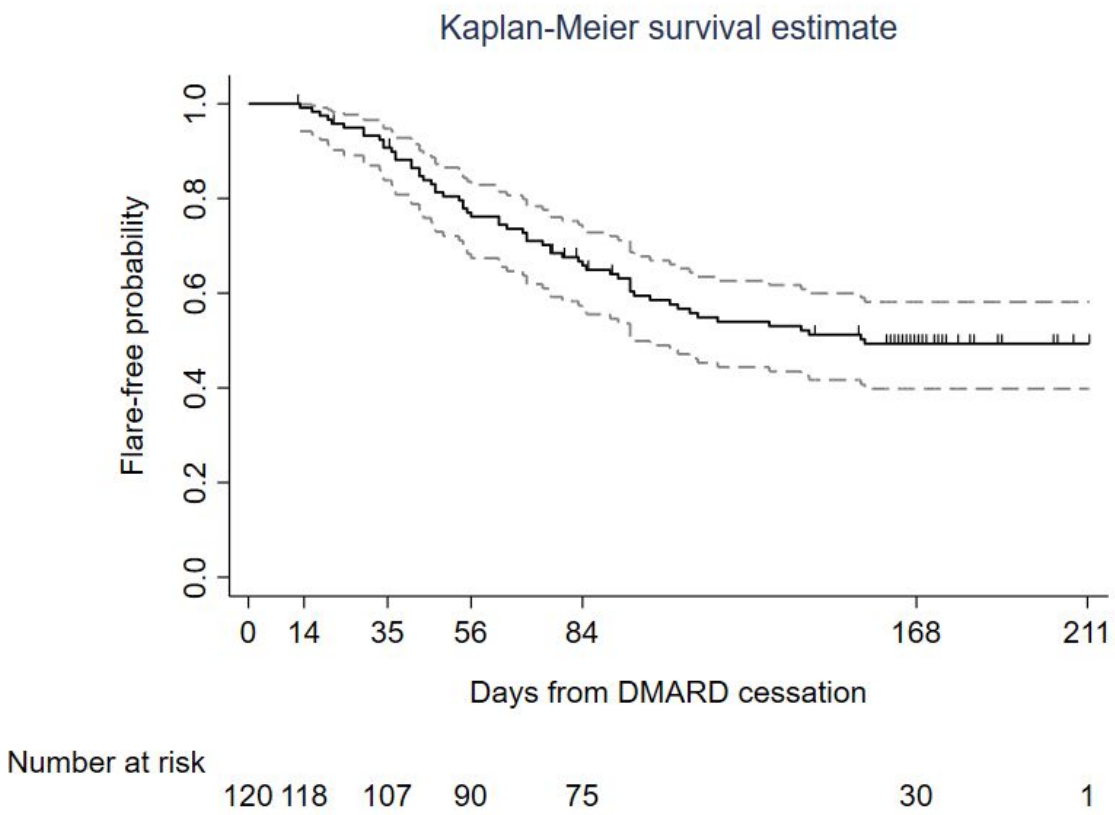


Figure 3: Kaplan-Meier plot of flare-free probability in the analysis cohort. Solid black line is the Kaplan-Meier estimate of the flare-free function, the grey dashed lines are the 95% CI, and black vertical marks indicate censoring. Outcomes defined as per primary analyses. A Kaplan-Meier plot including only data from final face-to-face study visits (sensitivity analysis) is included as *Supplementary Figure 1*. DMARD=Disease Modifying Anti-Rheumatic Drug.

		Complete case analysis			Multiple imputation with chained equations (MICE) (n=120)	
		Available n	HR [95% CI]	P-value	HR [95% CI]	P-value
	Reference					
Age		120	1.00 [0.98 to 1.02]	0.99		
Male sex	Female	120	0.54 [0.31 to 0.95]	0.03		
Years from diagnosis to baseline visit		118	1.01 [0.99 to 1.04]	0.37	1.01 [0.98 to 1.04]	0.47
Years from symptom onset to first DMARD		112	1.04 [1.00 to 1.07]	0.04	1.03 [1.00 to 1.07]	0.06
Methotrexate use at baseline	No	120	2.92 [1.16 to 7.31]	0.02		
RF level at baseline, per 10 IU/ml		118	1.03 [1.01 to 1.06]	0.001	1.03 [1.01 to 1.06]	0.001
ACPA level at baseline, per 10 U/ml		113	1.03 [1.01 to 1.04]	0.01	1.03 [1.01 to 1.04]	0.01
DAS28-CRP		120	1.14 [0.51 to 2.52]	0.75		
ACR/EULAR 2011 Boolean remission at baseline	Not in remission	120	1.16 [0.68 to 2.00]	0.59		
A-level and above education	GCSE and under	117	1.15 [0.68 to 1.94]	0.59	1.14 [0.67 to 1.94]	0.62
BMI		115	0.95 [0.90 to 1.01]	0.08	0.96 [0.90 to 1.01]	0.11
Current smoker	Never or ex-smoker	120	1.04 [0.41 to 2.60]	0.94		
Current alcohol use	No	119	0.63 [0.37 to 1.05]	0.07	0.63 [0.38 to 1.06]	0.09
Charlson comorbidity index		120	1.04 [0.88 to 1.22]	0.66		
Glucocorticoid use in past 12 months from baseline	No	120	0.78 [0.24 to 2.48]	0.67		

Table 2. Univariate analysis of candidate baseline variables predicting flare in analysis cohort. Employment variable was not included in the imputation model due to convergence issues. Variables with no missing data have empty rows under the MICE column because estimates will be identical to the complete case analysis. Hazard ratios for continuous variables are calculated per 1 unit increase unless otherwise stated. ACPA=Anti-citrullinated Peptide Antibody, BMI=Body Mass Index, RF=Rheumatoid Factor, GCSE=General Certificate of Secondary Education, HR=Hazard Ratio, NA=Not Applicable

Clinical predictors of flare and drug-free remission in rheumatoid arthritis: preliminary results from the prospective BIO-FLARE experimental medicine study

– Supplementary Material

Contents

Statistical analysis - Predictive model with baseline risk factors

The primary analysis of this study involved the building and presentation of a prediction model containing baseline risk factors of RA flare through three steps: 1) Variable selection for the prediction model; 2) Assessment of non-linear forms for continuous predictors; 3) Estimation of a shrinkage factor to reduce overfitting; 4) Internal validation of the prediction model; and 5) Presenting the predicted probability of flare as a function of the prognostic index and time. For ease of reporting, we denote lower case i to represent an imputed dataset ($i = 1, 2, \ldots, I$) and lower case j to denote a bootstrap resample of an imputed dataset ($j = 1, 2, \ldots, J$).

1. Variable selection. We computed the number of variables that may be included in our prediction model, under the assumption that there would be minimal overfitting (i.e. an expected shrinkage factor ≥0.90), using the R *pmsampsize* package (1, 2). We used the results from a previous baseline prediction model (3) for RA flare as the basis for our calculations: an adjusted Cox-Snell $R^2 = 0.473$ (derived from the lower limit of the 95% CI of their reported area under curve (0.91) which served as a proxy for the C statistic and using equations 5, 8, 18, 20, 21, and 22 from Riley et al); an overall flare rate of 23/2112 = 0.011, with an average follow-up of 48 days post DMARD cessation; a time point of interest of 168 days; and a sample size of n=121. The computation indicated an upper limit of nine predictor variables.

A subset of 16 baseline variables was first selected as risk factors for flare based on biological plausibility and previous literature.

- Age at baseline
- Sex (Female vs. Male)
- Time from diagnosis to baseline in years
- Time from symptom onset to DMARD commencement in years

- Methotrexate use (Yes vs. No)
- Rheumatoid factor (RF) in IU/ml
- Anti-citrullinated peptide antibody (ACPA) in U/ml
- DAS28-CRP
- ACR/EULAR Boolean Remission (In remission vs. Not in remission) (4)
- Education (GCSE and under vs. A-levels and above (including national vocational qualifications))
- Any employment (Unemployed vs. Employed or retired)*
- Body mass index in kg/m^2
- Current smoking (Never or ex-smoker vs. Current smoker)
- Any alcohol use (Yes vs. No)
- Charlson comorbidity index (5)
- Corticosteroid use

* Due to the problems with convergence for the employment variable observed in the univariate analyses, this variable was dropped from all multiple imputation and prediction modelling steps.

Owing to the presence of missing data, we performed all analyses with I=10 imputed datasets using multiple imputation by chained equations (MICE) (6). Predictive mean matching was used as the imputation method as it has been shown to produce the less biased estimates in and better predictive performance of Cox PH models than complete case analyses or single imputation methods when the missing covariate data rate is >10% (7). The 15 candidate predictors, the outcome indicator, and the Nelson-Aalen estimate of the cumulative baseline hazard were included in the imputation model (8).

For each imputed dataset i, J=200 datasets were generated via bootstrapping (resampling with replacement). The use of MI followed by bootstrapping allows one to account for uncertainty due to missing data and uncertainty due to sampling variability during variable selection (9). Within each MICE-cum-bootstrap dataset, the 15 variables were included in a penalised Cox proportional hazard model (PH) using elastic net penalty for further variable selection whilst addressing issues of multicollinearity. We used 10-fold cross validation to select an optimal mixing parameter α and tuning parameter λ. We varied α from 0.1 to 0.9 at increments of 0.1. At each value of α, a value of the tuning parameter λ was selected at one standard error from the value of that λ associated with the regularised model with the smallest out-of-fold concordance statistic C. From the nine sets of α_k, λ_k, and C_k $\{k = 1, 2,\ldots 9\}$, the optimal values of α and λ was obtained from searching for the largest C_k. We assessed the stability of the candidate predictors by inspecting their bootstrap inclusion frequencies (BIF) across all 2000 datasets. The idea is that if a baseline risk factor was associated with RA flare, it should consistently, or at least in a large number of times, be selected into the final model even under "perturbations" in the data. Bootstrap resampling has been found to be a useful method of mimicking these modifications for Cox PH models (10-12). The BIF of each candidate variable is the number of times it was included in the regularised model at the optimal values of α and λ across all bootstrap replications of an imputed dataset. We identified stable risk factors as predictors with >60% BIF across the average of all imputations (up to a limit of nine predictors). The value of 60% was decided a priori before analyses were conducted.

2. Assessment of non-linear forms for continuous covariates. For each imputed dataset, we conducted univariable fractional polynomials (FP) to explore the best-fitting non-linear functional form of the selected continuous predictors (RF and ACPA). For each covariate, we explored first-degree and second-degree FPs in a univariate Cox model using the RA2 closed test procedure with a nominal α value of 0.10. To avoid numerical issues, a constant of 0.1 was added to the continuous variables (13). For RF, all imputations suggested a first-degree fractional polynomial with a square root transformation. For ACPA, 4/10 imputations suggested second-degree fractional polynomials with inverse and negative square root transformations respectively, 3/10 suggested a first-degree fractional polynomial with a log-transformation respectively, and 3/10 suggested second-degree fractional polynomials with two terms with inverse transformations respectively.

3. Estimation of shrinkage factor. To reduce the effects of overfitting, for each imputation i we estimated a shrinkage factor S_i using bootstrap estimation with J=200 resamples. A recent study demonstrated that for studies with small sample sizes, bootstrapping may be preferred over the heuristic shrinkage or penalised regression methods to obtain a more reliable estimate of a shrinkage factor for small sample sizes (14). To illustrate the procedure, consider a particular bootstrap resample j for a particular imputed dataset i. The survival outcome of bootstrap dataset j is regressed on the stable risk factors in the bootstrap sample in a Cox regression model, and the coefficients are saved. A linear predictor (LP) is then calculated as the linear combination of the values of the stable risk factors in the imputed dataset i, weighted by the coefficients derived from the bootstrap sample earlier. The outcome of the imputed dataset i is regressed on the LP and the coefficient of the LP is

saved. The value of *Si* is the average of all the coefficients of LP across the *J* bootstraps. The estimated shrinkages factors ranged from 0.831 to 0.891, which were generally close to the assumed 0.90 uniform shrinkage assumed when computing the maximum number of predictors allowable in our prediction model.

Imputation	Estimated shrinkage factor
1	0.857
2	0.835
3	0.842
4	0.841
5	0.891
6	0.843
7	0.831
8	0.845
9	0.831
10	0.858
Average	**0.848**

4. Internal validation. The objective of this step is to evaluate the predictive performance of the model and derive optimism-corrected indices of discrimination (*C* index) and calibration (calibration slope and calibration-in-the-large).

To obtain estimates of optimism, we used bootstrap estimation with *J*=200 resamples. More details of this procedure can be found elsewhere (15). To illustrate the procedure, we describe the process for deriving the optimism-corrected *C* index but the process for the calibration slope follows a similar logic but using the coefficients of the linear predictors instead. Consider an imputed dataset *i*, we first regressed the survival outcome in imputed dataset *i* on the stable risk factors in imputed dataset *i*, and saved the coefficients. We then shrunk the coefficients by S_i to obtain shrunken coefficients. We computed the LP by taking the linear combination of the predictors weighted by the shrunken coefficients. We regressed the outcome in imputed dataset *i* on the LP and obtained an apparent C_i index.

Now consider a bootstrap resample *j* from the impute dataset *i*. We regressed the survival outcome in bootstrap dataset *j* on the stable risk factors in bootstrap dataset *j*, and saved the coefficients. We then shrunk the coefficients by S_i to obtain shrunken coefficients. We then computed two linear predictors: LP_{boot}, which is linear combination of the stable risk factors in the bootstrap sample *j* weighted by the shrunken coefficients; and LP_{test} which is linear combination of the stable risk factors in the imputed dataset *i* weighted by the shrunken coefficients. We regressed the survival outcome in bootstrap sample *j* on LP_{boot} and obtained C_{boot}. We regressed the survival outcome in imputed dataset *i* on LP_{test} and obtained C_{test}. We then subtracted C_{boot} from C_{test} to get an index of $optimism_j$. We averaged all $optimism_j$ across the *J* resamples to get a stable measure of $optimism_i$. An optimism-corrected C for imputation *i* was then optimism-corrected C_i = apparent C_i minus $optimism_i$. We then averaged all optimism-corrected C_i across *I*=10 imputations using Rubin's rules to obtain a single optimism-corrected *C* index (16).

Imputation	Optimism corrected C-index	Optimism-corrected calibration slope
1	0.711	1.005
2	0.701	1.021
3	0.708	1.010
4	0.710	1.000
5	0.704	0.944
6	0.715	1.014
7	0.707	1.019
8	0.714	1.003
9	0.707	0.998
10	0.708	0.991
Average	**0.709**	**1.000**

5. Presenting the predicted probability of RA flare. The final equation of the prediction model is obtained by first estimating the coefficients of the stable risk factors from a Cox model in each imputation *i,* performing shrinkage using S_i, and then pooling them using Rubin's rules.

We then appended the rows of all imputed datasets *i* to create a stacked dataset (15). Because each participant has *I* replications in this stacked dataset, we gave each observation a weight of 1/*I*. We then computed the *prognostic index* (PI) as the linear combination of the values of the stable risk factors in the stacked dataset, weighted by the coefficients in the final equation of the prediction model. We regressed the survival outcome in the stacked dataset on the PI in a weighted Cox PH model, and obtained the value of the baseline survival function $\hat{S}_0(t)$ (valued at PI equals zero) at 30, 60, 90, 120, and 168 days post-DMARD cessation (see below). The weighted stacked dataset was also used in the computation of calibration plots.

Days after cessation of DMARD	Baseline survival function
30	0.969
60	0.876
90	0.801
120	0.714
168	0.672

Procedures	Screening Visit	Day 0: Baseline (a) Telephone consultation	Day 0: Baseline (b) Synovial biopsy (OPTIONAL)	Day 14	Day 35	Day 56	Day 84	Day 168	Patient-requested ad-hoc study visits	Visit 2 weeks following ad-hoc study visit	Synovial biopsy assessment visit after flare confirmed
Discuss Study / confirm willingness to continue participation in study	X	X	X	X	X	X	X	X	X	X	X
Informed Consent for study	X										
Collect Demographics and medical history	X										
Record Current medication	X			X	X	X	X	X	X	X	
General Physical examination[1]	X										
Rheumatological Assessment - DAS28-CRP	X			X	X	X	X	X	X	X	
Instruction to discontinue DMARDs (if not opting for synovial Biopsy)		X									
Instruction to discontinue DMARDs (if opting for synovial biopsy)			X								
Patient Reported Outcome Measures / Questionnaires											
HAQ-DI	X							X	X	X	
RAPID-3	X			X	X	X	X	X	X	X	
EuroQol 5D-5L	X			X	X	X	X	X	X	X	
MFI	X							X	X	X	
RA-FQ	X			X	X	X	X	X	X	X	
FLARE-RA	X			X	X	X	X	X	X	X	
Blood tests											
Full Blood Count (FBC)	X			X	X	X	X	X	X	X	
Inflammatory markers (ESR & CRP)	X			X	X	X	X	X	X	X	
Antibodies (RF & ACPA)	X										
Other clinical bloods (UE, LFT & Clotting)	X								X		
Research blood tests (Serum, EDTA, Tempus and Heparinised samples)	X			X	X	X	X	X	X	X	
Other research tests											
Urine Sample	X			X	X	X	X	X	X	X	
Pregnancy test[2]	X										
Stool Sample (OPTIONAL)	X			X	X	X	X	X	X	X	
Ultrasound assessment for Synovial Biopsy (OPTIONAL AT BASELINE – additional consent required)			[X]								X
Accelerometer provided[3] (OPTIONAL)	X										
Activity diary provided (OPTIONAL)	X			X	X	X	X		X		

Supplementary Table 1: Schedule of events in the BIO-FLARE study.

[1] Depending on the circumstances of the consultation, physical examination may be indicated at any study visit to establish whether DAS28-CRP reflects arthritis activity or infection etc. General Physical Examination is only mandatory at Screening.
[2] Mandatory at Screening but should be performed at any visit subsequently if routine questioning suggests a participant may be pregnant. Serum or urine tests to be performed subsequently in line with local policy
[3] This may be provided after the study visit once eligibility confirmed, either by post, or at the optional Baseline Synovial Biopsy Visit (if applicable)

Adverse events

In total, 82 out of 121 participants (68%) experienced at least one adverse event (AE) in the sample. There were a total of 155 adverse events with a median of 1 (IQR: 1, 2) event per participant (range: 1 to 6). The breakdown of the number of participants reporting each type of AE is presented below, organised by their system organ class (Supplementary Table 5). Additionally, there were 4 serious adverse events (SAE) occurring over 4 participants (Supplementary Table 6).

		Modified per-protocol cohort (n=111)*	
	Study population (n=121)	Flared (n=58)	Remission at week 24 visit (n=53)
	N (%)	N (%)	N (%)
Blood and lymphatic system disorders			
Anaemia	1 (0.8)	1 (1.7)	0
Neutropenia	2 (1.7)	0	2 (3.8)
Thrombocytopenia	1 (0.8)	0	1 (1.9)
Ear and labyrinth disorders			
Excessive cerumen production	1 (0.8)	0	1 (1.9)
Vertigo	1 (0.8)	0	1 (1.9)
Endocrine disorders			
Hypothyroidism	1 (0.8)	0	1 (1.9)
Eye disorders			
Blepharitis	1 (0.8)	0	1 (1.9)
Cataract	1 (0.8)	0	1 (1.9)
Dry eye	1 (0.8)	1 (1.7)	0
Gastrointestinal disorders			
Constipation	1 (0.8)	0	1 (1.9)
Dyspepsia	1 (0.8)	0	1 (1.9)
Enteritis	1 (0.8)	1 (1.7)	0
Gastritis	1 (0.8)	0	1 (1.9)
Mouth ulceration	1 (0.8)	1 (1.7)	0
Pancreatic mass	1 (0.8)	0	1 (1.9)
Toothache	2 (1.7)	1 (1.7)	1 (1.9)
Vomiting	1 (0.8)	0	1 (1.9)
General disorders and administration site conditions			
Chest pain	1 (0.8)	0	1 (1.9)
Critical illness	1 (0.8)	1 (1.7)	0
Fatigue	2 (1.7)	2 (3.4)	0
Hernia	1 (0.8)	0	1 (1.9)
Malaise	1 (0.8)	0	1 (1.9)
Immune system disorders			
Hypersensitivity	1 (0.8)	0	1 (1.9)
Infections and infestations			
Cellulitis	1 (0.8)	1 (1.7)	0
Conjunctivitis viral	1 (0.8)	1 (1.7)	0
Coxsackie viral infection	1 (0.8)	0	1 (1.9)
Gastroenteritis	1 (0.8)	0	1 (1.9)
Gastroenteritis viral	1 (0.8)	0	1 (1.9)
Infected bite	1 (0.8)	1 (1.7)	0
Lower respiratory tract infection	5 (4.1)	1 (1.7)	4 (7.5)
Oral herpes	1 (0.8)	0	0
Otitis externa	1 (0.8)	1 (1.7)	1 (1.9)
Rash pustular	1 (0.8)	1 (1.7)	0
Rhinitis	1 (0.8)	0	1 (1.9)
Sinusitis	1 (0.8)	1 (1.7)	0
Tooth abscess	1 (0.8)	0	1 (1.9)
Tooth infection	1 (0.8)	0	1 (1.9)
Upper respiratory tract infection	20 (16.5)	9 (15.5)	8 (15.1)
Urinary tract infection	1 (0.8)	0	0

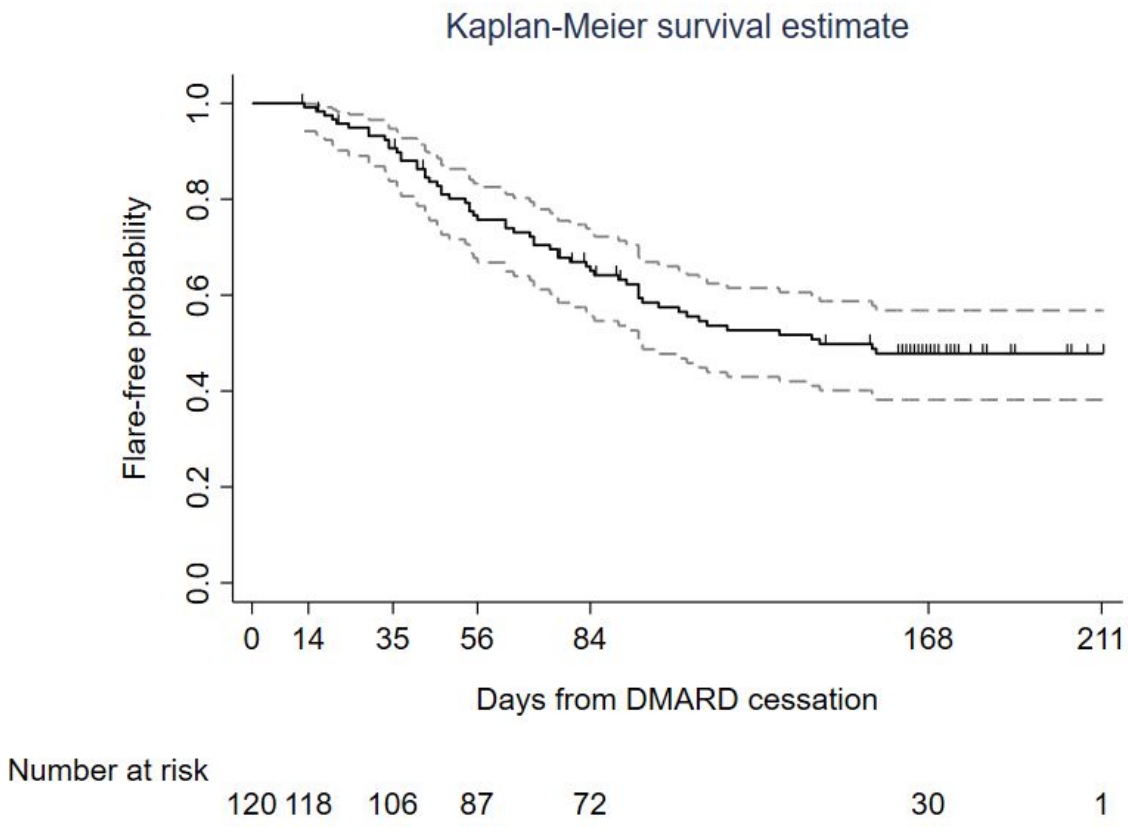
Viral infection	1 (0.8)	1 (1.7)	0
Viral upper respiratory tract infection	7 (5.8)	6 (10.3)	1 (1.9)
Injury, poisoning and procedural complications			
Arthropod bite	1 (0.8)	0	1 (1.9)
Avulsion fracture	1 (0.8)	0	1 (1.9)
Back injury	1 (0.8)	0	1 (1.9)
Contusion	1 (0.8)	1 (1.7)	0
Fall	2 (1.7)	1 (1.7)	1 (1.9)
Laceration	1 (0.8)	1 (1.7)	0
Limb injury	1 (0.8)	1 (1.7)	0
Spinal fracture	2 (1.7)	2 (3.4)	0
Wound	2 (1.7)	1 (1.7)	1 (1.9)
Investigations			
Blood glucose abnormal	1 (0.8)	1 (1.7)	0
C-reactive protein increased	1 (0.8)	1 (1.7)	0
Liver function test abnormal	1 (0.8)	1 (1.7)	0
Platelet count decreased	1 (0.8)	1 (1.7)	0
Musculoskeletal and connective tissue disorders			
Arthralgia	1 (0.8)	0	1 (1.9)
Fracture	1 (0.8)	0	1 (1.9)
Joint stiffness	1 (0.8)	1 (1.7)	0
Musculoskeletal pain	1 (0.8)	1 (1.7)	0
Myalgia	1 (0.8)	1 (1.7)	0
Pain in extremity	1 (0.8)	1 (1.7)	0
Periarthritis	1 (0.8)	1 (1.7)	0
Soft tissue swelling	1 (0.8)	0	1 (1.9)
Tendonitis	2 (1.7)	1 (1.7)	1 (1.9)
Tenosynovitis	2 (1.7)	2 (3.4)	0
Neoplasms benign, malignant and unspecified (incl cysts and polyps)			
Seborrhoeic keratosis	1 (0.8)	0	0
Nervous system disorders			
Cerebrovascular accident	1 (0.8)	0	0
Dizziness	1 (0.8)	1 (1.7)	0
Headache	3 (2.5)	1 (1.7)	2 (3.8)
Migraine	1 (0.8)	0	1 (1.9)
Neuralgia	1 (0.8)	1 (1.7)	0
Restless legs syndrome	1 (0.8)	0	1 (1.9)
Sciatica	4 (3.3)	2 (3.4)	2 (3.8)
Seizure	1 (0.8)	0	1 (1.9)
Syncope	1 (0.8)	0	1 (1.9)
Transient ischaemic attack	1 (0.8)	0	0
Psychiatric disorders			
Depressed mood	1 (0.8)	0	1 (1.9)
Emotional distress	1 (0.8)	1 (1.7)	0
Insomnia	1 (0.8)	1 (1.7)	0
Respiratory, thoracic and mediastinal disorders			
Cough	5 (4.1)	2 (3.4)	3 (5.7)
Nasal dryness	1 (0.8)	1 (1.7)	0
Oropharyngeal pain	6 (5)	4 (6.9)	2 (3.8)
Skin and subcutaneous tissue disorders			
Eczema	1 (0.8)	0	1 (1.9)
Neurodermatitis	1 (0.8)	0	1 (1.9)
Pruritus	1 (0.8)	1 (1.7)	0
Rash	1 (0.8)	1 (1.7)	1 (1.9)
Rash erythematous	2 (1.7)	1 (1.7)	0
Skin lesion	1 (0.8)	1 (1.7)	0
Transient acantholytic dermatosis	1 (0.8)	1 (1.7)	0
Surgical and medical procedures			

Medical device removal	1 (0.8)	0	1 (1.9)
Tooth extraction	2 (1.7)	1 (1.7)	1 (1.9)
Tooth repair	1 (0.8)	1 (1.7)	0
Vascular disorders			
Aneurysm	1 (0.8)	0	1 (1.9)
Hypertension	1 (0.8)	1 (1.7)	0
Temporal arteritis	1 (0.8)	0	0

Supplementary Table 2: All adverse events. *Discrepancy between study population versus modified per-protocol cohort is due to exclusion of participants who were lost to follow-up (n=7) or withdrawn (n=3) before week 24 visit.

Participant	Days from DMARD cessation to start of SAE	Duration of SAE (days)	SAE	Causality	Expected	Severity	Type of SAE / Action taken	Patient withdrawn from study?
1	174		Giant cell arteritis	Unrelated		Mild	Hospitalisation	Yes
2	176	2	Headache	Unrelated		Moderate	Hospitalisation	No
3	92	n/a	Incidental pancreatic body cystic mass	Unrelated	Unexpected	Severe	Other medically significant event – referred for urgent investigation	No
4	99	1	Brief hospital admission for atypical chest pain	Unrelated	Unexpected	Moderate	Hospitalisation	No

Supplementary Table 3: Serious adverse events



Supplementary Figure 1: Kaplan-Meier plot of flare-free survival based on face-to-face visits.
Solid line is the Kaplan-Meier estimate of the survival function, the grey dashed lines are the 95% CI, and vertical black marks indicate censoring. Outcomes defined as per sensitivity analysis, i.e. using last face-to-face visits / last available DAS28-CRP.

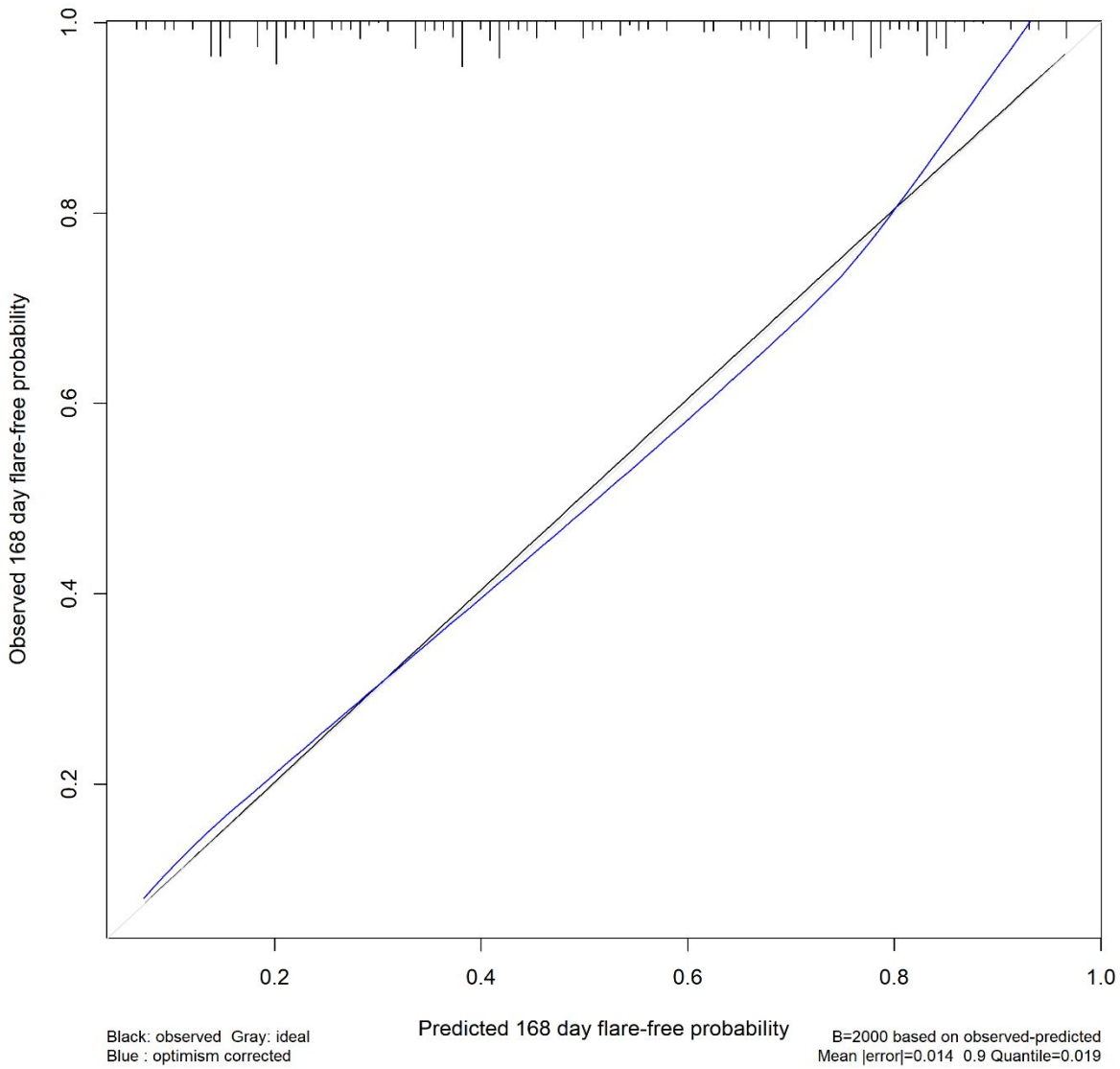
Patient	Description
1	1 tender joint (right wrist), patient VAS = 51/100, CRP = 8 giving a DAS28-CRP of 3.03. Clinical team and patient felt they were flaring so a shared decision was made to restart DMARDs rather than wait for a second ad-hoc appointment in 14 days to confirm flare
2	Ankle (tibialis posterior) tenosynovitis requiring treatment.
3	Bilateral knee synovitis; three swollen joints in total; clinicians and patient felt restarting DMARDs necessary.

Supplementary Table 4: Characteristics of patients adjudged to have flared based on clinician discretion (n=3)

	Imputation										Average BIF
	1	2	3	4	5	6	7	8	9	10	
Age	0.105	0.115	0.105	0.110	0.105	0.110	0.120	0.115	0.130	0.110	0.115
Sex	0.585	0.605	0.590	0.585	0.585	0.595	0.605	0.595	0.555	0.595	0.583
Time from diagnosis to baseline	0.170	0.175	0.155	0.215	0.205	0.240	0.205	0.180	0.200	0.155	0.187
Time from symptom onset to DMARD commencement	0.425	0.345	0.370	0.350	0.335	0.345	0.375	0.360	0.340	0.240	0.348
Methotrexate use	0.775	0.790	0.795	0.780	0.765	0.765	0.770	0.775	0.740	0.750	0.768
RF	0.865	0.885	0.870	0.865	0.890	0.900	0.840	0.925	0.925	0.920	0.888
ACPA	0.730	0.775	0.710	0.835	0.720	0.590	0.785	0.670	0.460	0.560	0.677
DAS28-CRP	0.155	0.150	0.155	0.155	0.150	0.160	0.155	0.150	0.160	0.145	0.153
ACR/EULAR Boolean Remission	0.200	0.185	0.185	0.195	0.175	0.165	0.180	0.185	0.165	0.190	0.182
Education	0.215	0.200	0.225	0.205	0.220	0.250	0.285	0.220	0.230	0.275	0.237
BMI	0.380	0.305	0.445	0.380	0.520	0.485	0.545	0.370	0.535	0.485	0.445
Current smoking	0.190	0.195	0.220	0.205	0.245	0.235	0.250	0.180	0.220	0.225	0.216
Any alcohol use	0.465	0.435	0.455	0.430	0.435	0.465	0.430	0.440	0.445	0.445	0.443
CCI	0.195	0.195	0.200	0.195	0.175	0.200	0.190	0.190	0.185	0.190	0.192
Corticosteroid use	0.260	0.240	0.235	0.240	0.240	0.255	0.245	0.255	0.260	0.255	0.247

Supplementary Table 5: Bootstrap inclusion frequencies

Sex, methotrexate use, RF, and ACPA were brought forward to the prediction model. Although sex did not cross the a priori 60% average BIF threshold, we included it in the prediction model as its average BIF was highly proximal to the threshold. RF=Rheumatoid Factor; ACPA=Anti-Citrullinated Peptide Antibody; BMI=Body Mass Index; CCI=Charlson Comorbidity Index. Bolded predictors are those that were included in the prediction model.



Supplementary Figure 2: Calibration plot at day 168

Dashed lines at the top represent a histogram of the predicted 168-day flare-free probabilities. The risk (or probability) of flare by 168 days may be taken as 1.0 minus the flare-free probability. The light grey diagonal line represents the line of perfect agreement between predicted and observed flare-free probabilities. The blue line indicates the optimism-corrected calibration curve, and the black line indicates the uncorrected calibration curve. The results suggest that the model produced predicted estimates of flare risk that had good agreement with observed risk. As a minor caveat, based on the calibration slopes for risk of flare by 168 days, the model slightly underestimated the predicted risk of flare for participants with an observed "moderate-high" risk (30–80% observed risk), and overestimated predicted risk of flare for participants at lower (≤30% observed risk) and higher observed risk (≥80% observed risk).

Supplementary References

1. Riley RD, Snell KI, Ensor J, Burke DL, Harrell Jr FE, Moons KG, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Statistics in Medicine. 2019;38(7):1276-96.
2. Riley RD. Correction to: Minimum sample size for developing a multivariable prediction model: Part II-binary and time-to-event outcomes by Riley RD, Snell KI, Ensor J, et al. Statistics in Medicine. 2019;38(30):5672-.
3. Baker KF, Skelton AJ, Lendrem DW, Scadeng A, Thompson B, Pratt AG, et al. Predicting drug-free remission in rheumatoid arthritis: A prospective interventional cohort study. Journal of autoimmunity. 2019;105:102298.
4. Bykerk VP, Massarotti EM. The new ACR/EULAR remission criteria: rationale for developing new criteria for remission. Rheumatology. 2012;51(suppl 6):vi16-vi20.
5. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. Journal of chronic diseases. 1987;40(5):373-83.
6. van Buuren SG-O, K. mice: Multivariate Imputation by Chained Equations in R. HJournal of Statistical Software. 2011;45(3):1-67.
7. Marshall A, Altman DG, Holder RL. Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: a resampling study. BMC Medical Research Methodology. 2010;10(1):112.
8. White IR, Royston P. Imputing missing covariate values for the Cox model. Statistics in Medicine. 2009;28(15):1982-98.
9. Heymans MW, Van Buuren S, Knol DL, Van Mechelen W, De Vet HC. Variable selection under multiple imputation using the bootstrap in a prognostic study. BMC Medical Research Methodology. 2007;7(1):33.
10. Sauerbrei W, Buchholz A, Boulesteix A-L, Binder H. On stability issues in deriving multivariable regression models. Biometrical Journal. 2015;57(4):531-55.
11. Sauerbrei W, Schumacher M. A bootstrap resampling procedure for model building: Application to the cox regression model. Statistics in medicine. 1992;11(16):2093-109.
12. Altman DG, Andersen PK. Bootstrap investigation of the stability of a cox regression model. Statistics in medicine. 1989;8(7):771-83.
13. Sauerbrei WM-H, C; Benner, A; Royston, P. Multivariable regression model building by using fractional polynomials: Description of SAS, STATA and R programs. Computational Statistics & Data Analysis. 2006;50:3464 - 85.
14. Riley RD, Snell KIE, Martin GP, Whittle R, Archer L, Sperrin M, et al. Penalization and shrinkage methods produced unreliable clinical prediction models especially when sample size was small. Journal of Clinical Epidemiology. 2021;132:88-96.
15. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating: Springer Cham; 2019. XXXIII, 558 p.
16. Marshall A, Altman DG, Royston P, Holder RL. Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. BMC Medical Research Methodology. 2010;10(1):7.

BMJ Open

Clinical predictors of flare and drug-free remission in rheumatoid arthritis: preliminary results from the prospective BIO-FLARE experimental medicine study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2024-092478.R1
Article Type:	Original research
Date Submitted by the Author:	18-Feb-2025
Complete List of Authors:	Rayner, Fiona; Newcastle University Translational and Clinical Research Institute; Newcastle Upon Tyne Hospitals NHS Trust, Rheumatology Hiu, Shaun; Newcastle University Population Health Sciences Institute, Newcastle University; University of Oxford Nuffield Department of Primary Care Health Sciences Melville, Andrew; University of Glasgow Institute of Infection Immunity and Inflammation Bigirumurame, Theophile; Newcastle University Population Health Sciences Institute Anderson, Amy; Newcastle University Translational and Clinical Research Institute Dyke, Bernard; NIHR Birmingham Biomedical Research Centre Kerrigan, Sean; University of Glasgow Institute of Infection Immunity and Inflammation McGucken, Andrew; University of Glasgow Institute of Infection Immunity and Inflammation Prichard, Jonathan; Newcastle University Population Health Sciences Institute Shahrokhabadi, Mohadeseh Shojaei; Newcastle University Population Health Sciences Institute Hilkens, Catharien M U; Newcastle University Translational and Clinical Research Institute Buckley, Christopher D; NIHR Birmingham Biomedical Research Centre; University of Oxford Kennedy Institute of Rheumatology McInnes, Iain; University of Glasgow Institute of Infection Immunity and Inflammation Ng, Wan-Fai; Newcastle University Translational and Clinical Research Institute; Newcastle Upon Tyne Hospitals NHS Foundation Trust, Musculoskeletal Unit Goodyear, Carl; University of Glasgow Institute of Infection Immunity and Inflammation Teare, Dawn; Newcastle University Population Health Sciences Institute Filer, Andrew; NIHR Birmingham Biomedical Research Centre, Institute of Inflammation and Ageing Siebert, Stefan; University of Glasgow, Institute of Infection, Immunity and Inflammation; NHS Greater Glasgow and Clyde, Raza, Karim; NIHR Birmingham Biomedical Research Centre, School of Immunity and Infection; Sandwell and West Birmingham Hospitals NHS Trust

	Pratt, Arthur; Newcastle University, Translational and Clinical Research Institute; Newcastle Upon Tyne Hospitals NHS Foundation Trust, Musculoskeletal Unit Baker, Kenneth F; Newcastle University Translational and Clinical Research Institute; Newcastle Upon Tyne Hospitals NHS Foundation Trust, Rheumatology Department Isaacs, John; Newcastle University Translational and Clinical Research Institute; Newcastle Upon Tyne Hospitals NHS Foundation Trust, Rheumatology Department
Primary Subject Heading:	Rheumatology
Secondary Subject Heading:	Rheumatology
Keywords:	Rheumatology < INTERNAL MEDICINE, Drug Therapy, IMMUNOLOGY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Clinical predictors of flare and drug-free remission in rheumatoid arthritis: preliminary results from the prospective BIO-FLARE experimental medicine study

Fiona Rayner[1,2], Shaun Hiu[3,4]*, Andrew Melville[5]*, Theophile Bigirumurame[3], Amy Anderson[1], Bernard Dyke[6], Sean Kerrigan[5], Andrew McGucken[5], Jonathan Prichard[3], Mohadeseh Shojaei Shahrokhabadi[3], Catharien M U Hilkens[1], Christopher D Buckley[6,7], Iain McInnes[5], Wan-Fai Ng[1,2], Carl Goodyear[5], M Dawn Teare[3], Andrew Filer[6], Stefan Siebert[5], Karim Raza[6,8], Arthur G Pratt[1,2], Kenneth F Baker[1,2], John D Isaacs[1,2].

*Contributed equally as second authors

Author Information

1. Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, UK
2. Rheumatology Department, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK
3. Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK
4. Nuffield Department of Primary Care Health Sciences, University of Oxford, UK
5. School of Infection and Immunity, University of Glasgow, Glasgow, UK
6. NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and Institute for Inflammation and Ageing, University of Birmingham, Birmingham, UK
7. Kennedy Institute of Rheumatology, University of Oxford, Oxford, UK
8. Department of Rheumatology, Sandwell and West Birmingham NHS Trust, Birmingham, UK

Corresponding author:

Professor John Isaacs

Translational and Clinical Research Institute, 4th Floor, William Leech Building, Framlington Place, Newcastle upon Tyne, NE2 4HH

john.isaacs@ncl.ac.uk

ORCiD: https://orcid.org/0000-0002-6103-7056

Keywords:

Rheumatoid arthritis, Flare, Remission, Drug-free, cessation, withdrawal, disease-modifying anti-rheumatic drug (DMARD)

Abstract

Objectives

Huge advances in rheumatoid arthritis (RA) treatment mean an increasing number of patients now achieve disease remission. However, long term treatments can carry side effects and associated financial costs. In addition, some patients still experience painful and debilitating disease flares, the mechanisms of which are poorly understood. High rates of flare and a lack of effective prediction tools can limit attempts at treatment withdrawal. The BIO-FLARE experimental medicine study was designed to study flare and remission immunobiology. Here we present the clinical outcomes and predictors of drug-free remission and flare, and develop a predication model to estimate flare risk.

Design, setting and participants

BIO-FLARE was a multicentre, prospective, single-arm, open-label experimental medicine study conducted across seven NHS Trusts in the UK. Participants had established RA in clinical remission (DAS28-CRP<2.4) and were receiving methotrexate, sulfasalazine or hydroxychloroquine (monotherapy or combination).

Interventions

The intervention was DMARD cessation, followed by observation for 24-weeks or until flare, with clinical and immune monitoring.

Outcome measures

The primary outcome measure was proportion of participants experiencing a confirmed flare, defined as DAS28-CRP ≥3.2, or DAS28-CRP ≥2.4 twice within two weeks, and time to flare.

Exploratory predictive modelling was also performed using multivariable Cox regression, to understand risk factors for flare.

Results

121 participants were recruited between September 2018 and December 2020. Flare rate by week 24 was 52.3% (95% CI 43.0 to 61.7) with a median (IQR) time to flare of 63 (41-96) days. Female sex, baseline methotrexate use, ACPA level and RF level were associated with flare. An exploratory prediction model incorporating these variables allowed estimation of flare risk, with acceptable classification (C index 0.709) and good calibration performance.

Conclusion

The rate of flare was approximately 50%. Several baseline clinical parameters were associated with flare. The BIO-FLARE study design provides a robust experimental medicine model for studying flare and remission immunobiology.

Strengths and Limitations

- Rheumatoid arthritis flare immunobiology is poorly understood. The BIO-FLARE study represents a robust experimental medicine model for the investigation of flare and remission immunobiology in RA.
- We have used routine baseline clinical parameters to develop an exploratory model for the prediction of flare following immunomodulatory drug cessation.
- Limitations include the open-label approach, which could allow for disease flares caused by the nocebo effect
- A short follow-up time of 6 months means flares after this time were not recorded
- The small sample size of 121 participants may limit generalisabilty, although is comparable with other published literature

1.0 Introduction

Rheumatoid arthritis (RA) is a chronic disease characterised by relapsing-remitting episodes of immune-mediated inflammation known as flares, which pose far-ranging negative consequences for patients [1]. RA flares have been associated with impaired physical function, increased fatigue, and reduced quality of life [2], as well as serious long-term sequelae including incremental joint damage [3] and increased risk of cardiovascular events [4]. Despite their importance, RA flares remain poorly understood at a mechanistic level, and are challenging to investigate scientifically because of their sporadic and unpredictable nature.

Historically most patients with RA suffered from frequent flares, though early diagnosis and rapid initiation of modern regimens of disease-modifying anti-rheumatic drugs (DMARDs) now mean that sustained remission is increasingly an achievable goal in around half of patients. Nevertheless, DMARDs carry risks of drug toxicity, are expensive to prescribe and monitor, and require regular blood testing. International guidelines now advocate consideration of DMARD dose reduction for patients in sustained remission [5], albeit with a risk of arthritis flare in around half of patients who attempt this [6, 7, 8]. DMARD cessation provides an experimental human model, acceptable to patients, by which to study the immunobiology of RA flare. In turn this could identify hitherto elusive biomarkers to guide individualised therapeutic decisions.

The BIOlogical Factors that Limit sustAined Remission in rhEumatoid arthritis (BIO-FLARE) study is an experimental medicine study in which patients with established RA in remission underwent complete DMARD cessation, with the over-arching aim of advancing understanding of the biological factors underpinning RA remission and flare through multi-parameter immune monitoring [9]. In this preliminary report we describe the clinical characteristics and outcomes of the BIO-FLARE cohort, and develop and internally validate an exploratory clinical model to predict risk of flare at the individual patient level. This model, based on clinical predictors alone, provides a baseline which we will subsequently strengthen by the addition of immune biomarkers, informed by our laboratory studies.

2.0 Methods and Materials

BIO-FLARE was a multi-centre, prospective, single-arm, open-label, experimental medicine study of complete DMARD cessation in RA patients who had achieved remission on conventional synthetic DMARDs (csDMARDs: methotrexate, sulfasalazine and hydroxychloroquine; either as monotherapy or in combination) [9]. All participants who fulfilled eligibility criteria stopped all DMARDs at enrolment without tapering. There was no randomisation or control arm, the comparator groups being those who flared versus those who remained in remission. Participants were followed up for 24 weeks, or until confirmed flare, whichever occurred earlier. The primary clinical outcomes were time to flare (in days) following DMARD cessation, and occurrence of flare (binary) during the 24-week study period. We adhered to the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) reporting guidelines [10].

2.1 Recruitment Criteria

Inclusion criteria included the following: 1) RA fulfilling the 1987 ACR or 2010 ACR/EULAR classification criteria, 2) stable dose csDMARDs, with no dose increase in the previous six months, and 3) clinical remission according to disease activity score in 28 joints (DAS28) with C-reactive protein (DAS28-CRP) <2.4 [11]. Exclusion criteria included current use of csDMARDs other than methotrexate, sulfasalazine or hydroxychloroquine; use of leflunomide within the previous 12 months (owing to its extended half-life due to enterohepatic recirculation); use of any biologic or targeted synthetic DMARDs in the previous 6 months; use of glucocorticoids in the previous 3 months (other than inhaled or topical forms); any previous ever use of cell-depleting therapies (e.g. rituximab). Potential participants were identified by their usual rheumatology teams across seven participating National Health Service (NHS) trusts in the United Kingdom (UK), between September 2018 and December 2020.

2.2 Procedures and Definitions

As shown in Figure 1, participants underwent a screening visit to confirm eligibility. Consenting participants stopped all DMARDs immediately once they were deemed eligible, with no dose tapering. An optional baseline ultrasound guided synovial biopsy was performed in consenting participants prior to DMARD cessation (within 14 days of screening visit). Ultrasound findings did not influence study eligibility. Subsequent study visits took place at weeks 2, 5, 8, 12 and 24 following DMARD cessation. Participant-initiated ad-hoc study visits could also be arranged at any time in response to suspected flare. At all study visits, participants underwent clinical assessment, including DAS28-CRP, adverse event (AE) and serious adverse event (SAE) recording, and blood and urine sampling.

Flare was defined as occurrence of any of the following: 1) DAS28-CRP ≥3.2 at any study visit, 2) DAS28-CRP ≥2.4 on two occasions within a 14-day period: if DAS28-CRP was ≥2.4 but <3.2 at any study visit, then another visit was arranged within 2 weeks, with flare confirmed if DAS28-CRP was ≥2.4 at second review, or 3) clinical indication for glucocorticoid rescue therapy and/or DMARD restart despite DAS28-CRP <2.4, e.g. for disease activity not captured by DAS28-CRP such as ankle or foot joint synovitis. Clinician discretion was permitted where DAS28-CRP ≥3.2 was felt to be driven by identifiable non-RA factors, e.g. concurrent infection. In such cases, an ad-hoc visit was arranged within 2 weeks and participants were considered to have remained in remission if subsequent DAS28-CRP was <2.4.

In the event of confirmed flare, an ultrasound guided synovial biopsy was performed within 7 days (if there was a joint deemed suitable for biopsy). Systemic or intra-articular glucocorticoid therapy could be administered immediately after biopsy, where indicated. Participants were then referred back to their usual rheumatology team for re-initiation of DMARDs.

2.3 Baseline Data Collection

Baseline data collected at the screening visit included participant demographics, RA history, current and previous treatments, medical history including significant co-morbidities (Charlson Comorbidity Index), and patient reported outcome measures including functional status (HAQ-DI) (Table 1). A full schedule of events is included in the *supplementary materials Table 1.*

2.4 Statistical analysis

The primary outcome for the current study was time to disease flare (in days). The Kaplan-Meier estimate of the survivor curve was computed along with numbers at risk at the scheduled visit dates (week 2, 5, 8, 12 and 24). Participants who were lost to follow-up or withdrawn from the study were censored at the last available visit.

Sixteen candidate baseline variables were considered for exploratory prediction model inclusion: age, sex, disease duration, time from symptom onset to first DMARD, baseline methotrexate use, glucocorticoids within 3-12 months of baseline visit, baseline rheumatoid factor (RF) level, baseline anti-citrullinated peptide antibody (ACPA) level, DAS28-CRP, ACR/EULAR Boolean remission status [12], education level, employment status, body mass index, smoking status, alcohol intake, and Charlson comorbidity index. These were chosen based on prior knowledge and before reviewing study data. Owing to the presence of some missing data points, analyses were performed with 10 imputed datasets using multiple imputation by chained equations (MICE) [13].

To provide clinical context, univariate analyses were performed to assess the strength of association between each candidate variable and time to flare, with hazard ratios (HR) and 95% confidence intervals (CI) determined.

A predictive model for flare containing baseline clinical variables was built using a Cox proportional hazards (PH) model following a sequential process of variable selection, estimation of shrinkage, and internal validation, described in detail in *supplementary materials*. Predictive performance was internally validated using bootstrapping and evaluated with optimism-corrected indices of

discrimination (*C* index) and calibration [14, 15, 16]. We report our predictive model as an equation for calculating the prognostic index (PI), representing an individual's "propensity" to flare, and a baseline survival function, which together allow calculation of estimated risk of flare by a given time following DMARD cessation.

2.5 COVID-19 mitigation and sensitivity analysis

The latter stages of the BIO-FLARE study overlapped with the onset of the COVID-19 pandemic, meaning some follow-up visits were disrupted. A mitigation strategy was adopted whereby affected participants received telephone consultations when their study visits were due, with assessments of flare/remission status based on participant-reported symptoms rather than DAS28-CRP, and face-to-face visits reserved for those with suspected flare. Seven participants were lost to follow-up during this period, while 4 had telephone consultations up to week 24. For our primary analyses, participants with telephone consultations up to week 24, and no symptoms of flare, were classified as having remained in remission. A sensitivity analysis of our predictive modelling process was conducted using last face-to-face study visits only (i.e. last available DAS28-CRP).

2.6 Study subpopulations

Overall baseline characteristics and adverse event data are described for all participants who stopped DMARDs (n=121, the total study population). Time-to-event analyses, including predictive modelling, were performed for participants with ≥1 follow-up visit (n=120, the analysis population which excludes 1 participant who withdrew soon after baseline because of the COVID-19 pandemic). Flare rate was calculated for participants with confirmed flare or remission status (n=111, which we term the 'modified per-protocol' population following the COVID-19 mitigation strategy), i.e. excluding 10 participants who did not experience flare but withdrew (n=3) or were lost to follow-up (n=7) before week 24.

2.7 Patient and Public Involvement

The Newcastle Patient and public Involvement in Musculoskeletal reSearch (PIMS) group were consulted at the planning stage of the project. The importance of the research topic and design of the study protocol was informed by their views and discussions. Clinical results from the study have been presented at national Versus Arthritis meetings with patient partners present. We will present more results of the study to local, regional and national PPIE groups as they become available.

3.0 Results

3.1 Baseline Characteristics and Adverse Events

One hundred and twenty-one participants met the inclusion criteria including DAS28-CRP <2.4 and stopped DMARD therapy (Figure 2). The overall baseline characteristics are presented in Table 1, along with the baseline characteristics for participants who flared and those who remained in remission at 24 weeks (n=111, the modified per-protocol population). For the total study population, mean (SD) age was 64.1 (11.9) years, 60.3% were female, and median (IQR) disease duration was 6.3 (4.5–12.3) years. 67/119 (56.3%) were RF positive and 76/114 (66.7%) were ACPA positive, with 64/113 (56.6%) double positive. Only 1 participant had previous biologic therapy (etanercept, stopped 7.5 years before study entry). 101/121 participants (83.5%) were treated with methotrexate at baseline (monotherapy or combination use) with a median (IQR) dose of 15 (12.5-20) mg weekly. Of 20 participants not on methotrexate at baseline, 7/20 had previously received methotrexate treatment. Mean (SD) baseline DAS28-CRP was 1.61 (0.32); 61.2%, 78.5% and 84.9% fulfilled ACR/EULAR Boolean remission criteria, Boolean 2.0 remission criteria [17], and simplified disease activity index (SDAI) remission criteria at baseline, respectively. There were 155 AEs (*Supplementary materials Table 2*), 4 SAEs relating to hospitalisations (*Supplementary materials Table 3*), and no deaths. The four SAEs were all considered to be unrelated to study participation or procedures.

		Total study population (n=121)	Modified per-protocol population (n=111)*		Missing data (n=121), N %
			Flare (n=58)	Remission up to week 24 (n=53)	
Age, years	Mean (SD)	64.07 (11.9)	64.76 (11.6)	64.68 (11.3)	0 (0.0)
Sex, Female	N (%)	73 (60.3)	41 (70.7)	27 (50.9)	0 (0.0)
Body mass index, kg/m^2	Mean (SD)	28.20 (5.7)	27.24 (5.3)	29.36 (5.9)	5 (4.1)
Charlson comorbidity index	Median (IQR)	2.00 (1.0, 3.0)	2.00 (1.0, 4.0)	3.00 (1.0, 3.0)	0 (0.0)
Tobacco smoking status	N (%)				0 (0.0)
Current		8 (6.6)	5 (8.6)	3 (5.7)	
Ex-smoker		64 (52.9)	32 (55.2)	26 (49.1)	
Never smoked		49 (40.5)	21 (36.2)	24 (45.3)	
Current alcohol use	N (%)	73 (60.8)	32 (55.2)	36 (67.9)	1 (0.8)
Ethnicity	N (%)				0 (0.0)
White British/Other White		113 (93.4)	55 (94.8)	51 (96.2)	
Asian/Asian British		6 (5.0)	3 (5.2)	1 (1.9)	
Black/Black British - Caribbean		2 (1.7)	0 (0.0)	1 (1.9)	
Highest educational qualification	N (%)				2 (1.7)
GCSEs or equivalent		33 (27.3)	12 (21.0)	18 (34.0)	
A-Level or equivalent		13 (10.7)	7 (12.1)	6 (11.3)	
Undergraduate		20 (16.5)	12 (20.7)	7 (13.2)	
Postgraduate		14 (11.6)	6 (10.3)	5 (9.4)	
NVQ or equivalent		14 (11.6)	5 (8.6)	6 (11.3)	
None of the above		24 (19.8)	15 (25.9)	9 (17.0)	
Not stated or missing		3 (2.5)	1 (1.7)	2 (3.8)	
Employment status	N (%)				0 (0.0)
Full-time		30 (24.8)	13 (22.4)	15 (28.3)	
Part-time		13 (10.7)	7 (12.1)	4 (7.6)	
Unemployed		3 (2.5)	0 (0.0)	0 (0.0)	
Self employed		2 (1.7)	1 (1.7)	1 (1.9)	
Retired		71 (58.7)	37 (63.8)	31 (58.5)	
Other		2 (1.7)	0 (0.0)	2 (3.8)	
Time from symptom onset to first DMARD, years	Median (IQR)	0.51 (0.3, 1.3)	0.54 (0.3, 2.1)	0.51 (0.3, 1.0)	4 (3.3)
Time from symptom onset to baseline, years	Median (IQR)	6.33 (4.5, 12.4)	6.34 (5.0, 13.7)	6.17 (3.9, 10.8)	4 (3.3)
Time from RA diagnosis to baseline, years	Median (IQR)	5.48 (3.7, 10.5)	5.48 (4.2, 10.7)	5.36 (3.3, 9.7)	2 (1.7)
MTX use at baseline	N (%)	101 (83.5)	53 (91.4)	39 (73.6)	0 (0.0)
MTX dose, mg/week	Median (IQR)	15 (12.5, 20)	15 (12.5, 20)	15 (12.5, 20)	0 (0.0)
MTX monotherapy	N (%)	72 (59.5)	37 (63.8)	28 (52.8)	0 (0.0)
MTZ + SZN	N (%)	5 (4.1)	3 (5.2)	2 (3.8)	0 (0.0)
MTX + HCQ	N (%)	22 (18.2)	11 (19.0)	9 (17.0)	0 (0.0)
MTX + SZN + HCQ	N (%)	2 (1.7)	2 (3.5)	0 (0.0)	0 (0.0)
SZN monotherapy	N (%)	10 (8.3)	3 (5.2)	7 (13.2)	0 (0.0)

		Total study population (n=121)	Modified per-protocol population (n=111)*		Missing data (n=121), N %
			Flare (n=58)	Remission up to week 24 (n=53)	
HCQ monotherapy	N (%)	8 (6.6)	2 (3.5)	5 (9.4)	0 (0.0)
SZN + HCQ	N (%)	2 (1.7)	0 (0.0)	2 (3.8)	0 (0.0)
Previous biologic therapy	N (%)	1 (0.8)	1 (1.7)	0 (0.0)	0 (0.0)
Corticosteroid use in past 12 months	N (%)	7 (5.8)	3 (5.2)	4 (7.6)	
Any		7 (5.8)	3 (5.2)	4 (7.6)	0 (0.0)
Oral		3 (2.5)	1 (1.7)	2 (3.8)	0 (0.0)
Intramuscular		1 (0.8)	0 (0.0)	1 (1.9)	0 (0.0)
Intra-articular		2 (1.7)	2 (3.5)	1 (1.9)	0 (0.0)
RF positive	N (%)	67 (56.3)	42 (72.4)	19 (36.5)	2 (1.7)
ACPA positive	N (%)	76 (66.7)	45 (77.6)	25 (49.0)	7 (5.8)
RF, IU/ml	Median (IQR)	32.00 (0.0, 94.1)	53.15 (14.0, 130.0)	12.65 (0.0, 40.1)	2 (1.7)
ACPA, U/ml	Median (IQR)	96.50 (1.1, 300.0)	207.00 (31.0, 306.5)	1.70 (0.8, 196.0)	7 (5.8)
DAS28-CRP	Mean (SD)	1.61 (0.3)	1.62 (0.3)	1.60 (0.3)	0 (0.0)
ACR/EULAR 2011 Boolean remission	N (%)	74 (61.2)	38 (65.5)	31 (58.5)	0 (0.0)
ACR/EULAR Boolean 2.0 remission	N (%)	95 (78.5)	48 (82.8)	41 (77.4)	0 (0.0)
SDAI remission	N (%)	101 (84.9)	49 (86.0)	45 (84.9)	2 (1.7)
HAQ-DI	Median (IQR)	0.00 (0.0, 0.6)	0.13 (0.0, 0.7)	0.00 (0.0, 0.4)	0 (0.0)
Follow-up time, days	Median (IQR)	115.5 (55.5, 167.5)	63.0 (41.0, 96.0)	168.0 (167.0, 174.0)	0 (0.0)

Table 1: Baseline characteristics. *The modified per-protocol population includes all participants with known outcome status and excludes those lost to follow-up (n=7) or withdrawn (n=3) before week 24 visit. ACR=American College of Rheumatology, ACPA=Anti-Citrullinated Peptide Antibody, DMARD=Disease Modifying Anti-Rheumatic Drug, EULAR=European Alliance of Associations for Rheumatology, HAQ-DI=Health Assessment Questionnaire Disability Index, HCQ=Hydroxychloroquine, IQR=Interquartile Range, MTX=Methotrexate, N=Number, NVQ=National Vocational Qualification, RF=Rheumatoid Factor, SD=Standard Deviation, SDAI=Simplified Disease Activity Index, SZN=Sulfasalazine.

3.2 Flare Characteristics

The flare rate at 24 weeks (168 days) was 52.3% (58/111, 95% CI 43.0 to 61.7). Flare-free probability is presented in a Kaplan-Meier plot in Figure 3 (Kaplan-Meir plot of flare-free probability based on face-to-face visits is available in *supplemental Figure 2*). For the 58 participants who experienced flare, median time to flare was 63 days (IQR 41–96 days, range 13–155 days).

Mean (SD) DAS28-CRP at time of flare was 3.81 (0.78). DAS28-CRP components at time of flare were as follows: median (IQR) tender joint count (TJC) 4 (1–5), swollen joint count (SJC) 2 (1–3), CRP 8.0 (4.5–14.4) mg/dL; mean (SD) patient global health VAS 48.4/100 (22.7).

Confirmation of flare was based on a single DAS28-CRP result ≥3.2 in 39/58 cases (of which 18 were scheduled study visits and 21 were ad-hoc visits), two DAS28-CRP results ≥2.4 within a 14-day period in 16/58 cases (of which 8 had DAS28-CRP ≥3.2 at the 2nd visit), and clinician discretion in 3/58 cases (described in *Supplementary materials Table 4*).

3.3 Univariate Cox proportional hazards models

Of the 16 variables considered in univariate analyses: female sex, methotrexate use at baseline, RF level, ACPA level, and longer time from symptom onset to first DMARD were statistically significantly associated with time to flare (Table 2).

		Complete case analysis			Multiple imputation with chained equations (MICE) (n=120)	
		Available n	HR [95% CI]	P-value	HR [95% CI]	P-value
	Reference					
Age		120	1.00 [0.98 to 1.02]	0.99		
Male sex	Female	120	0.54 [0.31 to 0.95]	0.03		
Years from diagnosis to baseline visit		118	1.01 [0.99 to 1.04]	0.37	1.01 [0.98 to 1.04]	0.47
Years from symptom onset to first DMARD		112	1.04 [1.00 to 1.07]	0.04	1.03 [1.00 to 1.07]	0.06
Methotrexate use at baseline	No	120	2.92 [1.16 to 7.31]	0.02		
RF level at baseline, per 10 IU/ml		118	1.03 [1.01 to 1.06]	0.001	1.03 [1.01 to 1.06]	0.001
ACPA level at baseline, per 10 U/ml		113	1.03 [1.01 to 1.04]	0.01	1.03 [1.01 to 1.04]	0.01
DAS28-CRP		120	1.14 [0.51 to 2.52]	0.75		
ACR/EULAR 2011 Boolean remission at baseline	Not in remission	120	1.16 [0.68 to 2.00]	0.59		
A-level and above education	GCSE and under	117	1.15 [0.68 to 1.94]	0.59	1.14 [0.67 to 1.94]	0.62
BMI		115	0.95 [0.90 to 1.01]	0.08	0.96 [0.90 to 1.01]	0.11
Current smoker	Never or ex-smoker	120	1.04 [0.41 to 2.60]	0.94		
Current alcohol use	No	119	0.63 [0.37 to 1.05]	0.07	0.63 [0.38 to 1.06]	0.09
Charlson comorbidity index		120	1.04 [0.88 to 1.22]	0.66		
Glucocorticoid use in past 12 months from baseline	No	120	0.78 [0.24 to 2.48]	0.67		

Table 2. Univariate analysis of candidate baseline variables predicting flare in analysis cohort. Employment variable was not included in the imputation model due to convergence issues due to low frequency in the unemployed subgroup. Variables with no missing data have empty rows under the MICE column because estimates will be identical to the complete case analysis. Hazard ratios for continuous variables are calculated per 1 unit increase unless otherwise stated. ACPA=Anti-citrullinated Peptide Antibody, BMI=Body Mass Index, RF=Rheumatoid Factor, GCSE=General Certificate of Secondary Education, HR=Hazard Ratio, NA=Not Applicable

3.4 Exploratory prediction model

Our prediction modelling procedure, including variable selection, resulted in the inclusion of sex, methotrexate use at baseline, RF level, and ACPA level into the prediction model (*see Supplementary*

materials Table 5). A square root transformation of RF, and two non-linear expressions of ACPA (inverse of ACPA and inverse square root of ACPA) were chosen as the best-fitting transformations. Thus, our prediction model consisted of five terms: sex, methotrexate use, $(RF + 0.1)^{0.5}$, $(ACPA + 0.1)^{-1}$, and $(ACPA + 0.1)^{-0.5}$.

The predicted probability of flare within *t* days after DMARD cessation can be computed as:

$$Predicted\ risk\ of\ flare\ by\ t\ days\ after\ DMARD\ cessation = 1 - \hat{S}_0(t)\text{^}\exp(PI)$$

Where $\hat{S}_0(t)$ is the estimated baseline survival function at time *t*, PI is the prognostic index, and *exp*(.) is the exponential function. The value of $\hat{S}_0(t)$ at t=168 days after DMARD cessation is 0.672. Additional values at 30, 60, 90 and 120 days are available in the *supplementary materials (statistical analysis section)*. The PI is computed as:

$$PI = (-0.55814869 \times Sex) + (1.05775338 \times Methotrexate\ use) + (0.03734463 \times \sqrt{RF + 0.1}) + f(ACPA)$$

Where $f(ACPA) = \left(\frac{0.55920681}{ACPA+0.1}\right) - \left(\frac{1.86737912}{\sqrt{ACPA+0.1}}\right)$, sex coded as female=0, male=1, methotrexate use coded as no=0, yes=1.

Thus, as an example, for a female patient, who was not taking methotrexate at baseline, has a RF measurement of 60 IU/ml, and an ACPA measurement of 150 IU/ml, the PI would be 0.141, and the predicted risk of flare by 168 days after DMARD cessation would be 36.7%.

The model had an optimism-corrected C index of 0.709 [95% CI 0.647 to 0.771] and calibration slope of 1.00 [95% CI 0.495 to 1.506], indicating acceptable classification performance and good agreement between estimates of flare risk and observed risk (see Statistical analysis section of *Supplementary materials and Supplementary Figure 1).* The sensitivity analysis of the prediction model, using last face-to-face study visits, demonstrated comparable predictive properties (C-index 0.707 [95% CI 0.643 to 0.771], calibration slope 0.996 [95% CI 0.494 to 1.497]).

4.0 Discussion

BIO-FLARE is an experimental medicine study designed to provide insights into the biological processes that trigger episodes of flare in patients with RA. The ability to compare patients who remain in remission upon DMARD cessation with those who flare provides a well-controlled biological model. In this current work we describe the clinical characteristics of the BIO-FLARE cohort, report the main clinical outcomes, and explore predictors of flare among routine baseline clinical parameters.

Approximately 50% of participants experienced a flare over the six-month study period, which is similar to the results of previous csDMARD withdrawal studies in RA [6, 18, 19, 20].

Among baseline parameters we identified methotrexate use, female sex, RF level, and ACPA level as significant predictors of flare following DMARD cessation. Higher RF and ACPA levels have been associated with adverse outcomes in RA, including radiographic progression, and may indicate a more aggressive disease phenotype [21], and seropositivity is associated with progression from pre-clinical to clinically apparent RA [22, 23], which might be analogous mechanistically to flare. Similarly, female sex has been associated with progression to RA from early undifferentiated arthritis [24].The increased risk of flare following methotrexate cessation might reflect more severe underlying disease, confounding by indication for other reasons (i.e. reasons for avoiding or previous discontinuation of methotrexate might be protective), and/or a particular pharmacodynamic mechanism of action that leads to highly effective suppression of disease activity but not true biological remission. Longer time from symptom onset to DMARD initiation had a borderline association with flare and was not selected for inclusion in the final model, but does hint at early and effective treatment modifying the probability of achieving drug-free remission, in line with the "window of opportunity" concept [25, 26].

The association between female sex, RF and ACPA positivity and flare has been noted in previous DMARD withdrawal studies and lends face validity to our results [18, 27]. In the BioRRA study, a precursor to BIO-FLARE, RF positivity and longer time from diagnosis to first DMARD were also associated with flare, while a borderline association was seen for baseline methotrexate use [6].

Unlike previous DMARD withdrawal studies, we adopted a predictive modelling approach towards our baseline clinical parameters and developed and internally validated an exploratory prediction model that allows estimation of risk of flare for an individual patient by a given time following csDMARD cessation. To our knowledge, no comparable clinical model has been described previously in this context. Our final model had acceptable performance in classifying flare versus remission, with good agreement overall between observed and predicted risks. An easy-to-use online version of the formula can be found at https://research.ncl.ac.uk/bioflare/outputs/. Using this tool, sex, methotrexate use, baseline RF and ACPA values can be entered and a predicted risk of flare at 90 or 168 days obtained. Given the lack of external validation, we do not recommend that this tool is used to guide clinical decisions. Nevertheless, in producing a predictive model using only routinely collected data, we present a benchmark against which future molecular or multimodal models can be compared.

Strengths of our study include the number of participants, which compares favourably with previous DMARD withdrawal studies, the prospective study design, and the minimal missing data among baseline parameters. Our predictive modelling followed a robust statistical approach, thereby reducing risk of bias from sensitivity to sampling variability through bootstrapping and overfitting through shrinkage. Nevertheless, our study does have some limitations. BIO-FLARE included participants on csDMARDs only, meaning the relevance of our findings to patients treated with biologic or targeted synthetic DMARDs is uncertain. However, recent studies suggest that up to 40 – 50% of real-world RA patients are treated with csDMARDs alone [28, 29], and it is possible that the immunobiological mechanisms underlying flare may be intrinsic to RA disease processes and thus

independent of DMARD treatment. The DAS28-CRP score, used in our study to define remission and flare, has been criticised in the past for being overly permissive of active inflammation [30]. However, we used a stringent cut-off of < 2.4, and found similar percentages of participants achieving Boolean and SDAI remission at baseline between subsequent flare and remission groups, suggesting flare was not simply driven by discrepancies in uncaptured initial disease activity. The open-label treatment withdrawal creates a risk of flares driven by the nocebo effect, but this was a pragmatic study design that reflects clinical practice. Musculoskeletal imaging was not performed at baseline, meaning the predictive potential of radiographic erosions, or ultrasonographic synovitis/tenosynovitis could not be assessed. The COVID-19 pandemic limited face-to-face assessments for a relatively small proportion of study participants, but the close similarity between our primary and sensitivity analyses suggests our mitigation strategy was valid, without an obvious impact on the performance of the prediction model. Finally, the six-month follow-up period means that longer term outcomes, such as occurrence of flares beyond 24 weeks, response to csDMARD re-initiation, and long term sequelae that might be associated with flares, were not captured by the current study. There was also no dedicated long-term follow up to identify any participants who did not quickly regain remission following flare in the study. This, along with longer term follow up of those that exited the study in remission, would be interesting further work to explore. Other published work has shown that remission is quickly regained in the majority of participants who experience mild disease flares when tapering or stopping DMARDs [31].

In conclusion, approximately half of RA patients in remission on csDMARDs experienced a flare within 6 months of stopping therapy, with a median time-to-flare of 9 weeks. Among baseline clinical parameters, RF and ACPA levels, female sex and methotrexate use were found to be predictive of flare. Our predictive model allows estimation of risk of flare at the individual level based on clinical parameters alone. We will subsequently strengthen this by the addition of immune biomarkers emerging from our BIO-FLARE laboratory analyses.

Contributors

JDI, AA, CMUH, CDB, IM, W-FN, CG, AF, SS, KR, AGP and KFB were involved in study conception, funding acquisition, study design and supervision. FR, BD, SK and AMc performed data collection and FR, BD and SK were involved in study design. FR, SH and AM directly accessed the full dataset and performed data analysis, interpretation and visualisation as well as performing a literature search, drafting the original text and reviewing and editing the text for final submission. SH, TB, MSS and MDT were involved in data analysis and validation. JP was responsible for data curation and software. All authors reviewed the final manuscript. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

The lead author (the manuscript's guarantor) JDI affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Funding

The study was funded by a Medical Research Council experimental medicine grant (reference number MR/N026977/1). The funder had no role in study design, data collection, data analysis, data interpretation or report writing.

Declarations of interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/disclosure-of-interest/ and declare: JDI discloses research grants from Pfizer, Janssen and GSK; conference support from Eli Lilly; and has consulted for: Abbvie, Anaptys Bio, AstraZeneca, BMS, Eli Lilly, Galapagos NV, Gilead Sciences Ltd, GSK, Istesso Ltd, Kira Biotech, Ono Pharma, Pfizer, Revelo Biotherapeutics, Roche, Sanofi. SS reports institutional research grants from Amgen (previously Celgene), Boehringer Ingelheim, Bristol Myers Squibb, Eli Lilly, GSK, Janssen and UCB; speaker/consulting fees from AbbVie, Amgen, AstraZeneca, Eli Lilly, GSK, Janssen, UCB. IM declares honoraria or research support from Abbvie, Janssen, Novartis, Eli Lilly, Astra Zeneca, GSK, BMS, Moonlake, Evelo, Causeway THerapeutics, Cabaletta, Roche, Pfizer and Compugen. CMUH declares research funding from GSK. KFB declares research support from Genentech, clinical improvement funding from Pfizer, and consulting fees from Modern Biosciences. AGP, KFB and JDI are named as inventors on a patent application by Newcastle University ("Prediction of Drug-Free Remission in Rheumatoid Arthritis"; International Patent Application Number PCT/GB2019/050902). KR has received research grant support from Bristol Myers Squibb and personal fees for lectures / consultancy from Abbvie and Sanofi. AF reports institutional research grants from BMS, Roche, UCB, Nascient, Mestag, GSK, Janssen and speaker/consulting fees from Roche, Janssen, Sonoma. All other authors have no completing interest to declare.

Ethics statements

Patient consent for publication

Not applicable

Ethics approval

The study was reviewed and approved by the North East – Newcastle and North Tyneside 1 Research Ethics Committee (REC) (reference 17/NE/0386). Written informed consent was provided by all participants before study enrolment.

Data availability statement

Data are available from the corresponding author upon reasonable request. All data relevant to the study are included in the article or uploaded as supplementary information.

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

Acknowledgements

This work was funded by the Medical Research Council and sponsored by the Newcastle upon Tyne NHS Foundation Trust. It was supported by the National Institute for Health and Care Research (NIHR) Newcastle and Birmingham Biomedical Research Centres and NIHR Newcastle and Birmingham Clinical Research Facilities, the Research into Inflammatory Arthritis Centre Versus Arthritis, and Rheuma Tolerance for Cure (European Union Innovative Medicines Initiative 2). JDI was supported by a NIHR Senior Investigator Award. KFB has received funding from a National Institute for Health and Care Research (NIHR) Clinical Lectureship (CL-2017-01-004 to KFB), an NIHR Advanced Fellowship (NIHR303620 to KFB), a Newcastle Health Innovation Partners Senior Clinical Fellowship, and the Newcastle Hospitals Charity (8033). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

References

1. McInnes IB, Schett G. The Pathogenesis of Rheumatoid Arthritis. New England Journal of Medicine. 2011;365(23):2205-19.
2. van Mulligen E, Weel A, Kuijper TM, Hazes JMW, van der Helm-van Mil AHM, de Jong PHP. The impact of a disease flare during tapering of DMARDs on the lives of rheumatoid arthritis patients. Seminars in arthritis and rheumatism. 2020;50(3):423-31.
3. Smolen JS, Pedersen R, Jones H, Mahgoub E, Marshall L. Impact of flare on radiographic progression after etanercept continuation, tapering or withdrawal in patients with rheumatoid arthritis. Rheumatology (Oxford). 2020;59(1):153-64.
4. Myasoedova E, Chandran A, Ilhan B, Major BT, Michet CJ, Matteson EL, et al. The role of rheumatoid arthritis (RA) flare and cumulative burden of RA severity in the risk of cardiovascular disease. Ann Rheum Dis. 2016;75(3):560-5.
5. Smolen JS, Landewé RBM, Bergstra SA, Kerschbaumer A, Sepriano A, Aletaha D, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2022 update. Annals of the rheumatic diseases. 2023;82(1):3-18.
6. Baker KF, Skelton AJ, Lendrem DW, Scadeng A, Thompson B, Pratt AG, et al. Predicting drug-free remission in rheumatoid arthritis: A prospective interventional cohort study. J Autoimmun. 2019;105:102298.
7. Tanaka Y, Hirata S, Kubo S, Fukuyo S, Hanami K, Sawamukai N, et al. Discontinuation of adalimumab after achieving remission in patients with established rheumatoid arthritis: 1-year outcome of the HONOR study. Ann Rheum Dis. 2015;74(2):389-95.
8. Henaux S, Ruyssen-Witrand A, Cantagrel A, Barnetche T, Fautrel B, Filippi N, et al. Risk of losing remission, low disease activity or radiographic progression in case of bDMARD discontinuation

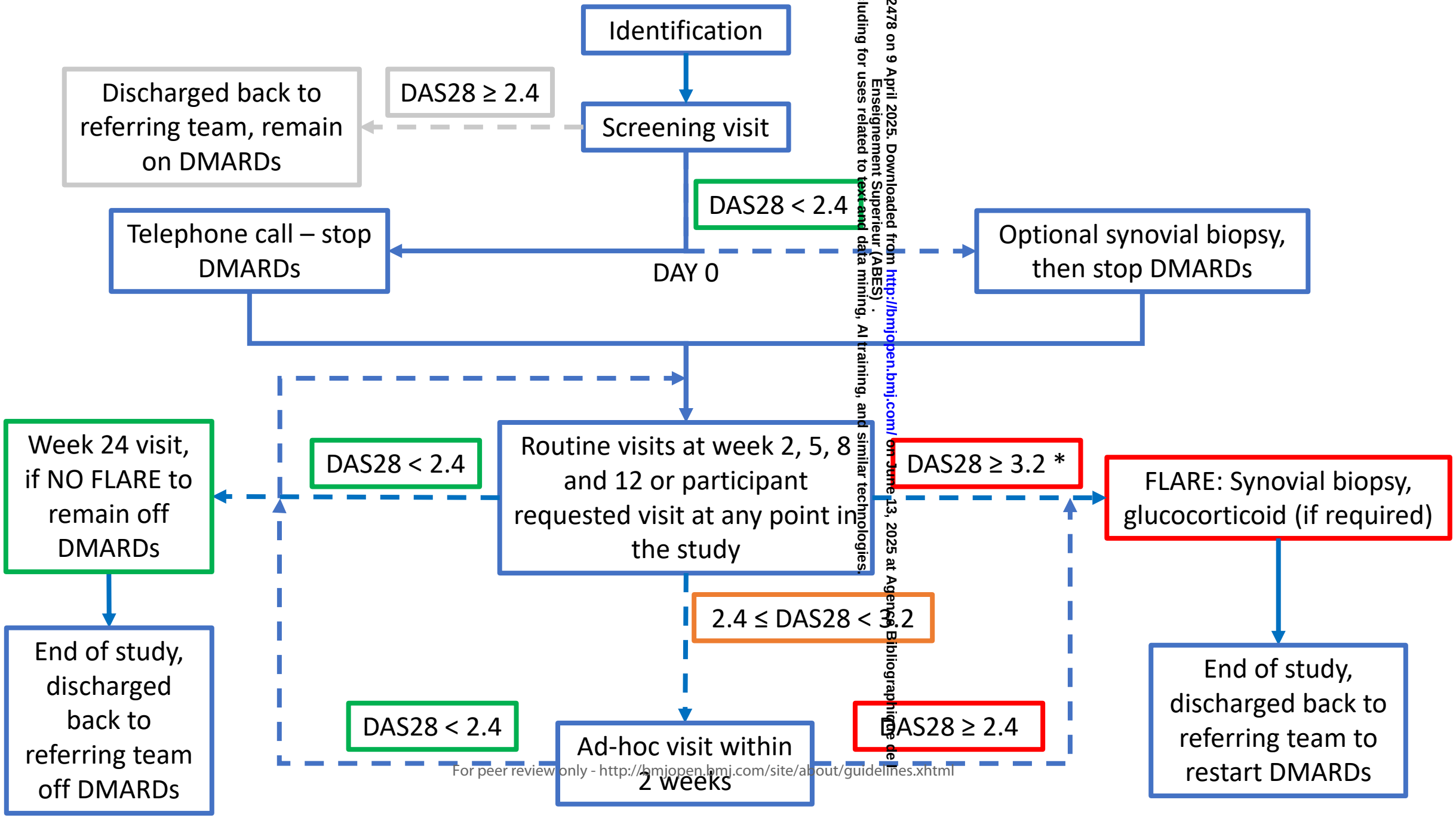
or tapering in rheumatoid arthritis: systematic analysis of the literature and meta-analysis. Annals of the rheumatic diseases. 2018;77(4):515-22.
9. Rayner F, Anderson AE, Baker KF, Buckley CD, Dyke B, Fenton S, et al. BIOlogical Factors that Limit sustAined Remission in rhEumatoid arthritis (the BIO-FLARE study): protocol for a non-randomised longitudinal cohort study. BMC rheumatology. 2021;5(1).
10. Collins GS, Reitsma KB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Medicine. 2015;13(1).
11. Fleischmann R, van der Heijde D, Koenig AS, Pedersen R, Szumski A, Marshall L, et al. How much does Disease Activity Score in 28 joints ESR and CRP calculations underestimate disease activity compared with the Simplified Disease Activity Index? Ann Rheum Dis. 2015;74(6):1132-7.
12. Bykerk VP, Massarotti EM. The new ACR/EULAR remission criteria: rationale for developing new criteria for remission. Rheumatology. 2012;51(suppl 6):vi16-vi20
13. van Buuren SG-O, K. mice: Multivariate Imputation by Chained Equations in R. HJournal of Statistical Software. 2011;45(3):1-67
14. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in medicine. 1996;15(4):361-87.
15. Stevens RJ, Poppe KK. Validation of clinical prediction models: what does the "calibration slope" really measure? Journal of Clinical Epidemiology. 2020;118:93-9.
16. Crowson CS, Atkinson EJ, Therneau TM. Assessing calibration of prognostic risk scores. Statistical Methods in Medical Research. 2016;25(4):1692-706.
17. Studenic P, Aletaha D, De Wit M, Stamm TA, Alasti F, Lacaille D, et al. American College of Rheumatology/EULAR Remission Criteria for Rheumatoid Arthritis: 2022 Revision. Arthritis & Rheumatology. 2023;75(1):15-22.
18. Verstappen M, Van Mulligen E, De Jong PHP, Van Der Helm-Van Mil AHM. DMARD-free remission as novel treatment target in rheumatoid arthritis: A systematic literature review of achievability and sustainability. RMD open. 2020;6(1):e001220.
19. Lee JS, Oh JS, Hong S, Kim YG, Lee CK, Yoo B. Six-month flare risk after discontinuing long-term methotrexate treatment in patients having rheumatoid arthritis with low disease activity. International journal of rheumatic diseases. 2020;23(8):1076-81.
20. Kuijper TM, Lamers-Karnebeek FBG, Jacobs JWG, Hazes JMW, Luime JJ. Flare Rate in Patients with Rheumatoid Arthritis in Low Disease Activity or Remission When Tapering or Stopping Synthetic or Biologic DMARD: A Systematic Review. The Journal of rheumatology. 2015;42(11):2012-22.
21. Syversen SW, Gaarder PI, Goll GL, Ødegård S, Haavardsholm EA, Mowinckel P, et al. High anti-cyclic citrullinated peptide levels and an algorithm of four variables predict radiographic progression in patients with rheumatoid arthritis: results from a 10-year longitudinal study. Annals of the rheumatic diseases. 2008;67(2):212-7.
22. van Steenbergen HW, Mangnus L, Reijnierse M, Huizinga TWJ, van der Helm-van Mil AHM. Clinical factors, anticitrullinated peptide antibodies and MRI-detected subclinical inflammation in relation to progression from clinically suspect arthralgia to arthritis. Annals of the rheumatic diseases. 2016;75(10):1824-30.
23. ten Brinck RM, van Steenbergen HW, van Delft MAM, Verheul MK, Toes REM, Trouw LA, et al. The risk of individual autoantibodies, autoantibody combinations and levels for arthritis development in clinically suspect arthralgia. Rheumatology. 2017;56(12):2145-53.
24. Van Der Helm-Vanmil AHM, Le Cessie S, Van Dongen H, Breedveld FC, Toes REM, Huizinga TWJ. A prediction rule for disease outcome in patients with Recent-onset undifferentiated arthritis: How to guide individual treatment decisions. Arthritis & Rheumatism. 2007;56(2):433-40.
25. Van Nies JAB, Tsonaka R, Gaujoux-Viala C, Fautrel B, Van Der Helm-Van Mil AHM. Evaluating relationships between symptom duration and persistence of rheumatoid arthritis: does a window of

opportunity exist? Results on the Leiden Early Arthritis Clinic and ESPOIR cohorts. Annals of the rheumatic diseases. 2015;74(5):806-12.
26. van der Linden MPM, le Cessie S, Raza K, van der Woude D, Knevel R, Huizinga TWJ, et al. Long-term impact of delay in assessment of patients with early arthritis. Arthritis & rheumatism. 2010;62(12):3537-46.
27. Vittecoq O, Desouches S, Kozyreff M, Nicolau J, Pouplin S, Rottenberg P, et al. Relapse in rheumatoid arthritis patients undergoing dose reduction and withdrawal of biologics: are predictable factors more relevant than predictive parameters? An observational prospective real-life study. BMJ open. 2019;9(12):e031467.
28. Yates M, Ledingham JM, Hatcher PA, Adas M, Hewitt S, Bartlett-Pestell S, et al. Disease activity and its predictors in early inflammatory arthritis: findings from a national cohort. Rheumatology. 2021;60(10):4811-20.
29. Mease PJ, Stryker S, Liu M, Salim B, Rebello S, Gharaibeh M, et al. Treatment patterns in rheumatoid arthritis patients newly initiated on biologic and conventional synthetic disease-modifying antirheumatic drug therapy and enrolled in a North American clinical registry. Arthritis research & therapy. 2021;23(1).
30. Aletaha D, Smolen JS. Joint damage in rheumatoid arthritis progresses in remission according to the Disease Activity Score in 28 joints and is driven by residual swollen joints. Arthritis Rheum. 2011 Dec;63(12):3702-11
31. Tascilar K, Hagen M, Kleyer A, Simon D, Reiser M, Hueber A et al. Treatment tapering and stopping in patients with rheumatoid arthritis in stable remission (RETRO): a multicentre, randomised, controlled, open-label, phase 3 trial. The Lancet Rheumatology. 2021;3(11):e767-e777.

Figure 1. Participant pathway through the study. *Flare defined as DAS28 ≥ 3.2 or flare based on clinical discretion

Figure 2: Participant flow diagram. * Participant discovered to have had an immunisation prior to screening at their week 2 visit; † n=2 participants flared based on clinical discretion at face-to-face visit; § n=1 flare based on clinical discretion at face-to-face visit; ** participant was censored at day 84 visit as discovered to have an intercurrent illness at week 24 visit and was withdrawn from the study; ¶ participants had last face-to-face visits at week 2 (n=1), week 5 (n=1), ad hoc visit after week 5 (n=1), and week 12 (n=1) visits.

Figure 3: Kaplan-Meier plot of flare-free probability in the analysis cohort. Solid black line is the Kaplan-Meier estimate of the flare-free function, the grey dashed lines are the 95% CI, and black vertical marks indicate censoring. Outcomes defined as per primary analyses. A Kaplan-Meier plot including only data from final face-to-face study visits (sensitivity analysis) is included as *Supplementary Figure 2*. DMARD=Disease Modifying Anti-Rheumatic Drug.



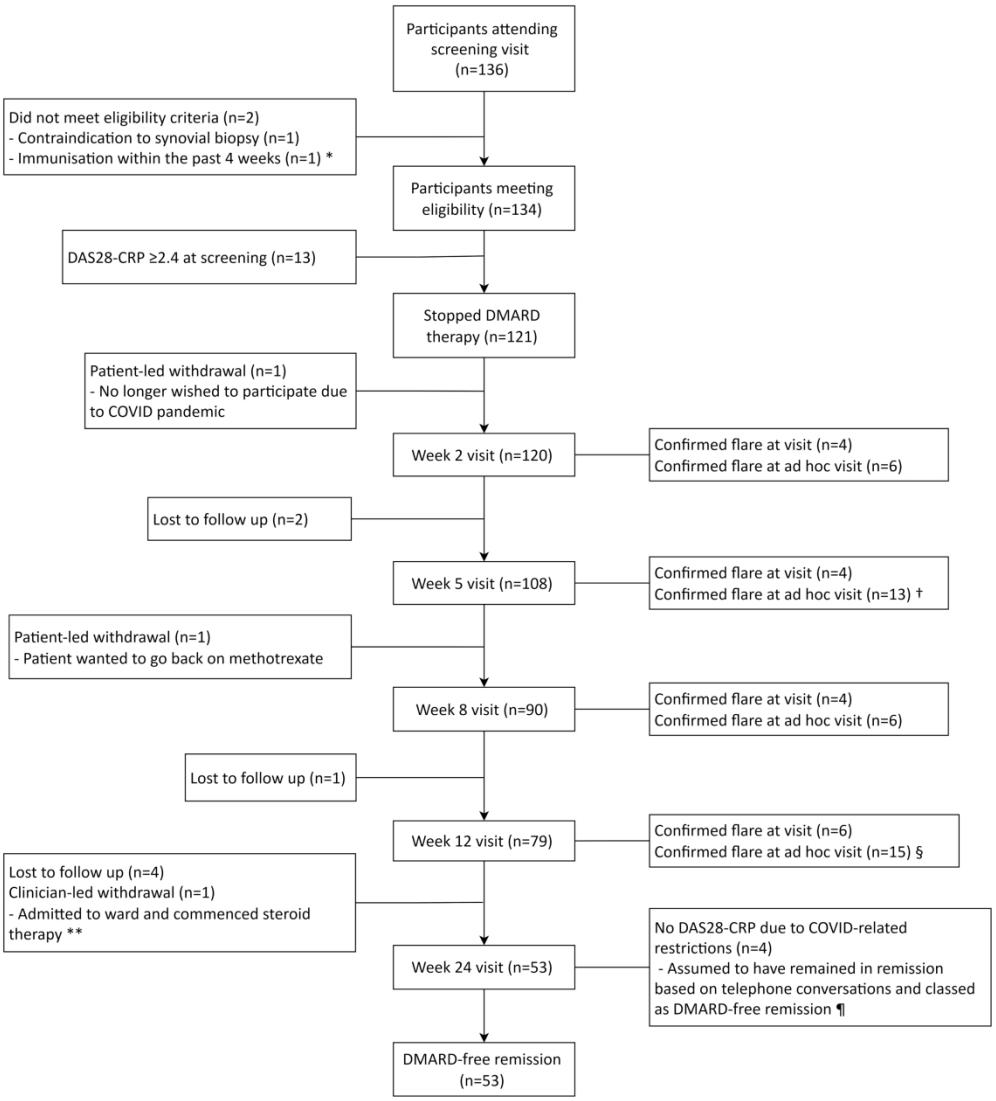


Figure 2: Participant flow diagram. * Participant discovered to have had an immunisation prior to screening at their week 2 visit; † n=2 participants flared based on clinical discretion at face-to-face visit; § n=1 flare based on clinical discretion at face-to-face visit; ** participant was censored at day 84 visit as discovered to have an intercurrent illness at week 24 visit and was withdrawn from the study; ¶ participants had last face-to-face visits at week 2 (n=1), week 5 (n=1), ad hoc visit after week 5 (n=1), and week 12 (n=1) visits.

268x296mm (300 x 300 DPI)

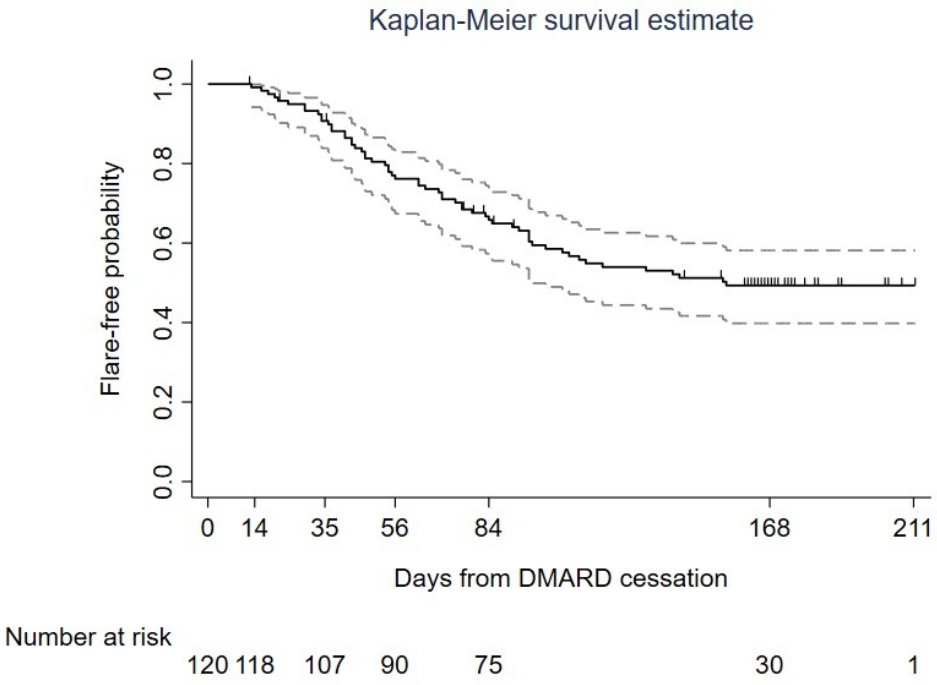


Figure 3: Kaplan-Meier plot of flare-free probability in the analysis cohort. Solid black line is the Kaplan-Meier estimate of the flare-free function, the grey dashed lines are the 95% CI, and black vertical marks indicate censoring. Outcomes defined as per primary analyses. A Kaplan-Meier plot including only data from final face-to-face study visits (sensitivity analysis) is included as Supplementary Figure 2. DMARD=Disease Modifying Anti-Rheumatic Drug.

69x50mm (300 x 300 DPI)

Clinical predictors of flare and drug-free remission in rheumatoid arthritis: preliminary results from the prospective BIO-FLARE experimental medicine study

– Supplementary Material

Contents

Statistical analysis - Predictive model with baseline risk factors

The primary analysis of this study involved the building and presentation of a prediction model containing baseline risk factors of RA flare through three steps: 1) Variable selection for the prediction model; 2) Assessment of non-linear forms for continuous predictors; 3) Estimation of a shrinkage factor to reduce overfitting; 4) Internal validation of the prediction model; and 5) Presenting the predicted probability of flare as a function of the prognostic index and time. For ease of reporting, we denote lower case i to represent an imputed dataset (i = 1, 2, …, I) and lower case j to denote a bootstrap resample of an imputed dataset (j = 1, 2, …, J).

0. Justification of maximum number of predictors in model. We computed the maximum number of variables that may be included in our prediction model, under the assumption that there would be minimal overfitting (i.e. an expected shrinkage factor ≥0.90) and a sample size of n=120 (1, 2). We used the lower bound of the 95% confidence interval of the area under curve statistic from a previous baseline prediction model (3) for RA flare as an approximation to the C index, forming the basis for our calculations. The computation indicated an upper limit of nine predictor variables.

We assumed that:

- There would be approximately E = 60 events based on the pre-specified assumption that half of participants will experience a flare (3) and the current sample size of approximately n = 120 (due to the early closure of the study); and
- A global shrinkage factor of SVH = 0.90, indicating that, given the number of events, a model with those p predictors but without shrinkage would only be slightly over-fitted to the data.

Let C $\approx$ 0.91. A value of Royston's D statistic (Equation 22) was computed as:

$$D = 5.50(C - 0.5) + 10.26(C - 0.5)^3 = 2.962$$

An apparent value of R^2_D (Equation 21) was computed as:

$$R^2_{D\_app} = \frac{\frac{\pi}{8}D^2}{\frac{\pi^2}{6} + \frac{\pi}{8}D^2} = 0.679$$

$R^2_{D\_app}$ was then used as a proxy for an apparent value of $R^2_{Royston}$ to compute an apparent value of $R^2_{O'Quigley}$ (Equation 20):

$$R^2_{O'Quigley\_app} = \frac{-\frac{\pi^2}{6}R^2_{Royston\_app}}{\left(1 - \frac{\pi^2}{6}\right)R^2_{Royston\_app} - 1} = 0.775$$

A likelihood ratio (LR) statistic was then computed (Equation 18) with $E = 60$:

$$LR = -E \times LN\left(1 - R^2_{O'Quigley\_app}\right) = 89.515$$

The number of predictors p was computed with a rearrangement of Equation 3 and $S_{VH} = 0.90$:

$$p = (1 - S_{VH}) \times LR = 8.952 \approx 9$$

* All figures are rounded to 3 d.p.

1. Variable selection.

A subset of 16 baseline variables was first selected as risk factors for flare based on biological plausibility and previous literature.

- Age at baseline
- Sex (Female vs. Male)
- Time from diagnosis to baseline in years
- Time from symptom onset to DMARD commencement in years
- Methotrexate use (Yes vs. No)
- Rheumatoid factor (RF) in IU/ml
- Anti-citrullinated peptide antibody (ACPA) in U/ml
- DAS28-CRP
- ACR/EULAR Boolean Remission (In remission vs. Not in remission) (4)
- Education (GCSE and under vs. A-levels and above (including national vocational qualifications))
- Any employment (Unemployed vs. Employed or retired)*
- Body mass index in kg/m^2
- Current smoking (Never or ex-smoker vs. Current smoker)
- Any alcohol use (Yes vs. No)
- Charlson comorbidity index (5)
- Corticosteroid use

* Due to the problems with convergence for the employment variable observed in the univariate analyses, this variable was dropped from all multiple imputation and prediction modelling steps.

Owing to the presence of missing data, we performed all analyses with I=10 imputed datasets using multiple imputation by chained equations (MICE) (6). Predictive mean matching was used as the imputation method as it has been shown to produce the less biased estimates in and better predictive performance of Cox PH models than complete case analyses or single imputation methods when the missing covariate data rate is >10% (7). The 15 candidate predictors, the outcome indicator, and the Nelson-Aalen estimate of the cumulative baseline hazard were included in the imputation model (8).

For each imputed dataset i, J=200 datasets were generated via bootstrapping (resampling with replacement). The use of MI followed by bootstrapping allows one to account for uncertainty due to missing data and uncertainty due to sampling variability during variable selection (9). Within each MICE-cum-bootstrap dataset, the 15

variables were included in a penalised Cox proportional hazard model (PH) using elastic net penalty for further variable selection whilst addressing issues of multicollinearity. We used 10-fold cross validation to select an optimal mixing parameter α and tuning parameter λ. We varied α from 0.1 to 0.9 at increments of 0.1. At each value of α, a value of the tuning parameter λ was selected at one standard error from the value of that λ associated with the regularised model with the smallest out-of-fold concordance statistic *C*. From the nine sets of α_k, λ_k, and C_k $\{k = 1, 2,\ldots 9\}$, the optimal values of α and λ was obtained from searching for the largest C_k. We assessed the stability of the candidate predictors by inspecting their bootstrap inclusion frequencies (BIF) across all 2000 datasets. The idea is that if a baseline risk factor was associated with RA flare, it should consistently, or at least in a large number of times, be selected into the final model even under "perturbations" in the data. Bootstrap resampling has been found to be a useful method of mimicking these modifications for Cox PH models (10-12). The BIF of each candidate variable is the number of times it was included in the regularised model at the optimal values of α and λ across all bootstrap replications of an imputed dataset. We identified stable risk factors as predictors with >60% BIF across the average of all imputations (up to a limit of nine predictors). The value of 60% was decided a priori before analyses were conducted.

2. Assessment of non-linear forms for continuous covariates. For each imputed dataset, we conducted univariable fractional polynomials (FP) to explore the best-fitting non-linear functional form of the selected continuous predictors (RF and ACPA). We only explored non-linear functional forms following the variable selection strategy. This was because we were not certain how to setup the elastic net variable selection process in a way such that if one of the non-linear terms for a given continuous variable (e.g., age) had its coefficient shrunk to zero during penalisation, we would also want the other non-linear terms to do the same. Thus to simplify the process, we decided to conduct the variable selection with linear terms only, then apply non-linear transformations on continuous variables that 'passed' variable selection.

For each covariate, we explored first-degree and second-degree FPs in a univariate Cox model using the RA2 closed test procedure with a nominal α value of 0.10. To avoid numerical issues, a constant of 0.1 was added to the continuous variables (13). For RF, all imputations suggested a first-degree fractional polynomial with a square root transformation. For ACPA, 4/10 imputations suggested second-degree fractional polynomials with inverse and negative square root transformations respectively, 3/10 suggested a first-degree fractional polynomial with a log-transformation respectively, and 3/10 suggested second-degree fractional polynomials with two terms with inverse transformations respectively. Thus, based on the transformation suggested most frequently across imputations, we decided on the second-degree fractional polynomials with inverse and negative square root transformations.

3. Estimation of shrinkage factor. To reduce the effects of overfitting, for each imputation *i* we estimated a shrinkage factor S_i using bootstrap estimation with *J*=200 resamples. A recent study demonstrated that for studies with small sample sizes, bootstrapping may be preferred over the heuristic shrinkage or penalised regression methods to obtain a more reliable estimate of a shrinkage factor for small sample sizes (14). To illustrate the procedure, consider a particular bootstrap resample *j* for a particular imputed dataset *i*. The survival outcome of bootstrap dataset *j* is regressed on the stable risk factors in the bootstrap sample in a Cox regression model, and the coefficients are saved. A linear predictor (LP) is then calculated as the linear combination of the values of the stable risk factors in the imputed dataset *i*, weighted by the coefficients derived from the bootstrap sample earlier. The outcome of the imputed dataset *i* is regressed on the LP and the coefficient of the LP is saved. The value of *Si* is the average of all the coefficients of LP across the *J* bootstraps. The estimated shrinkages factors ranged from 0.831 to 0.891, which were generally close to the assumed 0.90 uniform shrinkage assumed when computing the maximum number of predictors allowable in our prediction model.

Imputation	Estimated shrinkage factor
1	0.857
2	0.835
3	0.842
4	0.841
5	0.891
6	0.843
7	0.831
8	0.845

9	0.831
10	0.858
Average	**0.848**

4. Internal validation. The objective of this step is to evaluate the predictive performance of the model and derive optimism-corrected indices of discrimination (*C* index) and calibration (calibration slope and calibration-in-the-large).

To obtain estimates of optimism, we used bootstrap estimation with J=200 resamples. More details of this procedure can be found elsewhere (15). To illustrate the procedure, we describe the process for deriving the optimism-corrected *C* index but the process for the calibration slope follows a similar logic but using the coefficients of the linear predictors instead. Consider an imputed dataset i, we first regressed the survival outcome in imputed dataset i on the stable risk factors in imputed dataset i, and saved the coefficients. We then shrunk the coefficients by S_i to obtain shrunken coefficients. We computed the LP by taking the linear combination of the predictors weighted by the shrunken coefficients. We regressed the outcome in imputed dataset i on the LP and obtained an apparent C_i index.

Now consider a bootstrap resample j from the impute dataset i. We regressed the survival outcome in bootstrap dataset j on the stable risk factors in bootstrap dataset j, and saved the coefficients. We then shrunk the coefficients by S_i to obtain shrunken coefficients. We then computed two linear predictors: LP_{boot}, which is linear combination of the stable risk factors in the bootstrap sample j weighted by the shrunken coefficients; and LP_{test} which is linear combination of the stable risk factors in the imputed dataset i weighted by the shrunken coefficients. We regressed the survival outcome in bootstrap sample j on LP_{boot} and obtained C_{boot}. We regressed the survival outcome in imputed dataset i on LP_{test} and obtained C_{test}. We then subtracted C_{boot} from C_{test} to get an index of optimism$_j$. We averaged all optimism$_j$ across the J resamples to get a stable measure of optimism$_i$. An optimism-corrected C for imputation i was then optimism-corrected C_i = apparent C_i minus optimism$_i$. We then averaged all optimism-corrected C_i across I=10 imputations using Rubin's rules to obtain a single optimism-corrected *C* index (16).

Imputation	Optimism corrected C-index [95% CI]	Optimism-corrected calibration slope [95% CI]
1	0.711 [0.649, 0.774]	1.005 [0.530, 1.479]
2	0.701 [0.641, 0.761]	1.021 [0.516, 1.526]
3	0.708 [0.648, 0.767]	1.010 [0.498, 1.522]
4	0.710 [0.647, 0.773]	1.000 [0.503, 1.497]
5	0.704 [0.635, 0.774]	0.944 [0.386, 1.501]
6	0.715 [0.658, 0.771]	1.014 [0.537, 1.491]
7	0.707 [0.646, 0.768]	1.019 [0.564, 1.473]
8	0.714 [0.655, 0.773]	1.003 [0.490, 1.515]
9	0.707 [0.650, 0.765]	0.998 [0.568, 1.428]
10	0.708 [0.643, 0.773]	0.991 [0.397, 1.585]
Average	**0.709 [0.647, 0.771]**	**1.000 [0.495, 1.506]**

For the 95% confidence intervals (CI) for each interval, the bootstrap standard error is calculated as the standard deviation of the empirical distribution of bootstrap estimates. For the 95% CI for the average, we pooled the within-imputation standard errors using Rubin's rules to obtain a pooled standard error (17). Confidence intervals are calculated using the normal approximation.

5. Presenting the predicted probability of RA flare. The final equation of the prediction model is obtained by first estimating the coefficients of the stable risk factors from a Cox model in each imputation *i,* performing shrinkage using S_i, and then pooling them using Rubin's rules.

We then appended the rows of all imputed datasets *i* to create a stacked dataset (15). Because each participant has *I* replications in this stacked dataset, we gave each observation a weight of 1/*I*. We then computed the *prognostic index* (PI) as the linear combination of the values of the stable risk factors in the stacked dataset, weighted by the coefficients in the final equation of the prediction model. We regressed the survival outcome in the stacked dataset on the PI in a weighted Cox PH model, and obtained the value of the baseline survival function $\hat{S}_0(t)$ (valued at PI equals zero) at 30, 60, 90, 120, and 168 days post-DMARD cessation (see below). The weighted stacked dataset was also used in the computation of calibration plots.

Days after cessation of DMARD	Baseline survival function
30	0.969
60	0.876
90	0.801
120	0.714
168	0.672

Procedures	Screening Visit	Day 0: Baseline (a) Telephone consultation	Day 0: Baseline (b) Synovial biopsy (OPTIONAL)	Day 14	Day 35	Day 56	Day 84	Day 168	Patient-requested ad-hoc study visits	Visit 2 weeks following ad-hoc study visit	Synovial biopsy assessment visit after flare confirmed
Discuss Study / confirm willingness to continue participation in study	X	X	X	X	X	X	X	X	X	X	X
Informed Consent for study	X										
Collect Demographics and medical history	X										
Record Current medication	X			X	X	X	X	X	X	X	
General Physical examination[1]	X										
Rheumatological Assessment - DAS28-CRP	X			X	X	X	X	X	X	X	
Instruction to discontinue DMARDs (if not opting for synovial Biopsy)		X									
Instruction to discontinue DMARDs (if opting for synovial biopsy)			X								
Patient Reported Outcome Measures / Questionnaires											
HAQ-DI	X							X	X	X	
RAPID-3	X			X	X	X	X	X	X	X	
EuroQol 5D-5L	X			X	X	X	X	X	X	X	
MFI	X							X	X	X	
RA-FQ	X			X	X	X	X	X	X	X	
FLARE-RA	X			X	X	X	X	X	X	X	
Blood tests											
Full Blood Count (FBC)	X			X	X	X	X	X	X	X	
Inflammatory markers (ESR & CRP)	X			X	X	X	X	X	X	X	
Antibodies (RF & ACPA)	X										
Other clinical bloods (UE, LFT & Clotting)	X								X		
Research blood tests (Serum, EDTA, Tempus and Heparinised samples)	X			X	X	X	X	X	X	X	
Other research tests											
Urine Sample	X			X	X	X	X	X	X	X	
Pregnancy test[2]	X										
Stool Sample (OPTIONAL)	X			X	X	X	X	X	X	X	
Ultrasound assessment for Synovial Biopsy (OPTIONAL AT BASELINE – additional consent required)			[X]								X
Accelerometer provided[3] (OPTIONAL)	X										
Activity diary provided (OPTIONAL)	X			X	X	X	X		X		

Supplementary Table 1: Schedule of events in the BIO-FLARE study.

[1] Depending on the circumstances of the consultation, physical examination may be indicated at any study visit to establish whether DAS28-CRP reflects arthritis activity or infection etc. General Physical Examination is only mandatory at Screening.
[2] Mandatory at Screening but should be performed at any visit subsequently if routine questioning suggests a participant may be pregnant. Serum or urine tests to be performed subsequently in line with local policy
[3] This may be provided after the study visit once eligibility confirmed, either by post, or at the optional Baseline Synovial Biopsy Visit (if applicable)

Adverse events

In total, 82 out of 121 participants (68%) experienced at least one adverse event (AE) in the sample. There were a total of 155 adverse events with a median of 1 (IQR: 1, 2) event per participant (range: 1 to 6). The breakdown of the number of participants reporting each type of AE is presented below, organised by their system organ class. Additionally, there were 4 serious adverse events (SAE) occurring over 4 participants (Supplementary Table 3).

		Modified per-protocol cohort (n=111)*	
	Study population (n=121)	Flared (n=58)	Remission at week 24 visit (n=53)
	N (%)	N (%)	N (%)
Blood and lymphatic system disorders			
Anaemia	1 (0.8)	1 (1.7)	0
Neutropenia	2 (1.7)	0	2 (3.8)
Thrombocytopenia	1 (0.8)	0	1 (1.9)
Ear and labyrinth disorders			
Excessive cerumen production	1 (0.8)	0	1 (1.9)
Vertigo	1 (0.8)	0	1 (1.9)
Endocrine disorders			
Hypothyroidism	1 (0.8)	0	1 (1.9)
Eye disorders			
Blepharitis	1 (0.8)	0	1 (1.9)
Cataract	1 (0.8)	0	1 (1.9)
Dry eye	1 (0.8)	1 (1.7)	0
Gastrointestinal disorders			
Constipation	1 (0.8)	0	1 (1.9)
Dyspepsia	1 (0.8)	0	1 (1.9)
Enteritis	1 (0.8)	1 (1.7)	0
Gastritis	1 (0.8)	0	1 (1.9)
Mouth ulceration	1 (0.8)	1 (1.7)	0
Pancreatic mass	1 (0.8)	0	1 (1.9)
Toothache	2 (1.7)	1 (1.7)	1 (1.9)
Vomiting	1 (0.8)	0	1 (1.9)
General disorders and administration site conditions			
Chest pain	1 (0.8)	0	1 (1.9)
Critical illness	1 (0.8)	1 (1.7)	0
Fatigue	2 (1.7)	2 (3.4)	0
Hernia	1 (0.8)	0	1 (1.9)
Malaise	1 (0.8)	0	1 (1.9)
Immune system disorders			
Hypersensitivity	1 (0.8)	0	1 (1.9)
Infections and infestations			
Cellulitis	1 (0.8)	1 (1.7)	0
Conjunctivitis viral	1 (0.8)	1 (1.7)	0
Coxsackie viral infection	1 (0.8)	0	1 (1.9)
Gastroenteritis	1 (0.8)	0	1 (1.9)
Gastroenteritis viral	1 (0.8)	0	1 (1.9)
Infected bite	1 (0.8)	1 (1.7)	0
Lower respiratory tract infection	5 (4.1)	1 (1.7)	4 (7.5)
Oral herpes	1 (0.8)	0	0
Otitis externa	1 (0.8)	1 (1.7)	1 (1.9)
Rash pustular	1 (0.8)	1 (1.7)	0
Rhinitis	1 (0.8)	0	1 (1.9)
Sinusitis	1 (0.8)	1 (1.7)	0
Tooth abscess	1 (0.8)	0	1 (1.9)
Tooth infection	1 (0.8)	0	1 (1.9)
Upper respiratory tract infection	20 (16.5)	9 (15.5)	8 (15.1)
Urinary tract infection	1 (0.8)	0	0

Viral infection	1 (0.8)	1 (1.7)	0
Viral upper respiratory tract infection	7 (5.8)	6 (10.3)	1 (1.9)
Injury, poisoning and procedural complications			
Arthropod bite	1 (0.8)	0	1 (1.9)
Avulsion fracture	1 (0.8)	0	1 (1.9)
Back injury	1 (0.8)	0	1 (1.9)
Contusion	1 (0.8)	1 (1.7)	0
Fall	2 (1.7)	1 (1.7)	1 (1.9)
Laceration	1 (0.8)	1 (1.7)	0
Limb injury	1 (0.8)	1 (1.7)	0
Spinal fracture	2 (1.7)	2 (3.4)	0
Wound	2 (1.7)	1 (1.7)	1 (1.9)
Investigations			
Blood glucose abnormal	1 (0.8)	1 (1.7)	0
C-reactive protein increased	1 (0.8)	1 (1.7)	0
Liver function test abnormal	1 (0.8)	1 (1.7)	0
Platelet count decreased	1 (0.8)	1 (1.7)	0
Musculoskeletal and connective tissue disorders			
Arthralgia	1 (0.8)	0	1 (1.9)
Fracture	1 (0.8)	0	1 (1.9)
Joint stiffness	1 (0.8)	1 (1.7)	0
Musculoskeletal pain	1 (0.8)	1 (1.7)	0
Myalgia	1 (0.8)	1 (1.7)	0
Pain in extremity	1 (0.8)	1 (1.7)	0
Periarthritis	1 (0.8)	1 (1.7)	0
Soft tissue swelling	1 (0.8)	0	1 (1.9)
Tendonitis	2 (1.7)	1 (1.7)	1 (1.9)
Tenosynovitis	2 (1.7)	2 (3.4)	0
Neoplasms benign, malignant and unspecified (incl cysts and polyps)			
Seborrhoeic keratosis	1 (0.8)	0	0
Nervous system disorders			
Cerebrovascular accident	1 (0.8)	0	0
Dizziness	1 (0.8)	1 (1.7)	0
Headache	3 (2.5)	1 (1.7)	2 (3.8)
Migraine	1 (0.8)	0	1 (1.9)
Neuralgia	1 (0.8)	1 (1.7)	0
Restless legs syndrome	1 (0.8)	0	1 (1.9)
Sciatica	4 (3.3)	2 (3.4)	2 (3.8)
Seizure	1 (0.8)	0	1 (1.9)
Syncope	1 (0.8)	0	1 (1.9)
Transient ischaemic attack	1 (0.8)	0	0
Psychiatric disorders			
Depressed mood	1 (0.8)	0	1 (1.9)
Emotional distress	1 (0.8)	1 (1.7)	0
Insomnia	1 (0.8)	1 (1.7)	0
Respiratory, thoracic and mediastinal disorders			
Cough	5 (4.1)	2 (3.4)	3 (5.7)
Nasal dryness	1 (0.8)	1 (1.7)	0
Oropharyngeal pain	6 (5)	4 (6.9)	2 (3.8)
Skin and subcutaneous tissue disorders			
Eczema	1 (0.8)	0	1 (1.9)
Neurodermatitis	1 (0.8)	0	1 (1.9)
Pruritus	1 (0.8)	1 (1.7)	0
Rash	1 (0.8)	1 (1.7)	1 (1.9)
Rash erythematous	2 (1.7)	1 (1.7)	0
Skin lesion	1 (0.8)	1 (1.7)	0
Transient acantholytic dermatosis	1 (0.8)	1 (1.7)	0
Surgical and medical procedures			

Medical device removal	1 (0.8)	0	1 (1.9)
Tooth extraction	2 (1.7)	1 (1.7)	1 (1.9)
Tooth repair	1 (0.8)	1 (1.7)	0
Vascular disorders			
Aneurysm	1 (0.8)	0	1 (1.9)
Hypertension	1 (0.8)	1 (1.7)	0
Temporal arteritis	1 (0.8)	0	0

Supplementary Table 2: All adverse events. *Discrepancy between study population versus modified per-protocol cohort is due to exclusion of participants who were lost to follow-up (n=7) or withdrawn (n=3) before week 24 visit

Participant	Days from DMARD cessation to start of SAE	Duration of SAE (days)	SAE	Causality	Expected	Severity	Type of SAE / Action taken	Patient withdrawn from study?
1	174		Giant cell arteritis	Unrelated		Mild	Hospitalisation	Yes
2	176	2	Headache	Unrelated		Moderate	Hospitalisation	No
3	92	n/a	Incidental pancreatic body cystic mass	Unrelated	Unexpected	Severe	Other medically significant event – referred for urgent investigation	No
4	99	1	Brief hospital admission for atypical chest pain	Unrelated	Unexpected	Moderate	Hospitalisation	No

Supplementary Table 3: Serious adverse events

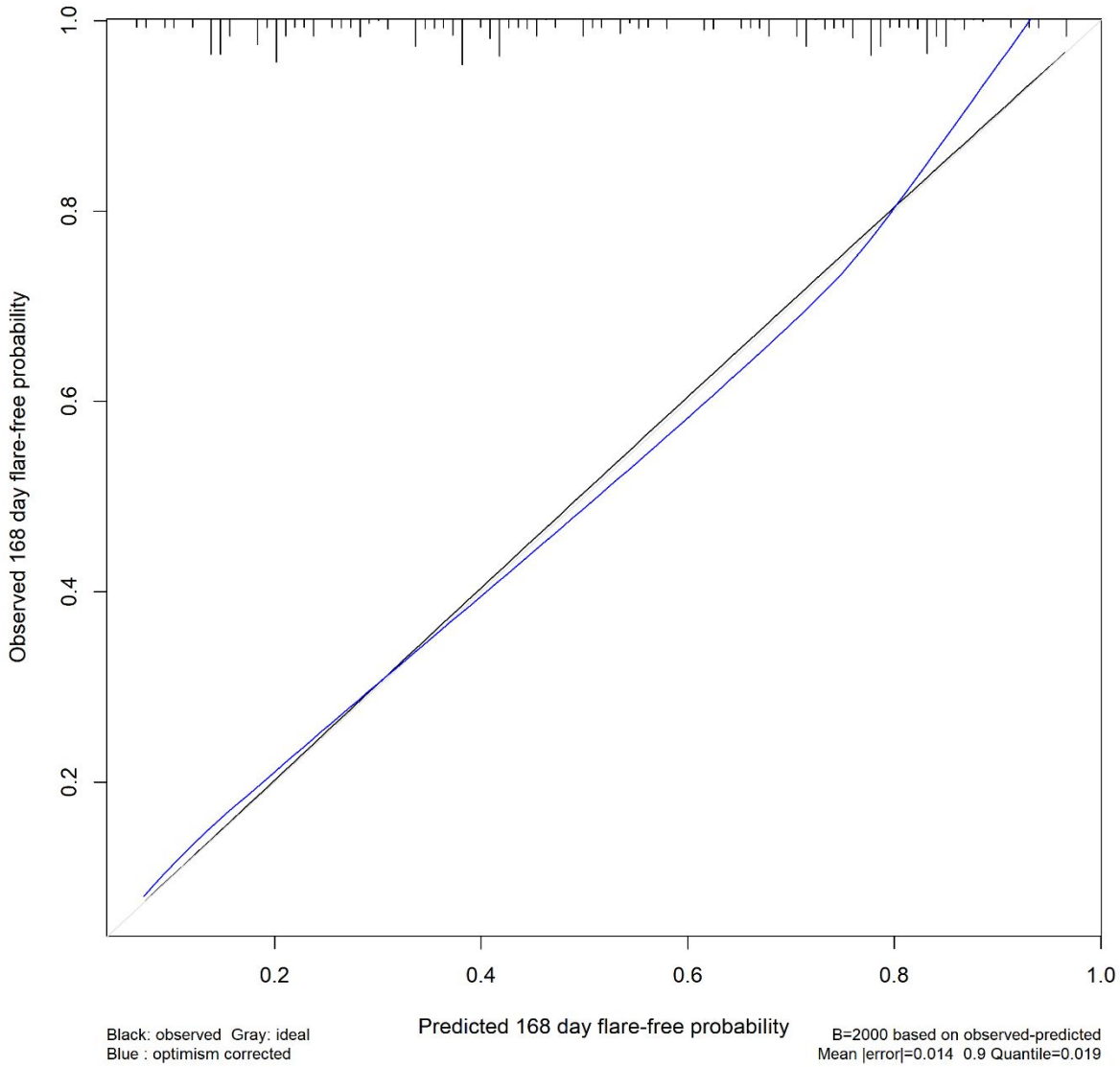
Patient	Description
1	1 tender joint (right wrist), patient VAS = 51/100, CRP = 8 giving a DAS28-CRP of 3.03. Clinical team and patient felt they were flaring so a shared decision was made to restart DMARDs rather than wait for a second ad-hoc appointment in 14 days to confirm flare
2	Ankle (tibialis posterior) tenosynovitis requiring treatment.
3	Bilateral knee synovitis; three swollen joints in total; clinicians and patient felt restarting DMARDs necessary.

Supplementary Table 4: Characteristics of patients adjudged to have flared based on clinician discretion (n=3).

	Imputation										Average BIF
	1	2	3	4	5	6	7	8	9	10	
Age	0.105	0.115	0.105	0.110	0.105	0.110	0.120	0.115	0.130	0.110	0.115
Sex	0.585	0.605	0.590	0.585	0.585	0.595	0.605	0.595	0.555	0.595	0.583
Time from diagnosis to baseline	0.170	0.175	0.155	0.215	0.205	0.240	0.205	0.190	0.200	0.155	0.187
Time from symptom onset to DMARD commencement	0.425	0.345	0.370	0.350	0.335	0.345	0.375	0.360	0.340	0.240	0.348
Methotrexate use	0.775	0.790	0.795	0.780	0.765	0.765	0.770	0.775	0.740	0.750	0.768
RF	0.865	0.885	0.870	0.865	0.890	0.900	0.840	0.925	0.925	0.920	0.888
ACPA	0.730	0.775	0.710	0.835	0.720	0.590	0.785	0.670	0.460	0.560	0.677
DAS28-CRP	0.155	0.150	0.155	0.155	0.150	0.160	0.155	0.150	0.160	0.145	0.153
ACR/EULAR Boolean Remission	0.200	0.185	0.185	0.195	0.175	0.165	0.180	0.185	0.165	0.190	0.182
Education	0.215	0.200	0.225	0.205	0.220	0.250	0.285	0.220	0.230	0.275	0.237
BMI	0.380	0.305	0.445	0.380	0.520	0.485	0.545	0.370	0.535	0.485	0.445
Current smoking	0.190	0.195	0.220	0.205	0.245	0.235	0.250	0.190	0.220	0.225	0.216
Any alcohol use	0.465	0.435	0.455	0.430	0.435	0.465	0.430	0.440	0.445	0.445	0.443
CCI	0.195	0.195	0.200	0.195	0.175	0.200	0.190	0.190	0.185	0.190	0.192
Corticosteroid use	0.260	0.240	0.235	0.240	0.240	0.255	0.245	0.255	0.260	0.255	0.247

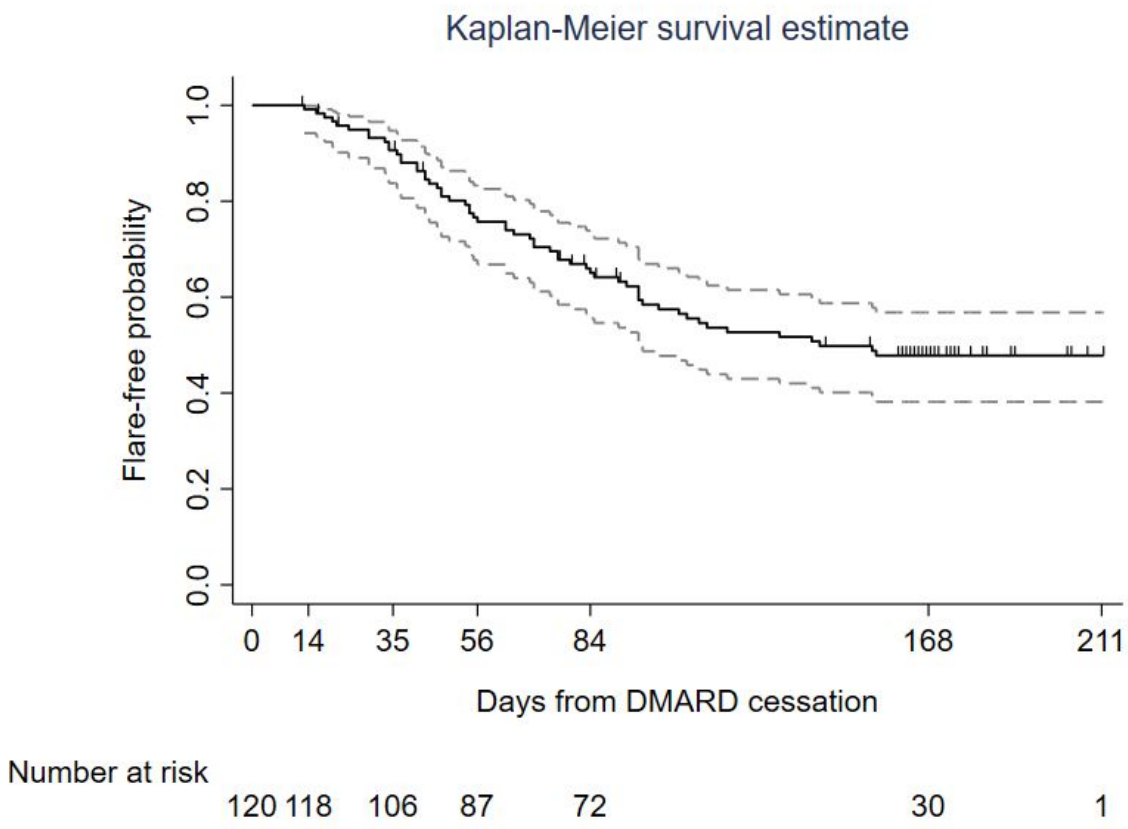
Supplementary Table 5: Bootstrap inclusion frequencies

Sex, methotrexate use, RF, and ACPA were brought forward to the prediction model. Although sex did not cross the a priori 60% average BIF threshold, we included it in the prediction model as its average BIF was highly proximal to the threshold. RF=Rheumatoid Factor; ACPA=Anti-Citrullinated Peptide Antibody; BMI=Body Mass Index; CCI=Charlson Comorbidity Index. Bolded predictors are those that were included in the prediction model



Supplementary Figure 1: Calibration plot at day 168

Dashed lines at the top represent a histogram of the predicted 168-day flare-free probabilities. The risk (or probability) of flare by 168 days may be taken as 1.0 minus the flare-free probability. The light grey diagonal line represents the line of perfect agreement between predicted and observed flare-free probabilities. The blue line indicates the optimism-corrected calibration curve, and the black line indicates the uncorrected calibration curve. The results suggest that the model produced predicted estimates of flare risk that had good agreement with observed risk. As a minor caveat, based on the calibration slopes for risk of flare by 168 days, the model slightly underestimated the predicted risk of flare for participants with an observed "moderate-high" risk (30–80% observed risk), and overestimated predicted risk of flare for participants at lower (≤30% observed risk) and higher observed risk (≥80% observed risk).



Supplementary Figure 2: Kaplan-Meier plot of flare-free survival based on face-to-face visits.
Solid line is the Kaplan-Meier estimate of the survival function, the grey dashed lines are the 95% CI, and vertical black marks indicate censoring. Outcomes defined as per sensitivity analysis, i.e. using last face-to-face visits / last available DAS28-CRP.

Supplementary References

1. Riley RD, Snell KI, Ensor J, Burke DL, Harrell Jr FE, Moons KG, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Statistics in Medicine. 2019;38(7):1276-96.
2. Riley RD. Correction to: Minimum sample size for developing a multivariable prediction model: Part II-binary and time-to-event outcomes by Riley RD, Snell KI, Ensor J, et al. Statistics in Medicine. 2019;38(30):5672-.
3. Baker KF, Skelton AJ, Lendrem DW, Scadeng A, Thompson B, Pratt AG, et al. Predicting drug-free remission in rheumatoid arthritis: A prospective interventional cohort study. Journal of autoimmunity. 2019;105:102298.
4. Bykerk VP, Massarotti EM. The new ACR/EULAR remission criteria: rationale for developing new criteria for remission. Rheumatology. 2012;51(suppl 6):vi16-vi20.
5. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. Journal of chronic diseases. 1987;40(5):373-83.
6. van Buuren SG-O, K. mice: Multivariate Imputation by Chained Equations in R. HJournal of Statistical Software. 2011;45(3):1-67.
7. Marshall A, Altman DG, Holder RL. Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: a resampling study. BMC Medical Research Methodology. 2010;10(1):112.
8. White IR, Royston P. Imputing missing covariate values for the Cox model. Statistics in Medicine. 2009;28(15):1982-98.
9. Heymans MW, Van Buuren S, Knol DL, Van Mechelen W, De Vet HC. Variable selection under multiple imputation using the bootstrap in a prognostic study. BMC Medical Research Methodology. 2007;7(1):33.
10. Sauerbrei W, Buchholz A, Boulesteix A-L, Binder H. On stability issues in deriving multivariable regression models. Biometrical Journal. 2015;57(4):531-55.
11. Sauerbrei W, Schumacher M. A bootstrap resampling procedure for model building: Application to the cox regression model. Statistics in medicine. 1992;11(16):2093-109.
12. Altman DG, Andersen PK. Bootstrap investigation of the stability of a cox regression model. Statistics in medicine. 1989;8(7):771-83.
13. Sauerbrei WM-H, C; Benner, A; Royston, P. Multivariable regression model building by using fractional polynomials: Description of SAS, STATA and R programs. Computational Statistics & Data Analysis. 2006;50:3464 - 85.
14. Riley RD, Snell KIE, Martin GP, Whittle R, Archer L, Sperrin M, et al. Penalization and shrinkage methods produced unreliable clinical prediction models especially when sample size was small. Journal of Clinical Epidemiology. 2021;132:88-96.
15. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating: Springer Cham; 2019. XXXIII, 558 p.
16. Marshall A, Altman DG, Royston P, Holder RL. Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. BMC Medical Research Methodology. 2010;10(1):7.
17. Bartlett JW, Hughes RA. Bootstrap inference for multiple imputation under uncongeniality and misspecification. Stat Methods Med Res. 2020 Dec;29(12):3533-3546.