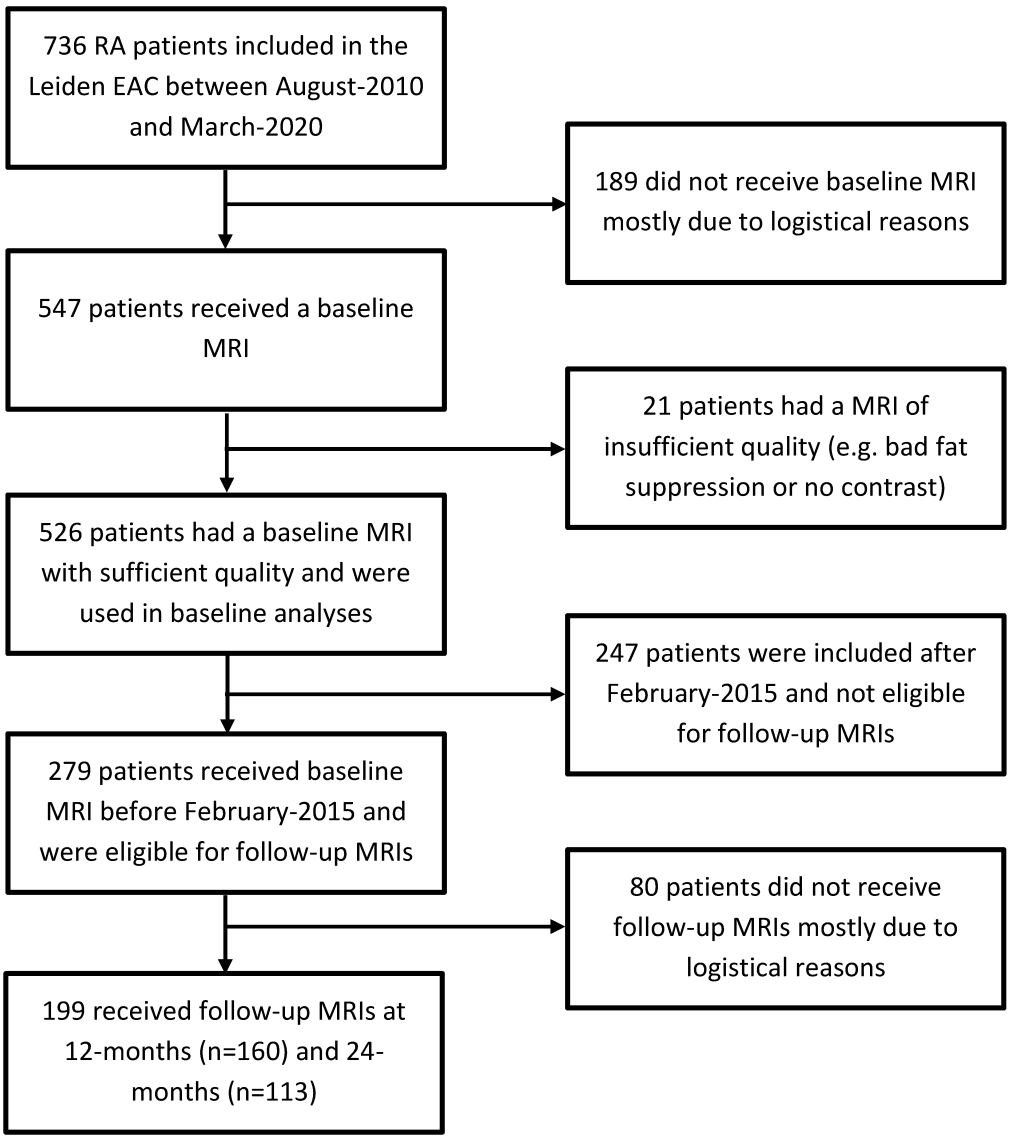
Supplementary File 1: Flow diagram of study inclusion



Supplementary File 2: Baseline characteristics of RA patients included between August-2010 and March-2020 with or without baseline MRI or baseline MRI of insufficient quality

	Patients with baseline MRI (n=526)	**Patients without baseline MRI (n=189)**	**Patients with baseline MRI of insufficient quality (n=21)**	**P-value**
Women, n (%)	339 (64)	118 (62)	13 (62)	0.87
Age in years, mean (SD)	59 (14)	59 (16)	57 (17)	0.77
Symptom duration, weeks median (IQR)	12 (6-28)	16 (8-37)	12 (5-52)	0.07
ACPA, n (%)	226 (45)	73 (42)	11 (52)	0.55
66-SJC, median (IQR)	5 (2-10)	6 (2-10)	4 (2-10)	0.76
68-TJC, median (IQR)	5 (3-7)	5 (3-9)	6 (3-9)	0.02
ESR, median (IQR)	28 (11-41)	29 (11-45)	17 (13-35)	0.62
NRS fatigue, median (IQR)	6 (2-7)	6 (3-8)	6 (4-7.5)	0.23
HAQ, median (IQR)	0.9 (0.5-1.5)	1.0 (0.5-1.5)	0.9 (0.4-1.4)	0.90

Legend: MRI, magnetic resonance imaging; n, number of patients; SD, standard deviation; IQR, inter quartile range; ACPA: anti-citrullinated protein antibodies; SJC, swollen joint count; TJC, tender joint count; ESR, erythrocyte sedimentation rate; NRS, numeric rating scale; HAQ, health assessment questionnaire

Supplementary File 3: Baseline characteristics of RA patients included between August-2010 and February-2015 with baseline MRI with or without follow-up MRIs

	Patients with follow-up MRI (n=199)	**Patients without follow-up MRI (n=80)**	**P-value**
Women, n (%)	130 (65)	60 (75)	0.12
Age in years, mean (SD)	56 (14)	58 (14)	0.44
Symptom duration, weeks median (IQR)	13 (7-29)	15 (7-33)	0.52
ACPA, n (%)	101 (51)	29 (36)	0.03
66-SJC, median (IQR)	5 (2-10)	5 (2-11)	0.96
68-TJC, median (IQR)	5 (3-7)	5 (3-8)	0.83
ESR, median (IQR)	25 (11-38)	22 (10-35)	0.40
NRS fatigue, median (IQR)	6 (2-7)	6 (3.3-7.3)	0.39
HAQ, median (IQR)	0.9 (0.5-1.5)	0.9 (0.6-1.3)	0.79
Total tenosynovitis, median (IQR)	4 (1-7)	4 (2-8)	0.28
Total synovitis, median (IQR)	5 (3-9)	6 (3-10)	0.21
Total osteitis, median (IQR)	4 (1-7)	3 (1-8)	0.73

Legend: MRI, magnetic resonance imaging; n, number of patients; SD, standard deviation; IQR, inter quartile range; ACPA: anti-citrullinated protein antibodies; SJC, swollen joint count; TJC, tender joint count; ESR, erythrocyte sedimentation rate; NRS, numeric rating scale; HAQ, health assessment questionnaire

Supplementary File 4: MRI scanning and scoring

MRI scanning

MRI was performed on a musculoskeletal 1.5T MRI system (GE, Wisconsin, USA) using a 145mm coil for the foot and a 100mm coil for the hand. The patient was positioned in a chair beside the scanner, with the hand or foot fixed in the coil with cushions.

In the hand (metacarpophalangeal (MCP)2-5 and wrist) the following sequence was acquired before contrast administration: T1-weighted fast spin-echo (FSE) sequence in the coronal plane (repetition time (TR) 575 ms, echo time (TE) 11.2 ms, acquisition matrix 388×288, echo train length (ETL) 2). After intravenous injection of gadolinium contrast (gadoteric acid, Guerbet, Paris, France, standard dose of 0.1 mmol/kg) the following sequences were obtained: T1-weighted FSE sequence with frequency selective fat saturation (fatsat) in the coronal plane (TR/TE 700/9.7ms, acquisition matrix 364×224, ETL 2), T1-weighted FSE fatsat sequence in the axial plane (wrist: TR/TE 540/7.7 ms; acquisition matrix 320x192; ETL 2 and MCP-joints: TR/TE 570/7.7 ms; acquisition matrix 320x192; ETL 2). No major adverse reactions to gadolinium were reported.

The obtained sequences of the forefoot (metatarsophalangeal (MTP)1-5 joints) were for the first 157 patients before contrast administration at baseline: T1-weighted FSE sequence in the axial plane (TR/TE 650/17ms; acquisition matrix 388x288, ETL 2); and T2-weighted FSE fatsat sequence in the axial plane (TR/TE 3000/61.8; acquisition matrix 300x224, ETL 7). Imaging of the foot was initially limited to pre-contrast axial sequences. For the other patients post-contrast sequences were included: T1-weighted FSE fatsat sequence in the axial plane (TR/TE 700/9.5ms; acquisition matrix 364x224, ETL 2) and: T1-weighted FSE fatsat sequence in the coronal plane (perpendicular to the axis of the metatarsals) (TR/TE

540/7.5ms; acquisition matrix 320x192, ETL 2). Field-of-view was 100mm for the hand and 140mm for the foot. Coronal sequences of the hand had 18 slices with a slice thickness of 2mm and a slice gap of 0.2mm. Coronal sequences of the foot had 20 slices with a slice thickness of 3mm and a slice gap of 0.3mm. All axial sequences had a slice thickness of 3mm and a slice gap of 0.3mm with 20 slices for the wrist, 16 for the MCP-joints and 14 for the foot.

We used the contrast enhanced T1-weighted fat suppressed sequence to assess osteitis in the MCP-joints of all patients. In the MTP-joints osteitis was assessed on T2-weighted fatsat sequences in the first 157 patients and on the contrast enhanced T1-weighted fat suppressed sequence in the latter 369 patients. According to the RAMRIS-method, T2-weighted fat suppressed sequences, or when this sequence is not available a short tau inversion recovery (STIR) sequence, should be used to assess osteitis. However, three previous studies have demonstrated that a contrast enhanced T1-weigthed fat suppressed sequence has a strong correlation with T2-weighted fat suppressed sequences.[1-3] Furthermore, the arthritis subcommittee of the European Society of Musculoskeletal Radiology (ESSR) also recommends the use of contrast enhanced T1-weighted fat suppressed sequences for depicting osteitis.[4] The T2-weighted image shows increased water signal and a contrast-enhanced T1-weighted sequence shows increased water content and the increased perfusion and interstitial leakage. A strong correlation has been shown in arthritis patients and in patients without inflammatory diseases such as bone bruises, intraosseous ganglions, bone infarcts and even nonspecific cases.[2,3] Based on these results osteitis was assessed on contrast enhanced T1-weighted fat suppressed sequences as it has a higher signal to noise ratio and allowed a shorter scan time for

patients. In addition, because T2-weighted fat suppressed sequences could be omitted, coronal sequences of the foot could be added. In total this resulted in a shorter total scan time and more information.

MRI scoring

All MRIs were scored for synovitis, tenosynovitis and osteitis. Synovitis was scored according to the Outcome Measures in Rheumatology Clinical Trials (OMERACT) rheumatoid arthritis MRI-scoring system (RAMRIS), which was adopted to also include the MTP's as well.[5] Tenosynovitis in the wrist and MCPs was scored as described by Haavaardsholm et al.[6]

Synovitis was assessed semi-quantitatively with a range of 0-3 based on the volume of enhancing tissue in the synovial compartment (none, mild, moderate, severe) in 12 joints: MCP 2-5, MTP 1-5 and in three regions of the wrist. Synovitis scores were summed for every patient. For missing synovitis scores, a 0 score was imputed.

Tenosynovitis was scored semi-quantitatively in 18 tendon-sheaths, ranging 0-3 based on the thickness of peritendinous effusion or synovial proliferation with contrast enhancement (normal, <2mm, 2-5mm, >5mm): in the wrist (10 tendons) and in the MCPs (8 tendons), separately. For missing tenosynovitis scores, a 0 score was imputed.

Osteitis was scored semi-quantitatively in 33 bones on a scale 0-3 based on the affected volume of the bone (no osteitis, >0-33%, >33-66%, >66%) in the proximal and distal MCP2-5 (8 bones) and proximal and distal MTP1-5 (10 bones) separately and in 15 bones in the wrist.

Synovitis, tenosynovitis and osteitis scores were averaged between two readers if appropriate and summed into a total inflammation score.

Supplementary File 5: Interreader reliability at baseline

Baseline MRI scans were scored by three pairs of readers. Average measures intraclass correlation coefficients (ICC) per inflammatory feature and for the summed total inflammation score are shown in the table below.

	Reader 1 vs Reader 2 (average measures)	Reader 3 vs Reader 4 (average measures)	Reader 5 vs Reader 6 (average measures)
Number of MRIs	598	229	399
Osteitis	0.86	0.91	0.96
Synovitis	0.93	0.96	0.95
Tenosynovitis	0.95	0.96	0.97
Total inflammation	0.95	0.97	0.98

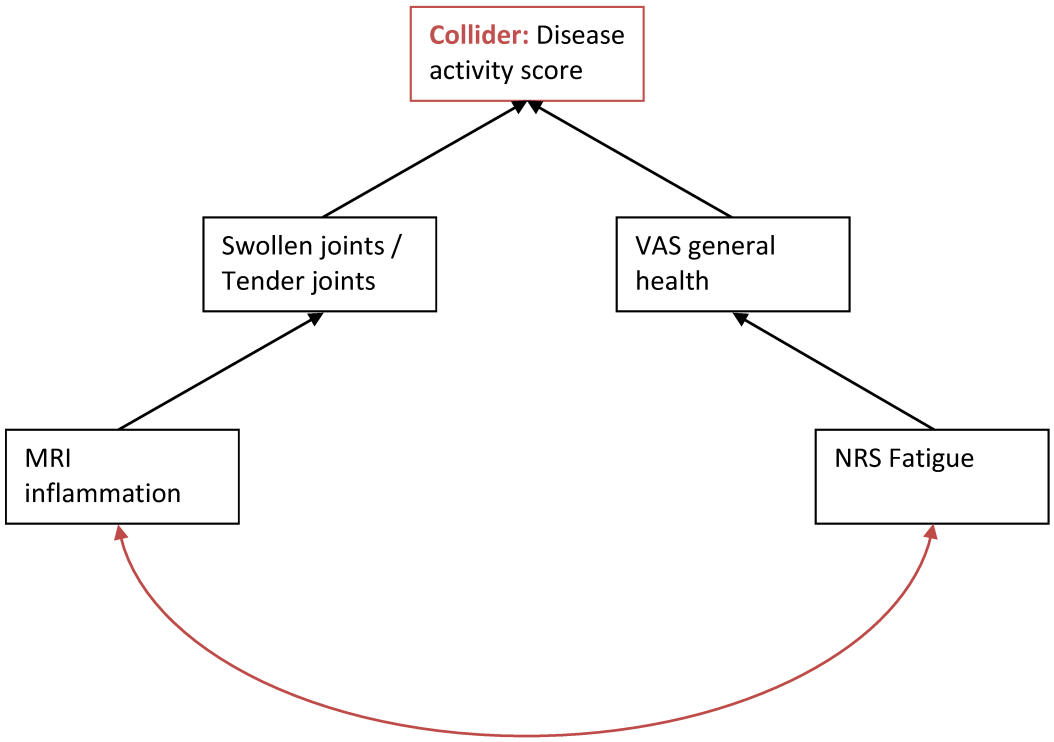
Supplementary File 6: Intrareader reliability over time

All MRI scans over time of 15 randomly selected patients were scored by the same reader after an interval of ≥2 months. Intrareader reliability was assessed at baseline and for the change between baseline and 1 year. Single measures intraclass correlation coefficients (ICC) per inflammatory feature and for the summed total inflammation score are shown in the table below.

	ICC baseline (single measures)	ICC delta baseline – 1 year (single measures)
Osteitis	0.95	0.92
Synovitis	0.95	0.96
Tenosynovitis	0.96	0.95
Total inflammation	0.98	0.97

Supplementary File 7: Explanation of collider effect

MRI inflammation is strongly associated with local clinically detectable inflammation in the form of swollen or tender joints which are included in the disease activity score.[7] In addition, the NRS fatigue is associated with the visual analogue scale (VAS) general health as included in the DAS.[8] Since increase in both fatigue and MRI inflammation is associated with increase in the DAS via different pathways, correction for disease activity distorts the association between MRI inflammation and fatigue. This is called a collider effect. This is visually depicted below.



Supplementary File 8: Association of individual DAS components with fatigue at baseline and over time

Univariable	Baseline	P-value	Over time	P-value
ESR	0.00 (-0.01;0.01)	0.47	**0.03 (0.01;0.04)**	**<0.001**
VAS general health	**0.05 (0.04;0.06)**	**<0.001**	**0.06 (0.05;0.07)**	**<0.001**
Swollen joints	0.04 (0.00;0.09)	0.064	**0.11 (0.06;0.16)**	**<0.001**
Tender joints	**0.13 (0.07;0.18)**	**<0.001**	**0.14 (0.09;0.19)**	**<0.001**
Multivariable	Baseline	P-value	Over time	P-value
ESR	0.00 (-0.01;0.01)	0.62	0.00 (-0.02;0.01)	0.89
VAS general health	**0.05 (0.04;0.06)**	**<0.001**	**0.06 (0.05;0.06)**	**<0.001**
Swollen joints	0.00 (-0.04;0.05)	0.89	-0.01 (-0.07;0.05)	0.77
Tender joints	**0.07 (0.01;0.12)**	**0.03**	**0.08 (0.02;0.13)**	**0.007**

Legend: Estimates and 95% confidence intervals are shown. Significant estimates are shown in bold. Multivariable analyses include the four individual components. DAS: disease activity score; VAS: visual analogue scale; ESR: erythrocyte sedimentation rate

Supplementary File 9: Analyses of association of MRI inflammation and DAS with fatigue, stratified for ACPA

Baseline	**ACPA-positive RA**	**P-value**	**ACPA-negative RA**	**P-value**
MRI inflammation <-> fatigue*	0.20 (-0.36;0.74)	0.49	-0.19 (-0.65;0.27)	0.42
DAS <-> fatigue*	**1.57 (0.98;2.17)**	**<0.001**	**0.94 (0.46;1.41)**	**<0.001**
Longitudinal				
MRI inflammation <-> fatigue*	-0.03 (-0.4;0.35)	0.89	0.01 (-0.44;0.45)	0.97
DAS <-> fatigue*	**1.14 (0.76;1.53)**	**<0.001**	**0.93 (0.61;1.25)**	**<0.001**
Subsequent change**				
MRI inflammation -> fatigue	**0.22 (0.01;0.44)**	**0.04**	0.15 (-0.07;0.36)	0.18
Fatigue -> MRI inflammation	-0.02 (-0.25;0.21)	0.84	0.10 (-0.16;0.36)	0.45
DAS -> fatigue	**0.29 (0.09;0.49)**	**0.004**	0.09 (-0.13;0.31)	0.42
Fatigue -> DAS	0.07 (-0.19;0.33)	0.61	0.16 (-0.05;0.38)	0.13

Legend: * Estimate (95% confidence interval) corrected for age and gender**. ****Standardized regression coefficients of change of one inflammatory feature to subsequent change in another inflammatory feature, corrected for the simultaneous pattern and previous values of those inflammatory features, with 95% confidence intervals. bold: significant estimate ($p<0.05$). RA: rheumatoid arthritis; DAS: disease activity score; ACPA: anti-citrullinated protein antibodies.

Supplementary File 10: Analyses of association of MRI inflammation and DAS with fatigue, with two questions from the SF-36 as outcome

Baseline	**"Did you feel tired over the last 4 weeks?"**	**P-value**	**"Did you feel worn out over the last 4 weeks?"**	**P-value**
MRI inflammation <-> fatigue*	-0.20 (-0.49;0.08)	0.16	-0.18 (-0.49;0.12)	0.24
DAS <-> fatigue*	**0.81 (0.49;1.12)**	**<0.001**	**1.1 (0.76;1.44)**	**<0.001**
Longitudinal				
MRI inflammation <-> fatigue*	0.11 (-0.10;0.32)	0.29	0.16 (-0.08;0.40)	0.19
DAS <-> fatigue*	**0.65 (0.48;0.82)**	**<0.001**	**0.75 (0.55;0.95)**	**<0.001**
Subsequent change**				
MRI inflammation -> fatigue	0.02 (-0.14;0.17)	0.84	0.07 (-0.1;0.25)	0.41
Fatigue -> MRI inflammation	0.04 (-0.18;0.26)	0.72	-0.08 (-0.25;0.09)	0.33
DAS -> fatigue	**0.13 (0.00;0.26)**	**0.046**	0.12 (-0.02;0.27)	0.10
Fatigue -> DAS	0.15 (-0.07;0.38)	0.19	0.17 (-0.02;0.35)	0.08

Legend: Answers from questions from the SF-36 were transformed to a scale of ranging from 0 (no fatigue) to 10 (extreme fatigue) to ensure comparability to the main results. * Estimate (95% confidence interval) corrected for age, gender and ACPA-status**.** **Standardized regression coefficients of change of one inflammatory feature to subsequent change in another inflammatory feature, corrected for the simultaneous pattern and previous values of those inflammatory features, with 95% confidence intervals. bold: significant estimate (p<0.05). RA: rheumatoid arthritis; DAS: disease activity score; ACPA: anti-citrullinated protein antibodies; SF-36 Short form 36 [9]

Supplementary File 11: Analyses of association of MRI inflammation and DAS with fatigue over time, in patients receiving initial MTX treatment within 100 days (n=137)

Longitudinal	**Estimate**	**P-value**
MRI inflammation <-> fatigue*	0.17 (-0.18;0.50)	0.33
DAS <-> fatigue*	**1.04 (0.75;1.34)**	**<0.001**
Subsequent change**		
MRI inflammation -> fatigue	0.14 (-0.06;0.33)	0.18
Fatigue -> MRI inflammation	0.06 (-0.13;0.25)	0.55
DAS -> fatigue	**0.23 (0.06;0.39)**	**0.007**
Fatigue -> DAS	0.01 (-0.19;0.21	0.90

Legend: *Estimate (95% confidence interval) corrected for age and gender**. ****Standardized regression coefficients of change of one inflammatory feature to subsequent change in another inflammatory feature, corrected for the simultaneous pattern and previous values of those inflammatory features, with 95% confidence intervals. bold: significant estimate ($p<0.05$). RA: rheumatoid arthritis; DAS: disease activity score; ACPA: anti-citrullinated protein antibodies; MTX: methotrexate.

SUPPLEMENTARY REFERENCES

1. Stomp W, Krabben A, van der Heijde D, Huizinga TW, Bloem JL, van der Helm-van Mil AH, et al. Aiming for a shorter rheumatoid arthritis MRI protocol: can contrast-enhanced MRI replace T2 for the detection of bone marrow oedema? European radiology. 2014;24(10):2614-22.
2. Schmid MR, Hodler J, Vienne P, Binkert CA, Zanetti M. Bone marrow abnormalities of foot and ankle: STIR versus T1-weighted contrast-enhanced fat-suppressed spin-echo MR imaging. Radiology. 2002;224(2):463-9.
3. Mayerhoefer ME, Breitenseher MJ, Kramer J, Aigner N, Norden C, Hofmann S. STIR vs. T1-weighted fat-suppressed gadolinium-enhanced MRI of bone marrow edema of the knee: computer-assisted quantitative comparison and influence of injected contrast media volume and acquisition parameters. Journal of magnetic resonance imaging : JMRI. 2005;22(6):788-93.
4. Sudol-Szopinska I, Jurik AG, Eshed I, Lennart J, Grainger A, Ostergaard M, et al. Recommendations of the ESSR Arthritis Subcommittee for the Use of Magnetic Resonance Imaging in Musculoskeletal Rheumatic Diseases. Seminars in musculoskeletal radiology. 2015;19(4):396-411.
5. Ostergaard M, Peterfy C, Conaghan P, McQueen F, Bird P, Ejbjerg B, et al. OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Studies. Core set of MRI acquisitions, joint pathology definitions, and the OMERACT RA-MRI scoring system. The Journal of rheumatology. 2003;30(6):1385-6.
6. Haavardsholm EA, Ostergaard M, Ejbjerg BJ, Kvan NP, Kvien TK. Introduction of a novel magnetic resonance imaging tenosynovitis score for rheumatoid arthritis: reliability in a multireader longitudinal study. Annals of the rheumatic diseases. 2007;66(9):1216-20.
7. Krabben A, Stomp W, Huizinga TW, van der Heijde D, Bloem JL, Reijnierse M, et al. Concordance between inflammation at physical examination and on MRI in patients with early arthritis. Annals of the rheumatic diseases. 2015;74(3):506-12.
8. Strand V, Kaine J, Alten R, Wallenstein G, Diehl A, Shi H, et al. Associations between Patient Global Assessment scores and pain, physical function, and fatigue in rheumatoid arthritis: a post hoc analysis of data from phase 3 trials of tofacitinib. Arthritis research & therapy. 2020;22(1):243.
9. Hewlett S, Dures E, Almeida C. Measures of fatigue: Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire (BRAF MDQ), Bristol Rheumatoid Arthritis Fatigue Numerical Rating Scales (BRAF NRS) for Severity, Effect, and Coping, Chalder Fatigue Questionnaire (CFQ), Checklist Individual Strength (CIS20R and CIS8R), Fatigue Severity Scale (FSS), Functional Assessment Chronic Illness Therapy (Fatigue) (FACIT-F), Multi-Dimensional Assessment of Fatigue (MAF), Multi-Dimensional Fatigue Inventory (MFI), Pediatric Quality Of Life (PedsQL) Multi-Dimensional Fatigue Scale, Profile of Fatigue (ProF), Short Form 36 Vitality Subscale (SF-36 VT), and Visual Analog Scales (VAS). 2011;63(S11):S263-S86.