Supplementary Table 1 Characteristics (baseline and primary endpoint) of the cohorts that shared data

	Amsterdam cohort	Erlangen cohort	Leeds cohort	Leiden CSA cohort	Rotterdam CSA cohort	Rotterdam SONAR cohort	TREAT EARLIER (placebo)	Vienna ASPRA cohort	Birmingham cohort	Rome SYNGem cohort
N	670	88	482	670	76	170	117	20	110	180
Baseline characteristics										
Female, n (%)	499 (75)	59 (67)	352 (73)	527 (79)	60 (79)	140 (82)	80 (68)	16 (80)	86 (78)	137 (76)
Age, mean in years (SD)	49.8 (11.7)	47.7 (12.6)	50.6 (13.5)	43.5 (12.6)	45.4 (13.0)	44.7 (11.3)	47.0 (10.8)	49.8 (11.2)	48.0 (13.3)	49.6 (12.8)
Symptom duration, med in weeks (IQR)	61 (30-156)	55 (28-182)	77 (37-165)	19 (10-44)	23 (12-41)	30 (19-43)	27 (16-51)	52 (8-52)		
Presence of hand symptoms, n (%)	NA	NA	295 (78)	530 (86)	51 (93)	NA	99 (87)	12 (63)	NA	129 (72)
Morning stiffness ≥60 min, n (%)	86 (18)	8 (9)	98 (20)	223 (35)	16 (21)	47 (34)	41 (35)	2 (10)	52 (47)	51 (28)
TJC44, med (IQR)	0 (0-3)	0 (0-2)	1 (0-3)	4 (2-9)	2 (0-5)	5 (3-8)	3 (1-8)	2 (0-11)	5 (1-10)	3 (2-4)
Increased CRP, n (%)	65 (10)	17 (28)	55 (12)	145 (22)	19 (25)	36 (23)	32 (27)	3 (15)	35 (32)	28 (16)
RF-positive, n (%)	448 (67)	47 (55)	188 (40)	136 (20)	21 (30)	46 (28)	35 (30)	16 (80)	42 (62)	37 (21)
ACPA-positive, n (%)	409 (61)	80 (93)	437 (91)	93 (14)	19 (26)	34 (21)	23 (20)	11 (55)	38 (35)	42 (23)

	Amsterdam cohort	Erlangen cohort	Leeds cohort	Leiden CSA cohort	Rotterdam CSA cohort	Rotterdam SONAR cohort	TREAT EARLIER (placebo)	Vienna ASPRA cohort	Birmingham cohort	Rome SYNGem cohort
Primary identification method for RA risk										
	Autoantibody-positive arthralgia	ACPA-positive arthralgia	Autoantibody-positive new non-specific MSK symptoms	CSA	CSA	CSA	CSA and subclinical MRI-inflammation	Autoantibody-positive arthralgia	CSA	CSA
Presence of imaging data										
US, n (%)	161 (24)	0	419 (87)	0	74 (97)	163 (96)	0	18 (90)	59 (54)	0
MRI n (%)	0	0	0	613 (92)	0	0	117 (100)	0	0	0
Part of persons started DMARDs in the phase of arthralgia/MSK-symptoms (without clinically apparent inflammatory arthritis)										
Yes/No	No	No	No	No	No	No	No	No	Yes	Yes
Primary endpoint										
Clinically apparent inflammatory arthritis within 1-year	116 (17)	20 (23)	76 (16)	101 (15)	17 (22)	41 (24)	16 (14)	2 (10)	37 (34)	46 (26)

CSA Clinically Suspect Arthralgia, N number, SD standard deviation, med median, IQR interquartile range, min minutes, NA not applicable, TJC tender joint count, CRP C-reactive protein, RF rheumatoid factor, ACPA anti-citrullinated peptide antibody, MSK musculoskeletal, US ultrasound, MRI magnetic resonance imaging.

Supplementary Table 2 Description of main inclusion and exclusion criteria of the different cohorts that were used for analyses

Cohort	Description
Amsterdam cohort [1]	ACPA and/or IgM-RF positive and (a history of) arthralgia, but not arthritis. Participants were recruited at rheumatology outpatient clinics in the Amsterdam area of the Netherlands and included between 2008 and 2013.
Birmingham cohort	Musculoskeletal symptoms that in the opinion of the managing rheumatologist put the patient at increased risk of progression to RA over time with no other more likely explanation.
Erlangen cohort [2]	ACPA-positive persons with joint symptoms and/or other symptoms for which ACPA was tested for clinical purpose. Exclusion criteria included clinically apparent synovitis
Leeds cohort [3]	Autoantibody-positive persons with new non-specific musculoskeletal (MSK) symptoms and no clinical synovitis. Participants were predominantly recruited in secondary care rheumatology clinics in Leeds and the Yorkshire region, UK. Some participants were identified in primary care by GPs checking ACPA in people presenting with non-specific MSK symptoms. Exclusion criteria included previous DMARD exposure and clinical synovitis.
Leiden CSA cohort [4]	Arthralgia of the small joints for <1 year that was, according to the clinical expertise of the rheumatologist, suspected to progress to RA over time (clinically suspect arthralgia, CSA). Importantly, CSA was not present if clinical arthritis was observed at physical examination or another explanation for the arthralgia was more likely (eg, osteoarthritis and fibromyalgia). Participants were recruited at the rheumatology outpatient clinic in the Leiden University Medical Center of the Netherlands. Studied individuals for this project were included between 2012 and 2022.
Rome SYNGem cohort	Arthralgia of the small and/or large joints for <1 year that was, according to the clinical expertise of the rheumatologist, suspected to progress to RA over time. Participants were referred for minimally invasive

	ultrasound guided synovial tissue biopsy at the Division of Rheumatology of the Fondazione Policlinico Universitario A. Gemelli IRCCS between 2017 and 2019.
Rotterdam CSA cohort [5]	Arthralgia of the small joints for <1 year that was (clinically suspect arthralgia, CSA), according to the clinical expertise of the rheumatologist, suspected to progress to RA over time. Importantly, CSA was not present if clinical arthritis was observed at physical examination or another explanation for the arthralgia was more likely (for example, osteoarthritis and fibromyalgia). Participants were recruited at 3 rheumatology outpatient clinics in the Rotterdam area of the Netherlands. Studied individuals for this project were included between 2017 and 2021.
Rotterdam SONAR cohort [6]	Inflammatory joint complaints of the hands, feet or shoulders without clinically apparent synovitis in any joint with symptom duration < 1 year which could not be explained by other conditions, such as inflammatory arthritis, fibromyalgia, overuse or trauma. To distinguish inflammatory arthralgia from other forms of arthralgia, patients had to have at least two painful joints in hands, feet or shoulders and two of the following criteria adapted from the Rotterdam Early Arthritis Cohort (REACH) trial: morning stiffness >1h, unable to clench a fist in the morning, pain when shaking someone's hand, pins and needles in the fingers, difficulties wearing rings or shoes, family history of RA and/or unexplained fatigue for < 1 year. Individuals were recruited at 3 rheumatology outpatient clinics in the Rotterdam area of the Netherlands and included between 2011 and 2014.
TREAT EARLIER trial, placebo-arm [7]	The inclusion criterium was presence of CSA (see defined as Leiden CSA cohort) and presence of MRI-detected subclinical inflammation in hand or forefoot (defined as present if at least one joint showed one or more inflammatory feature that was present in fewer than 5% of age-matched symptom-free volunteers at the same location). Participants were recruited from 13 rheumatology outpatient clinics in the southwest region of the Netherlands between 2015 and 2019. Only the individuals in the placebo-arm of this randomized, double-blind placebo-controlled trial (single i.m. glucocorticoid injection and 1-year course methotrexate versus placebo) were included in this project.

Vienna ASPRA cohort [2]	This is an observational cohort study at a tertiary center, including ACPA and/or RF positive individuals presenting with musculoskeletal symptoms suspicious of progression to RA, being follow-up over 5 years or until the presence of arthritis. Individuals included may not have clinical arthritis, or pretreatments with steroids or DMARDs.

ACPA anti-citrullinated peptide antibody, IgM Immunoglobulin, RF Rheumatoid Factors, RA rheumatoid arthritis, MSK musculoskeletal, UK United Kingdom, GP general practitioner, DMARD disease-modifying antirheumatic drugs, CSA Clinically Suspect Arthralgia

Supplementary Table 3 Overview of cohorts that performed ultrasound in the hands, wrist and feet and were studied

Cohort	Synovitis	Tenosynovitis
Amsterdam [8]	MCP2-3, PIP2-3, RC, IC and UC, MTP2-3 and 5; bilateral	No
Leeds CCP [9]	MCP1-5, PIP1-5, RC, IC, UC, MTP1-5; bilateral	Flexor dig 2-5, ECU
Rotterdam CSA [5]	MCP1-5, PIP1-5, RC, IC, DRU, MTP1-5; bilateral	Flexor dig 2-5, ECU
SONAR [6]	MCP2-5, PIP2-5 radiocarpal and intercarpal, MTP2-5; bilateral	Flexor dig 2-5
Vienna ASPRA [2]	MCP2-5, PIP2-5, RC, DRU; bilateral	No

MCP metacarpophalangeal, PIP proximal interphalangeal, RC radiocarpal, IC intercarpal, UC ulnocarpal, DRU distal radioulnar, MTP metatarsophalangeal, dig digitum, ECU extensor carpi ulnaris, GS grey scale, PD power Doppler

Supplementary Table 4 Results from lasso regression analysis on clinical & serological variables with 5-categories cohort variable instead of 2-categories cohort variable

	OR
Morning stiffness, 30-60 min	1.5
≥60 min	2.4
Patient-reported joint swelling	2.4
Difficulty making a fist	3.7
Increased CRP	1.3
RF, low-positive	1.5
high-positive	2.3
ACPA, low-positive	3.1
high-positive	9.4
5-categories cohort variable	
Group 1	ref
Group 2	1.9
Group 3	0.70
Group 4	1.4
Group 5	0.48
AUC (95% CI)	**0.81 (0.78;0.84)**

Legend. This analyses is performed in the total study population of 2,293 persons and is similar as presented in Table 2 but with a different cohort variable. Here the cohorts were grouped based on a combination of primary identification method for RA risk (autoantibody-positive arthralgia/MSK-symptoms or CSA) and geography: group 1) Leiden CSA and TREAT EARLIER (n=787); group 2) Rotterdam CSA and SONAR (n=246); group 3) Amsterdam (n=670); group 4) Erlangen and Vienna (n=108); group 5) Leeds (n=482).

OR odds ratio, Min minutes, CRP C-reactive protein, RF rheumatoid factor, ACPA anti-citrullinated peptide antibody, AUC area under the curve, ref reference, CI confidence interval

Supplementary Table 5 Results from Lasso regression analysis on aggregated ultrasound variables only

	OR
PD synovitis PIPs	3.6
PD tenosynovitis	1.9
PD synovitis MTPs	1.9
GS synovitis wrist	1.4
GS tenosynovitis	1.3
PD synovitis wrist	1.1
GS synovitis MTPs	0.98
GS synovitis MCPs	-
PD synovitis MCPs	-
GS synovitis PIPs	-
AUC (95% CI)	**0.63 (0.57;0.69)**

Legend. This analysis is performed in the population with ultrasound of 835 persons.
Variables not selected by lasso were GS synovitis MCPs, PD synovitis MCPs and GS synovitis PIPs.

OR odds ratio, PD power Doppler, GS grey scale, AUC area under the curve, CI confidence interval

Supplementary Table 6 Results from Lasso regression analysis on aggregated MRI variables only

	OR
Tenosynovitis flexors wrist	4.5
Tenosynovitis extensors MCP	4.4
Tenosynovitis extensors MTPs	2.6
Tenosynovitis extensors wrist	1.7
Synovitis MTPs	1.7
Osteitis wrist	1.1
Osteitis MCPs	1.1
Synovitis MCPs	-
Synovitis wrist	-
Tenosynovitis flexors MCPs	-
Tenosynovitis flexors MTPs	-
Osteitis MTPs	-
AUC (95% CI)	**0.76 (0.71;0.81)**

Legend. This analysis is performed in the population with MRI of 730 persons.
Variables not selected by lasso were synovitis MCPs, synovitis wrist, tenosynovitis flexors MCPs, tenosynovitis flexors MTPs and osteitis MTPs.

OR odds ratio, AUC area under the curve, CI confidence interval

Supplementary Table 7 Results of performance of clinical & serological variables in different groups of cohorts

Cohort group	AUC (95% CI)
Amsterdam (n=670)	0.78 (0.73;0.82)
Erlangen & Vienna (n=108)	0.78 (0.60;0.89)
Leeds (n=482)	0.82 (0.76;0.87)
Leiden CSA & TREAT EARLIER (n=787)	0.84 (0.80;0.88)
Rotterdam CSA & SONAR (n=246)	0.75 (0.66;0.82)

Legend. The eight cohorts were grouped based on a combination of primary identification method for RA risk (autoantibody-positive arthralgia/MSK-symptoms or CSA) and geography.

AUC area under the curve, n number, CI confidence interval, CSA clinically suspect arthralgia

Supplementary Table 8 Results of lasso regression with corresponding effect sizes, as performed for simplification

	Clinical & serological (n=2,293)		**Clinical & serological + MRI (n=730)**	
	Coef	**OR**	**Coef**	**OR**
Morning stiffness, 30-60 min	0.43	1.5	0.43	1.5
≥60 min	0.86	2.4	0.86	2.4
Patient-reported joint swelling	1.02	2.8	1.02	2.8
Difficulty making a fist	1.29	3.6	1.29	3.6
Increased CRP	0.34	1.4	0.34	1.4
RF, low-positive	0.42	1.5	0.42	1.5
high-positive	0.84	2.3	0.84	2.3
ACPA, low-positive	0.92	2.5	0.92	2.5
high-positive	1.84	6.3	1.84	6.3
MRI				
Tenosynovitis flexors wrist	-	-	1.26	3.5
Tenosynovitis extensors MCP	-	-	1.16	3.2
Tenosynovitis extensors MTPs	-	-	0.73	2.1
Tenosynovitis extensors wrist	-	-	0.38	1.5
AUC (95% CI)	**0.80 (0.77;0.82)**		**0.86 (0.82;0.90)**	

Legend:
To simplify risk stratification, the cohort variable was omitted from the analyses and for MRI only tenosynovitis variables were included in the analysis.

N number, Coef coefficient, OR odds ratio, min minutes, CRP C-reactive protein, RF rheumatoid factor, ACPA anti-citrullinated peptide antibody, MRI magnetic resonance imaging, MCP metacarpophalangeal joint, MTP metatarsophalangeal joint, AUC area under the curve, CI confidence interval

Supplementary Table 9 Test characteristics for different cutoffs in Risk-for-RA score with clinical & serological data only (A) and when also including MRI data (B)

A.

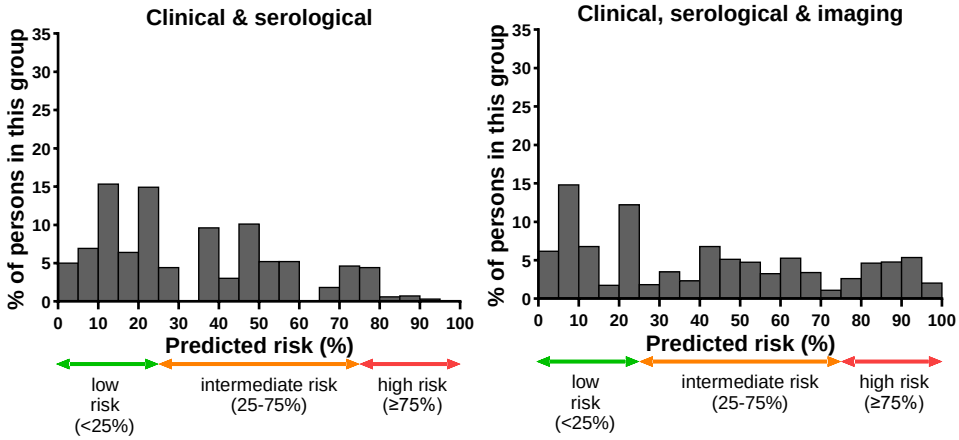
Cutoff risk score	Sensitivity	Specificity	PPV	NPV	% persons
≥2	99%	6%	19%	98%	95%
≥3	99%	13%	19%	98%	89%
≥4	95%	14%	23%	97%	88%
≥5	94%	35%	24%	97%	71%
≥6	90%	38%	26%	96%	68%
≥7	88%	47%	26%	95%	59%
≥8	78%	50%	33%	94%	57%
≥9	73%	67%	34%	93%	41%
≥10	66%	71%	38%	92%	36%
≥11	63%	78%	39%	91%	30%
≥12	51%	80%	46%	90%	28%
≥13	47%	88%	50%	89%	19%
≥14	37%	91%	53%	88%	16%
≥15	34%	93%	54%	88%	12%
≥16	24%	94%	60%	86%	11%
≥17	18%	97%	64%	85%	7%
≥18	13%	98%	64%	85%	5%
≥19	11%	99%	67%	85%	4%
≥20	7%	99%	71%	84%	3%
≥21	3%	99%	73%	83%	2%
≥22	2%	100%	81%	83%	1%

B.

Cutoff risk score	Sensitivity	Specificity	PPV	NPV	% persons
≥2	98%	14%	17%	98%	89%
≥3	98%	22%	17%	98%	82%

≥4	94%	24%	21%	98%	80%
≥5	94%	42%	22%	98%	64%
≥6	89%	47%	26%	97%	59%
≥7	89%	58%	27%	97%	49%
≥8	83%	61%	34%	96%	47%
≥9	79%	74%	38%	96%	35%
≥10	75%	79%	41%	95%	30%
≥11	72%	82%	44%	95%	27%
≥12	70%	85%	47%	95%	24%
≥13	65%	87%	53%	94%	22%
≥14	58%	91%	56%	93%	18%
≥15	56%	93%	61%	93%	15%
≥16	52%	94%	65%	93%	14%
≥17	50%	95%	68%	92%	12%
≥18	47%	96%	69%	92%	11%
≥19	38%	97%	74%	91%	9%
≥20	34%	98%	74%	90%	8%
≥21	30%	99%	78%	90%	7%
≥22	28%	99%	82%	89%	6%
≥23	21%	99%	83%	89%	5%
≥24	20%	99%	84%	88%	4%
≥25	17%	99%	83%	88%	4%
≥26	14%	99%	80%	88%	4%
≥27	12%	99%	83%	88%	3%
≥28	10%	99.6%	84%	87%	3%
≥29	8%	99.7%	100%	87%	3%
≥30	4%	100%	100%	87%	2%
≥31	3%	100%	100%	86%	2%
≥32	2%	100%	100%	86%	1%

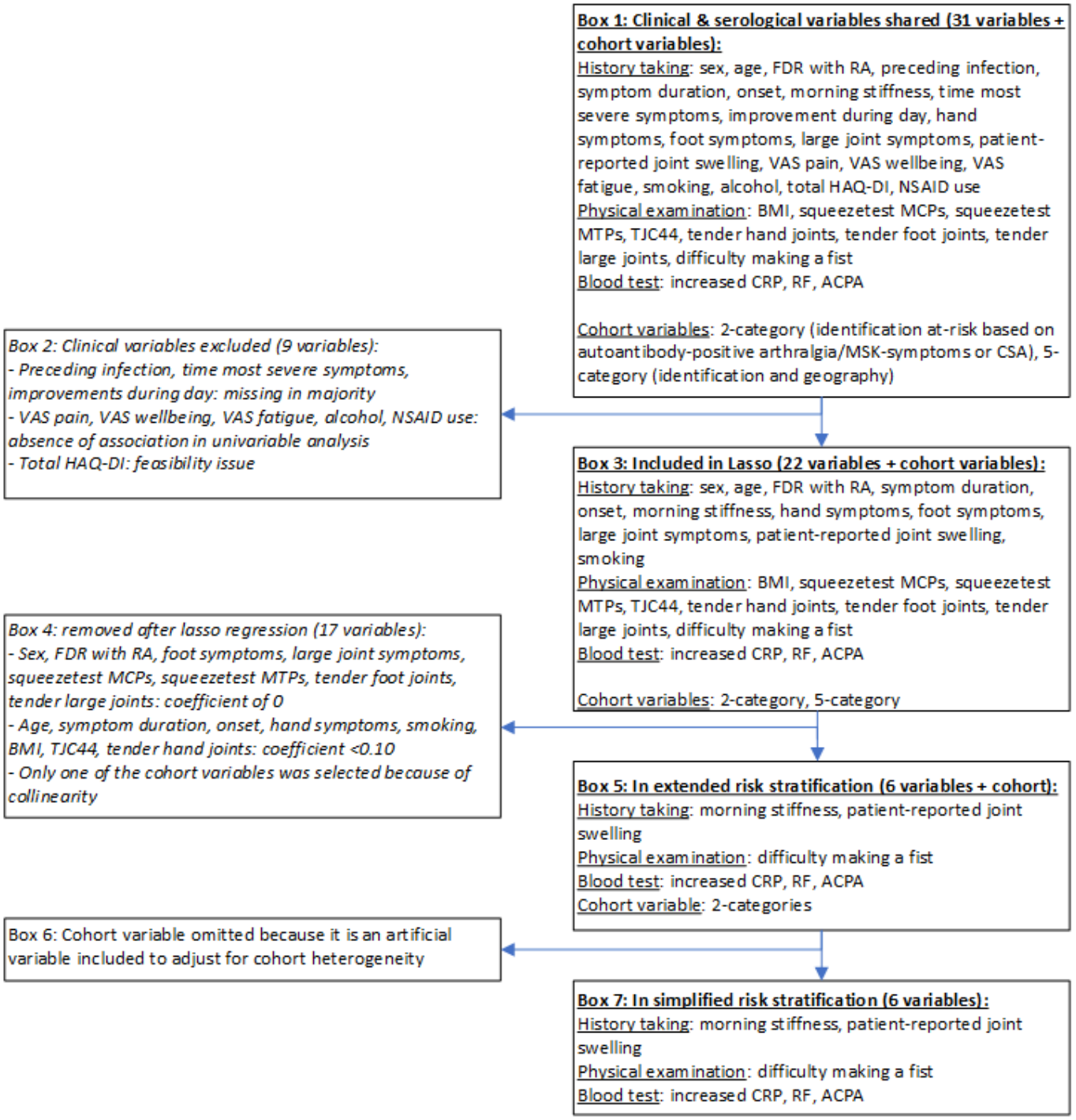
Supplementary Table 10. Distribution of predicted risks from participants **who reached the endpoint**; percentage of subjects categorized as low, intermediate and high risk based on the risk stratification criteria with clinical and serological data (left) and with also MRI-data (right) (upper part), and the characteristics from the individuals in these three risk groups (lower part)



	Low risk <25%	Intermed risk 25-75%	High risk ≥75%	Low risk <25%	Intermed risk 25-75%	High risk ≥75%
Morning stiffness ≥60min, %	21 %	43%	62%	34%	45%	59%
Patient-reported joint swelling, %	52%	80%	94 %	74%	81%	89%
Difficulty making a fist, %	13%	21%	70%	24%	20%	30%
Increased CRP, %	18%	29%	57%	24%	45%	46%
RF, Negative, %	**61%**	**20%**	**3%**	**80%**	**34%**	**0%**
ACPA, Negative, %	**47%**	**9%**	**0%**	**89%**	**34%**	**4%**
Tenosyn flex wrist, %	-	-	-	0.2%	47%	38%
Tenosyns ext wrist, %	-	-	-	12%	41%	63%
Tenosyn ext MCPs, %	-	-	-	14%	22%	57%
Tenosyn ext MTPs, %	-	-	-	0.5%	18%	58%

Legend. Individuals who reached the endpoint and were predicted as low risk in the clinical and serological model had few symptoms and signs, often normal CRP despite possible ACPA-positivity (left table). Individuals who reached the endpoint and were predicted as low risk in the model with MRI were mainly ACPA-negative and had little subclinical joint inflammation at presentation (right table).

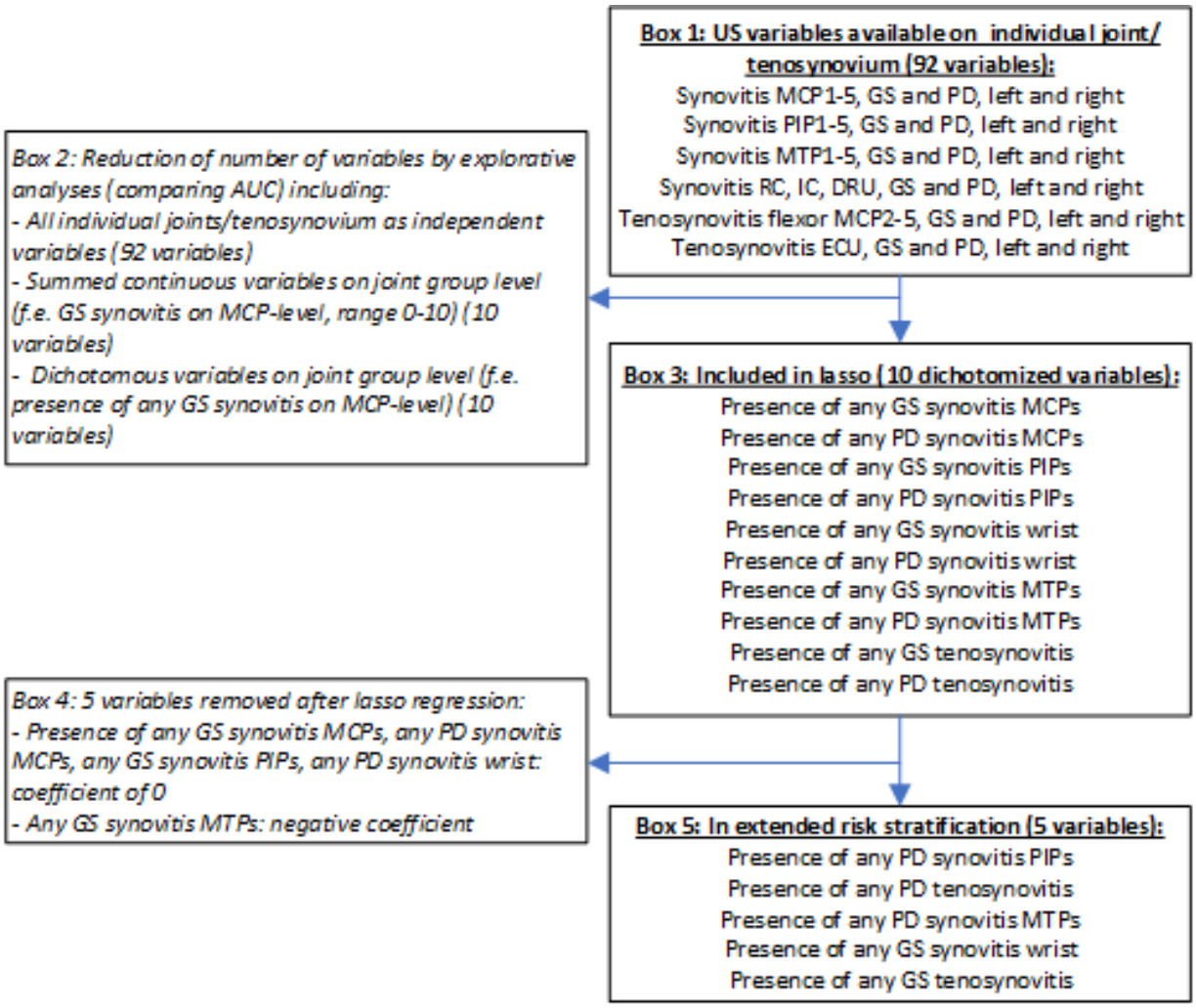
Supplementary Figure 1A Flowchart of clinical & serological variable reduction



FDR first-degree relative, RA rheumatoid arthritis, VAS visual analogue scale, HAQ-DI health assessment questionnaire, NSAID non-steroidal anti-inflammatory drug, BMI body mass index, MCPs metacarpophalangeal joints, MTP metatarsal phalangeal joints, TJC tender joint count, CRP C-reactive protein, RF rheumatoid factors, ACPA anti-citrullinated peptide antibodies, MSK musculoskeletal, CSA clinically suspect arthralgia

Legend:
After the inventory, data of 31 clinical & serological variables of potential interest were shared (box 1). These were based on history taking, physical examination or blood tests. In addition, to define cohort heterogeneity two cohort variables were made: a variable with two categories (identification of being at-risk based on autoantibody-positive arthralgia/MSK-symptoms or CSA), or five categories (combination of identification and geography). The five cohort groups were as following: 1) Leiden CSA and TREAT EARLIER; group 2) Rotterdam CSA and SONAR; group 3) Amsterdam; group 4) Erlangen and Vienna; group 5) Leeds. Because of the risk of overfitting due to the relatively low number of events in relation to predictors the number of variables had to be restricted. Nine were removed because of missingness or absence of association with the primary outcome (assessed in univariable logistic analyses with the clinical/serological variables as independent and clinically apparent inflammatory arthritis <1 year as dependent in available data (box 2)) and 22 were used for further analyses in (box 3). These 22 clinical & serological variables and one of the cohort variables were included in the lasso regression (after imputation of missingness) with clinical arthritis within 1 year as outcome (as the cohort-variables were collinear, one of them was included) (results of this analysis are presented in Table 2). Six clinical & serological variables and the cohort variables were identified by lasso as being important independent of the other included variables. The 2-category variable was used as cohort variable for the extended risk stratification (box 5), as the performance was similar including the 2-category variable and the 5-category variable (results in Table 2 and Suppl Table 4). Of the removed variables, 8 had a coefficient of 0 in the lasso and 8 had a coefficient <0.10 (e^coefficient is the odds ratio) (box 4). For the simplified risk stratification, the cohort variable was omitted (box 6) as this was an 'artificial variable'. The clinical & serological simplified risk stratification consisted of 6 variables (box 7, results in Suppl Table 8 and Box in manuscript).

Supplementary Figure 1B Flowchart of ultrasound variable reduction

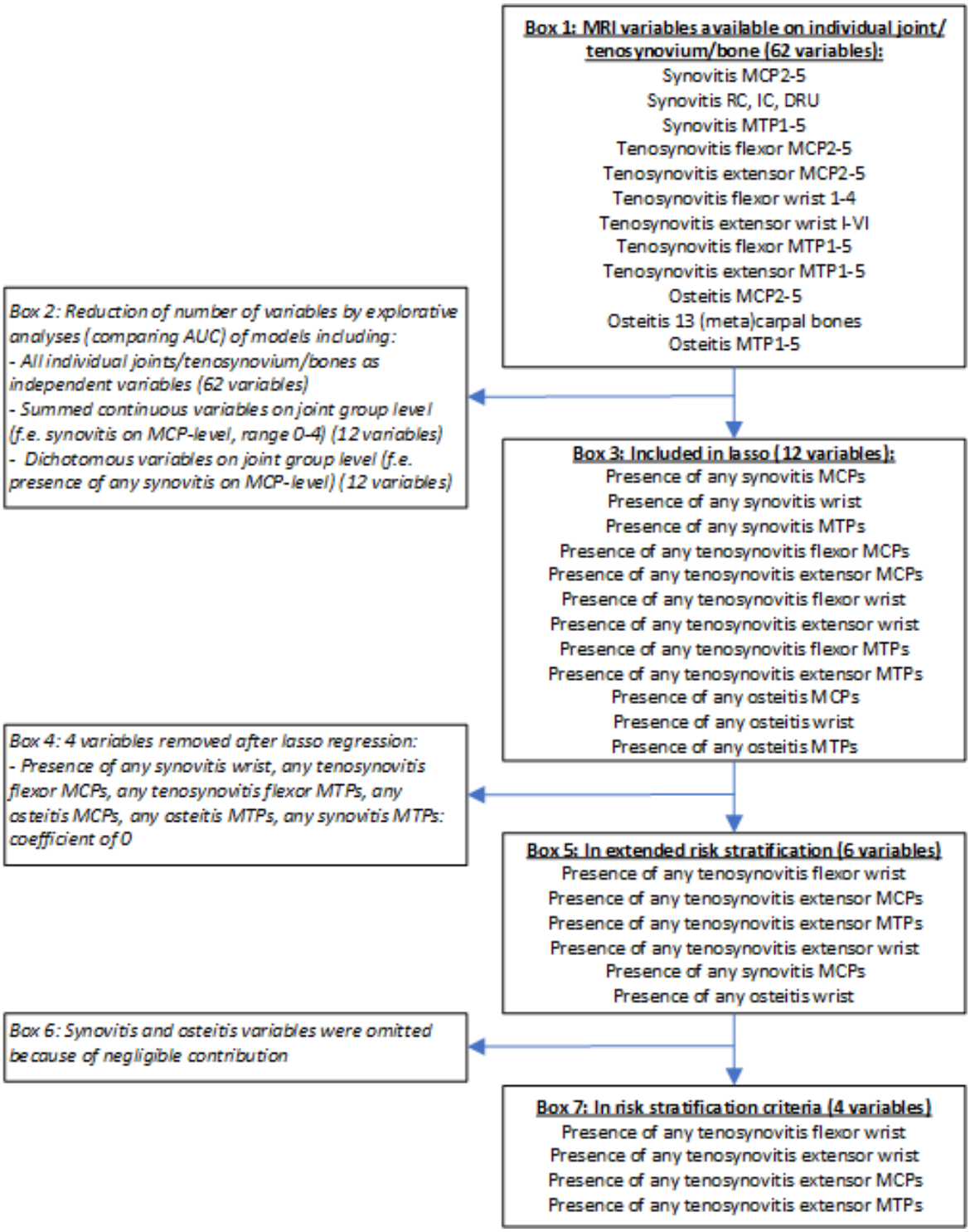


US ultrasound, MCP metacarpophalangeal, GS grey scale, PD power doppler, PIP proximal interphalangeal, MTP metatarsophalangeal, RC radiocarpal, IC intercarpal, DRU distal radioulnar, ECU extensor carpi ulnaris, AUC area under curve

Legend:
Despite differences in ultrasound (US)-protocols (Suppl Table 3), all cohorts that performed US assessed synovitis in at least part of the MCP-, PIP- and wrist-joints and the majority of cohorts in the MTP-joints. Tenosynovitis of the flexors of the fingers and extensor carpi ulnaris was assessed in the majority of patients. All cohorts graded grey scale (GS) and power Doppler (PD) according a 0-3 semi-quantitative scale, the majority according the OMERACT definitions.[10, 11] The presence of subclinical inflammation by US was defined on individual joint level. GS subclinical inflammation was present if GS≥2 (for MTP1-3 ≥3 because of prevalence of GS≤2 in healthy persons[12]), PD subclinical inflammation if PD≥1. In total 92 joints and tenosynovium locations were identified that were present across larger part of cohorts (box 1). To prevent the risk of overfitting when including 92 US-variables and to come to useful variables, explorative analyses were performed to aggregate US-variables. In the best possible manner it was considered important that the aggregated variable should contain information about the location of the joint with

subclinical inflammation (wrist, MCP, PIP, MTP), and the type of inflammation feature (synovitis or tenosynovitis, grey scale or power Doppler). Performance (AUC) of three different models including only US-variables as independent variables and clinical arthritis ≤1 year as outcome were compared (box 2). First, including all individual joint/tenosynovium variables (92 variables). Then, including combined variables in which the joints were grouped on joint level with a range of 0 to 10 involved joints at that joint level (for example 0 to 10 MCPs with GS synovitis or 0 to 10 locations with PD tenosynovitis (8 flexors of the fingers and 2 extensors carpi ulnaris), this resulted in 10 continuous combined variables. And finally, including combined variables similar as in previous analysis but then dichotomized for presence/absence of inflammation at joint group level (for example, presence/absence of any GS synovitis on MCP-level or presence/absence of any PD tenosynovitis of one of the ten locations), which resulted in 10 dichotomized variables. These three analyses revealed similar AUCs and the most easily applicable set of variables was chosen: presence/absence of inflammatory feature on joint group level (10 dichotomous variables) (box 3). These 10 aggregated variables were included in lasso regression with clinical arthritis ≤1 year as outcome (results are presented in Table 2). Five aggregated US-variables were identified by the lasso as being important independent of the other included variables (these concerned the other included US-variables but also the fixed clinical & serological variables from the analysis with clinical & serological variables) (box 5). Of the five removed variables, four were had a coefficient of zero in the lasso and one had a negative coefficient below 1 (box 4) (e^coefficient is the odds ratio). The five identified important US-variables were included in the extended risk stratification (box 5, results in Table 2, Fig 1).

Supplementary Figure 1C Flowchart of MRI variable reduction

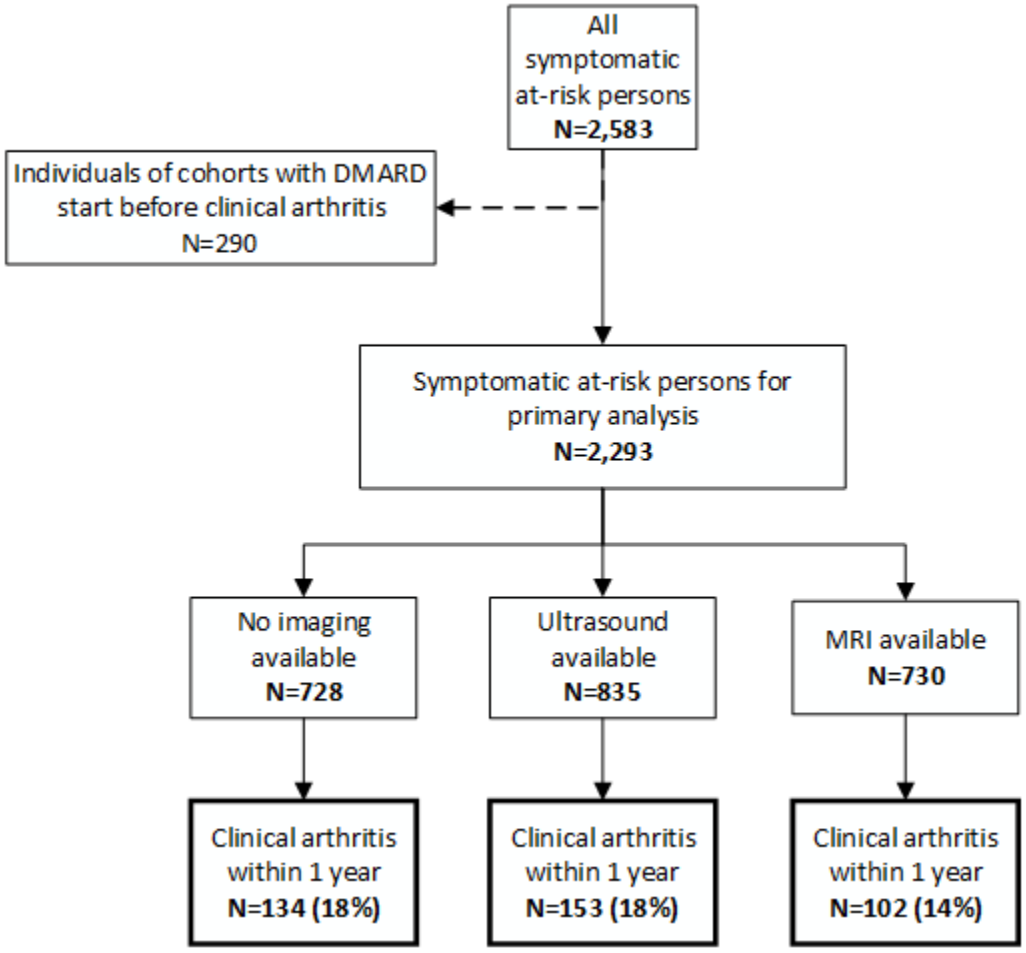


MRI magnetic resonance imaging, MCP metacarpophalangeal, RC radiocarpal, IC intercarpal, DRU distal radioulnar, MTP metatarsophalangeal, AUC area under curve

Legend:

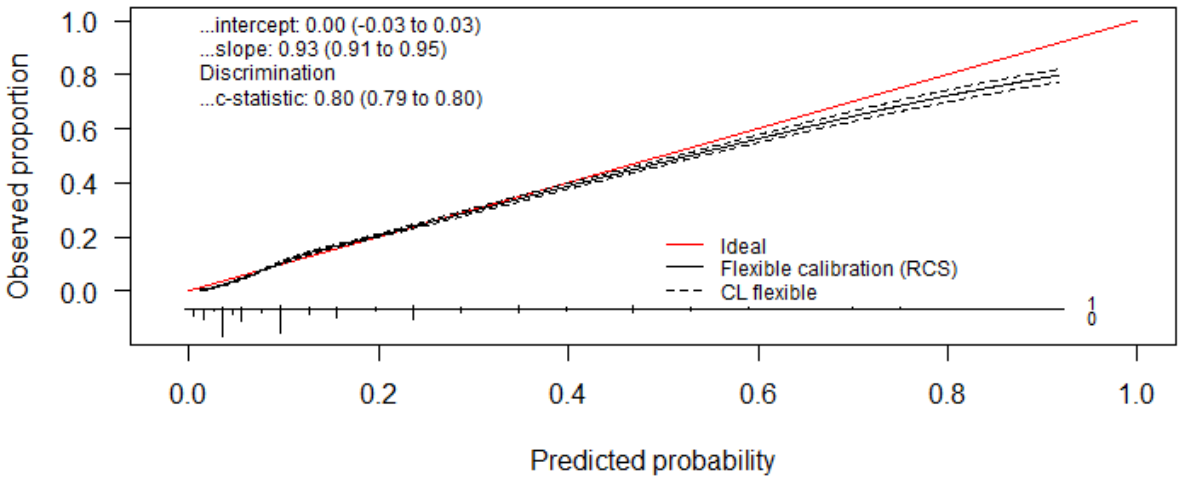
MRI was performed in two datasets, recognizing that the patients in the TREAT EARLIER dataset came from 13 different centers. MRI was made and scored using the RAMRIS protocol (Suppl File 1) [13, 14]. In short synovitis, tenosynovitis and osteitis were scored unilateral in the MCP-, wrist and MTP-joints (62 variables) (box 1). Subclinical (teno)synovitis or osteitis was defined at a location as an inflammatory feature was scored that was present in fewer than 5% of age-matched symptom-free volunteers at the same location.[15, 16] Also here, to prevent the risk of overfitting when including 62 MRI-variables and to come to clinical useful variables, analyses were performed to evaluate potential combined MRI-variables. Similar as for US, it was considered important that the combined variable should contain information about the location of the joint with subclinical inflammation (wrist, MCP, MTP), and the type of inflammation feature (synovitis, tenosynovitis, osteitis). For exploration, performance (AUC) of three different models including only MRI-variables as independent variables and clinical arthritis ≤1 year as outcome were compared (box 2). First, including all individual joint/tenosynovium/bone variables (62 variables). Then, including combined variables in which the joints were grouped on joint level with a continuous range the involved joints at that joint level (f.e. 0 to 4 for MCPs with synovitis and 0 to 13 for osteitis in the wrist), this resulted in 12 continuous aggregated variables. And finally, including combined variables similar as in previous analysis but then dichotomized for presence/absence of inflammation at joint group level (f.e. presence/absence of any synovitis on MCP-level), which resulted in 12 dichotomized aggregated variables. Similar as for US, these three analyses revealed comparable AUCs and the most easily applicable set of variables was chosen: presence/absence of inflammatory feature on joint group level (12 dichotomized variables) (box 3). These 12 variables were included in lasso regression with clinical arthritis ≤1 year as outcome (results are presented in Table 2). Six aggregated MRI-variables were identified by the lasso as being important independent of the other included variables (these concerned the other included MRI-variables but also the clinical & serological variables included as fixed linear predictor from the analysis with clinical & serological variables) (box 5). Of the five removed variables, five had a coefficient of 0 in the lasso (box 4). The six identified important MRI-variables were included in the extended risk stratification (box 5, Fig 1). For the simplified risk stratification, only the tenosynovitis variables were selected the other MRI-variables had negligible contribution (box 6). The simplified risk stratification consisted of 4 tenosynovitis variables (box 7, results in Suppl Table 8 and Box in manuscript).

Supplementary Figure 2 Flowchart of selection of study population and prevalence of primary outcome

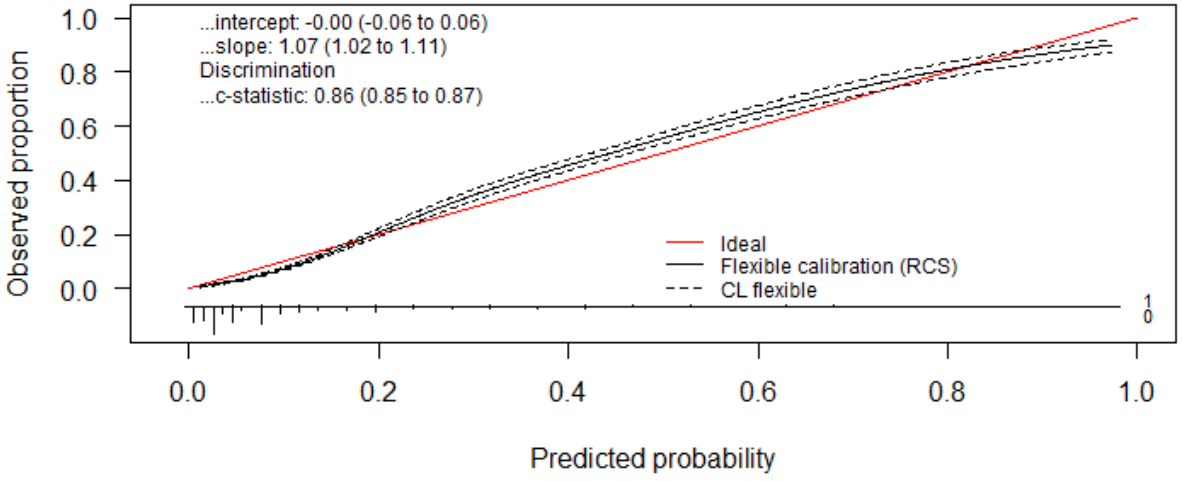


Supplementary File 3 Calibration graphs of risk stratification criteria with and without MRI.

Clinical & serological variables



Clinical, serological & imaging (MRI) variables



Supplementary File 1.

Protocol for Magnetic Resonance Imaging (MRI)

All patients who underwent MRI had the same protocol, i.e. a unilateral MRI of wrist, metacarpophalangeal (MCP) 2-5, and metatarsophalangeal (MTP)1-5 joints of the most painful side, or the dominant side in case of equally severe symptoms at both sides. The MRI was made with gadolinium contrast enhancement on an MSK-extreme 1.5T extremity MR imaging system (GE, Wisconsin, USA), using a 145mm coil for the foot and a 100mm coil for the hand. Patients were instructed not to use NSAIDs 24 hours prior to MRI. Patient were positioned in a chair beside the scanner, with the hand or foot fixed in the coil with cushions. In the hand (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) 575ms, echo time (TE) 11.2ms, acquisition matrix 388×288, echo train length (ETL) 2). After intravenous injection of gadolinium contrast (gadoteric acid, Guerbet, Paris, France, standard dose of 0.1mmol/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 sequence with frequency selective fat saturation in the axial plane (wrist: TR/TE 540/7.7ms; acquisition matrix 320x192; ETL 2 and MCP-joints: TR/TE 570/7.7ms; acquisition matrix 320x192; ETL 2). The obtained sequences of the forefoot (MTP 1-5 joints) concerned post-contrast images of the foot: 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 MTP-joints) (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.

Supplementary. MRI-inflammation; scoring and dichotomization

Synovitis was scored in line with the Outcome Measures in Rheumatology Clinical Trials (OMERACT) RA MRI scoring (RAMRIS)-method.[17] RAMRIS was not developed to score

MTP-joints, however others have previously adapted the RAMRIS to score MTP-joints as well.[18] Tenosynovitis was scored according to the method described by Haavardsholm (also applied at the flexor and extensor tendons at the 2-5 MCP-joints; range 0-3).[14] The synovitis score (range 0-3) was scored based on the volume of enhancing tissue in the synovial compartment (none, mild, moderate, severe) and the tenosynovitis-score (ranged 0-3) was based on the thickness of peritendinous effusion or synovial proliferation with contrast enhancement (normal, <2mm, 2-5mm, >5mm). The total sum of MRI-detected inflammation in wrist, MTP-joints and MCP-joints is maximal 213. MRIs were scored by two experienced readers, blinded to any clinical data. Inter- and intrareader intraclass correlation coefficients were ≥0.90, as published previously.[18] Presence of inflammation was dichotomized per feature of inflammation (synovitis, tenosynovitis, osteitis) and per location; if the inflammation-score of any feature was higher than present in <5% of age matched healthy controls at the same location, the joint was scored positive for inflammation.

Supplementary File 2. Statistics

Selection of risk variables before performing Lasso regression

Clinical & serological variables – First, univariable logistic regression analyses with one of the 31 available clinical/serological variables (Suppl Fig 1A, box 1) as independent and the primary outcome (clinically apparent arthritis ≤1 year) as dependent were performed. Variables with absence of univariable association or with too many missing were excluded. This resulted in clinical & serological risk variables to be included in further analyses (Suppl Fig 1A, box 3).

Ultrasound variables – For exploration three different logistic regression models including only US-variables as independent variables and clinical arthritis ≤1 year as outcome were compared on performance (AUC). First, a model (Lasso regression) including all individual joint/tenosynovium variables (92 variables, Suppl Fig 1B, box 1). Then, a model in which the joints were grouped on joint level with a range of 0 to 10 involved joints at that joint level (for example 0 to 10 MCPs with GS synovitis or 0 to 10 locations with PD tenosynovitis (8 flexors of the fingers and 2 extensors carpi ulnaris), this resulted in 10 continuous combined variables. And finally, the variables from the second model were dichotomized for presence/absence of inflammation at joint group level (for example, presence/absence of any GS synovitis on MCP-level or presence/absence of any PD tenosynovitis of one of the ten locations), which resulted in 10 dichotomized variables. These three analyses revealed similar AUCs. The variables of presence/absence of inflammatory feature on joint group level (10 dichotomous variables) were used in further analyses (Suppl Fig 1B, box 3) because this is the most easily applicable set of variables, and also because the continuous grouped variables (ranging 0-10) were extremely skewed to zero.

For MRI-variables – This was done similar as for US. Performance (AUC) of three different models including only MRI-variables as independent variables and clinically apparent inflammatory arthritis ≤1 year as outcome were compared. First, including all individual joint/tenosynovium/bone variables (62 variables, Suppl Fig 1C, box 2). Then, including combined variables in which the joints were grouped on joint level with a continuous range of the involved joints at that joint level (f.e. 0 to 4 for MCPs with synovitis and 0 to 13 for osteitis in the wrist), this resulted in 12 continuous aggregated variables. And finally, including combined variables similar as in previous analysis but then dichotomized for presence/absence of inflammation at joint

group level (f.e. presence/absence of any synovitis on MCP-level), which resulted in 12 dichotomized aggregated variables. These three analyses revealed comparable AUCs. The most easily applicable set of variables was used for further analyses: presence/absence of inflammatory feature on joint group level (12 dichotomized variables) (Suppl Fig 1C, box 3).

Imputation

Multiple imputation was performed using MICE [19] based on the selected 22 clinical & serological variables (see Suppl Fig 1A), 93 ultrasound variables of individual joints (see Suppl Fig 1B), 63 MRI variables of the individual joints (see Suppl Fig 1C), the cohort of origin and the outcome data on the primary (inflammatory arthritis ≤1 year) and secondary outcomes (fulfilling 2010 RA classification criteria ≤1 year and inflammatory arthritis ≤2 years). Twenty datasets were created with completed clinical & serological variables for all persons. Completed ultrasound or MRI data sets were only created for persons with imaging data available, as absence of imaging was considered not missing at random.

Lasso regression

Penalized regression (Least absolute shrinkage and selection operator, Lasso) was used to avoid overfitting, which was considered likely due to a relatively low number of events. Lasso shrinks unimportant variables to zero and thereby selects important variables for the outcome.

First, within the total dataset of all cohorts, a model was built with the 22 clinical and serological variables as independent variables and the primary endpoint (clinically apparent inflammatory arthritis ≤1 year) as dependent (Suppl Fig 1A, box 3). To adjust for cohort heterogeneity a cohort variable was added. For this, two variables for cohort were made: a variable with two categories (identification of being at-risk based on autoantibody-positive arthralgia/MSK-symptoms or CSA) and with five categories (combination of identification and geography). Two Lasso regression models were performed with both the 22 clinical and serological variables and one of the cohort variables. The cohort variable in the model with the best performance (as measured with AUC) was used for further analyses. A grouped Lasso regression was used to select variables as the cohort variable is a categorical variable. The clinical and serological variables selected by the Lasso here were used for the extended risk stratification (Suppl Fig 1A, box 5). The coefficients were obtained by refitting the selected variables with a

logistic regression model (Lasso regression was not possible for this due to computational efficiency).

Then, within the cohorts with available US-variables and MRI-variables other Lasso regression models were built (Suppl Fig 1B and 1C, box 3). To assess the incremental value of US/MRI-detected inflammation added to the information from the clinical and serological variables, we included a linear predictor of the previously developed model on the clinical and serological variables. This kept the coefficients for individual variables fixed, but allowed penalization of the linear predictor coefficient. The Lasso regression was done to select US/MRI-variables in the presence of the linear predictor. These analyses selected the US and MRI-variables for the extended risk stratification (Suppl Fig 1B and 1C, Box 5).

For easier clinical applicability, we opted to transform coefficients from the models into a sum score by scaling and rounding the coefficients. Starting with the model including only clinical & serological variables, we chose a scale factor of 4 (coefficients*4) and rounded thereafter, with the rationale of having countable sum scores without loss of interpretation or performance. For the US/MRI models, the scale factor of 4 was divided by the penalized coefficient of the included linear predictor of the clinical and serological variables (for both US and MRI this coefficient was ~0.8, which led to a scale factor of ~5). This allowed to keep counting the sum scores after including US or MRI information in the model. This resulted in risk stratification that consisted of a section with clinical & serological variables, and two imaging section (US or MRI-data) that could be added if available (Fig 1). The predicted risks were plotted against the risk scores. Test characteristics (sensitivity, specificity) and predictive values (positive and negative predictive values (PPV, NPV)) were determined for different cutoffs of risk scores.

Validation

We performed internal validation using bootstrapping to quantify the optimism of the prediction models. We used bootstrapping (resampling with replacement; 200 bootstrap replications) to assess the optimism of each model's performance in terms of area under the ROC-curve (AUC). The optimism-corrected AUC's are calculated with bootstrap samples within all imputed datasets and combined using Rubin's rules. Additionally, we performed a 2/3 and 1/3 set split. The split was performed at the cohort and outcome levels. AUC's are also calculated per cohort, to assess performance across different cohorts.

Performance

Discriminatory capacity of the three models (clinical and serological, +US, +MRI) was assessed primarily by the area under the ROC-curve (AUC), with higher AUCs indicating better performance. Calibration curves were generated to assess the relation between predicted and observed outcomes. Calibration intercept and slopes are estimated. The AUCs were also determined with the developed criteria tested against the secondary outcomes.

We evaluated whether a sensitivity and specific of ~80% (a priori selected criterion for good performance) was achieved for each of the derived models (with / without imaging, before and after simplification).

Finally, we studied the number of patients classified as low, intermediate and high risk for the risk stratification criteria with and without imaging. For this we arbitrarily defined low risk as <25% development of clinically apparent inflammatory arthritis; intermediate risk as 50-75%; and high risk as ≥75%. The goal was to have as few people as possible in the intermediate risk group, as the low and high risk groups were considered to be most clinically meaningful. We evaluated the percentage of patients in these risk categories when evaluating the total data set, but also when studying the patients that did and did not progress to the primary endpoint separately. Likewise, for the total population, we compared the percentage of patients that was correctly classified as low and high risk when using clinical and serological data only, and when MRI data was also used.

REFERENCES

1. van de Stadt LA, Witte BI, Bos WH, van Schaardenburg D. A prediction rule for the development of arthritis in seropositive arthralgia patients. Ann Rheum Dis. 2013 Dec; 72(12):1920-1926.
2. Studenic P, Hensvold A, Kleyer A, van der Helm-van Mil A, Pratt AG, Sieghart D, et al. Prospective Studies on the Risk of Rheumatoid Arthritis: The European Risk RA Registry. Front Med (Lausanne). 2022; 9:824501.
3. Duquenne L, Hensor EM, Wilson M, Garcia-Montoya L, Nam JL, Wu J, et al. Predicting Inflammatory Arthritis in At-Risk Persons: Development of Scores for Risk Stratification. Ann Intern Med. 2023 Aug; 176(8):1027-1036.
4. van Steenbergen HW, van Nies JA, Huizinga TW, Bloem JL, Reijnierse M, van der Helm-van Mil AH. Characterising arthralgia in the preclinical phase of rheumatoid arthritis using MRI. Ann Rheum Dis. 2015 Jun; 74(6):1225-1232.
5. Rogier C, Kortekaas MC, van der Helm-van Mil A, de Jong P, van Mulligen E. Sequential ultrasound in arthralgia patients at risk for inflammatory arthritis: is it of added value in risk stratification? Scand J Rheumatol. 2023 Jan; 52(1):96-98.
6. van der Ven M, van der Veer-Meerkerk M, Ten Cate DF, Rasappu N, Kok MR, Csakvari D, et al. Absence of ultrasound inflammation in patients presenting with arthralgia rules out the development of arthritis. Arthritis Res Ther. 2017 Sep 15; 19(1):202.
7. Krijbolder DI, Verstappen M, van Dijk BT, Dakkak YJ, Burgers LE, Boer AC, et al. Intervention with methotrexate in patients with arthralgia at risk of rheumatoid arthritis to reduce the development of persistent arthritis and its disease burden (TREAT EARLIER): a randomised, double-blind, placebo-controlled, proof-of-concept trial. Lancet. 2022 Jul 23; 400(10348):283-294.
8. van Beers-Tas MH, Blanken AB, Nielen MMJ, Turkstra F, van der Laken CJ, Meursinge Reynders M, et al. The value of joint ultrasonography in predicting arthritis in seropositive patients with arthralgia: a prospective cohort study. Arthritis Res Ther. 2018 Dec 19; 20(1):279.
9. Nam JL, Hensor EM, Hunt L, Conaghan PG, Wakefield RJ, Emery P. Ultrasound findings predict progression to inflammatory arthritis in anti-CCP antibody-positive patients without clinical synovitis. Ann Rheum Dis. 2016 Dec; 75(12):2060-2067.
10. D'Agostino MA, Terslev L, Aegerter P, Backhaus M, Balint P, Bruyn GA, et al. Scoring ultrasound synovitis in rheumatoid arthritis: a EULAR-OMERACT ultrasound taskforce-Part 1: definition and development of a standardised, consensus-based scoring system. RMD Open. 2017; 3(1):e000428.
11. Naredo E, D'Agostino MA, Wakefield RJ, Möller I, Balint PV, Filippucci E, et al. Reliability of a consensus-based ultrasound score for tenosynovitis in rheumatoid arthritis. Ann Rheum Dis. 2013 Aug; 72(8):1328-1334.
12. Padovano I, Costantino F, Breban M, D'Agostino MA. Prevalence of ultrasound synovial inflammatory findings in healthy subjects. Ann Rheum Dis. 2016 Oct; 75(10):1819-1823.
13. Østergaard M, Edmonds J, McQueen F, Peterfy C, Lassere M, Ejbjerg B, et al. An introduction to the EULAR-OMERACT rheumatoid arthritis MRI reference image atlas. Ann Rheum Dis. 2005 Feb; 64 Suppl 1(Suppl 1):i3-7.

14. Haavardsholm EA, Østergaard 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. Ann Rheum Dis. 2007 Sep; 66(9):1216-1220.
15. Mangnus L, van Steenbergen HW, Reijnierse M, van der Helm-van Mil AH. Magnetic Resonance Imaging-Detected Features of Inflammation and Erosions in Symptom-Free Persons From the General Population. Arthritis Rheumatol. 2016 Nov; 68(11):2593-2602.
16. Boer AC, Burgers LE, Mangnus L, Ten Brinck RM, Nieuwenhuis WP, van Steenbergen HW, et al. Using a reference when defining an abnormal MRI reduces false-positive MRI results-a longitudinal study in two cohorts at risk for rheumatoid arthritis. Rheumatology (Oxford). 2017 Oct 1; 56(10):1700-1706.
17. Østergaard 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. J Rheumatol. 2003 Jun; 30(6):1385-1386.
18. Dakkak YJ, Matthijssen XME, van der Heijde D, Reijnierse M, van der Helm-van Mil AHM. Reliability of Magnetic Resonance Imaging (MRI) Scoring of the Metatarsophalangeal Joints of the Foot according to the Rheumatoid Arthritis MRI Score. J Rheumatol. 2020 Aug 1; 47(8):1165-1173.