SUPPLEMENTARY INFORMATION

Supplementary methods

Protocol amendments

1. Classified primary and secondary objectives and endpoints.
   Rationale: Previous version listed all objectives and endpoints as exploratory.

2. Screening window extended to 6 weeks.
   Rationale: To facilitate scheduling of screening magnetic resonance imaging (MRI) scan, which should only be performed after other screening assessments have been passed.

3. Estimated number of screened subjects revised.
   Rationale: More conservative estimate based on recent clinical study experience.

4. Deleted ">15% relative decrease in diffusing capacity or transfer factor of the lung for carbon monoxide (D_{LCO}) from screening" as trigger for subject to be withdrawn from study drug.
   Rationale: Second D_{LCO} test is at week 12, after dosing of study drug has been completed.

5. Consolidated details of pulmonary function tests into additional guidance.
   Rationale: To avoid any misunderstanding of pulmonary function tests.

6. Addition of day 1 joint count, and clarification that chest high-resolution computed tomography (HSCT) must be performed if $D_{LCO} \geq 60\%$ - <70% predicted.
   Rationale: To ensure subjects still have active rheumatoid arthritis (RA) with the extended screening window, and to avoid any misunderstanding of requirement for chest HSCT.

7. Addition of history of sensitivity to gadolinium (Gd)-containing contrast agents to exclusion criteria.
   Rationale: To maintain safety.
8. Rewording so that re-screening is permitted.
   Rationale: Had previously stated that re-screening is required.

9. Revision that subjects must have passed all screening assessments, including laboratory tests, prior to undertaking MRI scanning, and that whole blood flow cytometry is scheduled at baseline (day 1).
   Rationale: It is more appropriate that these tests are done for patients likely to be randomised.

10. Clarification that RNA analysis is not part of the pharmacogenetics sub-study.
    Rationale: RNA transcriptomics/sequencing is not proposed for this study.

11. Removed "daily" from clinical assessments.
    Rationale: Error in frequency of assessment.

12. Re-wording to indicate that all MRI scans will be available on-site, and may be reported locally.
    Rationale: Local standard procedures for reporting MRI scans may differ.

13. Added missing abbreviations.
    Rationale: Some abbreviations had not been included.

14. Removed from eligibility criteria the requirement for two forms of complementary contraception, except when using hormonal contraceptives.
    Rationale: Due to potential drug interaction with CYP450 substrates, a second form of contraception is required for subjects using hormonal contraceptives. This requirement does not apply to subjects that are not using hormonal contraception, hence the contraception guidance has been corrected.

15. Removed "morning stiffness" assessment and added that the Borg dyspnoea scale will be completed on paper.
    Rationale: Morning stiffness is not being recorded in the study, and an electronic patient-reported outcome device not available for the Borg scale.

Protocol deviations

Of the 39 patients randomized in the intent-to-treat population, 8 (73%) patients in the placebo and 24 (86%) patients in the otilimab groups had major protocol deviations (see Table below). The majority were noncompliance with study procedures. No protocol violations that impacted the interpretation of the study results.

Table. Protocol deviations (intent-to-treat population)

Category, n (%)	Placebo (n=11)	Otilimab (n=28)
Patients with ≥1 minor protocol deviation	9 (82)	21 (75)
Patients with ≥1 major protocol deviation	8 (73)	24 (86)
Any major protocol deviation		
Assessments and/or procedures	5 (45)	17 (61)
Other protocol deviation category	2 (18)	5 (18)
Visit, assessment or timepoint window	3 (27)	9 (32)
Wrong administration date/Study drug not rotated/Study drug not administered	1 (9)	4 (14)

A subject may have had more than 1 protocol deviation.

Inclusion criteria

1. Age ≥18 years at the time of signing informed consent.
2. Met American College of Rheumatology (ACR)/European League Against Rheumatism (EULAR) 2010 RA Classification Criteria AND subject not diagnosed before age of 16 years.
3. Functional Class I, II or III defined by the 1992 ACR Classification of Functional Status in RA.
4. Active disease as defined by:
    a. Swollen joint count of ≥4 (66-joint count) and tender joint count of ≥4 (68-joint count) at screening and at day 1.

    AND

    b. Disease activity score for 28 different joints (DAS28) with C-reactive protein (CRP) (DAS28[CRP]) ≥3·2 at screening.

    AND

    c. CRP ≥3·0 mg/L.
5. Signs of inflammation such as synovitis in the MRI scan of the most-affected hand.
6. Must have been taking methotrexate (MTX) (15–25 mg weekly; oral/injected) for ≥12 weeks before screening, with no change in the route of administration, and with a stable and tolerated dose for ≥4 weeks prior to day 1. A stable dose of MTX ≥7·5 mg/week was acceptable, if the MTX dose had been reduced for reasons of documented intolerance to MTX, for example, hepatic or haematological toxicity, or per local requirement.
7. Body weight ≥45 kg.
8. Male or female subjects were eligible to participate so long as they met and agreed to abide by agreed contraceptive criteria.
9. Capable of giving signed informed consent, which included compliance with the requirements and restrictions listed in the consent form and the protocol.
10. Willing to continue or initiate treatment with oral folic acid (≥5 mg/week) or equivalent and be treated during the entire study (mandatory co-medication for MTX treatment).
11. D_{LCO} ≥60% predicted; forced expiratory volume in 1 second ≥70% predicted.
    a. For subjects with D_{LCO} values ≥60%–<70%, a baseline chest HRCT had to be performed during the screening period, and it was recommended that the subject be reviewed by a local pulmonologist to exclude significant pre-existing respiratory disease.
12. No evidence of active or latent infection with tuberculosis (TB), as defined by all of the following:
    a. No history of active or latent TB infection irrespective of treatment status.
    b. A negative diagnostic TB test at screening defined as:
        i. A negative QuantiFERON Gold test or T-spot test (two successive indeterminate QuantiFERON tests will be considered as a positive result).

    OR

ii. If QuantiFERON gold or T-spot test not approved or registered in country of participation, then a negative tuberculin skin test reaction as per local guidelines is required (it is strongly recommended that subjects with a history of Bacillus Calmette–Guérin (BCG) vaccination be tested with QuantiFERON gold test).
iii. Chest X-ray within 12 weeks of day 1 with no evidence of current or previous pulmonary TB, locally read by a radiologist.

NB: If there had been recent close contact with persons who have active TB prior to study enrolment the subject was referred to a TB physician to undergo additional evaluation.

Exclusion criteria

1. Pregnant or lactating, or women planning to become pregnant or initiating breastfeeding.
2. History of other inflammatory rheumatologic or autoimmune disorders, other than Sjögren's syndrome secondary to RA.
3. History of any respiratory disease that (in the opinion of the investigator) would compromise subject safety or the ability of the subject to complete the study (e.g. significant interstitial lung disease, such as pulmonary fibrosis, chronic obstructive pulmonary disease, moderate-severe asthma, bronchiectasis, previous pulmonary alveolar proteinosis).
4. Clinically significant (in the opinion of the investigator) persistent cough or clinically significant or unstable dyspnoea that is unexplained.
5. QT interval corrected for heart rate (QTc) >450 msec or QTc >480 msec for subjects with bundle branch block. The QTc is the QT interval corrected for heart rate according to Fridericia's formula (QTcF).
6. Liver function tests: alanine aminotransferase (ALT) >1·5x upper limit of normal (ULN); aspartate transaminase (AST) >1·5 upper limit of normal; alkaline phosphatase and bilirubin ≥1·5x ULN (isolated bilirubin >1·5x ULN is acceptable if bilirubin is fractionated and direct bilirubin <35%).
7. Current active liver or biliary disease (except for Gilbert's syndrome or asymptomatic gallstones or otherwise stable chronic liver disease per investigator assessment).
8. Significant unstable or uncontrolled acute or chronic disease (e.g. cardiovascular including uncompensated congestive cardiac failure New York Heart Association III or IV, myocardial infarction within 12 months, unstable angina pectoris, uncontrolled hypertension, uncontrolled hypercholesterolemia), and pulmonary, haematological, gastrointestinal (including Crohn's Disease or ulcerative colitis), hepatic, renal, neurological, psychiatric, malignancy, endocrinologic or infectious diseases, which, in the opinion of the investigator, could confound the results of the study or put the subject at undue risk.
9. A history of malignant neoplasm within the last 10 years or breast cancer within the last 20 years, except for non-melanoma skin cancers that had been excised and cured, or carcinoma in situ of the uterine cervix.
10. Current or history of renal disease, or estimated creatinine clearance <60 mL/min/1·73 m^2 or serum creatinine >1·5x ULN at screening.
11. Hereditary or acquired immunodeficiency disorder, including immunoglobulin deficiency.
12. History of infected joint prosthesis at any time, with the prosthesis still in situ. History of leg ulcers, catheters, chronic sinusitis or recurrent chest or urinary tract infections.
13. Active infections, or history of recurrent infections (excluding recurrent fungal infections of the nail bed), or has required management of acute or chronic infections, as follows:
    a. Currently taking any suppressive therapy for a chronic infection (such as TB, pneumocystis, cytomegalovirus, herpes simplex virus, herpes zoster and atypical mycobacteria), **OR**
    b. Hospitalisation for treatment of infection within 26 weeks of day 1, **OR**
    c. Use of parenteral (intravenous or intramuscular antimicrobials; antibacterials, antivirals, antifungals, or antiparasitic agents) within 26 weeks of day 1 or oral antimicrobials within 14 days of day 1.
14. A vaccination (live or attenuated) within 30 days of day 1 or BCG vaccination within 365 days of day 1, or a live vaccination planned during the study.
15. Any surgical procedure, including bone or joint surgery/synovectomy within 12 weeks prior to day 1 or any planned surgery within the duration of the study or follow-up period.
16. Contraindication to MRI scanning (as assessed by local MRI safety questionnaire), which includes but is not limited to:
    a. Intracranial aneurysm clips (except Sugita) or other metallic objects.

b. History of intra-orbital metal fragments that have not been removed by a medical professional.
c. Pacemakers or other implanted cardiac rhythm management devices and non-magnetic-resonance-compatible heart valves.
d. Inner ear implants.
e. History of claustrophobia, which may impact participation.
f. History of sensitivity to Gd-containing contrast agents.

17. Use of prohibited medications prior to AND throughout the study:
    - Any conventional disease-modifying antirheumatic drugs other than MTX (including hydroxychloroquine, sulphasalazine, minocycline and cyclosporin) had to be withdrawn ≥2 weeks prior to day 1.
        - Subjects may have required longer to discontinue leflunomide prior to randomisation:
            - Leflunomide must be discontinued ≥12 weeks prior to randomisation (or ≥14 days after 11 days of standard cholestyramine or activated charcoal washout).
        - For these subjects, written informed consent for the study was obtained prior to beginning the screening period. However, other screening assessments, other than consent, occurred within 42 days prior to randomisation.
    - Azathioprine was discontinued ≥28 days prior to randomisation.
    - Any alkylating agents (such as cyclophosphamide or chlorambucil).
    - Plasmapheresis or intravenous immunoglobulin or use of Staph protein A column (Prosorba) within 26 weeks of day 1.
    - Biologic agents were discontinued prior to day 1:
        - Anakinra or etanercept (4 weeks prior)
        - Adalimumab (6 weeks prior)
        - Infliximab (8 weeks prior)
        - Certolizumab pegol or golimumab (10 weeks prior)
        - Abatacept or tocilizumab (12 weeks prior)
        - Belimumab, rituximab or other selective B lymphocyte depleting agents (1 year prior, and if CD19/20+ counts are normal by fluorescence-activated cell sorter analysis).
    - Tofacitinib discontinued ≥2 weeks prior to day 1.
    - Corticosteroids:
        - Any intramuscular, intravenous or intra-articular corticosteroids within 8 weeks of day 1.
    - Oral corticosteroids:
        - Any treatment with >10 mg/day dose oral prednisolone (or equivalent) within 28 days of day 1.
        - New oral corticosteroid or changes in corticosteroid dose within the 28 days prior to day 1. New topical steroids and immunosuppressive agents (e.g. eye drops, creams) are permitted.
    - Non-steroidal anti-inflammatory drugs (NSAIDs):
        - New or change in dose of NSAID within 14 days of day 1.
    - Any prior investigational treatment discontinued for ≥4 weeks or ≥5 half-lives, whichever is longer, prior to day 1.
18. Current drug or alcohol abuse or dependence, or a history of drug or alcohol abuse or dependence within a year prior to day 1.
19. History of sensitivity to any of the study treatments, or components thereof or a history of drug or other allergy that, in the opinion of the investigator or Medical Monitor, contraindicates their participation.
20. Abnormal chest X-ray within 12 weeks of day 1 (locally read and reported by a radiologist) judged by the investigator as clinically significant.
21. Any Grade 3 or 4 haematology or clinical chemistry laboratory abnormality (common terminology criteria for adverse events (AEs), 2009 v4·0) at screening.
22. Haemoglobin ≤9 g/dL; white blood cell count ≤3·0 x 10^9/L; platelet count ≤100 x 10^9/L; absolute neutrophil count ≤1·5 x 10^9/L; lymphocyte count ≤0·5 x 10^9/L at screening.
23. Serologic evidence of current/previous hepatitis B virus infection based on the results of testing for hepatitis B surface antigen (HBsAg) and anti-hepatitis B core (anti-HBc) antibody as follows at screening:
    - Subjects positive for HBsAg and/or positive for anti-HBc antibody (regardless of anti-HBs antibody status) were excluded.
24. Positive test for hepatitis C virus antibody confirmed on a subsequent blood sample by RNA-polymerase chain reaction (PCR) assay at screening.

- Subjects who were positive for hepatitis C antibody and negative when the hepatitis C RNA-PCR assay was performed on a subsequent sample were eligible to participate.
- Subjects who were positive for hepatitis C antibody and had a positive result for HCV when the hepatitis C RNA-PCR assay was performed on the subsequent sample were not eligible to participate.

25. Positive serology for HIV 1 or 2 at screening.

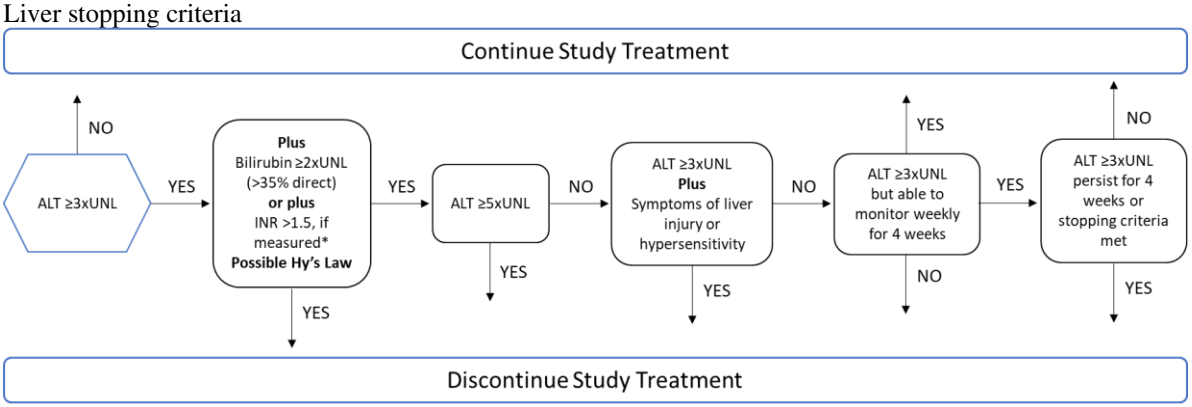
Treatment withdrawal or interruption

Stopping guidelines

Patients were to be withdrawn from the study if any of the following criteria were met:

- All serious infections.
- Pregnancy.
- Confirmed pulmonary alveolar proteinosis (PAP).
- Severe or serious hypersensitivity reactions (including anaphylaxis).
- Liver chemistry or QT stopping criteria.
- Persistent or recurrent haematological laboratory abnormalities.
- Other serious or severe AEs at the discretion of the investigator, after consultation with the Medical Monitor.

Liver stopping criteria



*International normalised rate value not applicable to patients on anticoagulants

QTc stopping criteria

A patient who meets either of the bulleted criteria were to be withdrawn from the study:

- QTc >500 msec OR uncorrected QT >600 msec
- Change from baseline (CFB) of QTc >60 msec

The QTc is the QT interval corrected for heart rate according to Fridericia's formula (QTcF).

Discontinuation criteria for patients with underlying **bundle branch block**

Baseline QTc with bundle branch block	**Discontinuation QTc with bundle branch block**
>450 msec 450−480 msec	>500 msec ≥530 msec

Treatment interruption

Study medications were to be temporarily suspended to allow investigation in the event of persistent cough (common terminology criteria [CTC] grade 2 or 3) or dyspnoea (Borg scale grade 3 or above) for three consecutive weeks. The patient would have been referred to a pulmonologist for further assessment. The study drug would have been suspended until the symptoms or signs that caused the referral had resolved and/or the diagnosis had been determined and clinically significant events had been excluded by the pulmonologist.

The following haematological laboratory abnormalities would have required temporary suspension of study medications and prompt retesting, ideally within 3–5 days:

- White blood cell count <2·0 x 10^9/L
- Absolute neutrophil count <1·0 x 10^9/L
- Lymphocyte count <0·5 x 10^9/L

Study medication would not have been restarted until the parameters were above these values, and patients followed as appropriate until resolution of the event. If these abnormalities were persistent (present on ≥two sequential tests), or occurred recurrently (on two separate occasions), study medications would have been permanently discontinued and the patient withdrawn from the study.

MRI methodology

MRI was performed at baseline, week 4, week 12 and week 22 (post-treatment follow-up). At each imaging visit, images were acquired for RA MRI scoring (RAMRIS)[1], Cartilage Loss Score (CARLOS)[2] and RA MRI quantitative (RAMRIQ) assessment[3] (contrast-enhanced MRI), and dynamic contrast-enhanced MRI (DCE-MRI) assessments in the most affected hand/wrist, as determined at screening.

Image acquisition

All images were acquired on 1·5T or 3T MRI scanners (Philips, Siemens or GE) with hardware and software capabilities deemed suitable to provide images of high quality for DCE-MRI, with imaging centres trained in the image acquisition procedures. Imaging was performed using either a multichannel knee coil or a dedicated hand/wrist MR coil (e.g. Siemens 15-channel hand coil). The same MRI scanner and coil were used at each visit for each patient.

Patients were positioned prone on the scanner bed and with the arm of the affected hand extended above their head. The hand was positioned in the centre of the coil, with the coil in the centre of the scanner magnet bore to ensure good signal in both the wrist and hand joints. The hand was positioned in the coil on an acrylic frame (M-frame, Spire Sciences, Inc.) to ensure reproducible bone and joint alignment between visits and definitive verification of the side (right or left) imaged, and biplanar slice alignment was used to ensure reproducible tomography of the anatomy.

Scanning protocol was standardised across sites and scanner platforms (Siemens, GE, Philips) to minimise variability in results between sites. The imaging protocol consisted of pre-contrast STIR (coronal short-tau inversion recovery) images followed by a pre-contrast fat-suppressed coronal T_1-weighted 3-dimensional (3D) gradient echo (3D GRE) acquisition for RAMRIS, CARLOS and RAMRIQ assessments. The DCE-MRI acquisition was based on a coronal 3D spoiled gradient echo acquisition (3D SPGR) sequence. 3D variable flip angle SPGR acquisitions were acquired first to allow baseline T_1-mapping, followed by a 7-minute dynamic 3D SPGR sequence (temporal resolution 9–13 s/volume, depending on scanner set-up), during which Gd contrast agent was administered. Post-contrast fat-suppressed coronal and axial T_1-weighted 3D GRE acquisitions for RAMRIS, CARLOS and RAMRIQ assessments were acquired.

Example imaging pulse sequence parameters are described in the table below. Field of view (FOV) for all coronal scans was set to 13·5 cm (L-R) by 18 cm (H-F) to allow good coverage the wrist and MCP joints, independent of hand size. Total scan time was approximately 28 minutes, with the imaging protocol designed to be limited to <30 minutes to allow for all necessary scans to be acquired while keeping the potential for patient discomfort to a minimum.

During the DCE-MRI acquisition, 0·1 mmol/kg (0·2 mL/kg) Gd contrast agent (e.g. gadobutrol (Gadovist), gadoterate (Dotarem)) was administered to the patient via a power injector. Rate of injection was 3 mls^{-1} and injection was timed to occur at the 6^{th} acquisition of the dynamic series (approx. 1 minute into the series).

Table. Imaging pulse sequence parameters

	2D SATIR	3D GRE+FS	3D DCE-MRI	3D GRE+FS
Scan plane	**Coronal**	**Coronal**	**Coronal**	**Axial**
TE (ms)	30	~4·6	~1·5	~4·6
TR (ms)	3000−5000	30	~4·0	30
TI (ms)	~150 (1·5T); ~220 (3T)	–	–	–
Flip angles (deg)	180	30	2, 6, 17 for T1-mapping; 17 for dynamic series	30
FOV (cm)	18	18	18	18
Phase FOV	0·75	0·75	0·75	~0·45
Slice thickness (mm)	3	1·5	1·5	2·0
Slices	20	40	40	72
Measurements	–	–	6 for T1-mapping; 35-45 for dynamic series	–
Acquisition matrix	384 x 256	512 x 256	180 x 136	256 x 112
Saturation	–	Fat suppression	–	Fat suppression

2D, two dimensional; 3D, three dimensional; DCE, dynamic contrast enhanced; deg, degree; FOV, field of view; FS, fast spin; GRE, gradient echo; MRI, magnetic resonance imaging; ms, millisecond; STIR, short tau inversion recovery; TE, time to echo; TI, inversion time; TR, repetition time.

Image analysis – image quality control

All images were centrally checked after acquisition to identify any technical, operator or subject-related errors. Data not considered analysable (e.g. significant patient motion or incorrect acquisition protocol used) resulted in a quality control failure and a re-scan request for that time point.

As part of a comprehensive quality control strategy for the study, imaging sites were required to periodically scan MRI T_1 phantoms (custom built using gel vials from the Eurospin TO5 Test Object (Diagnostic Sonar, Livingstone, Scotland)) specifically designed to monitor performance of both the scanner and the imaging coil over time. Erroneous phantom T_1 measurements and/or image artefacts indicated MRI or coil specific issues that were subsequently investigated and corrected. Phantom data also allowed for scanner specific T_1 correction of the patient data.

Image analysis – RAMRIS reads

Images of the hand and wrist acquired for RAMRIS and CARLOS reads were analysed centrally by an independent rheumatologist (Charles Peterfy) experienced in scoring using the OMERACT RAMRIS system[1,4] to evaluate CFB in synovitis, oedema/osteitis, and erosions , and CARLOS to evaluate CFB in articular cartilage. The reader was blinded to subject name, examination date, randomisation arm, investigator site identifiers, and time point. All time points for a single subject were read in one reading session, with the time points presented in random order. Images were semi-quantitatively assessed according to the RAMRIS system, assessing synovitis (scored 0–3), bone erosions (scores 0–10), and bone oedema/osteitis (scored 0–3), and according to the CARLOS method, assessing cartilage loss (scored 0–4).

Image analysis – RAMRIQ analysis

RAMRIQ was developed as a fully automated way to evaluate the features assessed in RAMRIS and was used here to attempt to quantify any change in these features with treatment. Active appearance models[5] were previously built from an independent training set from a different cohort of subjects (patients with RA and healthy volunteers) in which each bone and synovial capsule had been manually segmented.[6] Statistical shape modelling[3,6] was then used to automatically identify bones, soft tissues and related anatomical regions of interest in the fat-suppressed coronal 3D GRE images (pre- and post-contrast) for each subject. The model created for each subject was used to extract quantitative values for:

- Synovitis volume – calculated from the difference between the pre- and post-contrast images to identify regions of enhancement in the joint.
- Joint space width – the distance in 3D space between two bones, calculated from the mean of a series of measurements across the joint space.
- Volume of bone erosions – calculated by fitting a model of healthy/non-diseased bone to each bone in the image and calculating the volume of eroded bone.
- Volume of oedema – volume of bone inside each bone shape where a signal intensity higher than normal bone is identified.

The synovial capsule regions identified by the model were used to create a 3D region of interest mask for segmentation in the analysis of the DCE-MRI data.

Image analysis – DCE-MRI analysis

The DCE-MRI dynamic series were first transformed into contrast agent concentration units using the T_1 maps calculated from the baseline variable flip angle 3D SPGR acquisitions, after correcting for flip angle variation determined from site specific T_1 phantom data. Tracer kinetic analysis was performed using the extended Tofts model [7] and a population arterial input function,[8] which was corrected for each patient's haematocrit.[9] This allowed for voxelwise estimates of the volume transfer coefficient of the contrast agent K^{trans}, the extracellular extravascular volume v_e and the blood plasma volume v_p in the hand and wrist.

In addition to the tracer-kinetic analysis, two heuristic parameters were also measured – the initial rate of enhancement (*IRE*) and maximum enhancement (*ME*). Both *IRE* and *ME* were measured using voxelwise Tofts model fits of the contrast agent time course during the dynamic series, which allowed for more robust comparisons between scanning centres.

Image analysis – Summary statistics

Total RAMRIS and CARLOS scores for each feature were calculated for each time point by combining individual scores for each joint. For RAMRIQ analysis at each time point, bone marrow lesions and volume of erosions were both summed across all bones and normalised by the total volume of bone in the anatomy of interest, while volume of synovitis (VEP) was summed across all joints and normalised by the total volume of all joints of interest. RAMRIQ joint space width was calculated for each time point as the sum of the width for all joints.

DCE-MRI parametric maps were summarised from a region of interest based on the VEP. Each parameter was summarised as a median for each joint of interest, non-enhancing voxels and model-fits excluded. A total measure for the affected hand at each time point for each DCE-MRI parameter was calculated by summing across all joints and dividing by the total number of joints.

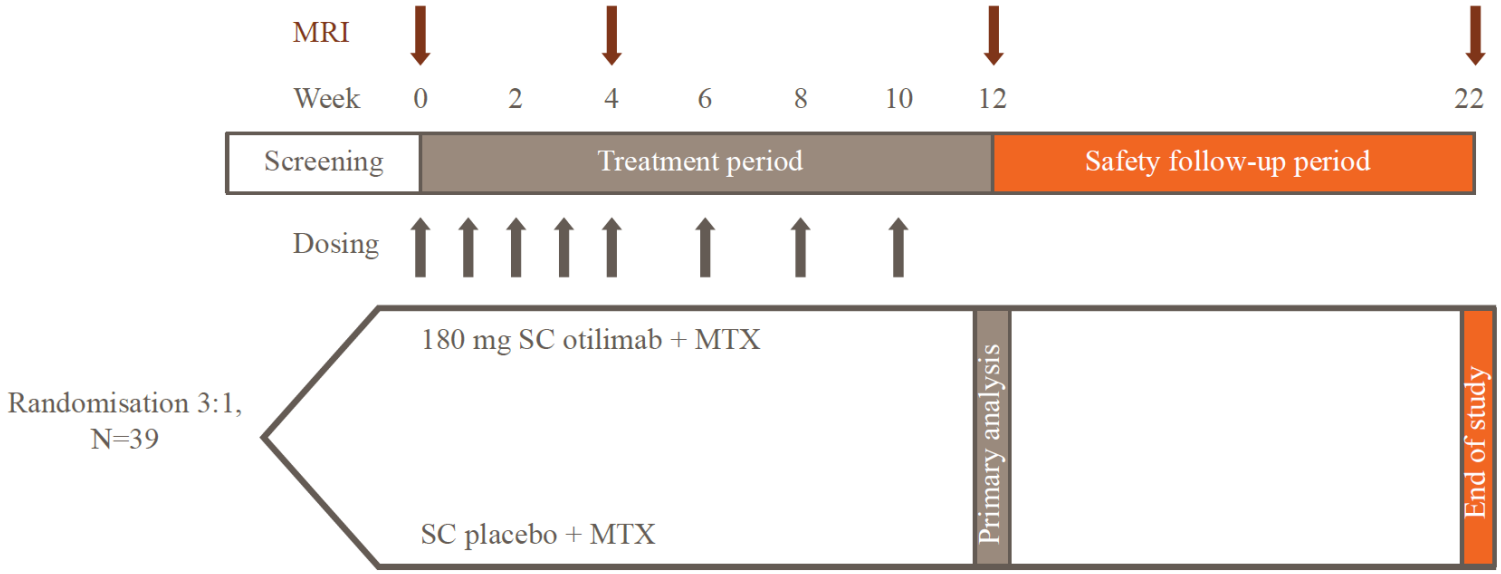
For all imaging endpoints, missing joint scores/results (due to e.g. surgically altered joint or poor signal in joint) were imputed as the mean of the non-missing joint values for that time point. If data for more than 50% of the joints was missing at the time of a given assessment, then the total count was set to missing for that visit.

References

1. Ostergaard M, Edmonds J, McQueen F, et al. An introduction to the EULAR-OMERACT rheumatoid arthritis MRI reference image atlas. *Ann Rheum Dis* 2005; **64 Suppl 1:** i3–7.
2. Peterfy CG, Olech E, DiCarlo JC, Merrill JT, Countryman PJ, Gaylis NB. Monitoring cartilage loss in the hands and wrists in rheumatoid arthritis with magnetic resonance imaging in a multi-center clinical trial: IMPRESS (NCT00425932). *Arthritis Res Ther* 2013; **15:** R44.
3. Conaghan PG, Ostergaard M, Bowes MA, et al. Comparing the effects of tofacitinib, methotrexate and the combination, on bone marrow oedema, synovitis and bone erosion in methotrexate-naive, early active rheumatoid arthritis: results of an exploratory randomised MRI study incorporating semiquantitative and quantitative techniques. *Ann Rheum Dis* 2016; **75:** 1024–33.
4. Ostergaard M, Peterfy CG, Bird P, et al. The OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging (MRI) Scoring System: Updated Recommendations by the OMERACT MRI in Arthritis Working Group. *J Rheumatol* 2017; **44:** 1706–12.
5. Cootes TF, Edwards GJ, CJ T. Active Appearance Models. *IEEE Transitions on Pattern Analysis and Machine Intelligence* 2001; **23:** 681–5.

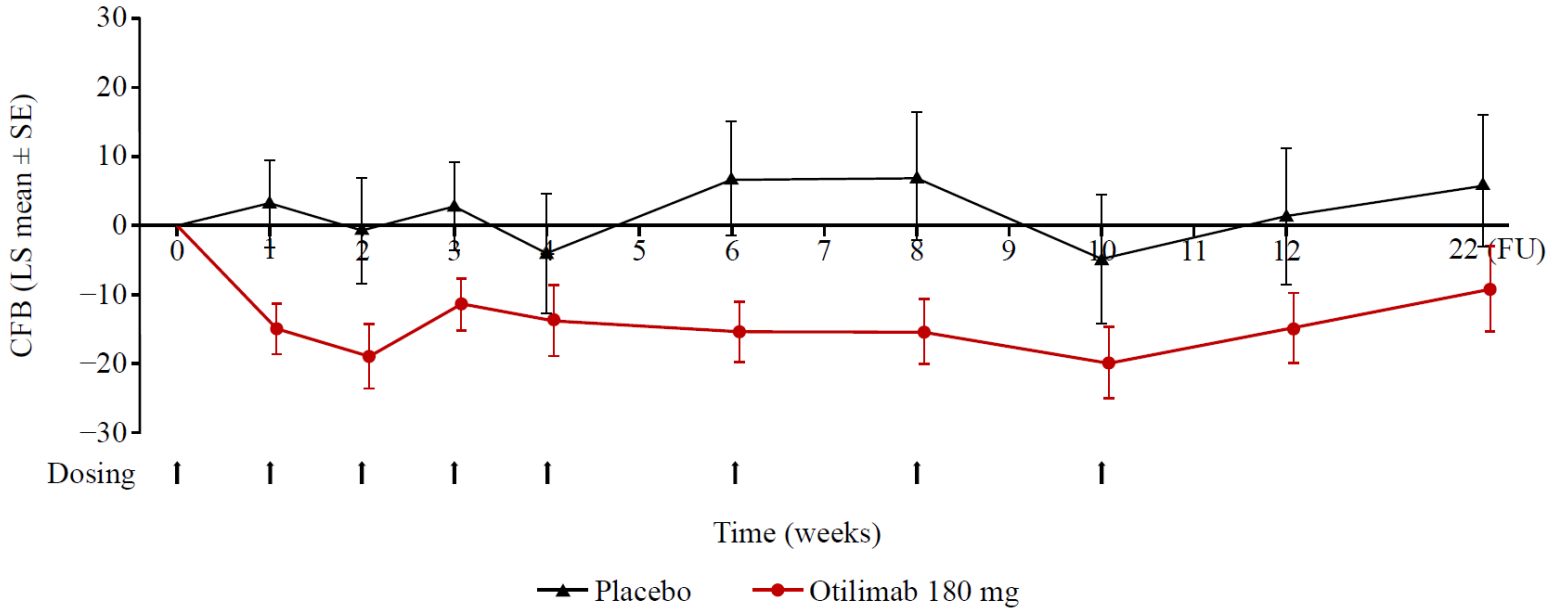
6. Waterton JC, Ho M, Nordenmark LH, et al. Repeatability and response to therapy of dynamic contrast-enhanced magnetic resonance imaging biomarkers in rheumatoid arthritis in a large multicentre trial setting. *Eur Radiol* 2017; **27:** 3662–8.
7. Tofts PS. Modeling tracer kinetics in dynamic Gd-DTPA MR imaging. *J Magn Reson Imaging* 1997; **7:** 91–101.
8. Parker GJ, Roberts C, Macdonald A, et al. Experimentally-derived functional form for a population-averaged high-temporal-resolution arterial input function for dynamic contrast-enhanced MRI. *Magn Reson Med* 2006; **56:** 993–1000.
9. Roberts C, Hughes S, Naish JH, et al. Use of An Individually Measured Hematocrit in DCE-MRI studies. *Proc Intl Soc Mag Reson Med* 2011; **19:** 1078.

Supplementary Figure 1. Study design



MRI, magnetic resonance imaging; MTX, methotrexate; RA, rheumatoid arthritis; SC, subcutaneous.

Supplementary Figure 2. LS mean CFB in pain VAS over time (ITT population)



CFB, change from baseline; CI, confidence interval; FU, follow-up; ITT, intent-to-treat; LS, least squares; VAS, visual analogue scale.

Supplementary Table 1. Biomarkers assessed during the study

Target Engagement	Free, soluble GM-CSF* GM-CSF: drug complex	
Putative biomarkers to assess and/or predict response to otilimab	CCL17 (TARC) CCL22 (MDC) TNFα IL-1β* IL-6* IL-8 IL-10* IL-15*	IL-17A* IL-17F* MMP-3 CXCL13 (BLC) sCD163 MRP8/14 complex 14-3-3η* YKL-40
Complement Biomarkers	C3 C4 C4a C5a TCC	
Extracellular matrix or aggrecan degradation biomarkers	ARGS Neo-epitope C1M C2M	C3M CRPM VICM
Flow Cytometry: 6 Color TBNK Panel	%CD3+ Absolute CD3+ %CD3+CD8+ Absolute CD3+CD8+ %CD3+CD4+ Absolute CD3+CD4+	%CD16+CD56+ Absolute CD16+CD56+ %CD19+ Absolute CD19+ cells Absolute lymphocytes Helper/Suppressor
Flow Cytometry: CD16+ Monocyte Panel	CD1br+CD16- / % of WBC CD14br+CD16- / % of monocytes CD14br+CD16- / Abs in cells/mL CD14br+CD16- / events CD14br+CD16-HLA-DR+CD200R1+ / % of classical monocytes CD14br+CD16-HLA-DR+CD200R1+ / events CD14lo+CD16br+ / % of WBC CD14lo+CD16br+ / % of monocytes CD14lo+CD16br+ / Abs in cells/mL CD14lo+CD16br+ / events CD14lo+CD16br+HLA-DR+CD200R1+ / % of non-classical monocytes CD14lo+CD16br+HLA-DR+CD200R1+ / events CD14br+CD16+ / % of WBC CD14br+CD16+ / % of monocytes CD14br+CD16+ / Abs in cells/mL CD14br+CD16+ / events CD14br+CD16+HLA-DR+CD200R1+ / % of Int monocytes CD14br+CD16+HLA-DR+CD200R1+ / events CD14-HLA-DR+CD11c-CD123br+ / % of monocytes CD14-HLA-DR+CD11c-CD123br+ / Abs in cells/mL CD14-HLA-DR+CD11c-CD123br+ / events CD14-HLA-DR+CD11c-CD123br+CD200R1+ / % of pDC CD14-HLA-DR+CD11c-CD123br+CD200R1+ / events CD14-HLA-DR+CD11c br+CD123+ / % of monocytes CD14-HLA-DR+CD11c br+CD123+ / Abs in cells/mL CD14-HLA-DR+CD11c br+CD123+ / events CD14-HLA-DR+CD11c br+CD123+CD200R1+ / % of mDC1 CD14-HLA-DR+CD11c br+CD123+CD200R1+ / events CD14-HLA-DR+CD11c br+CD123- / % of monocytes CD14-HLA-DR+CD11c br+CD123- / Abs in cells/mL CD14-HLA-DR+CD11c br+CD123- / events CD14-HLA-DR+CD11c br+CD123-CD200R1+ / % of mDC2 CD14-HLA-DR+CD11c br+CD123-CD200R1+ / events CD14-CD16+CD66b+ / % of WBC CD14-CD16+CD66b+ / per CMM CD14-CD16+CD66b+ / events CD14-CD16+CD66b+CD200R1+ / % of neutrophils CD14-CD16+CD66b+CD200R1+ / events	
Flow Cytometry: T Regulatory Cell FoxP3	%CD3+ %CD3+CD4+ %CD3+4+25+127- %CD3+4+FoxP3+25+127- %CD3+CD8+	CD3+A CD3+4+A CD3+4+25+127-A CD3+4+FoxP3+25+127-A CD3+8+A

T Helper Cell Panel	CD45+CD3+CD8-CD4+ / % of total helper T cells
	CD45+CD3+CD8-CD4+ / events total helper T cells
	CD45+CD3+CD8-CD4+CCR6+CXCR3-(Th17) / % of total helper T cells
	CD45+CD3+CD8-CD4+CCR6+CXCR3-(Th17) / events total helper T cells
	CD45+CD3+CD8-CD4+CCR6+CXCR3-CD38+HLA-DR+ / % of Th17 cells
	CD45+CD3+CD8-CD4+CCR6+CXCR3-CD38+HLA-DR+ / Events Th17 cells
	CD45+CD3+CD8-CD4+CCR6-CXCR3-(Th2) / % of total helper T cells
	CD45+CD3+CD8-CD4+CCR6-CXCR3-(Th2) / events total helper T cells
	CD45+CD3+CD8-CD4+CCR6-CXCR3-CD38+HLA-DR+ / % of Th2 cells
	CD45+CD3+CD8-CD4+CCR6-CXCR3-CD38+HLA-DR+ / events Th2 cells
	CD45+CD3+CD8-CD4+CCR6-CXCR3+(Th1) / % of Total helper T cells
	CD45+CD3+CD8-CD4+CCR6-CXCR3+(Th1) / events total helper T cells
	CD45+CD3+CD8-CD4+CCR6-CXCR3+CD38+HLA-DR+ / % of Th1 cells
	CD45+CD3+CD8-CD4+CCR6-CXCR3+CD38+HLA-DR+ / events Th1 cells
	CD45+CD3+CD8-CD4+CCR6+CXCR3+(Th proinflam) / % of total helper T cells
	CD45+CD3+CD8-CD4+CCR6+CXCR3+(Th proinflam) / events total helper T cells
	CD45+CD3+CD8-CD4+CCR6+CXCR3+CD38+HLA-DR+ / % of Th proinflammatory cells
	CD45+CD3+CD8-CD4+CCR6+CXCR3+CD38+HLA-DR+ / events proinflammatory cells
Safety Biomarkers	KL-6 antigen
	Surfactant protein D
	Amyloid A
	Anti-GM-CSF auto-antibodies
	3-beta-cholestenoic acid

*Biomarker interpretation compromised by assay sensitivity.

GM-CSF, granulocyte–macrophage colony-stimulating factor; KL-6, Krebs von den Lungen-6.

Supplementary Table 2. LS mean ratio to baseline in all assessed biomarkers at week 12.

	Placebo (n=11)	Otilimab 180 mg (n=28)	Ratio to placebo (95% CI)	p value (ratio to baseline)
Target engagement biomarkers, LS mean (%CV) ratio to baseline				
GM-CSF:drug complex ng/L	0·970 (44·95)	22·556 (24·38)	23·249 (8·579, 63·005)	<0·001
Putative predictive biomarkers, LS mean (%CV) ratio to baseline				
14-3-3 ETA protein, mg/L	1·131 (20·90)	0·897 (11·89)	0·793 (0·488, 1·290)	0·338
Amyloid A, ng/L	0·845 (49·31)	0·653 (23·45)	0·774 (0·261, 2·290)	0·632
CCL17, ng/L	1·711 (24·96)	0·890 (13·60)	0·520 (0·294, 0·922)	0·026
CXCL13, ng/L	0·924 (31·17)	1·077 (16·31)	1·165 (0·573, 2·369)	0·661
Chitinase 3 like 1, µg/L	1·066 (21·53)	1·071 (11·64)	1·005 (0·613, 1·645)	0·985
IL-6, ng/L	0·770 (21·89)	0·989 (12·35)	1·221 (0·734, 2·031)	0·432
Macrophage-derived chemokine, ng/L	1·245 (9·51)	1·058 (5·23)	0·849 (0·681, 1·059)	0·142
MMP-3, µg/L	1·000 (13·57)	0·951 (7·35)	0·951 (0·695, 1·301)	0·745
S100 CBP A8 and A9, mg/L	0·694 (21·61)	0·879 (12·05)	1·267 (0·769, 2·086)	0·342
Cartilage biomarkers, LS mean (%CV) ratio to baseline				
ARGS neo-epitope, µg/L	0·913 (13·97)	1·156 (7·78)	1·266 (0·916, 1·750)	0·147
Citrullinated MMP-Degraded Vimentin, µg/L	1·123 (29·34)	0·917 (16·00)	0·816 (0·417, 1·597)	0·541
MMP-degraded CRP, µg/L	1·006 (7·13)	1·108 (3·95)	1·102 (0·934, 1·300)	0·242
MMP-degraded Type I collagen, µg/L	1·148 (13·74)	1·023 (7·81)	0·891 (0·647, 1·228)	0·470
MMP-degraded Type II collagen, µg/L	1·174 (10·73)	1·072 (5·87)	0·913 (0·711, 1·173)	0·467
MMP-degraded Type III collagen, µg/L	0·929 (9·28)	0·910 (5·08)	0·979 (0·790, 1·214)	0·843
Flow cytometry 6 colour TBNK, LS mean (%CV) ratio to baseline				
CD16+CD56+, 10^9/L	1·193 (15·49)	0·911 (7·81)	0·763 (0·536, 1·087)	0·129
CD19, 10^9/L	1·054 (12·26)	1·003 (6·51)	0·952 (0·718, 1·262)	0·724

CD3, 10^9/L	0·912 (9·01)	0·992 (4·60)	1·088 (0·885, 1·336)	0·413
CD3+CD4+, 10^9/L	0·952 (9·69)	1·013 (4·94)	1·064 (0·853, 1·328)	0·573
CD3+CD8+, 10^9/L	–0·056 (0·0418)	–0·037 (0·0215)	0·020 (–0·076, 0·155)	0·677
Helper/Suppressor	1·040 (6·83)	1·088 (3·49)	1·046 (0·895, 1·222)	0·565
T cell, B cell, natural killer lymphocytes, 10^9/L	–0·093 (0·1792)	–0·006 (0·0925)	0·087 (–0·323, –0·498)	0·668
Flow Cytometry: T Regulatory Cell FoxP3, LS mean (%CV) ratio to baseline				
CD3+CD4+, 10^6/L	0·981 (10·55)	1·004 (5·31)	1·024 (0·804, 1·303)	0·844
CD3+CD8+, 10^6/L	0·878 (10·50)	0·945 (5·40)	1·076 (0·846, 1·370)	0·537
CD3+, 10^6/L	0·959 (9·62)	0·993 (4·90)	1·036 (0·831, 1·291)	0·748
Flow Cytometry: CD16+ Monocyte Panel, LS mean (SE) CFB				
CD14-CD16+CD66b+, 10^6/L	–125·9 (460·53)	–378·6 (229·47)	–252·6 (–1326·0, 820·8)	0·629
CD14-HLA-DR+CD11cbr+CD123-, 10^3/L	5730·4 (5085·86)	5453·0 (2399·11)	–277·4 (–12001·0, 11446·2)	0·961
CD14br+CD16+, 10^3/L	25351·6 (9433·73)	2668·0 (4589·23)	–22683·6 (–44298·5, –1068·8)	0·040
CD14br+CD16-, 10^3/L	–2981·8 (3778·70)	13201·2 (18684·22)	16182·9 (–70377·0, 102742·9)	0·704
CD14lo+CD16br+, 10^3/L	10938·4 (5088·36)	759·2 (2534·27)	–10179·1 (–21827·4, 1469·2	0·084

Note: Some of the biomarkers listing in **Supplementary Table 1** and on ClinicalTrials.gov were not assessed or did not have sufficient quantifiable data, so are not included in this table.

%CV, coefficient of variation; CFB, change from baseline; LS, least squares; SE, standard error.

Supplementary Table 3. LS mean CFB to weeks 4, 12 and 22 in MRI measures (ITT population)

CFB, LS mean (SE)	Placebo (n=11)	Otilimab 180 mg (n=28)	Difference from placebo (95% CI)	p value
RAMRIS synovitis score				
Week 4	0·0 (0·95)	−0·1 (0·59)	−0·1 (−2·5, 2·3)	0·932
Week 12	0·8 (1·17)	−1·3 (0·60)	−2·2 (−4·9, 0·5)	0·112
Week 22 (follow-up)	1·1 (1·17)	−1·1 (0·66)	−2·3 (−5·0, 0·5)	0·104
RAMRIS bone erosion score				
Week 4	0·2 (0·49)	0·3 (0·30)	0·0 (−1·1, 1·2)	0·945
Week 12	0·8 (0·50)	0·4 (0·26)	−0·4 (−1·5, 0·7)	0·475
Week 22 (follow-up)	1·5 (0·47)	0·5 (0·27)	−0·9 (−2·0, 0·1)	0·086
RAMRIS bone oedema/osteitis score				
Week 4	−0·1 (0·11)	−0·1 (0·06)	0·0 (−0·2, 0·3)	0·940
Week 12	0·0 (1·04)	−0·8 (0·56)	−0·8 (−3·2, 1·6)	0·521
Week 22 (follow-up)	0·5 (1·33)	−0·9 (0·74)	−1·3 (−4·4, 1·8)	0·396
CARLOS cartilage loss score				
Week 4	−0·09 (0·145)	−0·01 (0·105)	0·09 (−0·27, 0·45)	0·635
Week 12	−0·05 (0·163)	−0·07 (0·078)	−0·02 (−0·38, 0·34)	0·902
Week 22 (follow-up)	−0·08 (0·134)	0·02 (0·082)	0·10 (−0·22, 0·41)	0·531
RAMRIQ synovitis (µL)				
Week 4	34·1 (1052·52)	245·5 (776·24)	211·4 (−2589·2, 3012·0)	0·874
Week 12	−912·3 (1405·77)	−1417·0 (671·54)	−504·8 (−3730·4, 2720·9)	0·749
Week 22 (follow-up)	364·0 (1372·20)	−1172·1 (844·13)	−1536·0 (−4884·2, 1812·1)	0·352
RAMRIQ erosion damage (normalized)				
Week 4	0·0007 (0·0012)	0·0006 (0·0008)	−0·0002 (−0·0033, 0·0029)	0·915
Week 12	0·0003 (0·0011)	−0·0000 (0·0006)	−0·0004 (−0·0028, 0·0021)	0·771
Week 22 (follow-up)	−0·0002 (0·0023)	0·0003 (0·0013)	0·0005 (−0·0048, 0·0058)	0·850
RAMRIQ joint-space loss (mm)				
Week 4	0·527 (0·5387)	−0·536 (0·4030)	−1·063 (−2·505, 0·380)	0·138
Week 12	0·405 (0·4696)	0·345 (0·2211)	−0·060 (−1·131, 1·011)	0·909
Week 22 (follow-up)	0·677 (0·4920)	0·503 (0·3011)	−0·174 (−1·363, 1·015)	0·765
RAMRIQ bone marrow lesions (normalized)				
Week 4	−0·0045 (0·0102)	0·0084 (0·0057)	0·0128 (−0·0123, 0·0380)	0·291
Week 12	−0·0045 (0·0035)	−0·0009 (0·0019)	0·0036 (−0·0046, 0·0118)	0·375
Week 22 (follow-up)	−0·0038 (0·0016)	−0·0027 (0·0009)	0·0010 (−0·0028, 0·0049)	0·588
DCE-MRI K^{trans} (min^{-1})				
Week 4	0·0040 (0·0047)	0·0030 (0·0034)	−0·0010 (−0·0135, 0·0114)	0·862
Week 12	0·0081 (0·0055)	−0·0059 (0·0027)	−0·0140 (−0·0266, −0·0013)	0·031
Week 22 (follow-up)	0·0072 (0·0048)	−0·0047 (0·0029)	−0·0119 (−0·0234, −0·0004)	0·044
DCE-MRI extravascular extracellular space volume (v_e)				
Week 4	0·2088 (0·0955)	0·0756 (0·0453)	−0·1332 (−0·3636, 0·0972)	0·232
Week 12	–0·0216 (0·0691)	0·0427 (0·0377)	0·0643 (–0·1022, 0·2308)	0·428
Week 22 (follow-up)	0·0683 (0·0482)	−0·0423 (0·0303)	−0·1107 (−0·2319, 0·0106)	0·071
DCE-MRI fractional volume of blood plasma (v_p)				
Week 4	−0·0001 (0·0019)	0·0007 (0·0013)	0·0008 (−0·0040, 0·0056)	0·732

Week 12	0·0044 (0·0021)	0·0002 (0·0011)	−0·0042 (−0·0092, 0·0007)	0·092
Week 22 (follow-up)	0·0009 (0·0018)	−0·0013 (0·0011)	−0·0022 (−0·0066, 0·0021)	0·296
DCE-MRI *IRE* (μMs^{-1})				
Week 4	0·0001 (0·0002)	0·0001 (0·0001)	−0·0000 (−0·0006, 0·0005)	0·884
Week 12	0·0004 (0·0003)	−0·0002 (0·0001)	−0·0006 (−0·0012, 0·0000)	0·051
Week 22 (follow-up)	0·0003 (0·0002)	−0·0002 (0·0001)	−0·0005 (−0·0010, −0·0000)	0·042
DCE-MRI *ME* (mM)				
Week 4	0·0277 (0·0252)	0·0106 (0·0183)	−0·0171 (−0·0829, 0·0487)	0·591
Week 12	0·0357 (0·0282)	−0·0291 (0·0141)	−0·0648 (−0·1297, 0·0001)	0·050
Week 22 (follow-up)	0·0304 (0·0208)	−0·0226 (0·0122)	−0·0530 (−0·1030, −0·0031)	0·038

CARLOS, cartilage loss score; CFB, change from baseline; DCE-MRI, dynamic contrast enhanced magnetic resonance imaging; FU, follow-up; IRE, initial rate of enhancement; ITT, intent-to-treat; K^{trans}, rate of transfer of contrast agent from blood plasma to extracellular extravascular space; ME, maximal enhancement; LS, least squares; MRI, magnetic resonance imaging; RAMRIQ, rheumatoid arthritis MRI quantification; RAMRIS, rheumatoid arthritis MRI scoring system; SE, standard error.

Supplementary Table 4. Serum concentration (ng/mL) of otilimab (active arm only)

Visit	n	Geometric mean (95% CI)	CVb (%)	Median (min, max)
Baseline	28	20·0	–	20·0 (20, 20)
Week 1	27	1029·9 (612·8, 1730·8)	214·4	1320·0 (20, 5010)
Week 2	26	2162·0 (1746·8, 2675·9)	56·7	2335·0 (498, 4260)
Week 4	27	2789·6 (2203·6, 3531·4)	65·3	2710·0 (626, 5710)
Week 6	25	1396·1 (1099·1, 1773·3)	63·2	1420·0 (376, 3090)
Week 8	24	1105·1 (801·4, 1523·9)	88·6	1285·0 (76, 2990)
Week 12	24	990·9 (722·8, 1358·6)	86·5	1050·0 (251, 3400)
Week 22 (follow-up)	21	26·5 (19·6, 35·6)	73·1	20·0 (20, 235)

Baseline value is the latest pre-dose assessment. Values below the limit of quantification have been imputed by 20 ng/mL
CI, confidence interval; CVb, coefficient of biological variation; Max, maximum, Min, minimum.

Glossary of specialist terms

C1M: MMP-degraded Type I collagen,

CCL17: chemokine (C-C motif) ligand 17, (also called thymus and activation-regulated chemokine [TARC])

DCE-MRI: dynamic contrast-enhanced magnetic resonance imaging, measures T1 changes in tissues over time after administration of gadolinium

IRE: initial rate (gradient) of enhancement following administration of gadolinium contrast agents over 60 seconds post-arrival in tissue

K^{trans}: rate of transfer of contrast agent from blood plasma to extracellular extravascular space; K^{trans} repeatability is an indicator of reproducibility

LLoQ: lower limit of quantification

LS mean: least squares mean, a mean value for a group that is adjusted for the means of other factors in the model.

ME: maximum enhancement of the gadolinium contrast agent concentration curve during DCE-MRI series

PAP: pulmonary alveolar proteinosis, a rare lung disorder. As the presence of anti-GM-CSF antibodies has been observed in adult patients with PAP, and neutralisation of GM-CSF is hypothesised as a mechanism for PAP pathogenesis, this study included pulmonary safety monitoring