Appendix:

Ultrasound in the management of rheumatoid arthritis: ARCTIC randomized controlled strategy trial

Table of contents

Section 1. Study investigators in the ARCTIC trial

Study investigators in the ultrasound tight control arm:
Anna-Birgitte Aga; Hilde Berner Hammer; Inger Jorid Berg; Pernille Bolton-King; Åse Stavland Lexberg; Yngvill Hovde Bragnes; Hilde Stray; Merete Valen; Maria Karolina Jonsson; Solveig Hauge; Ellen Norli; Hilde Haukeland; Liv Elisabeth Kjustad; Shagaye Nabizadeh; Christian Høili; Ruth Stoklund Thomsen; Tina Pedersen; Vivi Bakkeheim; Gunnstein Bakland; Lucius Bader; Trude Jannecke Bruun; Hallvard Fremstad; Maud-Kristine Aga Ljoså; Inger Johanne Hansen; Helene Hetland; Anne Noraas Bendvold

Study investigators in the conventional tight control arm:
Till Uhlig; Liv Lefsaker; Antonela Botea; Elisabeth Langseth Esperø; Cecilie Kaufmann; Marianne Dalen; Navjot Grewal; Tove Borgen; Ada Wierød; Tor Magne Madland; Liv Turid Bertelsen; Anne Julsrud Haugen; Eva Purinszky; Per Jarle Tungevåg; Olav Bjørneboe; Anne Prøven; Erik Rødevand; Marianne Wallenius; Marit Seip; Synøve Kalstad; Carina Skorpen; Karen Irgens; Leif Kåre Haga; Geirmund Myklebust; Halvard Dovland.

Section 2. Full inclusion and exclusion criteria

Overview
Adult men and women with early RA according to the 2010 ACR/EULAR classification criteria with indication for DMARD therapy were eligible for participation in this study.

Inclusion criteria
To be eligible for this study, patients must meet all of the following criteria:

1. Male or non-pregnant, non-nursing female
2. > 18 years of age and < 75 years of age
3. Patients classified as having RA (according to new ACR/EULAR criteria)
4. Disease duration less than 2 years (defined as time from 1st joint swelling)
5. The treating rheumatologist decides the patient requires DMARD-treatment
6. The patient has taken no prior DMARD
7. Patients able and willing to give written informed consent and comply with the requirements of the study protocol

Exclusion criteria
Patients with any of the following criteria will not be eligible to participate in the study:

1. Abnormal renal function (serum creatinine > 142 μmol/L in female and > 168 μmol/L in male, or GFR < 40 mL/min/1.73 m^2).
2. Abnormal liver function (ASAT/ALAT > 3* normal), active or recent hepatitis, cirrhosis.
3. Major co-morbidities like severe malignancies, severe diabetic mellitus, severe infections, uncontrollable hypertension, severe cardiovascular disease (NYHA class 3-4) and/or severe respiratory diseases.
4. Leukopenia and/or thrombocytopenia.
5. Inadequate birth control conception, pregnancy, and/or breastfeeding.
6. Indications of active tuberculosis.
7. Psychiatric or mental disorders, alcohol abuse or other abuse of substances, language barriers or other factors which makes adherence to the study protocol impossible.

Section 3. Treatment regimen

Patients in both groups were treated according to the same fixed treatment algorithm, adhering to a treat-to-target strategy with DMARD escalation therapy if target was not met. The treatment adjustments (including i.a. injections) that could be made were defined in a pre-specified dosing regimen, outlined in table S1.

Treatment target
The treatment target in the ultrasound tight control strategy was clinical remission (defined as Disease Activity Score <1.6 and no swollen joints) and ultrasound imaging remission (defined as no power Doppler signal in any of the joints assessed by ultrasound). The treatment target in the conventional tight control strategy was clinical remission (Disease Activity Score <1.6 and no swollen joints). The ultrasound standardized score included assessments of the following 32 joints with both grey-scale and power Doppler (semi-quantitative score of 0-3 for all joints, with a reference atlas showing the different possible grades for all assessed joints): MCPs I-V, wrist (radio-carpal, radio-ulnar and inter-carpal), elbow, knee, talo-crural and MTP I-V bilaterally.[1]

Treatment adjustments
The decision of whether to adjust medication was based on change in and the level of the Disease Activity Score. If the patient does not respond as described in table S2, the treating physician immediately adjusted the therapy by proceeding to the next step in the treatment algorithm. If a patient responded or had reached the target, current medication was continued. In the ultrasound tight control group, the physician should overrule the decision based on the Disease Activity Score and proceed to the next step based on ultrasound findings, as described in table S2.

Intra-articular steroids
In both groups, clinically swollen joints were treated by intra-articular steroids when indicated. In the ultrasound tight control group an additional target was all joints with power Doppler signal, and all injections should be ultrasound guided. For both groups, intra-articular injections of only tender joints were not allowed. The maximum dosage of triamcinolone hexacetonid per visit was 80 mg which could be distributed within joints as decided by the treating rheumatologist.

NSAIDs, vitamin D and calcium
NSAIDs and coxibs were permitted. The choice and dosage of NSAIDs/coxibs was at the discretion of the treating rheumatologist. Analgesics up to the maximum recommended dose could be used for pain relief as required. Patients should avoid analgesics within 24 hours prior to a visit if possible.

All patients received vitamin D and calcium supplement during treatment with corticosteroids ≥ 7.5mg, and postmenopausal women and older men (>70 year) was considered for a bisphosphonate according to general guidelines. IV or IM corticosteroids were not allowed during the study. Oral corticosteroids were allowed as described in table S1. Other DMARDs than those described in table S1 was not allowed.

Section 4. Statistical analysis

The full analysis set for efficacy and safety included all patients randomly assigned to a treatment group and who started the allocated intervention defined as having completed at least one regular visit after the baseline visit.

The primary analysis on the primary endpoint and other binary endpoints were conducted using logistic regression models. The analyses were not adjusted for the stratification factors center and presence of anti-CCP due to low cell frequencies, but these variables were included in robustness analyses using exact logistic regression. Estimates of risk difference were calculated from the logistic regression parameters using the delta method to provide the confidence intervals.

Missing values of the primary endpoint were imputed using the following rule:
Radiographic score:

- If the radiographic score was missing at month 24, the patient was considered not to meet the primary endpoint (worst outcome)
- If a radiographic score was missing for visit 11 (16 months), we used last radiographic observation

Disease Activity Score (DAS):

- If unable to calculate DAS at visit 13 (month 24), the patient was considered not to meet the primary endpoint (worst outcome)
- If unable to calculate DAS at visit 11 or 12 (month 16/20), we used last DAS observation

Swollen Joint Count 44 (SJC44):

- If SJC44 at visit 13 (month 24) was missing, the patient was considered not to meet the primary endpoint (worst outcome)
- If SJC44 was missing at visit 11 or 12 (month 16/20), we used last SJC44 observation

Other binary endpoints were imputed with worst outcome.

The radiographic scores by the van der Heijde modified Sharp method (total, erosion and joint space narrowing) change from baseline was analyzed using median regression with baseline value, center and presence of anti-CCP as covariates. Estimates of treatment difference and corresponding confidence intervals were computed using 10 000 bootstrap replications. Missing values were imputed using the following rule:

- Imputation by linear interpolation was used when observations existed both before and after the missing value
- Imputation by linear extrapolation using the last two know observations was used when no later observation existed

Binary variables derived from the radiographic scores (e.g. progression or not progression) was derived from the imputed data. The imputation method for radiographic scores was changed from the original statistical analysis plan. The original plan was to handle missing data for radiographic scores using multiple imputations, similar to other continuous endpoints. The change of method was done in order to conform with the typical analyses performed for radiographic scores.

Other continuous variables were analyzed using analysis of covariance adjusted for baseline value in addition to center and presence of anti-CCP. Missing values were handled using multiple imputations with 10 imputations drawn from the observed distribution using the Markov-chain Monte Carlo method.

Post-hoc robustness analyses were performed on the primary and a selection of secondary endpoints. The first robustness analyses addressed the skewed distribution of women between the treatment groups by adding sex as a covariate in the logistic and median regression analyses (table S3). The second robustness analyses addressed if the handling of missing data for non-completers affected the results, by restricting the analyses to completers only (table S4).

Descriptive statistics are presented using imputed values (worst outcome) for dichotomous endpoints and non-imputed values for continuous endpoints.

All analyses were done using Stata version 14.0 (StataCorp. 2015. *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP.). All significance tests were two-sided, and we used 95% confidence limits. Secondary analyses were not adjusted for multiplicity.

Section 5. Summary narratives for malignancies (n=5)
1. A 66-year-old male in the conventional tight control group developed basal cell carcinoma an unknown date between four and 12 months after study start. During this period of time, she received triple synthetic DMARD therapy (methotrexate 20 mg weekly, salazopyrine 500-1000 mg twice daily, hydroxychloroquine 400 mg daily). Concomitant medication was folic acid. Medical history included unspecified cancer and osteoarthritis. The investigator considered the event not to be related to medications.

2. A 68-year-old male in the conventional tight control group was diagnosed with squamous cell carcinoma approximately 12 months after study start. Medical history included hypertension, diabetes, angina pectoris, myocardial infarction, cardiac surgery and lung disease. The patient was receiving 20 mg methotrexate weekly at the time of the event. Concomitant medications included metoprolol, lisinopril, atorvastatin, clopidogrel, acetylsalicylic acid, cetirizine, metformin, budesonide, mometasonefuroate, folic acid, vitamin B12 and B6. The patient was withdrawn from the study shortly after the occurrence of the event. He was treated with curative chemotherapy and radiation, and approximately four months after the occurrence of the event, the patient was reported recovered with no sequelae.

3. A 59-year-old female in the ultrasonography tight control group experienced a serious adverse event of cancer with liver metastases approximately two years after study start. The patient was receiving 20 mg methotrexate weekly at the time of the event. Concomitant medications included folic acid, colecalciferol and calcium. Medical history included pollen allergy, ankle fracture, periodical nausea, abdominal pain, elevated CRP and ESR. The patient was withdrawn from the study in response to the event and did not recover during the follow-up period. The investigator considered the event to be not related to the study drug.

4. A 49-year-old female in the conventional tight control group experienced a serious adverse event of breast cancer approximately four months after study start. Medical history included back pain. The patient was receiving 20 mg methotrexate weekly at the time of the event. Concomitant medications were folic acid, colecalciferol and calcium. The patient was withdrawn from the study five months after the onset of event and did not recover during the follow-up period. The investigator considered the event not likely related to the study drug.

5. A 70-year-old male in the ultrasonography tight control group experienced a serious adverse event of follicular lymphoma approximately two years after study start. Medical history included an unspecified type of cancer, cardiac disease, lung disease and arthrosis.
The patient was receiving tiotropium, formoterol, acetylsalicylic acid, metoprolol, atorvastatin, folic acid, colecalciferol and calcium. Methotrexate was stopped approximately one year prior to the onset of the event. The patient had not recovered at the time of reporting.

Section 6. Summary narratives for deaths (n=1)
1. A 66-year-old male with a medical history of hypertension in the ultrasonography tight control group died from pneumocystis jirovecii pneumonia approximately 10 months after study start. Suspect medications were methotrexate 22.5 mg weekly and etanercept 50 mg weekly. Concomitant medications included lisinopril, alendronate, folic acid, colecalciferol and calcium. He started treatment with methotrexate in February 2012 and etanercept in August 2012. He was in very good shape at the clinical visit in October, and had no side effects of the drugs. From November, however, he developed dyspnoea and was hospitalized. Pneumocystis jirovecii pneumonia was proven by PCR and widespread pulmonary fibrosis by computerised tomogram. The infection was treated, but due to the pulmonary fibrosis, further treatment was terminated. He died in January 2013. The national competent authority considered the event to be possibly related to etanercept.

Section 7. Statistical considerations: the conclusion of the ARCTIC study

The ARCTIC study did not reject the primary null hypothesis of the trial: "There is no difference in the probability of achieving complete DAS remission after 24 months of treatment between the two treatment regimens (applying vs not applying ultrasonography)". Failure to show an effect does not automatically imply a lack of effect of the intervention; the true effect might also be insufficiently large to be discovered by the trial. Negative trials can be divided into two categories: 1) True negative trials where the trial can rule out clinically important effects, and 2) Inconclusive trials where important clinical effects cannot be ruled out. In this section we discuss why we suggest that the ARCTIC trial is a true negative trial.

A central aspect in the assessment of a negative study is to consider the size of a clinically important potential effect. In our sample size calculations, we aimed to power the trial at 80% to detect a 20% difference between the interventions (Protocol). This was based on the remission rates in previous studies, in addition to discussions with clinicians and the study team regarding the effect size needed in order to introduce ultrasound in clinical practice. During the data collection in ARCTIC, two equivalence studies of biosimilar drugs in RA have been published, both with an equivalence margin of ±15%.[2 3] The setting in these studies were however very different, as the studies compared biological treatments, with response rates as the primary outcomes.

The estimated treatment difference of the primary endpoint was 3.3% with a 95% confidence interval of -7.1 to 13.7. The confidence interval is completely within both the ±20% and ±15% margin, ruling out a clinical important difference between the treatments according to both our estimate of an important clinical effect and the stricter definition used for assessment of biosimilar drugs. For the components of the primary endpoint, we can rule out a clinical important difference for the disease activity endpoints (no swollen joints and DAS remission) for both definitions, but the confidence interval for the difference in radiographic progression includes the ±15% margin.

If the study was to be repeated, the power to detect a 20% difference in the primary endpoint from 19% in the control group would have been 89%. The corresponding power with 15% difference would have been 68%. This further supports our conclusion.

Table S1. Treatment regimen in the ARCTIC trial

Visit (months)	**Treatment if no response** *(if response continue treatment at present step, see table S2)*
1 (0)	**A. Monotherapy* + Prednisolone:** 1. Methotrexate 15 mg/week, increase by 2.5 mg every 2nd week to target dose 20 mg/week, i.e. week 1+2 15mg, week 3+4 17.5 mg, week 5-8 20 mg (optional reduced dosage starting scheme for patients at risk for side effects: week 1 10 mg, week 2 12.5mg, week 3 15 mg, week 4 17.5mg, week 5-8 20 mg) 2. Concomitant folic acid 5 mg/week (1mg 5/7 days or 5 mg x 1/week) 3. Prednisolone 15 mg week 1, 10 mg week 2, 7.5 mg week 3, 5 mg week 4+5, 2.5 mg week 6+7 4. Calcium supplement 1000mg x 1 (while on prednisolone)
2 (1)	**A. Monitor start-up regimen (no changes in medication allowed unless due to AE)*** Joint injections allowed as indicated according to treatment arm.
3 (2)	**A. Optimize monotherapy*** Increase Methotrexate to 25-30 mg/week *Or increase SSZ/HCL/leflunomide dose*
4 (3)	**A. Monitor start-up regimen (no changes in medication allowed unless due to AE)*** Joint injections allowed as indicated according to treatment arm.
5 (4)	**B. Triple combination therapy (or other combination therapy if MTX not tolerated):**† 1. Add salazopyrine, step up over 4 weeks to 500mg 2 x 2 and 2. Add hydroxychlorochine 200mg 1 x 2
6 (6)	**B. Optimize triple combination therapy:**† Add Prednisolone 7.5 mg 1 x 1
7 (8)	**C. DMARD‡ and 1st biologic:**∫ 1. Highest tolerable dose MTX* and 2. Add 1st biologic (*according to current Norwegian guidelines)* **Or SSZ/HCL/leflunomide if MTX not tolerated*
8 (10)	**C. DMARD and 1st biologic:** Adjust dose/interval of 1st biologic
9 (12)	**D. DMARD‡ and 2nd biologic:** Switch to 2nd biologic (*according to current Norwegian guidelines)*
10 (14)	**D. DMARD‡ and 2nd biologic:** Adjust dose/interval of 2nd biologic
11 (16)	**E. DMARD‡ and 3rd biologic:** Switch to 3rd biologic (*according to current Norwegian guidelines)*
12 (20)	**E. Optimize DMARD and 3rd biologic plus prednisolone:** Adjust dose/interval of 3rd biologic and/or add prednisolone 7.5mg
13 (24)	**F.** Continue medication according to standard clinical care

***** *If MTX is not tolerated, switch to subcutaneous methotrexate), then continue according to scheme. In case of AE or not tolerated even in low dose subcutaneous, switch to salazopyrine or hydroxychlorochine monotherapy (standard dosage) if low disease activity, or leflunomide 20 mg in case of moderate or high disease activity (loading dose 40mg x 1 for 3 days, then 20 mg per day).*

† *In patients with high disease activity and risk factors for progressive joint destruction (ACPA or RF-positive and either erosions on CR or baseline RAMRIS bone marrow edema score >2) a rescue option is available which includes moving to the next step, i.e. introduce 1st biologic (treatment C at visit #5, without prescribing treatment B).*

‡ *In case of no tolerance for any conventional DMARD, this can be omitted if the biologic drug chosen has indication for monotherapy (e.g. tociluzimab).*

∫ *Requirement for adding biologic: There must be objective signs of ongoing inflammation, i.e. either elevated ESR/CRP (>UNL, and not due to other disease/infection) or SJC≥1 (or PD score >1 in US arm).*

Table S2. ARCTIC decision rules*

	Current DAS	No response†	Response‡	Reached target (DAS < 1.6)
Conventional tight control	≤ 2.4	Change of DAS < 0.6	Change of DAS ≥ 0.6	DAS < 1.6 and no swollen joints
	> 2.4	Change of DAS < 1.2	Change of DAS ≥ 1.2	
Ultrasound tight control	≤ 2.4	Change of DAS < 0.6 or <10% decrease of US total score	Change of DAS ≥ 0.6 and ≥ 10% decrease of US total score	DAS < 1.6 and no swollen joints and no joints with power Doppler synovitis
	> 2.4	Change of DAS < 1.2 or <20% decrease of US total score	Change of DAS ≥ 1.2 and ≥ 20% decrease of US total score	
Action		**Change therapy**	**Continue current medication**	**Continue current medication‡**

* *To be applied at all visits except visit 2 and visit 4.*

† *Both in cases of response and no response should clinically swollen joints be i.a. injected with steroids when indicated, up to the maximum allowed dosage per visit (80 mg triamcinolone hexacetonid). In group B joints with PD-signal on US is an additional target.*

‡ *If sustained remission ≥ 12 months, step-down to monotherapy MTX. If continued sustained response after this, decrease MTX by 2.5mg/week per 2 months.*

Table S3. Analyses of Primary and Key Secondary Endpoints Adjusted for Sex*

Variable	Ultrasound tight control (n=118)	Conventional tight control (n=112)	Difference (95% CI)	p-value
Primary endpoint – no. (%) †	26 (22.0)	21 (18.8)	2.9 (-7.7 to 13.6)	0.55
Components of primary endpoint				
No swollen joints – no. (%) ‡				
At 16, 20 and 24 months	62 (52.5)	61 (54.5)	-3.4 (-16.5 to 9.7)	0.61
DAS remission – no. (%) ‡				
At 16, 20 and 24 months	64 (54.2)	58 (51.8)	4.0 (-9.1 to 17.2)	0.55
No radiographic progression – no. (%) ‡				
Between 16-24 months	49 (41.5)	39 (34.8)	6.6 (-6.2 to 19.4)	0.32
Radiographic joint damage				
Δ Modified Sharp score at 24 months ǁ	1 (0 to 2.5)	1.5 (0.5 to 3)	-0.43 (-0.92 to 0.07)	0.09

* *All results were derived from the full analysis set, which included all randomized patients who underwent at least one visit after baseline. Median values are given with interquartile range (IQR). DAS=Disease Activity Score.*

† *The primary endpoint was the proportion of patients meeting all the 3 following criteria: 1) Sustained clinical remission, defined as DAS<1.6 at 16, 20 and 24 months 2) No swollen joints at 16, 20 and 24 months (44 Swollen Joint Count) and 3) No progression (<0.5 units) in van der Heijde-modified total Sharp Score between 16 and 24 months.*

‡ *Missing data before 24 months imputed using last observation carried forward, and missing data at 24 months imputed using worst outcome.*

ǁ *Values are observed, unadjusted median values given with interquartile range (IQR). Treatment difference is derived from a median regression model. Missing data were imputed using linear intra- and extrapolation.*

Table S4. Analyses of primary and key secondary endpoints, completer analysis set*

Variable	Ultrasound tight control (n=104)	Conventional tight control (n=100)	Difference (95% CI)	p-value
Primary endpoint – no. (%) †	26 (25.0)	21 (21.0)	4.0 (-7.5 to 15.5)	0.50
Components of primary endpoint				
No swollen joints – no. (%) ‡				
At 16, 20 and 24 months	62 (59.6)	61 (61.0)	-1.3 (-14.8 to 12.0)	0.84
DAS remission – no. (%) ‡				
At 16, 20 and 24 months	64 (61.5)	58 (58.0)	3.5 (-9.9 to 17.0)	0.52
No radiographic progression – no. (%) ‡				
Between 16-24 months	49 (47.1)	39 (39.0)	8.1 (-5.4 to 21.7)	0.24
Radiographic joint damage				
Δ Modified Sharp score at 24 months ǁ	1 (0 to 2.5)	1.5 (0.5 to 3)	-0.45 (-0.86 to -0.39)	0.03

* *All results were derived from the completer analysis set, which included all randomized patients who underwent at least one visit after baseline and who completed the study. Median values are given with interquartile range (IQR). DAS=Disease Activity Score.*

† *The primary endpoint was the proportion of patients meeting all the 3 following criteria: 1) Sustained clinical remission, defined as DAS<1.6 at 16, 20 and 24 months 2) No swollen joints at 16, 20 and 24 months (44 Swollen Joint Count) and 3) No progression (<0.5 units) in van der Heijde-modified total Sharp Score between 16 and 24 months.*

‡ *Missing data before 24 months imputed using last observation carried forward, and missing data at 24 months imputed using worst outcome.*

ǁ *Values are observed, unadjusted median values given with interquartile range (IQR). Treatment difference is derived from a median regression model. Missing data were imputed using linear intra- and extrapolation.*

Table S5. Gender distribution by centre and intervention

	Proportion of female patients (%)	
Centre	**Ultrasound tight control (n=118)**	**Conventional tight control (n=112)**
1	29/37 (78.4%)	21/35 (60.0%)
2	5/9 (55.6%)	3/8 (37.5%)
3	6/7 (85.7%)	5/6 (83.3%)
4	9/14 (64.3%)	5/12 (41.7%)
5	2/4 (50.0%)	1/5 (20.0%)
6	7/9 (77.8%)	4/9 (44.4%)
7	5/7 (71.4%)	4/7 (57.1%)
8	9/9 (100.0%)	4/7 (57.1%)
9	4/5 (80.0%)	1/3 (33.3%)
10	5/11 (45.5%)	5/15 (33.3%)
11	3/6 (50.0%)	4/5 (80.0%)

Table S6. Serious Adverse Events over 24 months (One Patient per Term)

MedDRA System Organ Class	Ultrasound tight control (n=118)	Conventional tight control (n=112)
Gastrointestinal disorders		Volvulus*
Infections and infestations	Abscess, bacterial* Pneumonia* Pneumocystis jirovecii pneumonia‡	Localised infection* Abscess*
Musculoskeletal and connective tissue disorders		Arthralgia*
Neoplasms benign, malignant and unspecified (incl cysts and polyps)	Follicle centre lymphoma, follicular grade I, II, III † Metastases to liver*	Breast cancer*
Nervous system disorders		Syncope*
Renal and urinary disorders	Nephrolithiasis*	
Surgical and medical procedures	Percutaneous coronary intervention†	Hospitalisation∫

* *The patient was receiving methotrexate.*

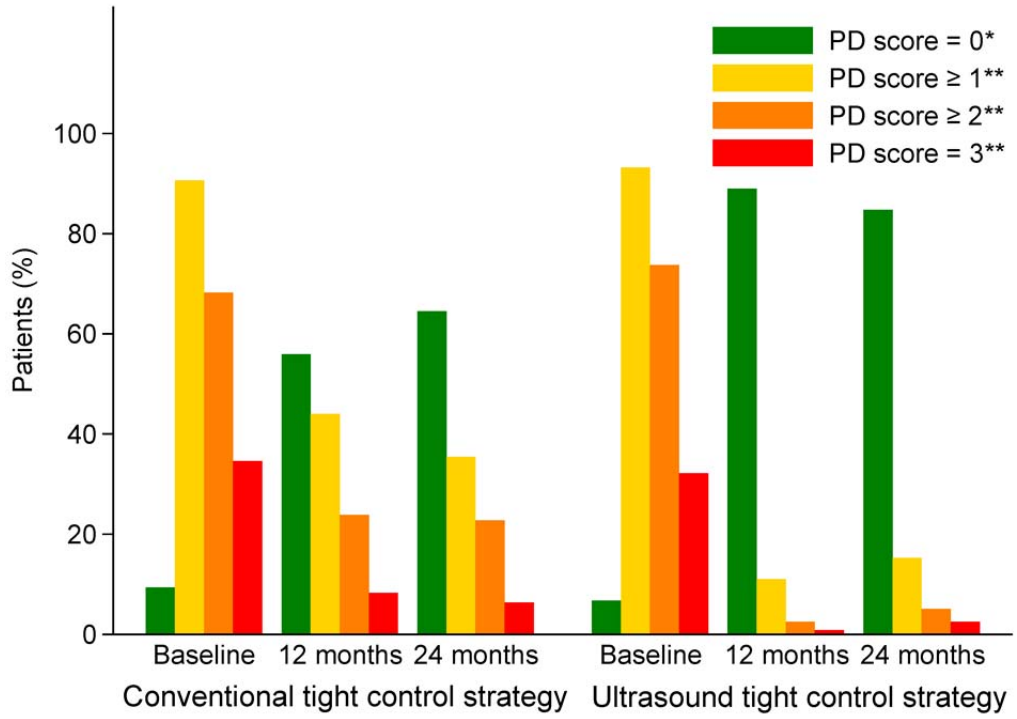
† *The patient did not receive any study medication, methotrexate was stopped one year prior to the diagnosis of the follicle centre lymphoma. The same patient received percutaneous coronary intervention four months after the baseline visit, and was then treated with methotrexate (this was prior to the diagnosis of lymphoma).*

‡ *The patient was receiving etanercept and methotrexate.*

∫ *The patient did not receive any study medication (methotrexate was stopped approximately three months prior to the onset of the event).*

Figure S1. Power Doppler activity at baseline, 12 months and 24 months

The histogram shows the proportion of patients with no power Doppler activity in any joint (PD score =0, colour green) and proportion of patients with at least one joint with power Doppler activity grade 1 (PD score ≥ 1, colour yellow), at least one joint with power Doppler activity grade 2 (PD score ≥ 2, colour orange) and at least one joint with power Doppler activity grade 3 (PD score = 3, colour red).



* in all joints ** in any joint

References

1. Hammer HB, Bolton-King P, Bakkeheim V, et al. Examination of intra and interrater reliability with a new ultrasonographic reference atlas for scoring of synovitis in patients with rheumatoid arthritis. AnnRheumDis 2011;**70**(11):1995-98.
2. Yoo DH, Hrycaj P, Miranda P, et al. A randomised, double-blind, parallel-group study to demonstrate equivalence in efficacy and safety of CT-P13 compared with innovator infliximab when coadministered with methotrexate in patients with active rheumatoid arthritis: the PLANETRA study. Ann Rheum Dis 2013;**72**(10):1613-20.
3. Emery P, Vencovsky J, Sylwestrzak A, et al. A phase III randomised, double-blind, parallel-group study comparing SB4 with etanercept reference product in patients with active rheumatoid arthritis despite methotrexate therapy. Ann Rheum Dis 2015.

Clinical Study Protocol

Full title of study:
Aiming for remission in rheumatoid arthritis: a randomized trial examining the benefit of ultrasonography in a clinical tight control regimen

Short title:
Aiming for Remission in RA - the ARCTIC trial

PROTOCOL APPROVAL

Protocol number / version:	DIA2010-1 / ver 1.7
Date:	See the latest date below
National Medical Leader: Date: 20/9 2010	Tore K. Kvien, professor PhD
Biostatistician: Date: 20/9-2010	Inge Christoffer Olsen, PhD
Special advisor imaging/design Date: 20-9-2010	Désirée van der Heijde, professor PhD

Contact information

Principal investigator/ project leader:	Espen A. Haavardsholm, MD PhD Diakonhjemmet Hospital Dept. of Rheumatology Box 23 Vinderen 0319 Oslo, Norway
Monitors:	Hege Øvergaard, Seksjon for GCP, Oslo University Hospital Ingunn Heie Anundskås, Innovest AS
Coordinating centre:	Diakonhjemmet Hospital, Dept. of Rheumatology (South-Eastern Health Authority Rheumatology Research Group and EULAR centre of excellence in Rheumatology), Oslo, Norway
Participating centres:	Universitetssykehuset i Nord-Norge, St. Olav Hospital, Helse Sunnmøre HF Ålesund, Haukeland Universitetssykehus Bergen, Sørlandet Sykehus Kristiansand HF, Haugesund Sanitetsforenings Revmatismesykehus AS , Martina Hansens Hospital, Betanien Hospital Skien, Sykehuset Østfold HF Moss, Lillehammer Revmatismesykehus, Vestre Viken HF Buskerud, Privatpraksis Bendvold/Dovland Kristiansand.

Table of Contents

Glossary of abbreviations

ACR	American College of Rheumatology
AE	Adverse event
ALT	Alanine aminotransferase
Anti-CCP	Antibodies to Cyclic Citrullinated Peptide
AST	Aspartate aminotransferase
AUC	Area under the curve
CDAI	Clinical Disease Activity Index
CI	Confidence Interval
CR	Conventional Radiography
CRP	C-Reactive Protein
CT	Computer Tomography
DAS	Disease Activity Score (in 44 joints)
DIP	Distal InterPhalangeal
DMARD	Disease Modifying Anti-Rheumatic Drug
ECG	Electrocardiogram
eCRF	Electronic case report form
ESR	Erythrocyte Sedimentation Ratio
Et al.	and others
EULAR	European League Against Rheumatism
FOV	Field Of View
GCP	Good Clinical Practice
GH	General Health
GI	Gastrointestinal
HAQ	Health Assessment Questionnaire
HBsAg	Hepatitis B surface antigen
HCV	Hepatitis C virus
HLA	Human Leukocyte Antigen
ICH	International Conference on Harmonization
IEC	Independent ethics committee
ILAR	International League of Associations for Rheumatology
IM	Intramuscular
ITT	Intent-to-treat
JSN	Joint Space Narrowing
LEF	Leflunomide
MCP	MetaCarpoPhalangeal
MRI	Magnetic Resonance Imaging
MTP	MetaTarsoPhalangeal
MTX	Methotrexate
NSAID	Non-steroidal anti-inflammatory drug
OMERACT	Outcome MEasures in Rheumatoid Arthritis Clinical Trials
RA	Rheumatoid Arthritis
RAMRIS	Rheumatoid Arthritis Magnetic Resonance Imaging Score
RF	Rheumatoid Factor
PIP	Proximal InterPhalangeal
PPD	Purified protein derivative tuberculin skin test
QoL	Quality of Life
SAE	Serious adverse event
SDAI	Simplified Diseases Activity Index

SD	Standard Deviation
SDD	Smallest Detectable Difference
SJC	Swollen Joint Count
SPSS	Statistical Package for the Social Sciences
SRM	Standardized Response Mean
STIR	Short Tau Inversion Recovery
TB	Tuberculosis
TJC	Tender Joint Count
TNF	Tumor Necrosis Factor
ULN	Upper limit of normal
US	UltraSonography
VAS	Visual Analogue Scale
vdHSS	van der Heijde modified Sharp Score
WBC	White blood cell
WHO	World Health Organization

1. Background and rationale

1.1 Introduction

This protocol describes a randomized, open, prospective, multi-centre, parallel-group, clinical study designed with a strategic treatment decision component to evaluate the efficacy and added value of applying ultrasonography (US) versus standard clinical assessment without the use of information from US in patients with early rheumatoid arthritis, with respect to achieving clinical remission and inhibition of radiographic progression.

1.2 Background

Rheumatoid arthritis (RA) is a chronic, systemic disease leading to joint damage and eventually loss of function. A commonsense approach to the management of a persistent, progressive, damaging condition such as RA would seem to intervene before the onset of damage, at a stage when the disease process still may be reversible. Such a phase of the disease has been described as a "window of opportunity" for intervention (1-3). A growing body of evidence has emphasized the consistent clinical and radiographic benefits of early, aggressive treatment of RA (4-6). These and other studies confirm that all therapies - monotherapy, combinations of disease modifying antirheumatic drugs (DMARD) and biologics - work better in early disease than in long-established RA.

Increasing research has focussed on treatment strategies to slow radiological progression and prevent loss of function. Several studies have shown that intensive patient management improves disease activity and halts radiographic progression better than routine patient management (7-13). A recent paradigm shift is to aim therapeutically for remission, instead of mere improvement in symptoms and signs (7;14-16). To reach this goal, treatment is intensified by combining various DMARDs, including biological agents, within a tight control regimen early in the disease (17). Methotrexate (MTX) is usually selected as initial therapy and as an anchor drug in combination therapies. TNF-inhibitors in combination with MTX are usually the first choice of biological therapy.

Tight control may be defined as a treatment strategy tailored to the disease activity of individual patients with RA with the aim of achieving a predefined level of low disease activity or preferably remission within a reasonable period of time (18). This strategy mimics the successful approach of treating to a pre-defined target used in other areas of medicine, such as cardiology and diabetes care and also includes a program for disease monitoring and adjustment of therapy if the pre-defined target is not reached (19).

In daily clinical practice, therapy choices and dose adjustments are based on the overall view of the individual rheumatologist. Disease activity is often determined ‘‘ad hoc’’ on the basis of a few objective and subjective clinical variables, because time at the outpatient clinic is limited. However, patients with rheumatoid arthritis who seem to be doing well at first glance often show more signs of disease activity when a full joint examination is performed. Recent studies suggest that validated composite disease activity measurements that include joint assessments should be mandatory in clinical practice to drive treatment decisions. The primary target should be the state of clinical remission, and if the desired target is not reached rapidly, treatment should be adjusted. The desired treatment target should be sustained over time.

Musculoskeletal ultrasonography is an imaging modality used for investigation and management of rheumatic diseases which application is rapidly growing. It has a number of practical advantages over other advanced imaging techniques such as magnetic resonance imaging (MRI), including low cost, good accessibility, and ability to scan multiple joints in a relatively short period of time. As well, it can simultaneously image bone and soft tissue. Pathologies that can potentially be visualised using US include bone erosions, cartilage, synovitis, tenosynovitis, and enthesopathies. The ability to perform

"dynamic" evaluation of tendons and help in guiding needle positioning in interventional manoeuvres are some of the other reasons for its success. More specifically, the power Doppler (PD) signal has proved to be a simple and promising tool for short-term monitoring of synovial vascularity changes in RA patients, and it is also a promising tool for prognostification in early RA (20;21).

Ultrasonography may also be an important assessment in patients in apparent clinical remission. In a study of patients in clinical remission by Brown et al, it was demonstrated that in patients who fulfilled the DAS28 remission criteria, 84% and 51% had synovial hypertrophy and increased PD signal, respectively, on musculoskeletal US, and 88% had demonstrable synovitis on MRI. Also the PD signal predicted radiographic progression, and thus the target of "imaging remission", i.e. clinical remission with also a lack of inflammatory activity assessed by imaging methods, may be an attractive target to stop radiographic deterioration. However, the additional benefit of this approach when used in combination with clinical assessments in terms of patient outcomes has not been demonstrated. Thus, clarification is needed if the use of ultrasonography is cost-effective in RA clinical practice.

1.3 Rationale for the study

There have been major advances in the treatment of RA over the last decade, with access to new therapeutic drugs, especially biologic agents. Other factors, including increased knowledge of prognostic factors and optimization of treatment regimens, have led to better care for patients with RA. However, despite these advances, approximately 30 to 40% of patients fail to achieve an optimal clinical response. The ultimate goal of RA treatment is remission, i.e. an absence of signs and symptoms of the disease. Remission may currently be an achievable goal if adequate treatment is started early in disease, and if the patient is monitored in a tight control regimen specifically aiming at remission (3;4;6;14) (7;8;13). Various definitions of remission have been proposed, including ACR remission criteria, cut-points for remission based on DAS and DAS28, as well as newer tools such as cut points for remission according to SDAI and CDAI. However, all available remission criteria may ignore important aspects of RA, especially sub-clinical inflammatory activity and radiographic progression.

In many other areas of medicine, such as diabetes care or cardiology, treatment targets have been defined to improve outcomes. In diabetes, targeting HbA1c levels of ≤7% and control of hypertension have led to better outcomes over the past decade (22;23). Likewise, aiming at predefined levels of blood pressure improves cardiovascular risk (24) and so does control of serum lipid levels according to established guidelines (25). This strategy significantly improves outcomes of these diseases. In rheumatology, this strategy has recently been adopted (16;19), and an international panel of experts has formulated recommendations on how to treat rheumatoid arthritis to the target of remission (16). Data from several studies suggest that a strategic approach to treating RA in the clinic targeting low disease activity may benefit a large proportion of patients and clearly improve the outcome of RA (7;8;13;26). However, data from studies targeting remission is still missing, and the optimal way of reaching this target is not known. There is also great debate as to what definition of remission should be used, i.e. what should be the optimal target.

Longitudinal disease monitoring is traditionally performed with conventional clinical measures reflecting the inflammatory activity. The DAS and DAS28 have been major advances in evaluation of disease activity in RA. However, patients in remission according to DAS28 (<2.6) may include patients with a considerable number of tender joints, and Landewe et al. concluded that that DAS28 remission at a cut-off level of 2.6 has insufficient construct validity and should be used with caution in clinical practice and clinical trials (27). In rheumatology clinical practice throughout Europe the use of ultrasonography (US) in regular clinical practice is growing and the use of this imaging modality has been implemented in clinical practice in several rheumatology departments in Norway. US provides information of sub-clinical synovitis, may reveal radiographically occult bone erosions and power-

Doppler (PD) signal may predict erosive progression (20;21;28;29) and disease flare (30). In a recent study by Brown et. al, 102 patients in clinical remission were assessed by various imaging modalities. Despite being in clinical remission, 19% of patients deteriorated radiographically, and this progression was largely explained by PD signal (with a 12 times higher odds ratio of progression in joints with increased PD signal) (31). Also, the use of US allows for direct visualization of the needle within the joint cavity during intra-articular joint injections, thus allowing for a more accurate procedure than blind injections.

Despite the apparent advantages of US described above, this new technology also has certain limitations, including large inter-observer variation, and undoubtedly the use of this tool is time-consuming. It is not known whether implementation of US in addition to traditional clinical measures in the disease monitoring will improve the outcome in patients with rheumatoid arthritis.

This study will assess if the use of a treatment strategy incorporating US information of sub-clinical synovitis (both grey-scale and power-Doppler synovitis) targeting both DAS-remission and imaging remission combined with more accurate i.a. injections will allow for better treatment results of patients with RA than a conventional treatment strategy based on clinical assessments alone and targeting DAS remission. We will also compare the strategies with regard to structural damage, as well as assess the cost-effectiveness of the interventions.

2. Objectives

2.1 Primary objective

The primary objective of this treatment strategy study is to assess the effect of applying ultrasonography versus not applying ultrasonography in a clinical tight control regimen in patients with early RA with respect to achieving clinical remission and non-progression of structural damage.

2.2 Secondary objectives

The secondary objectives of this study are to assess the two treatment regimens regarding:

- Efficacy of the two treatment regimens with respect to various measures of progression of structural damage
- Efficacy of the two treatment regimens with respect to various measures of low disease activity and remission
- Health economics (evaluation of direct and indirect costs in the treatment groups, cost-effectiveness of both treatment arms)
- Possible prognostic factors for structural damage/reaching remission, including proteomics, imaging and biomarkers
- Strength of associations over time both for damage, inflammation and physical function
- Time to reach various levels of disease activity/remission
- The burden of illness (disease activity, damage, disability, cost of illness) in patients with early arthritis
- Efficacy with respect to physical function, health related quality of life (HRQoL) and pain
- Aspects of imaging remission (especially if specific levels of MRI inflammation or US inflammation are related to non-progression of structural damage)
- Work performance (including aspects of "absenteeism" and "presenteeism")
- MRI measures of damage and inflammation in the two groups, and how MRI features evolve over time and how they relate to radiographic damage and US evaluation
- The feasibility of applying ultrasonography in a tight control dosing regimen in clinical practice

- The area under the curve of measures of disease activity/inflammation (clinical, US and MRI)
- The performance of various outcome measures in patients with early arthritis
- Patient satisfaction
- Number and kind of adverse events in both groups

3. Study design

3.1 Overview of study design and treatment regimen

This is an open, national, multicenter, 2-arm, randomized, parallel group, prospective clinical study of 2 years duration of a tight control regimen, with and without applying musculoskeletal ultrasonography, aiming at remission in patients with early RA. Patients in both groups are treated according to (the same) fixed treatment protocol. Treatment adjustments (including i.a. injections) may be made at every visit according to a pre-specified dosing regimen, described in table 1. The decision rules are described in table 2. The target is remission defined as DAS<1.6, plus the following criteria (different for the two treatment arms):

A. Non-US arm: No swollen joints.
B. US arm: No swollen joints and no joints with PD-signal.

The US standardized score has been developed in a pilot study (not published), and includes assessments of the following 32 joints: MCPs I-V, wrist (radio-carpal, radio-ulnar and inter-carpal), elbow, knee, talo-crural and MTP I-V bilaterally. The time for scoring and recording both a grey-scale and a power-Doppler semi-quantitative score of 0-3 for each joint takes 15 minutes (median).

The decision of whether to adjust medication is based on change in and the level of the Disease Activity Score (DAS) based on 44 joints, calculated by a trained research nurse. If the patient does not respond as described in figure 2, the treating physician immediately adjusts the therapy by proceeding to the next step in the treatment regimen. If a patient responds or has reached the target, current medication is continued. In the US group, the physician may overrule the decision based on DAS and proceed to the next step based on US findings (as described in figure 2).

In both groups, clinically swollen joints should be treated by i.a. steroids when indicated. In the US group an additional target is all joints with PD signal, and additionally all injections should be US-guided. For both groups, i.a. injection of only tender joints is not allowed. Likewise, inflamed tendon sheets may be injected in the non-US arm as clinically indicated – in the US arm tenosynovitis should be verified by US before injections are allowed. Maximum dosage triamcinolone hexacetonid per visit is 80 mg which can be distributed within joints and tendon sheets as decided by the treating rheumatologist.

Table 1. Treatment regimen

Visit (months)	Treatment if no response (if response continue treatment at present step, see fig 2)
1 (0)	**A. Monotherapy* + Prednisolone:** 1. Methotrexate 15 mg/week, increase by 2,5 mg every 2nd week to target dose 20 mg/week, i.e. week 1+2 15mg, week 3+4 17.5 mg, week 5-8 20 mg (optional reduced dosage starting scheme for patients at risk for side effects: week 1 10 mg, week 2 12,5mg, week 3 15 mg, week 4 17,5mg, week 5-8 20 mg) 2. Concomitant folic acid 5 mg/week (1mg 5/7 days or 5 mg x 1/week) 3. Prednisolone 15 mg week 1, 10 mg week 2, 7,5 mg week 3, 5 mg week 4+5, 2,5 mg week 6+7 4. Calcium supplement 1000mg x 1 (while on prednisolone)
2 (1)	**A. Monitor start-up regimen (no changes in medication allowed unless due to AE)*** Joint injections allowed as indicated according to treatment arm.
3 (2)	**A. Optimize monotherapy*** Increase Methotrexate to 25-30 mg/week *Or increase SSZ/HCL/leflunomide dose*
4 (3)	**A. Monitor start-up regimen (no changes in medication allowed unless due to AE)*** Joint injections allowed as indicated according to treatment arm.
5 (4)	**B.** Triple combination therapy (or other combination therapy if MTX not tolerated):** 1. Add salazopyrine, step up over 4 weeks to 500mg 2 x 2 and 2. Add hydroxychlorochine 200mg 1 x 2
6 (6)	**B. Optimize triple combination therapy:**** Add Prednisolone 7,5 mg 1 x 1
7 (8)	**C. DMARD*** and 1st biologic**** (according to LIS guidelines):** 1. Highest tolerable dose MTX * and 2. Add 1st biologic **Or SSZ/HCL/leflunomide if MTX not tolerated*
8 (10)	**C. DMARD and 1st biologic:** Adjust dose/interval of 1st biologic
9 (12)	**D. DMARD*** and 2nd biologic (according to LIS guidelines):** Switch to 2nd biologic (*according to current LIS-guidelines)*
10 (14)	**D. DMARD*** and 2nd biologic:** Adjust dose/interval of 2nd biologic
11 (16)	**E. DMARD*** and 3rd biologic (according to LIS guidelines):** Switch to 3rd biologic (*according to current LIS-guidelines)*
12 (20)	**E. Optimize DMARD and 3rd biologic plus prednisolon:** Adjust dose/interval of 3rd biologic and/or add prednisolon 7,5mg
13 (24)	**F.** Continue medication according to standard clinical care

***** *If MTX is not tolerated, switch to subcutaneous methotrexate (metoject), then continue according to scheme. In case of AE or not tolerated even in low dose subcutaneous, switch to salazopyrine or hydroxychlorochine monotherapy (standard dosage) if low disease activity, or leflunomide 20 mg in case of moderate or high disease activity (loading dose 40mg x 1 for 3 days, then 20 mg per day).*

*** In patients with high disease activity and risk factors for progressive joint destruction (ACPA or RF-positive and either erosions on CR or baseline RAMRIS bone marrow edema score >2) a rescue option is available which includes moving to the next step, i.e. introduce 1st biologic (treatment C at visit #5, without prescribing treatment B).*

**** In case of no tolerance for any conventional DMARD, this can be omitted if the biologic drug chosen has indication for monotherapy (e.g. tociluzimab)*

***** Requirement for adding biologic: There must be objective signs of ongoing inflammation, i.e. either elevated ESR/CRP (>UNL, and not due to other disease/infection) or SJC≥1 (or PD score >1 in US arm)*

Table 2. Definition of decision rules (to be applied at all visits except #2 and #4).

	Current DAS	No response*	Response*	Reached target (DAS < 1.6)
Conventional tight control	≤ 2.4	Change of DAS < 0.6	Change of DAS ≥ 0.6	DAS < 1.6 and no swollen joints
	> 2.4	Change of DAS < 1.2	Change of DAS ≥ 1.2	
US tight control	≤ 2.4	Change of DAS < 0.6 or <10% decrease of US total score	Change of DAS ≥ 0.6 and ≥ 10% decrease of US total score	DAS < 1.6 and no swollen joints and no joints with power Doppler synovitis
	> 2.4	Change of DAS < 1.2 or <20% decrease of US total score	Change of DAS ≥ 1.2 and ≥ 20% decrease of US total score	
Action		Change therapy	Continue current medication	Continue current medication**

** Both in cases of response and no response should clinically swollen joints be i.a. injected with steroids when indicated, up to the maximum allowed dosage per visit (80 mg triamcinolone hexacetonid) . In group B joints with PD-signal on US is an additional target.*
*** If sustained remission ≥ 12 months, step-down to monotherapy MTX. If continued sustained response after this, decrease MTX by 2,5mg/week per 2 months.*

3.1.1 Rationale for study design

The primary goal in RA therapy is to abrogate inflammation, prevent joint damage and normalize functioning. Abrogation of inflammation is usually associated with a halt of structural joint damage and maximum reversal of functional impairment. Several studies have shown that treatment by algorithm to reach a pre-defined target, applying regular measurements of disease activity and adjusting therapy accordingly, within a tight control management system, provides the best therapeutic strategy for delivering optimal care to patients with RA.

Tight control implies regular and frequent visits to a rheumatologist early in the disease. In this study, the patients will be assessed monthly the first 4 months, then bi-monthly for the next 12 months, then every 4 months the final 8 months of the study.

The study is designed to assess whether applying ultrasonography within a tight control scheme will lead to better outcome, both with regard to inflammation/disease activity (i.e. clinical remission without any swollen joints) and structural damage (i.e. non-progression of vdHSS).

3.1.2 Rationale for treatment algorithm

The treatment algorithm has been developed through a series of meetings with both researchers and clinicians, and is in line with current local clinical practice and evidence based medicine, as well as international guidelines for the management of early RA, and is designed to provide optimal care for both treatment arms (11;26;32-34) (8;9;13;15;35-41).

3.1.3 End of study

The study period is 24 months, with a possible extension to 5 years which will be decided by the study committee. The end of the trial is defined as the date of the last visit of last participating patient in this study. The possible extension to 5 years will be to further assess both primary and secondary outcomes, with a focus on structural damage, physical function and occurrence of extraarticular

manifestations, as well as economic evaluations, and will imply an amendment of the protocol which will be assessed by the regional ethics committee.

3.2 Number of patients and assignment to treatment groups

Approximately 240 patients with early RA will be recruited over an anticipated recruitment period of approximately 18-24 months (for power calculations see below).

3.3 Patient randomization

Patients will be randomly assigned (in a 1:1 ratio) to the treatment groups, applying a block randomization scheme stratified by participating centre and anti-CCP status at baseline (positive or negative).

3.4 Centers

Approximately 8-12 centres will participate in this study with approximately 10 -80 patients per centre (Based on historical data from the NOR-DMARD registry, approximately 40-50 patients will be recruited per year at Diakonhjemmet Hospital). The other centres taking part in this project will primarily be recruited from hospital based rheumatology centres in Helse-Sør Øst and in Helse-Midt, based on our previous successful cooperation with these rheumatology departments in the multi-centre studies NOR-DMARD and NOR-VEAC.

4. Study population

4.1 Overview

Adult men and women with early RA according to the 2010 ACR/EULAR classification criteria (42)(see appendix for details) in need of a DMARD will be included in this study.

4.2 Inclusion criteria

To be eligible for this study, patients must meet all of the following criteria:
1. Male or non-pregnant, non-nursing female
2. > 18 years of age and < 75 years of age
3. Patients classified as having RA (according to new ACR/EULAR criteria)
4. Disease duration less than 2 years (defined as time from 1st joint swelling)
5. The treating rheumatologist decides the patient requires DMARD-treatment
6. The patient has taken no prior DMARD
7. Patients able and willing to give written informed consent and comply with the requirements of the study protocol

4.3 Exclusion criteria

Patients with any of the following criteria will not be eligible to participate in the study:
1. Abnormal renal function (serum creatinine > 142 µmol/L in female and > 168 µmol/L in male, or GFR < 40 mL/min/1.73 m^2).
2. Abnormal liver function (ASAT/ALAT > 3* normal), active or recent hepatitis, cirrhosis.
3. Major co-morbidities like severe malignancies, severe diabetic mellitus, severe infections, uncontrollable hypertension, severe cardiovascular disease (NYHA class 3-4) and/or severe respiratory diseases.
4. Leukopenia and/or thrombocytopenia.
5. Inadequate birth control conception, pregnancy, and/or breastfeeding.
6. Indications of active TB

7. Psychiatric or mental disorders, alcohol abuse or other abuse of substances, language barriers or other factors which makes adherence to the study protocol impossible.

4.4 Concomitant medication and treatment

Concomitant medication will be recorded in the CRF, with particular attention on registering any concomitant antirheumatic and pain medications.

NSAIDs, coxibs as well as intra-articular injections with corticosteroids as described earlier are permitted. The choice and dosage of NSAIDs/coxibs will be at the discretion of the treating rheumatologist. Analgesics up to the maximum recommended dose may be used for pain relief as required. Patients should avoid analgesics within 24 hours prior to a visit if possible.

All patients will receive vitamin D and calcium supplement during treatment with corticosteroids ≥ 7,5mg, and postmenopausal women and older men (>70 year) will be considered for a bisphosphonate according to general guidelines. IV or IM corticosteroids should be avoided during the study. Oral corticosteroids are allowed as described in Table 1. Other DMARDs than those described in Table 1 should be avoided.

4.5 Criteria for premature withdrawals

Patients have the right to withdraw from the study at any time for any reason. In the case that a patient decides to prematurely withdraw from the study, he/she should be asked if they can still be contacted for further information, so that a final evaluation can be made with an explanation of why the patient is withdrawing from the study, including assessment of possible AE. Although a subject is not obliged to give his/her reason(s) for withdrawing prematurely from a trial, the investigator should make a reasonable effort to ascertain the reason(s), while fully respecting the subject's rights.

If possible, at the last visit of the patient, all assessments of the final visit will be done.

5. Schedule of assessments and procedures

The visits will be carried out according to the schedule shown in Table 3, with assessments as shown at baseline, then every month for 4 months, then every other month the next 12 months and for the remaining 8 months every 4 months, i.e. a total of 13 visits over a 24 month period.

The data collection will include assessments as set out in figure 3, including swollen joint assessments of 44 joints and Ritchie articular index, laboratory tests, vital signs, standardized assessments of health related quality of life and physical function and imaging procedures. Standardized ultrasonography examination in the US treatment arm will be made at every visit. Patients in the non-US group will be examined at baseline, 12 months and 24 months. The results from the baseline and 12 months examinations will be blinded to the patient and treating physician. MRI of metacarpo-phalangeal joints and wrist of dominant hand (if available at participating centre) will be performed at baseline and after 3, 6, 12, 16 and 24 months. Bone density measurements (DEXA hip and spine) will be performed at baseline and after 3, 12 and 24 months. Conventional radiographs (CR) of hands and feet will be taken at baseline and after 3, 6, 12, 16 and 24 months. Assessments of the imaging procedures except for US will be performed by trained experts at Diakonhjemmet Hospital or cooperating institutions based on methods which have met internationally validated standards.

The sequence of assessments and procedures at each visit is to be standardized as follows (at visits as specified in table 1):

1. Laboratory samples must be drawn at least 1 hour prior to physician assessments (or up to 7 days in advance)
2. Patient reported health outcomes assessments: VAS for pain, global assessment of diseases status and fatigue, HAQ-promis, SF-36, EQ-5D, RAID, WPAI
3. Study nurse/investigator assessments:
    - Joint counts, preferably done by the same assessor for consistency
    - Registration of co-medication, safety assessments (AEs, vital signs)
    - Nurse/investigator global assessment of disease activity (VAS)
4. CR, DEXA and MRI when applicable (this can also be done after 5.)
5. Treating physician:
    - Review of laboratory data
    - In US arm: 32 joint examination of gray scale and PD findings
    - Decide treatment strategy according to table 1 (including i.a. joint injections as indicated)

    Time spent by physician should be recorded (or estimated as follows: History/physical examination 20 minutes, US 20 minutes, i.a.injection procedures 5 minutes plus 3 minutes per joint area).

5.1 Screening examination and eligibility screening form

All patients must sign and date the written informed consent from before any study specific assessments or procedures are performed.

A screening examination should be performed before the start of the study (or at the baseline visit for patients that fulfil the entry criteria). The following procedures and assessments are to be completed:
- Physical examination including pulse rate and blood pressure
- Chest X-ray, posterior-anterior and lateral chest radiographs with formal readings (by radiologist)
- laboratory tests (CRP, ESR, haematology, blood chemistry, urine dipstick. HBsAg and HCV antibody screening may be done at screening or just prior to start of 1st biologic)
- PPD/Mantoux test

5.2 Procedures for enrolment of eligible patients

Eligible patients will be randomized through a pre-defined procedure, stratified blockwise for site and baseline anti-CCP level (positive or negative). Closed envelopes containing the allocated treatment group will be distributed and stored by ascending stratified randomization number in the participating centres.

5.3 Clinical assessments and procedures

5.3.1 Efficacy

Individual parameters of the ACR core data set, ACR response, DAS, EULAR response, the vdHSS on CR, the 32 joint PD and grey scale US score and RAMRIS (plus Haavardsholm tenosynovitis score) on MRI will be assessed for efficacy.

5.3.1.1 ACR core data set

The ACR definition of response includes tender and swollen joint counts, VAS scales for pain, patient and investigator global assessment of disease activity, patient-assessed disability (HAQ) and acute phase response (ESR/hsCRP).

5.3.1.2 ACR response

The ACR response rates ACR20, ACR50, ACR70 and ACR90 as well as ACR remission rates will be calculated.

5.3.1.3 DAS
DAS includes the Ritchie articular index, the 44 swollen joint counts, the Erythrocyte Sedimentation Rate and a general health assessment on a VAS.

The DAS is calculated as follows:
DAS = 0.54*sqrt(RAI) + 0.065*(swollen44) + 0.33*Ln(ESR) + 0.0072*GH

High disease activity >3.7, low disease activity <2.4, remission <1.6

5.3.1.4 EULAR response
Based on the DAS, response criteria have been developed: the EULAR response criteria. The EULAR response criteria include not only change in disease activity but also current disease activity. To be classified as responders, patients should have a significant change in DAS and also low current disease activity. Three categories are defined: good, moderate, and non-responders.

5.3.1.5 Sharp van der Heijde Score
Radiographs of hands (posterior/anterior) and foot (anterior/posterior) will be taken at baseline, 3, 6, 12, 16 and 24 months. The modified Sharp van der Heijde Score (vdHSS) will be calculated, including an erosion score and a joint space narrowing score.

According to the vdHSS, erosion is assessed in 16 joints (five MCP, four PIP, IP of the thumbs, 1st MCB, radius and ulna bones, trapezium and trapezoid as one unit (multangular), navicular, lunate) for each hand and wrist, and six joints (five MTP, IP) for each foot. One point is scored if erosions are discrete, rising to 2, 3, 4, or 5 depending on the amount of surface area affected. The score for erosion ranges from 0 to 160 in the hands and from 0 to 120 in the feet (the maximum erosion score for a joint in the foot is 10). JSN is assessed in 15 joints (five MCP, four PIP, CMC 3 to 5, multangular navicular-lunate, radiocarpal) for each hand and wrist, and six joints (five MTP, IP) for each foot. JSN is combined with a score for (sub)luxation and scored as follows: 0 = normal; 1 = focal or doubtful; 2 = generalised, less than 50% of the original joint space; 3 = generalised, more than 50% of the original joint space or subluxation; 4 = bony ankylosis or complete luxation. The score for JSN ranges from 0 to 120 in the hands and from 0 to 48 in the feet (43). The total score is the sum of scores of erosion and JSN, the maximum score being 448.

Original radiographs will be scored at a central reading site by two experienced readers blinded to patient identity, clinical information and timepoints of radiographs, grouped per patient.

5.3.1.6 Ultrasonography score
A Siemens Antares or machine with similar specifics will be used for the ultrasonography assessment, with the following hardware specifications:
PRF = 391
Frequency = 7.3 MHz
R/S = 5
Filter = 2
The following joints will be scored for grey scale synovitis and power Doppler signal, with the best possible match to a reference atlas developed by dr. Hilde Berner Hammer that is accessible within the GoTreatIT software. The scoring system is based on the OMERACT US recommendations, semi-quantitatively from 0 - 3 in 32 joints: bilateral MCP I-V, RCJ, DRUJ, intercarpal, elbow, knee, talo-crural joint and MTP I-V, giving a maximum total score of 64.

In the US group, patients will be examined at all visits. Patients in the non-US group will be examined at baseline, 12 months and 24 months. The results from the baseline and 12 months examinations will be blinded to the patient and treating physician.

In addition to the 32 joints the following will be scored, but not be part of the response algorithm: Bilateral PIP2 and 3 joints, extensor carpi ulnaris tendon and tibialis posterior tendon.

5.3.1.7 MRI (RAMRIS)
Magnetic resonance imaging of the dominant hand and wrist pre- and post-gadolinium will be undertaken at baseline, 3, 6, 12, 16 and 24 months. The same side will be assessed at all time-points.

Images will be read according to the RAMRIS score at a central reading site by two experienced readers. The RAMRIS consists of MRI definitions of important joint pathologies, a core set of MRI sequences and a semi-quantitative scoring system for erosions, bone marrow edema and synovitis (44). The RAMRIS core set of MRI sequences to assess inflammatory as well as destructive changes in RA joints includes: Imaging in 2 planes, with T1-weighted images before and after intravenous gadolinium-contrast to assess synovitis and erosions; plus a T2-weighted fat saturated sequence or a STIR sequence to assess bone marrow edema.

Both flexor and extensor tenosynovitis will also be evaluated, according to the scoring system described by Haavardsholm et al, at the level between the radioulnar joint and the hook of the hamate, thus including both wrist and finger tendons (45). Flexor and extensor tenosynovitis are evaluated semi-quantitatively in 10 different anatomical areas, graded from grade 0 to 3 (total score 0–30). The grading is based on the maximum width (in mm) of post-contrast enhancement within each anatomical area on axial T1-weighted MR images.

5.3.1.8 Remission

ACR remission criteria and remission criteria based on DAS and DAS28 as well as the US Food and Drug Administration (FDA) definition of remission will be calculated (46).

ACR criteria
ACR clinical remission, a minimum of five of the following items must be present for at least two subsequent months:

- Morning stiffness <15 minutes
- No fatigue
- No joint pain by history
- No joint tenderness or pain on motion
- No soft-tissue swelling in joints or tender sheats
- ESR < 30 mm/1st hour in women or < 20 mm/1st hour in men

Disease activity score criteria
DAS remission defined as a score <1.6 using a composite index of the following measures:

- Ritchie articular index of tender joints
- 44 swollen joint count
- ESR
- Patient's assessment of general health (measured on a 100 mm visual analogue scale)

DAS28 remission defined as a score <2.6 using a composite index of the following measures:

- 28-joint count for tender and swollen joints
- ESR

- Patient's assessment of general health

FDA criteria:
Remission
Requires achieving ACR clinical remission and absence of radiological progression (Larsen or Sharp-van der Heijde method) over a continuous 6 month period in the absence of DMARDs.

Complete clinical remission
Same as FDA remission, but while continuing DMARD therapy

Major clinical response
Requires achieving ACR70 response for at least 6 subsequent months (ACR70 response means 70% improvement of tender and swollen joint count coupled with 70% improvement in 3 of 5 of the following: patient's assessment, physician's assessment, ESR or CRP, pain scale, Health Assessment Questionnaire).

Other remission criteria
In addition, a "complete clinical" DAS remission will be calculated (similar to the FDA complete clinical remission, but for DAS) and is defined as follows:
Same as "DAS-remission" (i.e. DAS<1.6), but sustained for > 6 months while at the same time requires absence of clinically swollen joints & no radiographic progression > 6 months period. This will be the primary endpoint of the study.

There is an ongoing initiative to develop new remission criteria, and the new ACR/EULAR remission criteria will also be applied in this trial (47).

5.3.2 Safety

5.3.2.1 Physical examination
A general physical examination (including the cardiovascular, respiratory, GI and neurological systems) should be performed at each visit, and recorded as normal or abnormal, with a description of abnormalities. Diagnosis of new abnormalities, or worsening of abnormalities, should be recorded as an AE if appropriate.

5.3.2.2 Vital signs
Vital signs including pulse rate, systolic and diastolic blood pressure and body weight and height will be assessed at times indicated in table 3.

5.3.2.3 Chest x-ray
PA and lateral chest radiographs with formal readings should be obtained at screening/baseline or up to 4 weeks prior to baseline.

5.3.2.4 Mantoux
A mantoux test will be performed at screening/baseline, according to local guidelines. A positive test will be followed-up as set forth in local guidelines.

5.4 Laboratory assessments

The following laboratory tests will be recorded at the time points indicated in table 3.
Hematology/CBC: Hb, hct, erythrocytes, white blood cells with differentials, platelet counts.
Blood chemistry: AST, ALP, albumin, creatinine, random glucose, potassium, sodium, chloride, calcium, phosphorus.

Urinanalyses: Dipstick for blood, protein and glucose (with microscopic examination as clinically indicated).
Acute phase reactants: CRP is measured by high sensitivity CRP nefelometry (mg/dL) and ESR by the Westergren method (mm/hr), according to local practice.
HBsAg and HCV at screening or just prior to start of 1st biologic.

5.5 Quality of life, disability and utility assessments

5.5.1 SF-36

The SF-36 is a multi-purpose, short-form health survey with only 36 questions. It yields an 8-scale profile of functional health and well-being scores as well as psychometrically-based physical and mental health summary measures and a preference-based health utility index. It is a generic measure, as opposed to one that targets a specific age, disease, or treatment group. Accordingly, the SF-36 has proven useful in surveys of general and specific populations, comparing the relative burden of diseases, and in differentiating the health benefits produced by a wide range of different treatments.

5.5.2 Health Assessment Questionnaire

The Stanford Health Assessment Questionnaire was introduced in the 1980s and is now widely used in evaluation of physical function in patients with RA (48). The disability index of this instrument includes questions concerning the ability of patients to perform 20 activities of daily living, and is most commonly referred to as the HAQ questionnaire, and sometimes as the HAQ disability index (HAQ-DI) (49). A new version has recently been developed, the Patient-Reported Outcomes Measurement Information (PROMIS) HAQ, including a 20-item short form that will be used in this study. While the original HAQ had 4 response categories, this new version includes a fifth response option, “with a little bit of difficulty”. Also, the scoring algorithm was changed from the HAQ’s 0–3 unit scale to a 0–100 unit scale. Completion time was reduced by over one-third compared to the original HAQ.

5.5.3 EQ-5D

EQ-5D is a standardised instrument for use as a measure of health outcome. Applicable to a wide range of health conditions and treatments, it provides a simple descriptive profile and a single index value for health status.

5.5.4 RAID

The RA Impact of Disease (RAID) score is a patient-derived composite response index for use in clinical trials in RA, includes seven domains with the following relative weights: pain (21%), functional disability (16%), fatigue (15%), emotional well-being (12%), sleep (12%), coping (12%) and physical well-being (12%).

5.6 Genetic markers and biomarkers

Samples (including serum, plasma, full blood and urine) for biomarker or DNA/RNA discovery and validation will be collected and stored in a freezer at -70 C at visits shown in table 3. These samples will be used for research purposes only, and may include measurement of cytokines and other known or potential new markers of inflammation or damage, such as interleukins, interferons, metalloproteases, transforming growth factor, TNFs, adhesion molecules etc as well as DNA/RNA analyses (genomics and proteomics). All samples will be stored in a certified biobank.

5.7 Worker productivity

Worker productivity is generally subdivided into 2 components: absenteeism and presenteeism. The concept of absenteeism has been defined as productivity loss due to health-related absence from work,

while presenteeism refers to reduced performance or productivity while at work due to health reasons. Absenteeism may include personal time off, sick days off work, time on short and/or longterm work disability, or time on worker's-compensated days; and presenteeism could be characterized as the time not being on the task, or decreased work quality and quantity. Patients will be asked to answer the Work Productivity and Activity Impairment Questionnaire: Rheumatoid arthritis V2.0 (WPAI:RA).

The WPAI yields four types of scores:
1. Absenteeism (work time missed)
2. Presenteeism (impairment at work / reduced on-the-job effectiveness)
3. Work productivity loss (overall work impairment / absenteeism plus presenteeism)
4. Activity Impairment

6. Investigational procedure

6.1 Schedule of the investigational procedure and comparator(s)

The duration of the study will be 24 months, starting from baseline. Patients will receive treatment according to the treatment schedule in figure 1, according to the decision rules in figure 2.

6.2 Dose modifications, interruptions and delays

The following criteria for side effects will result in dose modifications, replacement of MTX by other DMARD or discontinuation of MTX:

1. Anaemia:	policy after analyses according to the rheumatologists own insight.
2. Leukopenia	number of white blood cells between 2.5 and 3.5 G/l: **Stop MTX until recovery. Thereafter, halve dose MTX.** number of white blood cells $\leq$ 2.5 G/l: S**top MTX**.
3. Thrombocytopenia	number of thrombocytes between 100 en 150 G/l: **Stop MTX until recovery. Thereafter, halve dose MTX.** number of thrombo's $\leq$ 100 G/l: **Stop MTX.**
4. Pancytopenia	number of white blood cells $\leq$ 2.5 G/l and number of thrombo's $\leq$ 100 G/l: **stop MTX**.
5. Severe nausea, dyspepsia	**change to MTX subcutaneaously.**
6. Oral ulcera	very severe : **stop MTX until recovery, thereafter, halve dose MTX.**
7. Transaminase increment	$\geq$ 2x upper normal level: **do not increase MTX.** $\geq$ 3x upper normal level : **stop MTX until recovery, thereafter, halve dose MTX.** $\geq$ 3x upper normal level and additional signs of liver toxicity (bilirubin > 2x ULN, INR > 1.5x ULN, alkaline phosphatase > 2 x ULN, presence of worsening fatigue, nausea, vomiting, fever, rash or eosinophilia): **Stop MTX**
8. Renal function	GFR decrease > 25%: **stop MTX until recovery, thereafter administration half dose MTX**.
9. Pneunonitis	in case of pneumonitis because of MTX: **stop MTX**.

Possible side effects for other DMARDs and biologicals will be at the treating rheumatologist's discretion, according to established guidelines.

7. Safety

7.1 Adverse events and laboratory abnormalities

7.1.1 Clinical adverse events and serious adverse events

Clinical adverse events (AE) and serious adverse event (SAE) encountered during the clinical study will be reported in the eCRF (see appendix for definitions and clarification of what should be reported).

7.1.2 Treatment and follow-up of adverse events

AEs should be followed up as clinically indicated until they have returned to baseline status or stabilized.

7.1.3 Laboratory test abnormalities

Laboratory test result will be recorded in the eCRF, or appear on electronically produced laboratory reports, and should not be recorded as a AE unless there is an associated clinical condition for which the patient is given treatment or current treatment is altered.

7.1.3 Follow-up of abnormal laboratory test values

In the event of medically significant unexplained abnormal laboratory test values, the tests should be repeated and followed up until they have returned to the normal range and/or an adequate explanation of the abnormality is found.

7.2 Pregnancy

A female patient must be instructed to immediately inform the treating rheumatologist if she becomes pregnant during the study, which should then counsel the patient according to current guidelines.

8. Statistical considerations and analytical plan

8.1 Primary and secondary study endpoints

8.1.1 Primary endpoints

The primary endpoint is the proportion of patients reaching “complete clinical DAS-remission” (i.e. DAS < 1.6, absence of swollen joints & no radiological progression during the last 8 months) at the end of the study.

Specifically, the primary endpoint will be the proportion of patients with all the following criteria met:

- DAS score < 1.6 at visits 11, 12 and 13 (after 16, 20 and 24 months)
- Absence of swollen joints at visits 11, 12 and 13 (after 16, 20 and 24 months)
- No radiological progression (absolute change in vdHSS $\leq$ 0.5 unit) between visit 11 (16 months) and visit 13 (24 months)

8.1.2 Secondary endpoints/exploratory endpoints

- ACR20, ACR50, ACR70 and ACR90 response at each visit
- Proportion of patients reaching various remission criteria at each visit
- Proportion of patients with radiologic progression between baseline and 12 and 24 months (absolute change in vdHSS $\leq$ 1 unit/year).
- LDAS
- Change from baseline in individual parameters of the ACR core set

- AUC from baseline to 12 and 24 months for measures of disease activity (e.g. DAS, SDAI, CDAI)
- Time to ACR response
- Time to EULAR response
- Change from baseline in DAS at each scheduled visit
- Rate of withdrawals
- Change from baseline in HAQ at each visit
- Change from baseline in SF-36 at each visit
- Change from baseline to 12 and 24 months for RAMRIS synovitis, erosions, bone marrow edema and tenosynovitis, as well as MRI total inflammation score
- Change from baseline to 12 and 24 months for US scores of grey scale and PD synovitis
- Number of patients started with anti-TNF
- DMARD/NSAID/pain medication use
- Infections which need antibiotics or antiviral medication
- Infections which need hospitalisation
- All serious adverse events
- Work performance and status
- QUALYs
- Exploratory/secondary endpoints will not be limited to those mentioned above, and will include endpoints as necessary to explore the secondary objectives of the study as described in section 2.2.

8.1.3 Safety

Safety of the treatment will be evaluated based on reported AEs, laboratory test, vital signs and performance status.

8.2 Statistical and analytical methods

8.2.1 Statistical model

This randomized (1:1), two-parallel-arm clinical trial of a treatment strategy with US versus treatment strategy without US aims primarily to describe and estimate efficacy parameters and test pre-specified statistical hypothesis.

The primary variable will be analysed using a logistic regression model, including, in addition to the treatment group, the stratification factors used at randomization. Because we want to assess the treatment effect across centres, centre will be included as a covariate in the analysis. Anti-CCP status (positive/negative) is regarded as a strong predictive factor for the primary variable outcome, and is thus included as a stratification factor (and covariate in the statistical model) to increase the power and the coverage probability of the confidence intervals. (50) The SAP will detail these procedures, as well as alternative and further supportive evaluations, such as analyses including unbalanced baseline predictors or modifications of the logistic regression model in case validity assumptions are not met.

8.2.1.1 Primary variable

Statistical hypothesis (superiority test):
Null hypothesis: Efficacy (primary endpoint) is identical in the two groups
Alternate hypothesis: Efficacy (as defined above) is not identical in the two groups, i.e. a two-sided test will be employed.

8.2.1.2 Secondary/exploratory variables

Exploratory between-group comparisons will be performed for secondary efficacy endpoints.

8.2.2 Sample size

The sample size has been estimated on the basis of the primary variable proportion of patients reaching "complete clinical DAS-remission" analysed using a logistic regression model. The assumed rate of complete clinical DAS-remission is 45% in the non-US group and 65% in the US group, giving a treatment difference of 20%. See Table 1 for a rationale on these assumptions. On basis of the assumptions, a total sample size (study completers) of 198 (99 in each group) is needed to achieve 80% power to detect a difference of 20% between the groups in a two-sided test at 5% significance level. To compensate for withdrawals/loss to follow-up, a total enrolment of 240 patients will be targeted.

Table 1. Rationale for sample size calculations.

Study	**Remission rate**	
	Aggressive strategy	**Routine strategy**
FIN-RACo		
ACR-remission	14%	3%
DAS28-remission	51%	16%
Good treatment response	67%	27%
TICORA		
DAS28 remission	65%	16%
Good response	82%	44%
CAMERA		
Clinical remission	50%	37%
CIMESTRA		
DAS28 remission	43/34%	n/a

8.2.3 Hypotheses testing

The primary efficacy analyses will be performed using a logistic regression model (see section 8.2.1 above for details on statistical model).

The between-group comparisons for secondary variables will be tested as for the primary variable where applicable and additional analyses will be performed based on the following methods (but not limited to):
- Changes in secondary variables will be subject to analyses of covariance or an appropriate non-parametric alternative
- Binary response variables will be analyzed using logistic regression or chi-square/Mantel-Haenszel test
- Time to event variables will be analyzed using the Kaplan-Meier method and comparisons between the two groups will be performed using the log rank test or Cox regression analyses.

Unless otherwise specified, all statistical hypotheses will be tested at the 5% significance level ($\alpha = 0.05$) against two-sided alternatives, and corresponding intervals will be reported as appropriate. No correction of possible type I error for multiple comparisons will be performed for exploratory analyses of secondary variables.

8.2.4 Efficacy analyses

All patients randomized into the study will be included in the ITT population which will form the primary analysis population of the study. All efficacy analyses will be presented with the results from

the hypothesis testing (by p-value) in addition to estimates and 95% confidence limits of the treatment effect. For the primary variable specifically, this will be the estimated odds ratio and relative risk with corresponding 95% confidence limits.

8.2.5 Safety analyses

The safety analyses population will include all patients who completed at least one follow-up visit. Safety analyses will be descriptive and presented as summary tables by treatment group and (if applicable) by visit.

8.2.6 Quality of life and disability analyses

QoL and disability will be assessed using SF-36 and HAQ. These scores will be summarized by descriptive summary tables at baseline and over time, and at the end of study. Missing data at end of study will be replaced by the last valid post-baseline assessment.

8.2.7 Other analyses/subanalyses

Exploratory subgroup analyses of primary, secondary and exploratory efficacy variables may be performed if appropriate. The decision to include such analyses will be made on basis of the collected data.

8.2.8 Interim analyses

Interim analyses of selected efficacy variables will be performed after 12 months, but will not be made public/published until the end of the 24 month visit of the last included patient.

8.2.9 Missing data

Missing data for the primary endpoints will be imputed with a negative outcome (i.e. not reaching complete clinical DAS-remission). If the handling of missing data is regarded as having a significant effect on the conclusions of the trial, sensitivity analyses with different methods for handling missing data will be included. Such methods may include complete case analyses, last observation carried forward, positive outcome imputation and multiple imputation techniques.

9 Data quality assurance

Accurate and reliable data collection will be assured by verification and cross-checking of the CRF/eCRF by the study nurse, study physician and external monitor. An eCRF software solution that adheres to GCP will be used to collect study data (Viedoc™, Uppsala, Sweden). The GO-TreatIT software package (DiagraphIT, Kristiansand, Norway) may be used as source document for the e-CRF. If a centre does not have access to the GOTreatIT software package, or the GOTreatIT program do not offer the necessary specifications, paper files/paper CRF combined with the patient's electronic medical journal will be used as source documents for the eCRF when applicable.

10 References

(1) Cush JJ. Early rheumatoid arthritis - is there a window of opportunity? J Rheumatol Suppl 2007 Nov;80:1-7.:1-7.

(2) Boers M. Understanding the window of opportunity concept in early rheumatoid arthritis. Arthritis Rheum 2003 Jul;48(7):1771-4.

(3) Furst DE. Window of opportunity. J Rheumatol 2004 Sep;31(9):1677-9.

(4) Bukhari MA, Wiles NJ, Lunt M, Harrison BJ, Scott DG, Symmons DP, et al. Influence of disease-modifying therapy on radiographic outcome in inflammatory polyarthritis at five years: results from a large observational inception study. Arthritis Rheum 2003 Jan;48(1):46-53.

(5) Lard LR, Visser H, Speyer I, vander Horst-Bruinsma IE, Zwinderman AH, Breedveld FC, et al. Early versus delayed treatment in patients with recent-onset rheumatoid arthritis: comparison of two cohorts who received different treatment strategies. Am J Med 2001 Oct 15;111(6):446-51.

(6) Nell VP, Machold KP, Eberl G, Stamm TA, Uffmann M, Smolen JS. Benefit of very early referral and very early therapy with disease-modifying anti-rheumatic drugs in patients with early rheumatoid arthritis. Rheumatology (Oxford) 2004 Jul;43(7):906-14.

(7) Allaart CF, Goekoop-Ruiterman YP, de Vries-Bouwstra JK, Breedveld FC, Dijkmans BA. Aiming at low disease activity in rheumatoid arthritis with initial combination therapy or initial monotherapy strategies: the BeSt study. Clin Exp Rheumatol 2006 Nov;24(6 Suppl 43):S-82.

(8) Verstappen SM, Jacobs JW, Van D, V, Heurkens AH, Schenk Y, ter Borg EJ, et al. Intensive treatment with methotrexate in early rheumatoid arthritis: aiming for remission. Computer Assisted Management in Early Rheumatoid Arthritis (CAMERA, an open-label strategy trial). Ann Rheum Dis 2007 Nov;66(11):1443-9.

(9) Mottonen T, Hannonen P, Leirisalo-Repo M, Nissila M, Kautiainen H, Korpela M, et al. Comparison of combination therapy with single-drug therapy in early rheumatoid arthritis: a randomised trial. FIN-RACo trial group. Lancet 1999 May 8;353(9164):1568-73.

(10) Goekoop-Ruiterman YP, de Vries-Bouwstra JK, Allaart CF, Van ZD, Kerstens PJ, Hazes JM, et al. Comparison of treatment strategies in early rheumatoid arthritis: a randomized trial. Ann Intern Med 2007 Mar;%20;146(6):406-15.

(11) Bijlsma JW, Weinblatt ME. Optimal use of methotrexate: the advantages of tight control. Ann Rheum Dis 2007 Nov;66(11):1409-10.

(12) Saunders SA, Capell HA, Stirling A, Vallance R, Kincaid W, McMahon AD, et al. Triple therapy in early active rheumatoid arthritis: a randomized, single-blind, controlled trial comparing step-up and parallel treatment strategies. Arthritis Rheum 2008 May;58(5):1310-7.

(13) Grigor C, Capell H, Stirling A, McMahon AD, Lock P, Vallance R, et al. Effect of a treatment strategy of tight control for rheumatoid arthritis (the TICORA study): a single-blind randomised controlled trial. Lancet 2004 Jul 17;364(9430):263-9.

(14) Smolen JS, Aletaha D. What should be our treatment goal in rheumatoid arthritis today? Clin Exp Rheumatol 2006 Nov;24(6 Suppl 43):S-13.

(15) Sokka T, Makinen H, Puolakka K, Mottonen T, Hannonen P. Remission as the treatment goal--the FIN-RACo trial. Clin Exp Rheumatol 2006 Nov;24(6 Suppl 43):S-6.

(16) Smolen JS, Aletaha D, Bijlsma JW, Breedveld FC, Boumpas D, Burmester G, et al. Treating rheumatoid arthritis to target: recommendations of an international task force. Ann Rheum Dis 2010 Apr;69(4):631-7.

(17) Smolen JS, Landewe R, Breedveld FC, Dougados M, Emery P, Gaujoux-Viala C, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs. Ann Rheum Dis 2010 Jun;69(6):964-75.

(18) Bakker MF, Jacobs JW, Verstappen SM, Bijlsma JW. Tight control in the treatment of rheumatoid arthritis: efficacy and feasibility. Ann Rheum Dis 2007 Nov;66 Suppl 3:iii56-60.

(19) Atar D, Birkeland KI, Uhlig T. 'Treat to target': moving targets from hypertension, hyperlipidaemia and diabetes to rheumatoid arthritis. Ann.Rheum.Dis. 69[4], 629-630. 2010.

(20) Naredo E, Collado P, Cruz A, Palop MJ, Cabero F, Richi P, et al. Longitudinal power Doppler ultrasonographic assessment of joint inflammatory activity in early rheumatoid arthritis: predictive value in disease activity and radiologic progression. Arthritis Rheum 2007 Feb 15;57(1):116-24.

(21) Freeston JE, Wakefield RJ, Conaghan PG, Hensor EM, Stewart SP, Emery P. A diagnostic algorithm for persistence of very early inflammatory arthritis: the utility of power doppler ultrasound when added to conventional assessment tools. Ann Rheum Dis 2009 Apr 27.

(22) Patel A, MacMahon S, Chalmers J, Neal B, Woodward M, Billot L, et al. Effects of a fixed combination of perindopril and indapamide on macrovascular and microvascular outcomes in patients with type 2 diabetes mellitus (the ADVANCE trial): a randomised controlled trial. Lancet 2007 Sep 8;370(9590):829-40.

(23) Eeg-Olofsson K, Cederholm J, Nilsson PM, Gudbjornsdottir S, Eliasson B. Glycemic and risk factor control in type 1 diabetes: results from 13,612 patients in a national diabetes register. Diabetes Care 2007 Mar;30(3):496-502.

(24) Egan BM, Lackland DT, Cutler NE. Awareness, knowledge, and attitudes of older americans about high blood pressure: implications for health care policy, education, and research. Arch Intern Med 2003 Mar 24;163(6):681-7.

(25) Pearson TA. The epidemiologic basis for population-wide cholesterol reduction in the primary prevention of coronary artery disease. Am J Cardiol 2004 Nov 4;94(9A):4F-8F.

(26) Goekoop-Ruiterman YP, de Vries-Bouwstra JK, Allaart CF, Van ZD, Kerstens PJ, Hazes JM, et al. Clinical and radiographic outcomes of four different treatment strategies in patients with early rheumatoid arthritis (the BeSt study): A randomized, controlled trial. Arthritis Rheum 2008 Feb;58(2 Suppl):S126-S135.

(27) Landewe R, Van Der HD, van Der LS, Boers M. Twenty-eight-joint counts invalidate the DAS28 remission definition owing to the omission of the lower extremity joints: a comparison with the original DAS remission. Ann Rheum Dis 2006 May;65(5):637-41.

(28) Szkudlarek M, Court-Payen, Strandberg C, Klarlund M, Klausen T, Ostergaard M. Power Doppler ultrasonography for assessment of synovitis in the metacarpophalangeal joints of patients with rheumatoid arthritis: a comparison with dynamic magnetic resonance imaging. Arthritis Rheum 2001 Sep;44(9):2018-23.

(29) Szkudlarek M, Klarlund M, Narvestad E, Court-Payen, Strandberg C, Jensen KE, et al. Ultrasonography of the metacarpophalangeal and proximal interphalangeal joints in rheumatoid arthritis: a comparison with magnetic resonance imaging, conventional radiography and clinical examination. Arthritis Res Ther 2006 Mar 6;8(2):R52.

(30) Scire CA, Montecucco C, Codullo V, Epis O, Todoerti M, Caporali R. Ultrasonographic evaluation of joint involvement in early rheumatoid arthritis in clinical remission: power Doppler signal predicts short-term relapse. Rheumatology (Oxford) 2009 Sep;48(9):1092-7.

(31) Brown AK, Conaghan PG, Karim Z, Quinn MA, Ikeda K, Peterfy CG, et al. An explanation for the apparent dissociation between clinical remission and continued structural deterioration in rheumatoid arthritis. Arthritis Rheum 2008 Oct;58(10):2958-67.

(32) Combe B, Landewe R, Lukas C, Bolosiu HD, Breedveld F, Dougados M, et al. EULAR recommendations for the management of early arthritis: report of a task force of the European Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). Ann Rheum Dis 2007 Jan;66(1):34-45.

(33) Goekoop-Ruiterman YP, de Vries-Bouwstra JK, Allaart CF, Van ZD, Kerstens PJ, Hazes JM, et al. Clinical and radiographic outcomes of four different treatment strategies in patients with early rheumatoid arthritis (the BeSt study): a randomized, controlled trial. Arthritis Rheum 2005 Nov;52(11):3381-90.

(34) Goekoop-Ruiterman YP, de Vries-Bouwstra JK, Kerstens PJ, Nielen MM, Vos K, van SD, et al. DAS-driven therapy versus routine care in patients with recent-onset active Rheumatoid Arthritis. Ann Rheum Dis 2009 Jan;20.

(35) Hetland ML, Stengaard-Pedersen K, Junker P, Lottenburger T, Hansen I, Andersen LS, et al. Aggressive combination therapy with intra-articular glucocorticoid injections and conventional disease-modifying anti-rheumatic drugs in early rheumatoid arthritis: second-year clinical and radiographic results from the CIMESTRA study. Ann Rheum Dis 2008 Jun;67(6):815-22.

(36) Kiely PD, Brown AK, Edwards CJ, O'Reilly DT, Ostor AJ, Quinn M, et al. Contemporary treatment principles for early rheumatoid arthritis: a consensus statement. Rheumatology (Oxford) 2009 Jul;48(7):765-72.

(37) Korpela M, Laasonen L, Hannonen P, Kautiainen H, Leirisalo-Repo M, Hakala M, et al. Retardation of joint damage in patients with early rheumatoid arthritis by initial aggressive treatment with disease-modifying antirheumatic drugs: five-year experience from the FIN-RACo study. Arthritis Rheum 2004 Jul;50(7):2072-81.

(38) Landewe RB, Boers M, Verhoeven AC, Westhovens R, van de Laar MA, Markusse HM, et al. COBRA combination therapy in patients with early rheumatoid arthritis: long-term structural benefits of a brief intervention. Arthritis Rheum 2002 Feb;46(2):347-56.

(39) O'Dell JR. The BeSt way to treat early rheumatoid arthritis? Ann Intern Med 2007 Mar;20;146(6):459-60.

(40) van der Kooij SM, Allaart CF, Dijkmans BA, Breedveld FC. Innovative treatment strategies for patients with rheumatoid arthritis. Curr Opin Rheumatol 2008 May;20(3):287-94.

(41) van Tuyl LH, Lems WF, Voskuyl AE, Kerstens PJ, Garnero P, Dijkmans BA, et al. Tight control and intensified COBRA combination treatment in early rheumatoid arthritis: 90% remission in a pilot trial. Ann Rheum Dis 2008 Nov;67(11):1574-7.

(42) Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, III, et al. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Ann Rheum Dis 2010 Sep;69(9):1580-8.

(43) Van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol 1999 Mar;26(3):743-5.

(44) 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):1385-6.

(45) 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):1216-20.

(46) van der Helm-van Mil AH, Breedveld FC, Huizinga TW. Aspects of early arthritis. Definition of disease states in early arthritis: remission versus minimal disease activity. Arthritis Res Ther 2006;8(4):216.

(47) van Tuyl LH, Vlad SC, Felson DT, Wells G, Boers M. Defining remission in rheumatoid arthritis: results of an initial American College of Rheumatology/European League Against Rheumatism consensus conference. Arthritis Rheum 2009 May 15;61(5):704-10.

(48) Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: the health assessment questionnaire, disability and pain scales. J Rheumatol 1982 Sep;9(5):789-93.

(49) Lillegraven S, Kvien TK. Measuring disability and quality of life in established rheumatoid arthritis. Best Pract Res Clin Rheumatol 2007 Oct;21(5):827-40.

(50) Negassa A, Hanley JA. The effect of omitted covariates on confidence interval and study power in binary outcome analysis: a simulation study. Contemp Clin Trials 2007 May;28(3):242-8.

11 Ethical aspects

11.1 Local regulations/Declaration of Helsinki

The study will be reviewed by the local "personombud", and permission to store sensitive data will be obtained from the Norwegian Data Inspectorate. A biobank will be established according to Norwegian regulations. The study will be conducted in full conformance with the principles of the "Declaration of Helsinki". The study will adhere to the principles outlined in "Guideline for Good Clinical Practice" ICH Tripartite Guideline (January 1997).

11.2 Informed consent

It is the responsibility of the treating rheumatologist (or a person designated by him/her, i.e. a study nurse) to obtain signed informed consent from each patient prior to participating in this study after adequate explanation of the aims, methods, anticipated benefits and potential hazards of the study.

11.3 Independent ethics committees

The study protocol and any accompanying material provided to the patients will be submitted to the regional committee for medical and health research ethics ("REK").

12 Study documentation and eCRF/CRF

Adequate and accurate records of each study visit will be maintained to enable the conduct of the study to be fully documented and the study data to be subsequently verified.

13 Study committee

A study committee will be established, and will primarily review safety data as needed, as well as make recommendations regarding continuation, termination or modification of the study.

The study committee will include the national medical leader, the principal investigator, the biostatistician, an internationally recognized medical professor/epidemiologist, a local investigator, a patient representative and a representative from Diakonhjemmet Sykehus AS.

14 Conditions for modifying the protocol

Protocol modifications to an ongoing study must be approved by the study committee, and must be approved by the National Medical Leader and biostatistician. Protocol modifications will be submitted to the REK for approval and to the regulatory agencies (SLV) as required.

15 Conditions for terminating the study

The study committee reserves the right to terminate the study at any time. This may be due to safety reasons or if new knowledge arises that invalidates the study (including results from interim analyses). Other reasons that may have a major impact on the study, including ethical and financial aspects, may also lead to termination of the study. In terminating the study, the study committee and investigators will assure that adequate consideration is given to the protections of patients' interests.

16 Monitoring the study

This study will be monitored by the "Seksjon for GCP, Oslo University Hospital" and "Innovest AS". Monitoring is the act of overseeing the progress of a clinical trial, and of ensuring that it is conducted, recorded, and reported in accordance with the protocol, Standard Operating Procedures (SOPs), Good Clinical Practice (GCP), and the applicable regulatory requirement(s).

It is understood that the responsible monitor will contact and visit the trial sites regularly, and will be allowed, on request, to inspect the various records of the trial (eCRF and other records/source material), provided that patient confidentiality is maintained in accord with local requirements.

17 Confidentiality of trial documents and patient records

All trial documents and patient records will be stored as required by the law, and protected from unauthorized parties.

18 Publication of data

The results of this study will be published and/or presented at scientific meetings and in international peer-reviewed journals. Authorship will be based on scientific contribution and enrolment, according to the guidelines set forth in the Vancouver protocol.

19 Financial aspects, plan for insurance

The medical treatment will be covered as for “usual care”, i.e. by “Folketrygden”/”NAV” and patients own payment “egenandel”. Procedures/examinations that are not part of “usual care” will be covered by the study, and there will be no own payments.

All patients will receive a reimbursement of NOK 100 per visit, to cover parking fees as the study implies longer visits than “usual care”.

All centres will receive a reimbursement of NOK 500 per visit (with payments after final visit), to cover extra resources allocated to the study, including study nurses and administration of the logistics of the study.

Diakonhjemmet Sykehus AS and participating centres will cover necessary insurance in accordance with Norwegian law.

Funding will partly be from Helse Sør-Øst research grant to Diakonhjemmet Hospital (Helse Sør-Øst research Group in rheumatology) and other public funding schemes, partly by unrestricted grants from pharmaceutical companies and support from the research foundation of the Norwegian Rheumatism Association.

20 Appendix

20.1 Appendix 1 Overview of visits

Before or at the baseline visit the following screening procedure is employed:
- Physical examination including pulse rate and blood pressure (systolic and diastolic)
- Chest X-ray, posterior-anterior and lateral chest radiographs with formal readings (by radiologist)
- laboratory tests (CRP, ESR, hematology, blood chemistry,urine dipstick).
- Mantoux/PPD
- cross-check of inclusion- and exclusion criteria

Table 3. Overview of visits.

Visits	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
Months	**0**	**1**	**2**	**3**	**4**	**6**	**8**	**10**	**12**	**14**	**16**	**20**	**24**
History													
Informed consent	X												
Medical History	X												
Comorbidity	X	X	X	X	X	X	X	X	X	X	X	X	X
Medication	X	X	X	X	X	X	X	X	X	X	X	X	X
AEs	X	X	X	X	X	X	X	X	X	X	X	X	X
Work performance (WPAI)	X	X	X	X	X	X	X	X	X	X	X	X	X
Smoking, coffee, alcohol	X					X			X				X
Clinical assessments by nurse/physician													
Assessment of disease activity	X	X	X	X	X	X	X	X	X	X	X	X	X
Swollen joints	X	X	X	X	X	X	X	X	X	X	X	X	X
Tender joints	X	X	X	X	X	X	X	X	X	X	X	X	X
Vital signs	X	X	X	X	X	X	X	X	X	X	X	X	X
US 32 joints	X	X	X	X	X	X	X	X	X	X	X	X	X

Patient Assessments													
Pain/fatigue/ assessment of disease activity (VAS)	X	X	X	X	X	X	X	X	X	X	X	X	X
HAQ-promis	X			X		X			X		X		X
EQ-5D	X			X		X			X		X		X
SF 36	X			X		X			X		X		X
RAID	X			X		X			X		X		X
Biochemical Assessments													
ACPA/IgA-RF/IgM-RF	X								X				X
ESR	X	X	X	X	X	X	X	X	X	X	X	X	X
CRP	X	X	X	X	X	X	X	X	X	X	X	X	X
Hematology/ ALT/Creatinine/ GFR/blood chemistry (revma kontroll)	X	X	X	X	X	X	X	X	X	X	X	X	X
BioBank (DNA, serum, plasma, urine)	X	X	X	X		X			X		X		X
Radiological Assessments													
MRI of dominant hand	X			X		X			X		X		X
DEXA	X			X					X				X
CR of hands and feet	X			X		X			X		X		X

20.2 ACR/EULAR Classification Criteria for Rheumatoid Arthritis

ACR/EULAR Classification Criteria for Rheumatoid Arthritis: score based algorithm for classification in an eligible patient (i.e. clinical synovitis not related to a specific aetiology). **Cutpoint for RA: ≥6/10**.

Table 4. ACR/EULAR Classification Criteria for Rheumatoid Arthritis

JOINT INVOLVEMENT[1] (0-5)	
1 medium-large[2] joint	0
2-10 medium-large joints	1
1-3 small[3] joints (with or without involvement of large joints)	2
4-10 small joints (with or without involvement of large joints)	3
>10 joints[4] (at least one small joint)	5
*SEROLOGY[5] (0-3)	
Negative RF AND negative ACPA	0
Low positive RF OR low positive ACPA	2
High positive RF OR high positive ACPA	3
*ACUTE PHASE REACTANTS[6] (0-1)	
Normal CRP AND normal ESR	0
Abnormal CRP OR abnormal ESR	1
DURATION OF SYMPTOMS[7] (0-1)	
<6 weeks	0
≥6 weeks	1

[1] Joint involvement refers to any *swollen* or *tender* joint on examination, or evidence of synovitis on magnetic resonance imaging or ultrosonography. Distal interphalangeal joints (DIPs), 1st carpo-metacarpal (CMC) joint, and 1st metatarso-phalangeal (MTP) joint are ***excluded from assessment.*** Categories of joint distribution are classified according to the location and number of the involved joints, with placement into the highest category possible based on the pattern of joint involvement

[2] Medium to large joints refer to shoulders, elbows, hips, knees, and ankles

[3] Small joints refer to the metacarpo-phalangeal (MCP) joints, proximal interphalangeal (PIP) joints, metatarso-phalangeal (MTP) joints 2-5, thumb interphalangeal (IP) joints, and wrists

[4] In this category, at least one of the involved joints must be a small joint; the other joints can include any combination of large and additional small joints, as well as other joints not specifically listed elsewhere (*e.g.*, temporomandibular, acromioclavicular, sternoclavicular, *etc.*)

[5] Negative refers to international unit (IU) values that are ≤upper limit of normal (ULN) for the lab and assay; low-positive refers to IU values that are >ULN but ≤3x ULN for the lab and assay; high-positive referes to IU values that are >3x ULN for the lab and assay. Where RF is only available as positive or negative, a positive result should be scored as "low-positive" for RF.

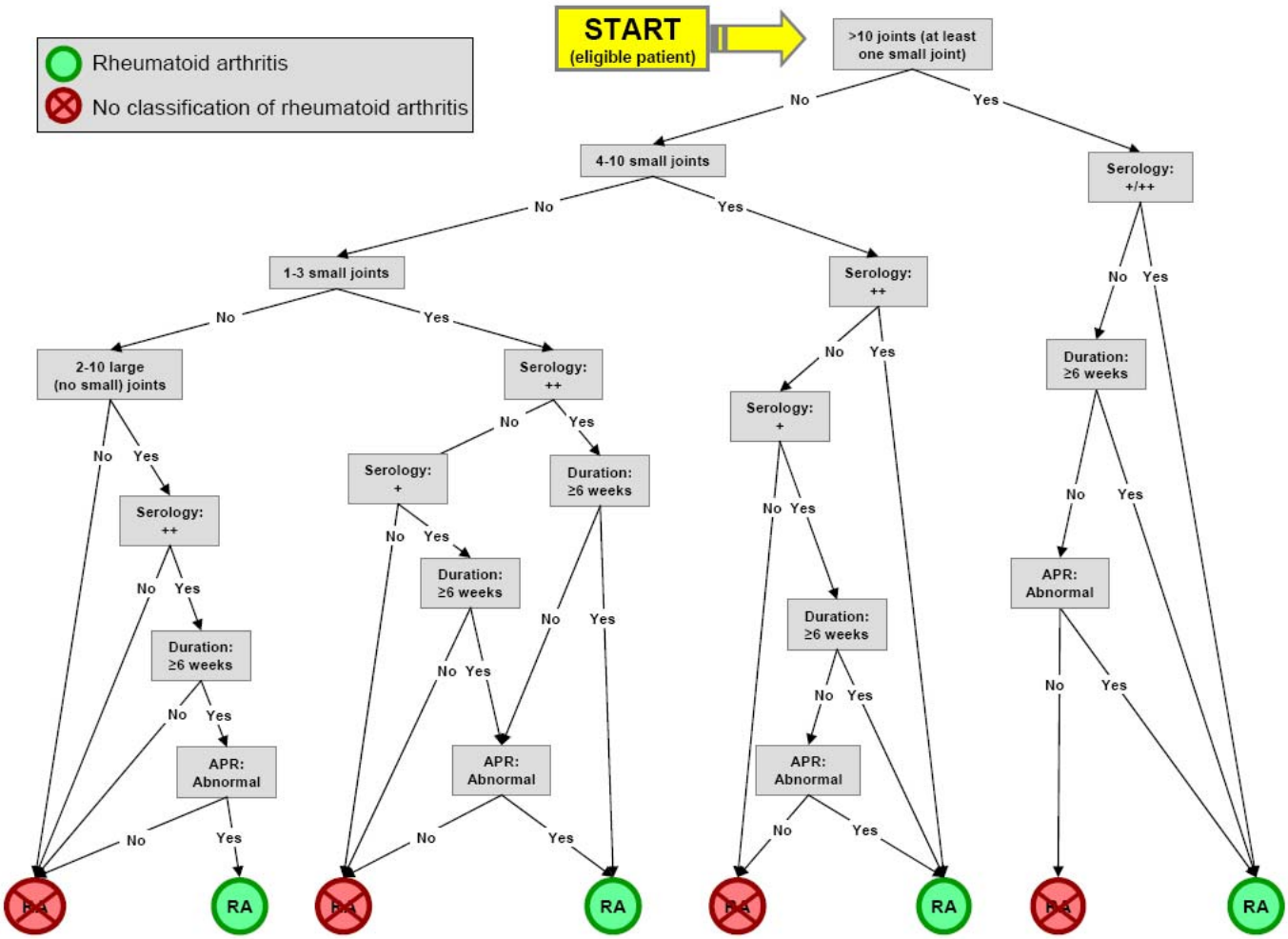
[6] Normal/abnormal is determined by local laboratory standards.

[7] Duration of symptoms refers to patient self-report of the duration of signs or symptoms of synovitis (*e.g.*, pain, swelling, tenderness) of joints that are clinically involved at the time of assessment.

* Individuals should only be scored by these criteria if at least one serologic test and at least one acute phase reactant test result is available. Where a value for a serologic test or acute phase reactant is not available, that test should be considered as 'negative/normal'.

Abbreviations: RF = rheumatoid factor; ACPA = anti-citrullinated protein/peptide antibodies; ULN = upper limit of normal; ESR = erythrocyte sedimentation rate; CRP = C-reactive protein.

Figure 1. Tree algorithm to classify definite rheumatoid arthritis (green circles) or to exclude its current presence (red circles) among those who are eligible to be assessed by the new criteria. For definitions of categories (e.g. serology: +, ++, or joint regions) please see footnotes to Table 4.



20.3 Appendix 3 ACR core set

1. Patient's visual analog scale of pain (100mm horizontal VAS)
2. Patient's global assessment of disease status (100mm horizontal VAS)
3. Physician's global assessment of disease status (100mm horizontal VAS)
4. Swollen joints
5. Tender joints
6. ESR/hsCRP
7. Patient's HAQ.

20.4 Appendix 4 Joints to assessed for swelling and tenderness

Ritchie Articular Index
This is a long-standing approach to doing a graded assessment of the tenderness of 26 joint regions, based on summation of joint responses after applying firm digital pressure.[24] Four grades can be used: 0, patient reported no tenderness; +1, patient complained of pain; +2, patient complained of pain and winced; and +3, patient complained of pain, winced, and withdrew. Thus, the index ranges from 0 to 3 for individual measures and 0 to 78 overall, with higher scores being worse tenderness.

Certain joints are treated as a single unit, such as the metacarpal-phalangeal and proximal interphalangeal joints of each hand and the metatarsal-phalangeal joints of each foot. For example, the maximum score for the five metacarpal-phalangeal joints of the right hand would be 3, not 15.

44 Swollen joint count
A 44 swollen joint count (see table below) is included in the DAS and includes the sternoclavicular and acromioclavicular joints, the shoulders, elbows, wrists, knees, ankles, and MCP, PIP, and MTP joints.

Figure 2, overview of joint assessments, from Sokka and Pincus, Clin Exp Rheumatol 2005; 23 (Suppl. 39):S58-S62.

Joint	66/68 joints (15)	Ritchie Index (16)	44 joints (22)	36 joints (23)	28 joints (19)	42 joints (24)
Temporomandibular	+	+*				
Sternoclavicular	+	+*	+			
Acromioclavicular	+	+*	+			
Shoulder	+	+	+		+	+
Elbow	+	+	+	+	+	+
Wrist	+	+	+	+	+	+
Metacarpophalangeal		+				
First	+		+		+	+
Second	+		+	+	+	+
Third	+		+	+	+	+
Fourth	+		+	+	+	+
Fifth	+		+	+	+	+
Proximal interphalangeal		+				
First	+		+	+	+	+
Second	+		+	+	+	+
Third	+		+	+	+	+
Fourth	+		+	+	+	+
Fifth	+		+	+	+	+
Distal interphalangeal						
Second	+					
Third	+					
Fourth	+					
Fifth	+					
Hip	+#	+				+
Knee	+	+	+	+	+	+
Ankle	+	+	+	+		+
Talocalcaneal		+				
Tarsus	+	+				
Metatarsophalangeal		+				
First	+		+	+		+
Second	+		+	+		+
Third	+		+	+		+
Fourth	+		+	+		+
Fifth	+		+	+		+
Proximal interphalangeal (toe)						
First	+					
Second	+					
Third	+					
Fourth	+					
Fifth	+					

#Assessed for tenderness only; *right and left joints assessed together.

20.5 Appendix 5 SF-36

The SF-36 is a multi-purpose, short-form health survey with 36 questions. It yields an 8-scale profile of functional health and well-being scores as well as psychometrically-based physical and mental health summary measures and a preference-based health utility index. It is a generic measure, as opposed to one that targets a specific age, disease, or treatment group. Accordingly, the SF-36 has proven useful in surveys of general and specific populations, comparing the relative burden of diseases, and in differentiating the health benefits produced by a wide range of different treatments.

See CRF for the wording of the questionnaire.

20.6 Appendix 6 EQ-5D

See CRF for the wording of the questionnaire.

20.7 Appendix 7 HAQ-promis

The Health Assessment Questionnaire (HAQ) was originally developed in 1978 by James F. Fries, MD, and colleagues at Stanford University. It was one of the first self-report functional status (disability) measures and has become the dominant instrument in many disease areas, including arthritis. It is widely used throughout the world and has become a mandated outcome measure for clinical trials in rheumatoid arthritis and some other diseases. In this trial we will apply the updated version called HAQ-promis, which has one extra response category compared to the original HAQ.

Patients usually find the instruments self-explanatory. Clarification is seldom required.

PROMIS 20-item Physical Function Short Form

Response pattern: Without any difficulty - With a little difficulty - With some difficulty - With much difficulty - Unable to do

Are you able to do chores such as vacuuming or yard work?
Are you able to run a short distance, such as to catch a bus?
Are you able to get in and out of a car?
Are you able to push open a heavy door?
Are you able to dry your back with a towel?
Are you able to change a light bulb overhead?
Are you able to hold a plate full of food?
Are you able to transfer from a bed to chair and back?
Are you able to dress yourself, including tying shoelaces and doing buttons?
Are you able to wash and dry your body?
Are you able to shampoo your hair?
Are you able to squeeze a new tube of toothpaste?
Are you able to sit on the edge of a bed?
Are you able to get on and off the toilet?

Response pattern: Not at all - Very little - Somewhat - Quite a lot - Cannot do

Does your health now limit you in lifting or carrying groceries?
Does your health now limit you in doing two hours of physical labor?
Does your health now limit you in doing vigorous activities, such as running, lifting heavy objects, participating in strenuous sports?
Does your health now limit you in climbing one flight of stairs?
Does your health now limit you in bending, kneeling, or stooping?
Does your health now limit you in walking more than a mile?

20.8 Appendix 8 RAID

See CRF for the wording of the questionnaire.

20.9 Appendix *9 EULAR response*

Based on the DAS, response criteria have been developed: the EULAR response criteria. The EULAR response criteria include not only change in disease activity but also current disease activity. To be classified as responders, patients should have a significant change in DAS and also low current disease activity. Three categories are defined: good, moderate, and non-responders.

20.10 Appendix 10 Sharp van der Heijde Score

Radiographs of hands (posterior/anterior) and foot (anterior/posterior) will be taken at baseline, 3, 6, 12, 16 and 24 months. The modified Sharp van der Heijde Score (vdHSS) will be calculated, including an erosion score and a joint space narrowing score.

According to the vdHSS, erosion is assessed in 16 joints (five MCP, four PIP, IP of the thumbs, 1st MCB, radius and ulna bones, trapezium and trapezoid as one unit (multangular), navicular, lunate) for each hand and wrist, and six joints (five MTP, IP) for each foot. One point is scored if erosions are discrete, rising to 2, 3, 4, or 5 depending on the amount of surface area affected. The score for erosion ranges from 0 to 160 in the hands and from 0 to 120 in the feet (the maximum erosion score for a joint in the foot is 10). JSN is assessed in 15 joints (five MCP, four PIP, CMC 3 to 5, multangular navicular-lunate, radiocarpal) for each hand and wrist, and six joints (five MTP, IP) for each foot. JSN is combined with a score for (sub)luxation and scored as follows: 0 = normal; 1 = focal or doubtful; 2 = generalised, less than 50% of the original joint space; 3 = generalised, more than 50% of the original joint space or subluxation; 4 = bony ankylosis or complete luxation. The score for JSN ranges from 0 to 120 in the hands and from 0 to 48 in the feet (43). The total score is the sum of scores of erosion and JSN, the maximum score being 448.

Original radiographs will be scored at a central reading site by two experienced readers.

20.11 Appendix 11 Ultrasonography score

A Siemens Antares or machine with similar qualities will be used for the ultrasonographic assessments, with the following (preliminary) specifications:
PRF = 391
Frequency = 7.3 MHz
R/S = 5
Filter = 2
The following joints will be scored for grey scale synovitis and power Doppler signal, with the best possible match to a reference atlas developed by HBH that is accessible within the GoTreatIT software. The scoring system is based on the OMERACT US recommendations, semi-quantitatively from 0 - 3 in 32 joints: bilateral MCP I-V, RCJ, DRUJ, intercarpal, elbow, knee, talocrural joints and MTP I-V, giving a maximum total score of 96 for both grey scale and power Doppler.

In addition to the 32 joints the following will be scored, but not be part of the response algorithm: Bilateral PIP2 and 3 joints, extensor carpi ulnaris tendon and tibialis posterior tendon.

20.12 Appendix 12 RAMRIS

Magnetic resonance imaging of the dominant hand and wrist with and without gadolineum will be undertaken at baseline, 3, 6, 12, 16 and 24 months. The same side will be assessed at all time-points.

Images will be read according to the RAMRIS score at a central reading site by two experienced readers. The RAMRIS consists of MRI definitions of important joint pathologies, a core set of MRI sequences and a semi-quantitative scoring system for erosions, bone marrow edema and synovitis (44). The RAMRIS core set of MRI sequences to assess inflammatory as well as destructive changes in RA joints includes: Imaging in 2 planes, with T1-weighted images before and after intravenous gadolinium-contrast to assess synovitis and erosions; plus a T2-weighted fat saturated sequence or a STIR sequence to assess bone marrow edema.

Both flexor and extensor tenosynovitis will also be evaluated, according to the scoring system described by Haavardsholm et al, at the level between the radioulnar joint and the hook of the hamate, thus including both wrist and finger tendons (45). Flexor and extensor tenosynovitis are evaluated semi-quantitatively in 10 different anatomical areas, graded from grade 0 to 3 (total score 0–30). The grading is based on the maximum width (in mm) of post-contrast enhancement within each anatomical area on axial T1-weighted MR images.

20.13 Appendix 13 WPAI (Norwegian translations available)

Work Productivity and Activity Impairment Questionnaire:
Rheumatoid arthritis V2.0 (WPAI:RA)

The following questions ask about the effect of your rheumatoid arthritis on your ability to work and perform normal daily activities. *Please fill in the blanks or circle a number, as indicated.*

1. Are you currently in paid employment? _____ NO ___ YES
   If 'NO', tick "NO" and skip to question 6.

The next questions refer to the **past seven days**, not including today.

2. During the past seven days, how many hours did you miss from work because of problems associated with your rheumatoid arthritis? *Include hours you missed on sick days, times you went in late, left early, etc., because of your rheumatoid arthritis. Do not include time you missed to participate in this study.*

   _____ HOURS

3. During the past seven days, how many hours did you miss from work because of any other reason, such as annual leave, holidays, time off to participate in this study?

   _____ HOURS

4. During the past seven days, how many hours did you actually work?

   _____ HOURS *(If "0", skip to question 6.)*

5. During the past seven days, how much did your rheumatoid arthritis affect your productivity while you were working?

   Think about days you were limited in the amount or kind of work you could do, days you accomplished less than you would like, or days you could not do your work as carefully as usual. If rheumatoid arthritis affected your work only a little, choose a low number. Choose a high number if rheumatoid arthritis affected your work a great deal.

   Consider only how much rheumatoid arthritis affected productivity while you were working.

Rheumatoid arthritis had no effect on my work	0	1	2	3	4	5	6	7	8	9	10	Rheumatoid arthritis completely prevented me from working

   CIRCLE A NUMBER

6. During the past seven days, how much did your rheumatoid arthritis problems affect your ability to perform your normal daily activities, excluding your job?

   By normal activities, we mean the usual activities you perform, such as working around the house, shopping, childcare, exercising, studying, etc. Think about times you were limited in the amount or kind of activities you could perform and times you accomplished less than you would like. If rheumatoid arthritis affected your activities only a little, choose a low number. Choose a high number if rheumatoid arthritis affected your activities a great deal.

Consider only how much <u>rheumatoid arthritis</u> affected your ability to perform your normal daily activities, excluding your job.

Rheumatoid arthritis had no effect on my daily activities	0	1	2	3	4	5	6	7	8	9	10	Rheumatoid arthritis completely prevented me from performing my daily activities

CIRCLE A NUMBER

WPAI:RA V2.0 (English for the UK)

20.14 Appendix 14 Adverse events

An **Adverse Event** (AE) is defined as: Any untoward medical occurrence in a patient or clinical investigation subject administered a pharmaceutical product and which does not necessarily have a causal relationship with this treatment. An adverse event (AE) can therefore be any unfavourable and unintended sign (including an abnormal laboratory finding), symptom, or disease temporally associated with the use of a medicinal (investigational) product, whether or not related to the medicinal (investigational) product (see the ICH Guideline for Clinical Safety Data Management: Definitions and Standards for Expedited Reporting).

A **Serious Adverse Event** (SAE) is defined as: Any untoward medical occurrence that at any dose:
• results in death,
• is life-threatening,
• requires inpatient hospitalization or prolongation of existing hospitalization,
• results in persistent or significant disability/incapacity,
or
• is a congenital anomaly/birth defect
(see the ICH Guideline for Clinical Safety Data Management: Definitions and Standards for Expedited Reporting).

The **severity/intensity** of adverse events is graded on a 3-point scale as follows:

MILD: Discomfort noted, but no disruption to normal daily activities.
MODERATE: Discomfort sufficient to reduce or affect normal daily activities.
SEVERE: Inability to perform normal daily activities.

The following AEs should be reported in the eCRF:

- Any Adverse Event (including laboratory abnormalities) that leads to a clinical condition requiring medical treatment or leads to a change of medication.
- In addition, all Adverse Events that may have a relationship (probable, possible or remote) to treatment with intra-articular steroid injections should be reported, even if the above criterion is not met.

All SAEs encountered during the clinical trial will be reported in the eCRF, and reported to RELIS.

The investigator's assessment of the event's relationship to the trial treatment should be defined according to the following standard statement:

Categories for Determining Relationship to Trial Treatment:

PROBABLE (must have first three)

This category applies to those adverse events which are considered, with a high degree of certainty, to be related to the trial treatment. An adverse event may be considered **probable**, if:

1. It follows a reasonable temporal sequence from administration of the drug.
2. It cannot be reasonably explained by the known characteristics of the subject's clinical state, environmental or toxic factors, or other modes of therapy administered to the subject.

3. It disappears or decreases on cessation or reduction in dose. (There are important exceptions when an adverse event does not disappear upon discontinuation of the drug, yet drug-relatedness clearly exists; e.g. [1] bone marrow depression, [2] tardive dyskinesias.)
4. It follows a known pattern of response to the suspected drug.
5. It reappears upon re-challenge.

POSSIBLE (must have the first two)

This category applies to those adverse events in which the connection with the test drug administration appears unlikely, but cannot be ruled out with certainty. An adverse event may be considered **possible** if, or when:

1. It follows a reasonable temporal sequence from administration of the drug.
2. It may have been produced by the subject's clinical state, environmental or toxic factors, or other modes of therapy administered to the subject.
3. It follows a known pattern of response to the suspected drug.

REMOTE (must have the first two)

In general, this category is applicable to an adverse event which meets the following criteria:
1. It does **not** follow a reasonable temporal sequence from administration of the drug.
2. It may readily have been produced by the subject's clinical state, environmental or toxic factors, or other modes of therapy administered to the subject.
3. It does not follow a known pattern of response to the suspected drug.
4. It does not reappear or worsen when the drug is re-administered.

UNRELATED

This category is applicable to those adverse events which are judged to be clearly and incontrovertibly due only to extraneous causes (disease, environment, etc.) and do not meet the criteria for drug relationship listed under REMOTE, POSSIBLE or PROBABLE.

Immediately Reportable Adverse Events

Any adverse event that is considered SERIOUS (as defined below) must be reported IMMEDIATELY by the investigator, ie. within one working day of becoming aware of the event, and a report should be sent to RELIS.

An **Immediately Reportable Adverse Event (IRAE)** is any serious adverse event or abnormal laboratory test value that occurs during the defined treatment period, and which suggests a significant hazard, contraindication, side effect or precaution. An IRAE **must** be immediately reported to the clinical monitors **within one working day**.

Immediately Reportable adverse events include any event or experience that is:
- fatal;
- life threatening;
- permanently disabling (ie. severely incapacitating or interfering with the ability to resume usual life patterns);
- requires inpatient hospitalisation or prolonged hospitalisation;
- an overdose (ie. a deliberate or inadvertent administration of a treatment at a dose higher than that specified in the protocol *and* higher than known therapeutic doses for that specific indication)

The definition of an IRAE includes any event which is expected or unexpected, related or unrelated to the drug.

All IRAEs must also be reported in the eCRF and must be assessed for severity and the relationship to the trial treatment. The actions taken by the investigator and the outcome of the event must be reported.

IRAEs must be reported to the appropriate ethics committee, if requested by the committee and/or according to local legal requirements.

It is important that the SEVERITY of an adverse event is not confused with the SERIOUSNESS of the event. For example, vomiting which persists for many hours may be severe, but is not necessarily a serious adverse event. On the other hand, stroke which results in only a limited degree of disability may be considered a mild stroke, but would be a serious adverse event.

Drug Administration and Follow-Up of Adverse Events
All adverse events MUST be documented and followed up until the event is either resolved or adequately explained, even after the subject has completed their trial treatment.

21 Norwegian summary

Norsk protokollsammendrag ARCTIC

Norsk kort tittel: Remisjon ved tidlig RA - ARCTIC-studien

Deltakende sentra: Revmatologisk avdeling ved: Diakonhjemmet Sykehus AS (koordinerende senter), Universitetssykehuset i Nord-Norge, St. Olav Hospital, Haugesund Sanitetsforenings Revmatismesykehus AS, Helse Sunnmøre HF Ålesund, Haukeland Universitetssykehus Bergen, Sørlandet Sykehus Kristiansand HF, Martina Hansens Hospital, Betanien Hospital Skien, Sykehuset Østfold HF Moss, Lillehammer Revmatismesykehus, Vestre Viken HF Buskerud, Privat spesialist praksis Bendvold/Dovland Kristiansand

Prosjektleder: Lege Ph.D. Espen A. Haavardsholm , Diakonhjemmet Sykehus.

Klinisk ansvarlig: Avdelingssjef professor dr. med. Tore K. Kvien, Diakonhjemmet Sykehus.

Rådgiver studie-design og imaging: Professor Ph.D Désirée van der Heijde, Diakonhjemmet Sykehus og Leiden University Medical Centre.

Medisinsk statistiker: Inge C. Olsen, Ph.D., Smerud Medical AS.

Monitor (GCP): Seksjon for GCP, OUS (Helse Sør-Øst) og Innovest AS (øvrige helseregioner).

Fase: IV.

Populasjon: Pasienter med revmatoid artritt i henhold til ACR/EULAR 2010 kriterier, symptomvarighet < 2 år og indikasjon for DMARD behandling.

Målsetting: Hensikten med denne studien er å undersøke hvilken behandlingsstrategi som gir best resultater ved tidlig RA med tanke på å oppnå remisjon og hindre røntgenologisk leddskade: Enten tett oppfølgning med ultralyd kombinert med injeksjonsbehandling av subklinisk inflammasjon (ledd med Power Doppler signal), eller tett konvensjonell klinisk oppfølgning (uten bruk av ultralyd) kombinert med injeksjonsbehandling av klinisk hovne ledd.

Design: Multisenter, randomisert, åpen, to-armet parallell-gruppe studie av to års varighet.

Antall pasienter: 240.

Tidsplan: Inklusjonsperiode: 06.09.10 til 05.09.12, deretter 24 måneders oppfølgning. Interimanalyse etter 12 måneder.

Behandlingsregime: Den medikamentelle behandlingen er identisk i de to armene, men vurderingen av respons er noe ulik, da man i ultralydarmen også tar hensyn til endring i UL-score. Ved oppnådd respons eller nådd målsetting om remisjon kontinueres det medikamentelle regimet. Ved manglende respons er det først doseendring, hvis ikke dette fører til respons går man videre til neste nivå i tabellen under. På hver visitt injiseres hovne ledd/ledd med PD-signal, og det kan injiseres inntil 80mg Triamcinolon hexacetonid per visitt.

Nedenfor er en skjematisk oversikt over behandlingsregimet:

Nivå	**Medikamentregime**
1	Behandling med **methotrexate** (initialt 15mg/uke, opptrapping til 20 mg/uke), med innledningsvis prednisolonkur (imitialt 15 mg, trappes ned over 7 uker). Eventuelt doseøkning til 20mg/uke methotrexate ved manglende respons etter 2 måneder.
2	Trippelbehandling med **methotrexate**, **plaquenil** og **salazopyrin**, eventuelt med tillegg av prednisolon ved manglende respons etter 2 måneder.
3	Behandling med kombinasjon av **methotrexate** og **biologisk legemiddel** (1. valg*), eventuelt med dosejustering ved manglende respons etter 2 måneder.
4	Behandling med kombinasjon av **methotrexate** og **biologisk legemiddel** (2. valg*), eventuelt med dosejustering ved manglende respons etter 2 måneder.
5	Behandling med kombinasjon av **methotrexate** og **biologisk legemiddel** (3. valg*), eventuelt med dosejustering og tillegg av prednisolon ved manglende respons etter 4 måneder.

* *I henhold til gjeldene LIS-anbefalinger.*

Effektmål: Primærendepunkt:
Komplett klinisk remisjon siste 8 måneder av studien (definert som DAS < 1.6, ingen hovne ledd og ingen røntgenprogresjon).

Sekundærendepunkter (ikke fullstendig, se engelsk fullstendig protokoll):

- EULAR respons kriterier
- ACR respons (ACR20/50/70/90)
- Kliniske remisjonskriterier (ACR/EULAR)
- RAMRIS score (MR av dominant hånd)
- Mod. Sharp van der Heijde Score (konvensjonell røntgen hender/føtter)
- HAQ-PROMIS
- SF-36
- WPAI

Bivirkninger: DEXA av hofte/rygg, puls/blodtrykk, hematologi, klin. kjemi, før øvrig AE/SAE rapportering etter ICH Guidelines (med noen modifikasjoner).

November 30th 2015

Dear Editor,

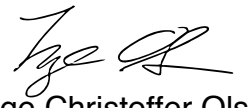
Attached is the final version of the Statistical Analysis Plan for the ARCTIC clinical trial (Aiming for remission in rheumatoid arthritis: a randomized trial examining the benefit of ultrasonography in a clinical tight control regimen, ClinicalTrials.gov NCT01205854).

The original version of the SAP was approved on September 9th 2015, with final approval of the current version on October 29th 2015. Section 1 of the SAP documents the changes made in the amended version, and confirms that no changes were made to the primary analysis of the trial. The amendment was approved after the database was locked and after data review.

The changes made in the amendment are summarized below. We have no indication that the changes to the planned analyses after review of the data could compromise the validity of the results:
- An additional analysis set (completer analysis set) was added. The addition of the completer analysis set was done prior to analysis of any sub-group of the primary full analysis set and should thus not affect the validity of the results.
- The handling of missing data for radiographic scores was changed from multiple imputation to linear intra- and extrapolation. This change was mainly to conform with typical analyses performed for radiographic scores, in addition to convergence problems in the multiple imputation method. As linear intra- and extrapolation imputation is a commonly used method for radiographic imputation, this change should not be considered controversial.
- A specification of the median regression analysis, including bootstrapping for inference. This is not a controversial methodological approach.

Overall, the analyses were done as pre-specified, with a few and relatively small alterations after data base lock.

Sincerely yours,



Inge Christoffer Olsen, PhD
Statistician
Department of Rheumatology
Diakonhjemmet Hospital
Oslo, Norway

Statistical Analysis Plan

Aiming for remission in rheumatoid arthritis: a randomized trial examining the benefit of ultrasonography in a clinical tight control regimen

The ARCTIC trial

Protocol DIA2010-1

Final Protocol Version 1.7: 20 September 2010

SAP Version 2.0 (Amendment 1 of the SAP)

Date: 29 October 2015

Approval Page

Biostatistician Diakonhjemmet Hospital

Signature _____

Inge Christoffer Olsen, PhD

Date: November 29th, 2015

Principal investigator/Project leader

Signature _____

Espen A. Haavardsholm, PhD MD

Date November 29th, 2015

Abbreviations

ARD	Adjusted Risk Difference
ARR	Adjusted Relative Risk
ATC	Anatomical Therapeutic Chemical
AUC	Area under the curve
BMD	Bone mineral density
BME	Bone Marrow Edema
BMI	Body Mass Index
CDAI	Clinical Disease Activity Index
CI	Confidence Interval
CRP	C-Reactive Protein
CSR	Clinical Study Report
DAS	Disease Activity Score
DMARD	Disease modifying anti-rheumatic drug
eCRF	Electronic Case Report Form
EOT	End of Treatment
ESR	Erythrocyte Sedimentation Rate
EULAR	European League Against Rheumatism
FAS	Full analysis Set
FDA	U.S. Food and Drug Administration
GS	Grey Scale
HAQ	Health Assessment Questionnaire
HR	Hazard Ratio
HRQoL	Health related quality of life
IDMC	Independent Data Monitoring Committee
ITT	Intention to treat
JSN	Joint Space Narrowing
KM	Kaplan Meier
MCP	Metacarpophalangeal
MedDRA	Medical Dictionary for Regulatory Activities
MRI	Magnetic resonance image
MTP	Metatarsophalangeal
MTX	Methotrexate
NRS	Numeric rating scale
PD	Power Doppler
PGA	Patient's Global Assessment of disease activity
PH	Proportional Hazards
PhGA	Physician's Global Assessment of disease activity
PIP	Proximal interphalangeal
PP	Per Protocol
PT	Preferred Term
QoL	Quality of Life
RA	Rheumatoid Arthritis

RAI	Ritchie Articular Index
RAID	The Rheumatoid Index of Disease
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SC	Sternaclavicular
SD	Standard Deviation
SDAI	Simplified Disease Activity Index
SE	Standard Error
SJC	Swollen Joint Count
SOC	System Organ Class
TEAE	Treatment Emerging Adverse Events
TFL	Tables Figures Listings
TJC	Tender Joint Count
US	Ultrasound
VAS	Visual Analogue Scale
vdHSS	Sharp van derHeijde Score
WHO	World Health Organisation
WPAI	Work Productivity and Activity Impairment Questionnaire

TABLE OF CONTENTS

1. Amendments from previous version

Note: This Statistical Analysis Plan was amended after database lock and after initiation of the analyses.

1.1 Amendment 1

Changes from the previous version (version 1.0, September 9, 2015) have been made:

- The Completer Analysis Set was included as Section 4.4.
- Specification of the median regression analysis method used (Section 8.1.1.2)
- In order to confirm with the typical analyses performed for radiographic total scores and its components, missing values will be imputed using linear intra- and extrapolation (Section 8.2.3).

2. INTRODUCTION

This document describes the planned data summaries and statistical analyses to be performed for the Clinical Trial Protocol ARCTIC (DIA2010-1): Aiming for remission in rheumatoid arthritis: a randomized trial examining the benefit of ultrasonography in a clinical tight control regimen - The ARCTIC trial. It is intended to supplement the study protocol, which contains details regarding the objectives and design of the study.

2.1 Study Objectives

2.1.1 Primary Objective

The primary objective of this treatment strategy study is to assess the effect of applying ultrasonography versus not applying ultrasonography in a clinical tight control regimen in patients with early RA with respect to achieving clinical remission and non-progression of structural damage.

2.1.2 Secondary Objectives

The secondary objectives of this study are to assess the two treatment regimens regarding:

- Efficacy of the two treatment regimens with respect to various measures of progression of structural damage
- Efficacy of the two treatment regimens with respect to various measures of low disease activity and remission
- Health economics (evaluation of direct and indirect costs in the treatment groups, cost- effectiveness of both treatment arms)
- Possible prognostic factors for structural damage/reaching remission, including proteomics, imaging and biomarkers
- Strength of associations over time both for damage, inflammation and physical function
- Time to reach various levels of disease activity/remission
- The burden of illness (disease activity, damage, disability, cost of illness) in patients with early arthritis
- Efficacy with respect to physical function, health related quality of life (HRQoL) and pain
- Aspects of imaging remission (especially if specific levels of MRI inflammation or US inflammation are related to non-progression of structural damage)
- Work performance (including aspects of "absenteeism" and "presenteeism")
- MRI measures of damage and inflammation in the two groups, and how MRI features evolve over time and how they relate to radiographic damage and US evaluation
- The feasibility of applying ultrasonography in a tight control dosing regimen in clinical practice
- The area under the curve of measures of disease activity/inflammation (clinical, US and MRI)
- The performance of various outcome measures in patients with early arthritis
- Patient satisfaction
- Number and kind of adverse events in both groups

2.2 Study Design

This is an open, national, multicentre, 2-arm, randomized, parallel group, prospective clinical study of 2 years duration of a tight control regimen with and without application of musculoskeletal ultrasonography, aiming at remission in patients with early RA. Patients in both groups are treated according to the same treatment protocol as outlined in Table 1. Treatment adjustments (including i.a. glucocorticoid injections) may be made at every visit according to a pre-specified dosing regimen, described in table 1. The decision rules are described in table 2. The treatment target is remission defined as DAS <1.6, plus the following criteria (different for the two treatment arms):

A. Non-US arm: No swollen joints.

B. US arm: No swollen joints and no joints with PD-signal.

The standardized US scoring system used has been developed in a pilot study (Hammer et al., 2011), and includes assessments of the following 32 joints: MCPs I-V, wrist (radio-carpal, radio-ulnar and inter-carpal), elbow, knee, talo-crural and MTP I-V bilaterally. In addition, the PIP II-III and extensor carpi ulnaris tendon and the tibialis posterior tendon are scored bilaterally. The time for scoring and recording both a grey-scale and a power-Doppler semi-quantitative score of 0-3 for each joint is estimated to be approximately 15 minutes (median).

The decision of whether to adjust medication is based on change in and the level of the Disease Activity Score (DAS). If the patient does not respond as described in Table 2, the treating physician immediately adjusts the therapy by proceeding to the next step in the treatment regimen. If a patient responds or has reached the target, current medication is continued. In the US group, the physician may overrule the decision based on DAS and proceed to the next step based on US findings (as described in Table 2).

In both groups, clinically swollen joints were treated by i.a. steroids when indicated. In the US group an additional target were all joints with PD signal, and additionally all injections were performed US-guided. For both groups, i.a. injections of only tender joints were not allowed. Likewise, inflamed tendon sheets were injected in the non-US arm as clinically indicated – in the US arm tenosynovitis was verified by US before injections were allowed. Maximum dosage triamcinolone hexacetonid per visit was 80 mg which could be distributed within joints and tendon sheets as decided by the treating rheumatologist.

Table 2.1 Treatment regimen

Visit (months)	Treatment if no response (if response continue treatment at present step, see fig 2)
1 (0)	**A. Monotherapy* + Prednisolone:** 1. Methotrexate 15 mg/week, increase by 2,5 mg every 2nd week to target dose 20 mg/week, i.e. week 1+2 15mg, week 3+4 17.5 mg, week 5-8 20 mg (optional reduced dosage starting scheme for patients at risk for side effects: week 1 10 mg, week 2 12,5mg, week 3 15 mg, week 4 17,5mg, week 5-8 20 mg) 2. Concomitant folic acid 5 mg/week (1mg 5/7 days or 5 mg x 1/week) 3. Prednisolone 15 mg week 1, 10 mg week 2, 7,5 mg week 3, 5 mg week 4+5, 2,5 mg week 6+7 4. Calcium supplement 1000mg x 1 (while on prednisolone)
2 (1)	**A. Monitor start-up regimen (no changes in medication allowed unless due to AE)*** Joint injections allowed as indicated according to treatment arm.
3 (2)	**A. Optimize monotherapy*** Increase Methotrexate to 25-30 mg/week *Or increase SSZ/HCL/leflunomide dose*
4 (3)	**A. Monitor start-up regimen (no changes in medication allowed unless due to AE)*** Joint injections allowed as indicated according to treatment arm.
5 (4)	**B.** Triple combination therapy (or other combination therapy if MTX not tolerated):** 1. Add salazopyrine, step up over 4 weeks to 500mg 2 x 2 and 2. Add hydroxychlorochine 200mg 1 x 2
6 (6)	**B. Optimize triple combination therapy:**** Add Prednisolone 7,5 mg 1 x 1
7 (8)	**C. DMARD*** and 1st biologic**** (according to LIS guidelines):** 1. Highest tolerable dose MTX * and 2. Add 1st biologic **Or SSZ/HCL/leflunomide if MTX not tolerated*
8 (10)	**C. DMARD and 1st biologic:** Adjust dose/interval of 1st biologic
9 (12)	**D. DMARD*** and 2nd biologic (according to LIS guidelines):** Switch to 2nd biologic (*according to current LIS-guidelines)*
10 (14)	**D. DMARD*** and 2nd biologic:** Adjust dose/interval of 2nd biologic
11 (16)	**E. DMARD*** and 3rd biologic (according to LIS guidelines):** Switch to 3rd biologic (*according to current LIS-guidelines)*
12 (20)	**E. Optimize DMARD and 3rd biologic plus prednisolone:** Adjust dose/interval of 3rd biologic and/or add prednisolone 7,5mg
13 (24)	**F.** Continue medication according to standard clinical care

***** *If MTX is not tolerated, switch to subcutaneous methotrexate (metoject), then continue according to scheme. In case of AE or not tolerated even in low dose subcutaneous, switch to salazopyrine or hydroxychlorochine monotherapy (standard dosage) if low disease activity, or leflunomide 20 mg in case of moderate or high disease activity (loading dose 40mg x 1 for 3 days, then 20 mg per day).*

*** In patients with high disease activity and risk factors for progressive joint destruction (ACPA or RF-positive and either erosions on CR or baseline RAMRIS bone marrow edema score >2) a rescue option is available which includes moving to the next step, i.e. introduce 1st biologic (treatment C at visit #5, without prescribing treatment B).*

**** In case of no tolerance for any conventional DMARD, this can be omitted if the biologic drug chosen has indication for monotherapy (e.g. tociluzimab)*

***** Requirement for adding biologic: There must be objective signs of ongoing inflammation, i.e. either elevated ESR/CRP (>UNL, and not due to other disease/infection) or SJC≥1 (or PD score >1 in US arm)*

Table 2.2 Definition of decision rules (to be applied at all visits except #2 and #4)

	Current DAS	No response*	Response*	Reached target (DAS < 1.6)
Conventional tight control	≤ 2.4	Change of DAS < 0.6	Change of DAS ≥ 0.6	DAS < 1.6 and no swollen joints
	> 2.4	Change of DAS < 1.2	Change of DAS ≥ 1.2	
US tight control	≤ 2.4	Change of DAS < 0.6 or <10% decrease of US total score	Change of DAS ≥ 0.6 and ≥ 10% decrease of US total score	DAS < 1.6 and no swollen joints and no joints with power Doppler synovitis
	> 2.4	Change of DAS < 1.2 or <20% decrease of US total score	Change of DAS ≥ 1.2 and ≥ 20% decrease of US total score	
Action		Change therapy	Continue current medication	Continue current medication**

** Both in cases of response and no response should clinically swollen joints be i.a. injected with steroids when indicated, up to the maximum allowed dosage per visit (80 mg triamcinolone hexacetonid) . In group B joints with PD-signal on US is an additional target.*
*** If sustained remission ≥ 12 months, step-down to monotherapy MTX. If continued sustained response after this, decrease MTX by 2,5mg/week per 2 months.*

3. Hypotheses and decision rules

3.1 Statistical Hypotheses

This protocol is designed to establish the superiority of applying ultrasonography to not applying ultrasonography in a clinical tight control regimen in patients with early RA for the primary endpoint complete DAS remission after 24 months of treatment. The null hypothesis is that there is no difference in the probability of achieving complete DAS remission after 24 months of treatment between the two treatment regimes (applying vs not applying ultrasonography). The alternative hypothesis is that there is a difference in the probability of achieving complete DAS remission after 24 months of treatment between the two treatment regimes (applying vs not applying ultrasonography).

3.2 Statistical Decision Rule

This protocol is designed to address a single primary endpoint. Statistical significance is claimed if the null hypothesis is rejected on the significance level (alpha) of 0.05 (two-sided). That is, if the p-value of the null hypothesis test is less than or equal to 0.05.

4. ANALYSIS SETS

4.1 Enrolled

The Enrolled set will include all patients who have provided informed consent and have been included into the study data base.

4.2 Full Analysis Set

The Full Analysis Set (FAS) will be defined as all patients randomly assigned to a treatment group, and have started the allocated intervention defined as having completed at least one regular visit after the baseline visit. The FAS will form the primary analysis set of the study, and used for all primary and secondary endpoints.

4.3 Safety Analysis Set

The Safety Set will include all patients who completed at least one regular visit after the baseline visit.

4.4 Completer Analysis Set
The Completer Analysis Set will include all randomised patients having started the allocated intervention and not withdrawn during during the study.

4.5 Per Protocol Analysis Set
The Per Protocol Analysis Set (PPS) will include all randomised patients meeting the study entry criteria and with no major protocol deviations (see SAP section 4.7).

4.6 Treatment Misallocation
If patients were:

- randomized but not treated: patient will appear on the patient evaluation table as randomized but not treated; this is the extent of how much the patient will be reported.
- treated but not randomized: then by definition the patient will be excluded from both the efficacy and safety analyses since randomized treatment is missing
- randomized but did not follow protocol according to allocation: then they will be reported under their randomized treatment group for all efficacy and safety analyses as part of the FAS and safety analyses, but omitted from the PPS

Note that treated in this section is defined as having completed at least one post-baseline regular visit.

4.7 Protocol Deviation
The following sections describe any protocol deviations that relate to the statistical analyses and forms the requirement for exclusion from the PPS.

4.7.1 Deviations to inclusion and/or exclusion criteria
Any patient who enters the study when the inclusion or exclusion criteria would have prevented entry will be considered to have had a protocol deviation.

4.7.2 Deviations assessed Post-randomization
Only protocol deviations thought to affect the efficacy of application of ultrasonography will be considered in the SAP. Each of these cases will be reviewed by the team and a clinical judgment made in each particular circumstance as to whether efficacy would have been affected in the case of these specific classes of protocol deviations assessed post randomization:

- Patients who did not follow the tight control regime and treatment decision rules outlined in section 1.2
- Patients who withdrew or was withdrawn during the study

5. DEFINITIONS AND DERIVED VARIABLES

For all clinically planned measures, visits should occur within a window of the scheduled visit, see table below for definitions.

Visit Label	Target Day	Definition (Day window)
Screening	-1	Prior to Day 0
V1. Baseline	Day 0 (Randomization)	Day 0
V2. Month 1	28	Days 1 to 42
V3. Month 2	56	Days 43 to 73
V4. Month 3	91	Days 74 to 105
V5. Month 4	119	Days 106 to 151
V6. Month 6	182	Days 152 to 213
V7. Month 8	245	Days 214 to 273
V8. Month 10	301	Days 274 to 332
V9. Month 12	364	Days 333 to 395
V10. Month 14	427	Days 396 to 458
V11. Month 16	490	Days 459 to 549
V12. Month 20	609	Days 550 to 668
V13. Month 24	728	Days 669 to 816

If two or more visits fall into the same window, keep the one closest to the Target Day. If two visits are equal distant from the Target Day in absolute value, the later visit should be used.

5.1 Change from baseline
Change from baseline (Δ) = time-point value - baseline value.
% change from baseline (%Δ) = [(time-point value – baseline value) / baseline value] *100%

5.2 Inflammation parameters
Inflammation parameters include the Erythrocyte Sedimentation Rate (ESR) in mm/h and high sensitivity C-reactive protein (CRP) in mg/L. ESR is assessed by the Westergren method.

5.3 Disease Modifying Anti-Rheumatic Drugs
Disease Modifying Anti-Rheumatic Drugs (DMARDs) include the following:

5.3.1 Biologic drugs:
Including anakinra (Kineret), etanercept (Enbrel), adalimumab (Humira), infliximab (Remicade/Remsima/Inflectra), abatacept (Orencia), tocilizumab (RoActemra), golimumab (Simponi), certolizumab pegol (Cimzia), rituximab (Mabthera)

5.3.2 Synthetic drugs:
Including methotrexate, auranofin (oral gold), aurothiomalate (injectable gold), sulfasalazine, d-penicillamine, azathioprine, cyclosporine, tacrolimus, leflunomide (Arava), tetracycline (used as a DMARD), hydroxychlorochin. sulf. (Plaquenil)

5.4 Joint Counts
The Ritchie Articular Index (RAI) is a long-standing approach to a graded assessment of the tenderness of 26 joint regions, based on summation of joint responses after applying firm digital pressure. Four grades can be used: 0, patient reported no tenderness; +1, patient complained of pain; +2, patient complained of pain and winced; and +3, patient complained of pain, winced, and withdrew. Thus, the index ranges from 0 to 3 for individual measures and the sum 0 to 78 overall, with higher scores indicating more tenderness. Certain joints are treated as a single unit, such as the metacarpal-phalangeal and proximal interphalangeal joints of each hand and the metatarsophalangeal joints of each foot. For example, the maximum score for the five metacarpophalangeal joints of the right hand would be 3, not 15.

Swollen joint counts are performed on 44 joints, with total joint count ranging from 0 to 44. This is denoted the 44 swollen joint count (SJC44).

In addition, tender and swollen joint counts on 28 joints are denoted TJC28 and SJC28. A tender joint is defined as a joint with RAI > 0. For the MCP and IP/PIP regions we use the following definition to estimate TJC28: if RAI = 0, TJC(region) = 0; if RAI = 1, TJC(region) = 2; if RAI = 2, TJC(region) = 3: if RAI = 3, TJC(region) = 4.

See Table 5.1 Overview of Joint CountsTable 5.1 for an overview of joints and their count.

Table 5.1 Overview of Joint Counts

Joints	RAI left	RAI right	SJC44 left	SJC44 right	SJC28 left	SJC28 right
Cervical spine (Neck)	0-3		NA	NA	NA	NA
Temporomandibular (Jaws)	0-3		NA	NA	NA	NA
Sternoclavicular (SC)	0-3		0-1	0-1	NA	NA
Acromioclavicular (AC)	0-3		0-1	0-1	NA	NA
Shoulder*	0-3	0-3	0-1	0-1	0-1	0-1
Elbow*	0-3	0-3	0-1	0-1	0-1	0-1
Wrist*	0-3	0-3	0-1	0-1	0-1	0-1
Metacarpophalangeal (MCP)*						
- First (MCP1)	0-3	0-3	0-1	0-1	0-1	0-1
- Second (MCP2)			0-1	0-1	0-1	0-1
- Third (MCP3)			0-1	0-1	0-1	0-1
- Fourth (MCP4)			0-1	0-1	0-1	0-1

- Fifth (MCP5)			0-1	0-1	0-1	0-1
Proximal interphalangeal (IP/PIP)*						
- First (IP1)			0-1	0-1	0-1	0-1
- Second (PIP2)			0-1	0-1	0-1	0-1
- Third (PIP3)	0-3	0-3	0-1	0-1	0-1	0-1
- Fourth (PIP4)			0-1	0-1	0-1	0-1
- Fifth (PIP5)			0-1	0-1	0-1	0-1
Hip	0-3	0-3	NA	NA	NA	NA
Knee*	0-3	0-3	0-1	0-1	0-1	0-1
Ankle	0-3	0-3	0-1	0-1	NA	NA
Talocalcaneal	0-3	0-3	NA	NA	NA	NA
Tarsus	0-3	0-3	NA	NA	NA	NA
Metatarsophalangeal (MTP)						
- First (MTP1)			0-1	0-1	NA	NA
- Second (MTP2)			0-1	0-1	NA	NA
- Third (MTP3)	0-3	0-3	0-1	0-1	NA	NA
- Fourth (MTP4)			0-1	0-1	NA	NA
- Fifth (MTP5)			0-1	0-1	NA	NA

NA: Not assessed; *: Included in the 28-joint Disease Activity Score (DAS28)

5.5 ACR core data set

The American College of Rheumatology (ACR) definition of response includes tender and swollen joint counts, visual analogue scales (VAS) for pain, patient and investigator global assessment of disease activity, patient-assessed disability by the Health Assessment Questionnaire (HAQ) and acute phase response (ESR or high sensitivity CRP).

5.6 ACR response

The ACR response rates ACR20, ACR50, ACR70 and ACR90 as well as ACR remission rates will be calculated.

An ACR20 response is defined if the following criteria are fulfilled:

- 20% improvement in RAI, AND
- 20% improvement in swollen joint count 44, AND
- 20% improvement in at least 3 of 5 other core set items

The other core set items consist of:

- Investigator global assessment of disease activity
- Patient global assessment of disease activity
- Patient pain
- Disability
- ESR/hsCRP

ACR50, ACR70 and ACR90 are defined in a similar manner with 50%, 70% and 90% improvement, respectively. In the ARCTIC study, VAS will be used to assess pain and patient/investigator global assessment of disease activity, and PROMIS Physical Function raw score will be used to assess disability. High sensitivity CRP will be used as primary measure of inflammation, while ESR will be used if hsCRP is not available. All improvements will be % change from baseline.

Time to ACR20/50/70/90 response = Date of first visit with ACR20/50/70/90 response – date of randomisation +1

5.6.1 US Food and Drug Administration (FDA) major clinical response

Requires achieving ACR70 response at the current visit and at each visit within the previous 6 months.

5.7 ACR remission

A patient is regarded as in ACR remission if at least 5 of the following criteria are present for at least two consecutive months:

- Morning stiffness ≤ 15 minutes; self-reported
- No fatigue; self-reported fatigue ≤ 14 on VAS (100mm)
- No joint pain; self-reported joint pain ≤ 14 on VAS (100mm)
- No joint tenderness or pain on motion; RAI = 0
- No soft tissue swelling in joints or tendon sheaths; swollen joint count = 0
- ESR ≤ 30 mm/h for female or ≤ 20 mm/h for male

5.8 Disease Activity

5.8.1 DAS

Disease Activity Score (DAS) includes the RAI, the 44 swollen joint counts, the Erythrocyte Sedimentation Rate (ESR) and the Patient's Global Assessment of disease activity on a VAS 0-100 mm (PGA).

The DAS is calculated as follows:
DAS = 0.54*sqrt(RAI) + 0.065*(SJC44) + 0.33*Ln(ESR) + 0.0072*PGA
With missing values of ESR and/or PGA, the following formulas are used
DAS = 0.54*sqrt(RAI) + 0.065*(SJC44) + 0.33*Ln(ESR) + 0.22
DAS = 0.54*sqrt(RAI) + 0.065*(SJC44) + 0.17*Ln(CRP+1) + 0.0072*PGA + 0.45
DAS = 0.54*sqrt(RAI) + 0.065*(SJC44) + 0.17*Ln(CRP+1) + 0.65

According to DAS, the following cut-points are used:
High disease activity: DAS > 3.7
Moderate disease activity: 3.7 ≥ DAS>2.4
Low disease activity: 2.4 ≥ DAS ≥ 1.6
In remission: DAS < 1.6

5.8.2 DAS28

The 28-joint Disease Activity Score (DAS28) includes TJC28, SJC28, ESR and PGA.
The DAS28 is calculated as follows:
DAS28 = 0.56*sqrt(TJC28) + 0.28*sqrt(SJC28) + 0.70*Ln(ESR) + 0.014*PGA
With missing values of ESR and/or PGA, the following formulas are used:
DAS28 = [0.56*sqrt(TJC28) + 0.28*sqrt(SJC28) + 0.70*Ln(ESR)]*1.08 + 0.016
DAS28 = 0.56*sqrt(TJC28) + 0.28*sqrt(SJC28) + 0.36*Ln(CRP+1) + 0.014*PGA + 0.96
DAS28 = [0.56*sqrt(TJC288) + 0.28*sqrt(SJC28) + 0.36*Ln(CRP+1)]*1.10 + 1.15

According to DAS28, the following cut-points are used:
High disease activity: DAS28 > 5.1
Moderate disease activity: 5.1 ≥ DAS28>3.2
Low disease activity: 3.2 ≥ DAS28 ≥ 2.6
In remission: DAS28 < 2.6

5.8.3 CDAI

The Clinical Disease Activity Index (CDAI) includes TCJ28, SJC28, PGA in addition to the treating Physician's Global Assessment of disease activity on a VAS 0-100 mm (PhGA).

The CDAI is calculated as follows:
CDAI=TCJ28 + SJC28 + PGA/10 + PhGA/10

According to CDAI, the following cut-points are used:
High disease activity: CDAI > 22.0
Moderate disease activity: 22.0 ≥ CDAI>10.0
Low disease activity: 10.0 ≥ CDAI > 2.8
In remission: CDAI ≤ 2.8

5.8.4 SDAI

The Simplified Disease Activity Index (SDAI) includes TCJ28, SJC28, PGA, PhGA and CRP.

The SDAI is calculated as follows:

SDAI=TCJ28 + SJC28 + PGA/10 + PhGA/10 + CRP/10

According to SDAI, the following cut-points are used:
High disease activity: SDAI> 26.0
Moderate disease activity: 26.0 ≥ SDAI>11.0
Low disease activity: 11.0 ≥ SDAI > 3.3
In remission: SDAI ≤ 3.3

5.9 EULAR response

The European League Against Rheumatism (EULAR) response rates will be calculated. A EULAR response is defined by the state and change in DAS and DAS28, and categorized into good, moderate and none using the following definitions:

Table 5.2 EULAR DAS response

	Change from baseline in DAS		
DAS at time-point	ΔDAS ≤ - 1.2	-1.2 < DAS < -0.6	DAS ≥ 0.6
DAS ≤ 2.4	Good	Moderate	None
2.4 < DAS ≤ 3.7	Moderate	Moderate	None
DAS > 3.7	Moderate	None	Noen

Table 5.3 EULAR DAS28 response

	Change from baseline in DAS28		
DAS28 at time-point	ΔDAS28 ≤ - 1.2	-1.2 < DAS28 < -0.6	DAS28 ≥ 0.6
DAS28 ≤ 3.2	Good	Moderate	None
3.2 < DAS28 ≤ 5.1	Moderate	Moderate	None
DAS28 > 5.1	Moderate	None	Noen

Time to EULAR response = Date of first visit with EULAR response – date of randomisation +1

5.10 Sharp van der Heijde Score

The Sharp van der Heijde Score (vdHSS) is a score of erosion and joint space narrowing (JSN) based on radiographs of hands and feet. The score for JSN ranges from 0 to 120 in the hands and from 0 to 48 in the feet. The total vdHSS score is the sum of scores of erosion and JSN, the maximum score being 448.

Radiographic progression is defined as a ΔvdHSS of ≥ 1 unit/year. I.e. a radiographic progression after one year (at visit 9) is defined as a ΔvdHSS of ≥ 1 unit, a radiographic progression after 2 years (visit 13) is defined as a ΔvdHSS of ≥ 2 units. Rate of progression is defined as ΔvdHSS divided by time (years). Progression rates will in addition be dichotomized according to ≤ 0.5, ≤ 2.0 and ≤ 5.0 units per year.

Table 5.4 Overview of vdHSS

Area	Joints	Erosion left	Erosion right	JSN left	JSN right
Hand	Metacarpophalangeal (MCP)				
	- First (MCP1)	0-5	0-5	0-4	0-4
	- Second (MCP2)	0-5	0-5	0-4	0-4
	- Third (MCP3)	0-5	0-5	0-4	0-4
	- Fourth (MCP4)	0-5	0-5	0-4	0-4
	- Fifth (MCP5)	0-5	0-5	0-4	0-4
	Proximal interphalangeal (IP/PIP)				
	- First (IP1)	0-5	0-5	NA	NA
	- Second (PIP2)	0-5	0-5	0-4	0-4
	- Third (PIP3)	0-5	0-5	0-4	0-4
	- Fourth (PIP4)	0-5	0-5	0-4	0-4
	- Fifth (PIP5)	0-5	0-5	0-4	0-4
	Carpometacarpal (CMC)				
	- Third (CMC3)	NA	NA	0-4	0-4
	- Fourth (CMC4)	NA	NA	0-4	0-4
	- Fifth (CMC5)	NA	NA	0-4	0-4
Wrist	First metacarpal base (MCB)	0-5	0-5	NA	NA

	Radius bone	0-5	0-5	NA	NA
	Ulna bone	0-5	0-5	NA	NA
	Trapezium/trapezoid (multangular)	0-5	0-5	NA	NA
	Navicula	0-5	0-5	NA	NA
	Lunate	0-5	0-5	0-4	0-4
	Multangular navivular-lunate	NA	NA	0-4	0-4
	radiocarpal	NA	NA	0-4	0-4
Foot	Metatarsophalangeal (MTP)				
	- First (MTP1)	0-10	0-10	0-4	0-4
	- Second (MTP2)	0-10	0-10	0-4	0-4
	- Third (MTP3)	0-10	0-10	0-4	0-4
	- Fourth (MTP4)	0-10	0-10	0-4	0-4
	- Fifth (MTP5)	0-10	0-10	0-4	0-4
	Interphalangeal (IP)	0-10	0-10	0-4	0-4

5.11 Ultrasonography score

The Ultrasonography (US) score will be based on US of 32 joints scored 0-3 for both grey scale (GS) synovitis and power Doppler signal. In addition, two joints and two tendons will be scored but excluded from the US score. The total GS and PD score will range from 0 to 96, while the total US score will range from 0 to 192.

Part of the score	Joints	GS left	GS right	PD left	PD right
Yes	Metacarpophalangeal (MCP)				
	- First (MCP1)	0-3	0-3	0-3	0-3
	- Second (MCP2)	0-3	0-3	0-3	0-3
	- Third (MCP3)	0-3	0-3	0-3	0-3
	- Fourth (MCP4)	0-3	0-3	0-3	0-3
	- Fifth (MCP5)	0-3	0-3	0-3	0-3
	Radio-carpal				
	Inter-carpal	0-3	0-3	0-3	0-3
	Radio-ulnar	0-3	0-3	0-3	0-3
	Elbow	0-3	0-3	0-3	0-3
	Knee	0-3	0-3	0-3	0-3
	Talocrural	0-3	0-3	0-3	0-3
	Metatarsophalangeal (MTP)				
	- First (MTP1)	0-3	0-3	0-3	0-3
	- Second (MTP2)	0-3	0-3	0-3	0-3
	- Third (MTP3)	0-3	0-3	0-3	0-3
	- Fourth (MTP4)	0-3	0-3	0-3	0-3
	- Fifth (MTP5)	0-3	0-3	0-3	0-3
	Interphalangeal (IP)	0-3	0-3	0-3	0-3
No	*Extensor carpi ulnaris*	*0-3*	*0-3*	*0-3*	*0-3*
	Tibialis posterior tendon	*0-3*	*0-3*	*0-3*	*0-3*
	PIP2	*0-3*	*0-3*	*0-3*	*0-3*
	PIP3	*0-3*	*0-3*	*0-3*	*0-3*

5.12 Magnetic resonance imaging

Magnetic resonance imaging of the dominant hand and wrist pre- and post-gadolinium will be undertaken at baseline, 3, 6, 12, 16 and 24 months. The same side will be assessed at all time-points.

Images will be read according to the OMERACT Rheumatoid Arthritis MRI Scoring (RAMRIS) criteria. The original RAMRIS consists of MRI definitions of important joint pathologies, a core set of MRI sequences and a semi-quantitative scoring system for erosions, bone marrow edema and synovitis. In addition, a semi-quantitative JSN score has been developed. Both flexor and extensor tenosynovitis will also be evaluated, according to the scoring system described by Haavardsholm et al, at the level between the radioulnar joint and the hook of the hamate, thus including both wrist and finger tendons.

The RAMRIS core set of MRI sequences to assess inflammatory as well as destructive changes in RA joints includes: Imaging in 2 planes, with T1-weighted images before and after intravenous gadolinium-contrast to assess synovitis and erosions; plus a T2-weighted fat saturated sequence or a STIR sequence to assess bone marrow edema.

5.12.1 RAMRIS erosion score

Erosion is defined as a sharply marginated bone lesion, with correct juxta-articular localization and typical signal characteristics that is visible in two planes with a cortical break seen in at least one plane. Each bone is scored on a scale of 0-10 based on the proportion of eroded bone compared to the "assessed bone volume" on all available images. For long bones, the "assessed bone volume" is from the articular surface (or its best estimated position, if absent) to a depth of 1 cm, and in carpal bones it is the whole bone.

- No erosion = 0
- 1-10% of bone eroded = 1
- 11-20% of bone eroded = 2
- 21-30% of bone eroded = 3
- 31-40% of bone eroded = 4
- 41-50% of bone eroded = 5
- 51-60% of bone eroded = 6
- 61-70% of bone eroded = 7
- 71-80% of bone eroded = 8
- 81-90% of bone eroded = 9
- 91-100% of bone eroded = 10 (*)
- Joint unable to be scored =U

* When scoring the wrist area, if the bones are fused, score erosions as 10.

The following bones are scored:

Wrist Bones

- Distal radius
- Distal ulna
- Scaphoid
- Lunate
- Triquetrum
- Pisiform
- Trapezium
- Trapezoid
- Capitate
- Hamate
- Proximal metacarpal 1
- Proximal metacarpal 2
- Proximal metacarpal 3
- Proximal metacarpal 4
- Proximal metacarpal 5

MCP Bones

- Proximal proximal phalanx 1
- Proximal proximal phalanx 2
- Proximal proximal phalanx 3
- Proximal proximal phalanx 4
- Proximal proximal phalanx 5
- Distal metacarpal 1
- Distal metacarpal 2
- Distal metacarpal 3
- Distal metacarpal 4

- Distal metacarpal 5

Without the first digit (which excludes trapezium and proximal metacarpal 1), the maximum score for the wrist is 130. The maximum score for the MCP bones is 80. With the first digit included (which includes the trapezium and proximal metacarpal 1), the maximum score for the wrist is 150. The maximum score for the MCP bones is 100. The definition of no progression of erosions from baseline at to end of study will be an average change in RAMRIS erosion score ≤0.

5.12.2 Bone Marrow Edema (BME, Osteitis)

Bone marrow edema is defined as a lesion, that may occur alone or surrounding an erosion or other bone abnormalities, within the trabecular bone, with ill-defined margins and signal characteristics consistent with increased water contents (high signal intensity on STIR coronal images and low signal intensity on T1 weighted images). Bone Edema is assessed in each bone of the Wrist and MCP bones. Each bone is scored on a scale of 0-3 based on the proportion of bone with edema.

- No edema = 0
- 1-33% of bone edematous = 1
- 34-66% of bone edematous = 2
- 67-100% of bone edematous = 3
- Joint unable to be scored = U

The following bones are scored:

Wrist

- Distal radius
- Distal ulna
- Scaphoid
- Lunate
- Triquetrum
- Pisiform
- Trapezium
- Trapezoid
- Capitate
- Hamate
- Proximal metacarpal 1
- Proximal metacarpal 2
- Proximal metacarpal 3
- Proximal metacarpal 4
- Proximal metacarpal 5

MCP bones

- Proximal proximal phalanx 1
- Proximal proximal phalanx 2
- Proximal proximal phalanx 3
- Proximal proximal phalanx 4
- Proximal proximal phalanx 5
- Distal metacarpal 1
- Distal metacarpal 2
- Distal metacarpal 3
- Distal metacarpal 4
- Distal metacarpal 5

Without the first digit (which excludes trapezium and proximal metacarpal 1), the maximum score for the wrist is 39. The maximum score for the MCP bones is 24. With the first digit is included (which includes trapezium and metacarpal base 1), the maximum score for the wrist is 45. The maximum score for the MCP bones is 30.

5.12.3 Synovitis

Synovitis is defined as an area in the synovial compartment that shows above normal post-gadolinium enhancement of a thickness greater than the width of the normal synovium. Enhancement is judged by comparison of T1 weighted images obtained before and after intravenous gadolinium contrast.
The synovitis scale is 0-3. A score of 0 is normal, and 1-3 (mild, moderate, severe) are by thirds of the presumed maximum volume of enhancing tissue in the synovial compartment.

- No synovitis = 0
- Mild synovitis = 1
- Moderate synovitis = 2
- Severe synovitis = 3
- Joint unable to be scored = U

Synovitis is scored 0-3 as described in the EULAR OMERACT RAMRIS reference atlas, with best possible match to the reference images".
The following joints and bones are scored:
Wrist

- Distal Radioulnar
- Radiocarpal
- Intercarpal-Carpometacarpal 2-5
- CMC-1

MCP joints

- MCP-1
- MCP-2
- MCP-3
- MCP-4
- MCP-5

Without the first digit, the maximum score for synovitis for the wrist is 9. The maximum score for the MCP joints is 12. With the first digit included, maximum score for synovitis for the wrist is 12. The maximum score for the MCP joints is 15.

5.12.4 Joint Space Narrowing (JSN, Cartilage loss)

JSN is defined as reduced joint space width compared to normal, as assessed in a slice perpendicular to the joint surface. The finding is scored on coronal images at the narrowest point. On coronal T1-weighted images score “white (bone marrow fat) to white”. If surface is eroded, ignore erosions and measure “width if there had been no erosion”. The Cartilage scale is 0-4, with 0 indicating normal to 4 indicating complete ankylosis.

- No Narrowing = 0
- Focal or mild (<33%) narrowing = 1
- Moderate (34% - 66%) narrowing = 2
- Moderate to severe (67% - 99%) narrowing = 3
- Ankylosis = 4
- Joint unable to be scored = U

The following joints are scored:
Wrist

- Radius-scaphoid
- Radius-lunate
- Scaphoid-Lunate
- Lunate-triquetrum
- Scaphoid-trapezium
- Scaphoid-trapezoid
- Capitate-scaphoid
- Capitate-lunate
- Hamate-triquetrum
- Trapezoid-trapezium

- Capitate-trapezoid
- Capitate-hamate
- Carpometacarpal 1
- Carpometacarpal 2
- Carpometacarpal 3
- Carpometacarpal 4
- Carpometacarpal 5

MCP joints

- Metacarpophalangeal 1
- Metacarpophalangeal 2
- Metacarpophalangeal 3
- Metacarpophalangeal 4
- Metacarpophalangeal 5

Without the first digit (excluding trapezium-metacarpal base 1), the maximum score for the wrist is 60. The maximum score for the MCP joints is 16. With the first digit included (including trapezium-metacarpal base 1), maximum score for the wrist is 68. The maximum score for the MCP joints is 20.

5.12.5 Tenosynovitis

Tenosynovitis on MRI is defined as tendon sheath fluid, sheath thickening and enhancement after
intravenous contrast injection. As small amounts of fluid can be seen in normal tendon sheets, it is essential that the tenosynovitis is visible in at least two consecutive axial slices within the tendon sheet to be scored as abnormal. Tendon sheath abnormalities are graded semi-quantitatively from grade
0 to grade 3, reflecting the maximum width (in mm) of enhancement within each anatomical area as described below:

- Grade 0 (normal): no peritendinous effusion or synovial proliferation with enhancement.
- Grade 1: ,2 mm peritendinous effusion and/or synovial proliferation with enhancement.
- Grade 2: >2 and ,5 mm peritendinous effusion and/or synovial proliferation with enhancement.
- Grade 3: >5 mm peritendinous effusion and/or synovial proliferation with enhancement.

The extent of the synovial enhancement is measured at the point of maximal thickness, perpendicular to the tendon surface.

Flexor and extensor tenosynovitis are evaluated semi-quantitatively in 10 different anatomical areas.
Dorsally:

- extensor pollicis brevis, abductor pollicis longus
- extensor carpi radialis brevis, extensor carpi radialis longus
- extensor pollicis longus
- extensor digitorum communis, extensor indicus proprius
- extensor digiti quinti proprius
- extensor carpi ulnaris.

On the volar side:

- the flexor carpi ulnaris tendon (located ulnar to the carpal tunnel)
- the flexor digitorum superficialis and profundus tendons (in the carpal tunnel, enclosed in a common sheath—the ulnar bursa)
- the flexor pollicis longus tendon (located dorsally and radially to the median nerve as it passes through the carpal tunnel, and entering a continuous sheath that becomes the radial bursa)
- the flexor carpi radialis (localised radially to the tendons enclosed in the ulnar bursa)

The maximum tenosynovitis score is 30.

5.12.6 Total MRI inflammation score
The total MRI inflammation score will be the sum of the synovitis score, the osteitis score and the tenosynovitis score.

5.13 Remission
Remission status is calculated at each visit. In addition to remission according to ACR, DAS, DAS28, CDAI and SDAI defined previously, the following remission criteria are defined:

5.13.1 FDA remission
Requires achieving ACR clinical remission at the current visit and no radiological progression (change in vdHSS =0) and no use of DMARDs at each visit within the previous 6 months.

5.13.2 FDA complete clinical response
Same as FDA remission, but with no requirement on the use of no DMARDs.

5.13.3 Complete DAS remission
Based on the FDA's definition of complete clinical response, a complete DAS remission will be defined as: DAS remission (i.e. DAS<1.6) and no swollen joints (i.e. SJC44 = 0) and no radiographic progression (i.e. change in vdHSS =0) for > 6 months. Complete DAS remission after 24 months is the primary endpoint of the ARCTIC study.

5.13.4 ACR/EULAR remission
The patient must satisfy all of the following in order to achieve ACR/EULAR remission:

- RAI ≤ 1
- SJC44 ≤ 1
- CRP ≤ 1
- PGA ≤ 1 (on a scale 0-10, in this study ≤ 14 on a scale 0-100)

5.14 SF-36
The SF-36 is a multi-purpose, short-form health survey with 36 questions. The SF-36 will be scored according to RAND 36-Item Health Survey 1.0 (http://www.rand.org/health/surveys_tools/mos/mos_core_36item_scoring.html) to form eight measures scores 0-100: physical functioning, bodily pain, role limitations due to physical health problems, role limitations due to personal or emotional problems, emotional well-being, social functioning, energy/fatigue, and general health perceptions. In addition, composite scores for physical and mental health summary measures are calculated according to the New England Medical Centre scoring instructions.(Ware, Kosinski, & Keller, 1994) The composite scores are computed according to the 1998 US general population means and standard deviations.

5.15 HAQ
The Stanford Health Assessment Questionnaire was introduced in the 1980s and is now widely used in evaluation of physical function in patients with RA. The disability index of this instrument includes questions concerning the ability of patients to perform 20 activities of daily living, and is most commonly referred to as the HAQ questionnaire, and sometimes as the HAQ disability index (HAQ-DI). A new version has recently been developed, the Patient-Reported Outcomes Measurement Information (PROMIS) HAQ, including a 20-item short form used in this study (adult physical function version 1.0 20-item PROMIS short form – 20a). While the original HAQ had 4 response categories, this new version includes a fifth response option.Each question has thus five response options, ranging in value from one to five. To find the total raw score, the sum of the values of the response to each question is calculated, giving a range in scores from 20 to 100 if all questions are answered. If at least 50% of the questions are answered, the form can be scored, according to the following formula:

$$\frac{\text{(Raw sum x number of items on the short form)}}{\text{Number of items that were actually answered}}$$

The total raw score should be translated into a T-score for each participant (either by standardized conversion tables or using item-level calibrations), which rescales the raw score into a standardized score with a mean of 50 and a standard deviation (SD) of 10. Therefore, a person with a T-score of 40 is one SD worse than average. The standardized T-score is reported as the final score for each participant.

5.16 RAID

The Rheumatoid Arthritis Index of Disease (RAID) is calculated based on seven numerical rating scales (NRS) questions. Each NRS is assessed as a number between 0 and 10. The seven NRS correspond to pain, function, fatigue, sleep, emotional wellbeing, physical wellbeing and coping/self-efficacy.

1. Calculation

RAID final value = (pain NRS value (range 0–10) × 0.21) + (function NRS value (range 0–10) × 0.16) + (fatigue NRS value (range 0–10) × 0.15) + (physical wellbeing NRS value (range 0–10) × 0.12) + (sleep NRS value (range 0–10) × 0.12) + (emotional wellbeing NRS value (range 0–10) × 0.12) + (coping NRS value (range 0–10) × 0.12).
Thus, the range of the final RAID value is 0–10 where higher figures indicate worse status.

2. Missing data imputation

If one of the seven NRS values composing the RAID is missing, the imputation is as follows:

a Calculate the mean value of the six other (non-missing) NRS (range 0–10)
b Impute this value for the missing NRS
c Then, calculate the RAID as explained above.

If two or more of the NRS are missing, the RAID is considered as missing value (no imputation).

5.17 EQ-5D

EQ-5D is a standardised instrument for use as a measure of health outcome. Applicable to a wide range of health conditions and treatments, it provides a simple descriptive profile and a single index value for health status. The EQ-5D index values are calculated according to the EQ-5D UK Time Trade-Off (TTO) value set.

5.18 WPAI

Worker productivity is generally subdivided into 2 components: absenteeism and presenteeism. The worker productivity in this study is based on the Work Productivity and Activity Impairment Questionnaire: Rheumatoid arthritis V2.0 (WPAI:RA).

The WPAI yields four types of scores:
1. Absenteeism (work time missed)
2. Presenteeism (impairment at work / reduced on-the-job effectiveness)
3. Work productivity loss (overall work impairment / absenteeism plus presenteeism)
4. Activity Impairment

The scores are based on the following questions:

Q1= currently employed
Q2 = hours missed due to specified problem
Q3 = hours missed other reasons
Q4 = hours actually worked
Q5 = degree problem affected productivity while working
Q6 = degree problem affected regular activities

Scores:
Multiply scores by 100 to express in percentages.

Percent work time missed due to RA (Absenteeism): $\frac{Q2}{Q2+Q4}$

Percent impairment while working due to RA (Presenteeism): $\frac{Q5}{10}$

Percent overall work impairment due to RA (Work productivity loss):

$$\frac{Q2}{(Q2+Q4)} + \left[1 - \frac{Q2}{Q2+Q4}\right] \cdot \frac{Q5}{10}$$

Percent activity impairment due to problem: $\frac{Q6}{10}$

5.19 Other calculations

Age (years) = [(date of baseline – date of birth)/365.25].

BMI = weight in kilograms / (height in metres) x (height in metres)
BMI will be categorized according to the WHO definitions for underweight, normal, overweight and obese.

Area under the curve (AUC) will be calculated as the integral under the measure curve using trapezoids

Time of withdrawal = date of withdrawal – date of randomization +1

5.20 Safety definitions

5.20.1 Treatment emerging adverse events

Treatment emerging adverse events (TEAEs) are defined as AEs with a start date on or after the randomization date.

5.20.2 Past disease and concomitant disease

Past disease/condition
A disease/condition is considered as past disease/condition if it is not ongoing at screening visit.

Concomitant disease
A disease/condition is considered as concomitant disease/condition if it is ongoing at screening visit.

Previous and Concomitant medications

- previous medication (start date < date of randomisation);
- concomitant medication (start date ≥ date of randomisation);

In case of missing or incomplete dates/times not directly allowing allocation to any of the two categories of medications, a worst-case allocation was performed according to the available parts of the start and the end dates. The medication was allocated to the first category allowed by the available data, according to the following order:

- concomitant medication;
- previous medication.

6. Efficacy Endpoints

6.1 Primary endpoint

The primary endpoint is the proportion of patients reaching “complete clinical DAS-remission” (i.e. DAS < 1.6, absence of swollen joints & no radiological progression during the last 8 months) at the end of the study (visit 13, 24 months).
Specifically, the primary endpoint will be the proportion of patients meeting the three following criteria:

- DAS score < 1.6 at visits 11, 12 and 13 (after 16, 20 and 24 months)
- SJC44 = 0 at visits 11, 12 and 13 (after 16, 20 and 24 months)
- Absolute change in vdHSS < 0.5 units between visit 11 (16 months) and visit 13 (24 months)

6.2 Secondary endpoints

6.2.1 Efficacy endpoints

Group	Endpoint	Assessment time	Type
Response	ACR20/50/70/90	All post-baseline visits	Dichotomous, Time to event
	FDA major clinical response	Visit 13 (24 months)	Dichotomous
	EULAR good response	All post-baseline visits	Dichotomous, Time to event
Remission	ACR remission	All post-baseline visits	Dichotomous,

			Time to event
	DAS remission	All post-baseline visits	Dichotomous, Time to event
	DAS28 remission	All post-baseline visits	Dichotomous, Time to event
	CDAI remission	All post-baseline visits	Dichotomous, Time to event
	SDAI remission	All post-baseline visits	Dichotomous, Time to event
	FDA remission	Visit 13 (24 months)	Dichotomous, Time to event
	FDA complete clinical remission	Visit 13 (24 months)	Dichotomous, Time to event
	ACR/EULAR remission	All post-baseline visits	Dichotomous, Time to event
Radiology	Radiographic progression	Visit 9 and 13 (after 12 and 24 months)	Dichotomous
	ΔvdHSS	Visits 4, 6, 9, 11, 13 (after 3, 6, 12, 16 and 24 months)	Continuous Δ
	ΔErosion score	Visits 4, 6, 9 11, 13 (after 3, 6, 12, 16 and 24 months)	Continuous Δ
	ΔJSN	Visits 4, 6, 9 11, 13 (after 3, 6, 12, 16 and 24 months)	Continuous Δ
	Yearly progression rate in vdHSS $\leq$ 0.5	Visit 9 and 13 (after 12 and 24 months)	Dichotomous
	Yearly progression rate in vdHSS $\leq$ 2.0	Visit 9 and 13 (after 12 and 24 months)	Dichotomous
	Yearly progression rate in vdHSS $\leq$ 5.0	Visit 9 and 13 (after 12 and 24 months)	Dichotomous
Disease Activity	ΔDAS	All post-baseline visits	Continuous Δ
	DAS AUC from baseline	Visit 9 and 13 (after 12 and 24 months)	Continuous AUC
	ΔDAS28	All post-baseline visits	Continuous Δ
	DAS28 AUC from baseline	Visit 9 and 13 (after 12 and 24 months)	Continuous AUC
	ΔSDAI	All post-baseline visits	Continuous Δ
	SDAI AUC from baseline	Visit 9 and 13 (after 12 and 24 months)	Continuous AUC
	ΔCDAI	All post-baseline visits	Continuous Δ
	CDAI AUC from baseline	Visit 9 and 13 (after 12 and 24 months)	Continuous AUC
ACR core set	ΔRAI	All post-baseline visits	Continuous Δ
	ΔSJC44	All post-baseline visits	Continuous Δ
	ΔPhGA	All post-baseline visits	Continuous Δ
	ΔPGA	All post-baseline visits	Continuous Δ
	ΔJointPain	All post-baseline visits	Continuous Δ
	ΔHAQ	V4, V6, V9, V11, V13	Continuous Δ
	ΔESR	All post-baseline visits	Continuous Δ
	ΔCRP	All post-baseline visits	Continuous Δ
RAMRIS	ΔSynovitis	Visits 4, 6, 9 11, 13 (after 3, 6, 12, 16 and 24 months)	Continuous Δ
	ΔErosions	Visits 4, 6, 9 11, 13 (after 3, 6, 12, 16 and 24 months)	Continuous Δ
	ΔBone marrow edema	Visits 4, 6, 9 11, 13 (after 3, 6, 12, 16 and 24 months)	Continuous Δ
	ΔTenosynovitis	Visits 4, 6, 9 11, 13 (after 3, 6, 12, 16 and 24 months)	Continuous Δ
	ΔMRI total inflammation score	Visits 4, 6, 9 11, 13 (after 3, 6, 12, 16 and 24 months)	Continuous Δ
	ΔJoint space narrowing	Visits 4, 6, 9 11, 13 (after 3, 6, 12, 16 and 24 months)	Continuous Δ

US scores	ΔGrey scale synovitis	Visit 9 and 13 (after 12 and 24 months)	Continuous Δ
	ΔPD synovitis	Visit 9 and 13 (after 12 and 24 months)	Continuous Δ
		Visit 9 and 13 (after 12 and 24 months)	Continuous Δ
Medication	On anti-TNF treatment	Visit 13 (24 months)	Dichotomous

6.2.2 Quality of life endpoints

Group	Endpoint	Assessment time	Type
SF-36	ΔPhysical functioning	V4, V6, V9, V11, V13	Continuous Δ
	ΔBodily pain	V4, V6, V9, V11, V13	Continuous Δ
	ΔRole limitations due to physical health problems	V4, V6, V9, V11, V13	Continuous Δ
	ΔRole limitations due to personal or emotional problems	V4, V6, V9, V11, V13	Continuous Δ
	ΔEmotional well-being	V4, V6, V9, V11, V13	Continuous Δ
	ΔSocial functioning	V4, V6, V9, V11, V13	Continuous Δ
	ΔEnergy/fatigue	V4, V6, V9, V11, V13	Continuous Δ
	ΔGeneral health perception	V4, V6, V9, V11, V13	Continuous Δ
	ΔPhysical health composite score	V4, V6, V9, V11, V13	Continuous Δ
	ΔMental health composite score	V4, V6, V9, V11, V13	Continuous Δ
RAID	ΔRAID total score	V4, V6, V9, V11, V13	Continuous Δ
EQ5D	ΔEQ5D index value	V4, V6, V9, V11, V13	Continuous Δ
WPAI	ΔAbsenteeism	V4, V6, V9, V11, V13	Continuous Δ
	ΔPresenteeism	V4, V6, V9, V11, V13	Continuous Δ
	ΔWork productivity loss	V4, V6, V9, V11, V13	Continuous Δ
	ΔActivity impairment	V4, V6, V9, V11, V13	Continuous Δ

7. SAFETY PARAMETERS

Measures of safety will include the following:

- Clinical and laboratory adverse events (AEs) and coding of AEs performed (using the [Medical Dictionary for Regulatory Activities] MedDRA, v.13.0).
- Clinical laboratory data
- Vital signs

8. STATISTICAL METHODOLOGY

8.1 Statistical and Analytical Issues

8.1.1 Statistical Methods

The primary efficacy analyses will be based on the FAS and PPS. Secondary efficacy analyses will be based on the FAS alone. As there is only one identified primary analysis, there will be no adjustments for multiple testing in the secondary analyses.

All categorical (binary and ordinal) data will be summarised using frequency counts and percentages of patient incidence. Percentages will be calculated using the study population (FAS); any exceptions to this will be highlighted in the table footnote. The continuous variables will be summarised using number of patients (N), mean, standard deviation (SD), median, 25/75 percentile and range (minimum/maximum). In general, minimum and maximum will be presented to the same degree of precision as data is recorded, with mean and median having 1 additional place after the decimal and standard deviation having 2 additional places after the decimal. Percentages less than 100 will be displayed to 1 place after the decimal, where space permits.

All efficacy analyses will be presented by the size (point estimate) of the difference between the treatments, the associated 95% confidence interval and the result of the hypothesis test by two-sided p-value.

All statistical analyses will be done in Stata v14 (StataCorp. 2015. *Stata Statistical Software: Release 14*. College Station, TX, USA) with the exception of the exponential time-response non-linear mixed model and time to event analyses which will be analysed using SAS v9.3 (SAS Institute Inc., Cary, NC, USA).

8.1.1.1 Analyses of dichotomous endpoints

Dichotomous efficacy endpoints (including the primary and secondary endpoints of section 6.2.1) at 12 (if available) and 24 months will be analysed using logistic regression adjusted for centre and anti-CCP status at baseline (stratification factors) providing estimates of adjusted risk difference (ARD) and adjusted relative risk (ARR) for the treatment effect. If any cell frequencies are less than 5, centre will be dropped as adjustment factor. If any cell frequencies are still less than 5, anti-CCP status at baseline will be dropped from the model, and unadjusted analyses will be performed. In any case, exact logistic regression adjusted for centre and anti-CCP at baseline will be included as sensitivity analyses, providing estimates of odds ratio for the treatment effect. The mid-p value will be presented. Analyses of efficacy over time for dichotomous endpoints will include linear mixed models with patient-specific random intercept and treatment, visit, treatment-visit interaction, centre and anti-CCP status as fixed factors. In addition we will fit an exponential time-response non-linear mixed model, adjusted for centre and anti-CCP status to assess the treatment effect over time (Reeve et al., 2013).

The linear mixed model is specified as

$$p_{ij} = \alpha + \beta_t X_t + \beta_j X_j + \beta_{tj} X_t X_j + \beta_c X_c + \beta_a X_a + A_i + \epsilon_{ij}, \qquad i \in [1,n], j \in [1,13]$$

where p_{ij} is the probability of positive outcome for patient i at visit j, α is the overall intercept, $\beta_t,\ \beta_j,\ \beta_{tj},\ \beta_c,\ \beta_a$ are the fixed regression coefficients for treatment (t), visit j, treatment-visit interaction, centre (c) and anti-CCP status (a), while $X_t, X_j, X_c,\ X_a$ are indicators for treatment (0=no ultrasonography, 1=ultrasonography), visit, centre and anti-CCP status (0=negative, 1=positive). $A_i \sim N(0, \sigma_A^2)$ is the patient-specific random intercept and $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ is the residual. Standard error will be estimated using robust methodology (sandwich estimator).

The exponential time-response non-linear mixed model is specified as

$$p_i(t) = (\alpha + \beta_t X_t + \beta_c X_c + \beta_a X_a + A_i)\left(1 - e^{-(\theta_0 + \theta_t X_t)t}\right) + \epsilon_{ij}, \qquad i \in [1,n], j \in [1,13]$$

where $p_i(t)$ is the probability of a positive outcome for patient i at time t, while θ_0, θ_t are the intercept and treatment coefficient for the time-response slope. The other parameters are defined as above. Estimates from the linear mixed models and exponential time-response

non-linear mixed models will be presented by time/visit-response plots of marginal means in addition to estimates of β_t and θ_t.

8.1.1.2 Analyses of change from baseline and AUC of continuous endpoints

Change from baseline and AUC of continuous endpoints (see section 6.2.1) at 12 and 24 months will be analysed using analysis of covariance (ANCOVA) models adjusted for centre, anti-CCP status at baseline as well as baseline value. Model assumptions (normal residuals) will be assessed using quantile-quantile plots of the residuals. The ANCOVA analysis will provide estimates of adjusted mean difference between treatments. Sensitivity analyses will be performed using bootstrapped median regression adjusted for the same independent variables as for the ANCOVA analysis. The median regression analysis will provide estimate of adjusted median difference between treatments, using 10 000 bootstrap replications. Seed will be set to the date of the database lock (9092015). Analyses of change from baseline over time for continuous endpoints will include linear mixed models with patient-specific random intercept and treatment, visit, treatment-visit interaction, centre and anti-CCP status as fixed factors. Time/visit-response plots of marginal means will present estimates from the linear mixed models.

8.1.1.3 Analyses of time to event endpoints

Time to event endpoints (see section 6.2.1) will be analysed using a Weibull regression model adjusted for centre and anti-CCP status at baseline. Weibull regression and not Cox regression will be used because of the interval-censored nature of the data. We will not know the exact time of the event, but the interval where the event occurred. Sensitivity analyses will be performed using Cox regression and Kaplan-Meier product-limit analysis. Estimates of the hazard ratio will be presented in addition to Kaplan-Meier plots.

8.2 Handling of Dropouts and Missing Data

In general missing values will not be imputed for descriptive statistics.

8.2.1 Primary endpoint

For the analysis of the primary endpoint, consisting of DAS and radiographic scores, missing data will be handled as follows:

Radiographic score:

- If the radiographic score is missing at month 24, the patient will be considered not in complete DAS remission (failure)
- If a radiographic score is missing for visit 11 (16 months), use last radiographic observation

DAS:

- If unable to calculate DAS at visit 13 (month 24) by any of the expressions in section 5.8.1, the patient will be considered not in complete DAS remission (failure)
-
- If unable to calculate DAS at visit 11 or 12 (month 16/20) by any of the expressions in section 5.8.1, use last DAS observation

SJC44:

- If SJC44 at visit 13 (month 24) is missing, the patient will be considered not in complete DAS remission (failure)
- If SJC44 is missing at visit 11 or 12 (month 16/20), use last SJC44 observation

8.2.2 ACR20, ACR50 and ACR70

Because the ACR20, ACR50, ACR70 variables are based on several component variables, it is possible that the values may still be calculated even if the component variables have some missing values. In this case, no imputation method is needed. If the ACR value is still missing, an imputation method will be applied. If the ACR value is missing due to missing values in any of the components, while the patient is still enrolled, the method of last observation carried forward (LOCF) will be used to carry forward any of the missing components, and from that mix of actual and carried-forward values, the values of ACR20, ACR50 and ACR70 will be determined.

After the LOCF imputation has been applied, missing values due to a patient dropping from the study for any reason (e.g. lack of efficacy or adverse event) will be handled by setting the

ACR value (ACR20, ACR50 and ACR70) to nonresponsive (that is, baseline observation carried forward, BOCF) from that visit onward.

8.2.3 Radiographic endpoints

For the analysis of radiographic scores, missing values will be imputed using the following rule:

- Imputation by linear interpolation will be used when observations exists both before and after the missing value
- Imputation by linear extrapolation using the last two know observations will be used when no later observation exists

Binary variables derived from the radiographic scores (e.g. progression or not progression) will be derived from the imputed data.

8.2.4 Other dichotomous endpoints

For the analyses of dichotomous endpoints ((including the primary and secondary endpoints of section 6.2.1) at 12 and 24 months not already mentioned, all missing observations will be imputed with worst outcome. In the analysis using all assessments, analysed using linear and non-linear mixed models based on observed case data, missing values will be assumed to be missing at random

8.2.5 Other continuous endpoints

For the analyses of change from baseline and AUC of continuous endpoints at 12 and 24 months using ANCOVA, missing data will be handled using multiple imputations. Sensitivity analyses will be analysed using complete cases. There will be no imputation in the linear mixed models, and missing values will be assumed missing at random.

8.2.6 Time to event endpoints

For the time to event analysis, all patients who withdraw from follow-up will be censored on the withdrawal date.

8.2.7 Pooling of Investigator Sites

There will be no pooling of investigator sites. Handling of centre-effects is described previously.

8.2.8 Determination of Sample Size

The sample size has been estimated on the basis of the primary variable proportion of patients reaching “complete clinical DAS-remission” analysed using a logistic regression model. The assumed rate of complete clinical DAS-remission is 45% in the non-US group and 65% in the US group, giving a treatment difference of 20%. See Table 1 for a rationale on these assumptions. On basis of the assumptions, a total sample size (study completers) of 198 (99 in each group) was deemed necessary to achieve 80% power to detect a difference of 20% between the groups in a two-sided test at 5% significance level. To compensate for withdrawals/loss to follow-up, a total enrolment of 240 patients was targeted.

Table 8.1 Rationale for sample size calculations

Study	Remission rate	
	Aggressive strategy	Routine strategy
FIN-RACo		
ACR-remission	14%	3%
DAS28-remission	51%	16%
Good treatment response	67%	27%
TICORA		
DAS28 remission	65%	16%
Good response	82%	44%
CAMERA		
Clinical remission	50%	37%
CIMESTRA		
DAS28 remission	43/34%	n/a

8.2.9 Timing of Main Analysis
The main analysis is planned when all patients have concluded 24 months of treatment, all data up to 24 months have been entered, verified and validated and the primary database has been locked. This will include the radiographic and US scores. Analysis on MRI data (RAMRIS-scores) is planned when all images are received, scored and entered into the MRI part of the database. This will be done after the main analysis.

8.3 Patient Characteristics

8.3.1 Patient Disposition
The disposition of all patients will be listed and summarised by treatment arm. The number and percentage of patients who are randomised, received any study treatment, prematurely discontinued from treatment and lost to follow-up will be summarised.

The number and percentage of patients will be categorized by the reason(s) for

1) End of study treatment/withdrawal from study treatment: This is when it is decided that a patient will not receive further treatment according to study protocol and enters the survival follow-up phase. Reasons can be: adverse event, patient withdrawal of consent, investigator decision, death, lost to follow-up, wrong diagnosis, major protocol deviation, unknown, other.

8.3.2 Protocol Deviations
Major protocol deviators will be determined and summarised by type (Major/Minor) and treatment group.

Major protocol deviations are:
- Informed consent not dated and signed
- Eligibility, according to Inclusion / Exclusion criteria, not met
- Serious adverse event (SAE) reporting requirements not met; seriousness criteria misinterpreted, timelines not respected
- Treatment regimen deviation: if the patient for some reason does not comply to the tight control management system defined in the protocol.
- Randomisation non-compliance; patient starts other treatment strategy than allocated

8.3.3 Background and Demographic Characteristics
Patient demographics and baseline characteristics will be summarised for the ITT population.

Patient demographics and baseline characteristics will be summarised by randomised treatment arm and overall using descriptive statistics (N, mean, standard deviation, median, 25/75 percentiles, minimum, and maximum) for continuous variables, and number and percentages of patients for categorical variables. The patient demographics and baseline characteristics to be summarised include age in years, gender, symptom duration, anti-CCP status, weight, height, BMI, education, work status, smoking status, consumption of coffee, consumption of alcohol, CRP, TJC, SJC, RAI, physicians global VAS, patient global VAS, patient VAS pain, DAS, vdHSS, erosion score $\geq$1, HAQ, US-score.

Demographics and baseline characteristics will also be summarised by randomisation strata (anti-CCP status).

Medical history will be coded using the MedDRA dictionary (v17.0E) and will be summarised. Concomitant medication will be coded using the ATC coding system and summarised.

8.3.4 Treatment Exposure
Treatment exposure will be summarised by the highest reached step in the treatment regimen according to Table 8.2 (refer to Table 2.1)

Table 8.2 Treatment regimen

Notation	Description
A1	Monotherapy + Prednisolone:

A2	Optimize monotherapy
B1	Triple combination therapy (or other combination therapy if MTX not tolerated):
B2	Optimize triple combination therapy:
C1	DMARD and 1st biologic
C2	DMARD and 1st biologic (adjust)
D1	DMARD and 2nd biologic
D2	DMARD and 2nd biologic (adjust)
E1	DMARD and 3rd biologic
E2	Optimize DMARD and 3rd biologic plus prednisolone:

8.3.5 Treatment Compliance
Data summarizing the proportions of patients complying with the treatment regimen according to protocol will be presented by treatment arm. Treatment arm difference will be tested using the chi-square test.

8.3.6 Concomitant Medications and Other Therapies
Concomitant medication information collected will be coded by the Anatomical Therapeutic Chemical (ATC) classification system. Concomitant medications taken during the study (including any prior medication that is either continuing at the time of first dose, or that finished within 2 weeks before the first dose) will be summarised by generic name. The number and percentage of patients who took at least one drug within each specific preferred term will be presented. Patients will only be counted once if they are taking more than one medication (within the same code) or take the same generic medication more than once. If it cannot be determined whether a medication is concomitant (based on stop date or, if the stop date is missing, start date), then the medication will be considered to be concomitant.

8.3.7 Quality of Life data
Analyses of Quality of Life (QoL) data will be done using the procedures described for change from baseline continuous endpoints (section 8.1.1.2).

8.3.8 Exploratory Analysis
Samples (including serum, plasma, full blood and urine) for biomarker or DNA/RNA discovery and validation have been collected and stored in a freezer at -70 C. These samples will be used for exploratory analyses, and may include measurement of cytokines and other known or potential new markers of inflammation or damage, such as interleukins, interferons, metalloproteases, transforming growth factor, TNFs, adhesion molecules etc. as well as DNA/RNA analyses (genomics and proteomics).

Exploratory endpoints will not be limited to those mentioned above, and will include variables/endpoints and statistical methods/modelling as necessary to explore the secondary objectives of the study as described in section 2.2 of the protocol.

9. Safety Analysis
General safety evaluations will be based on the incidence, intensity, and type of AEs, and clinically significant changes in the patient's physical examination findings, vital signs and clinical laboratory results. Safety variables will be tabulated and presented for all patients in the safety set.

9.1.1 Adverse Events
Adverse events will be coded using MedDRA, version 17.0E. The investigator records the maximum intensity of each AE using the levels mild, moderate and severe. Adverse events with missing intensity will be considered to be severe.

The number (%) of subjects with any adverse event, with 1, 2 or $\geq$ 3 adverse events, with treatment related adverse events, and with serious adverse events (SAE) will be summarised by treatment group. The number of events and number (%) of subjects with adverse events by system organ class (SOC) and preferred term (PT) will be summarised by treatment group, overall, for severe AEs and for AEs leading to study discontinuation. In addition, a summary table of AEs reported by >= 20% of all patients will be presented by SOC and PT. A detailed patient narrative will be given for any serious adverse event in the clinical study report in addition to listing.

9.1.2 Clinical Laboratory Parameters
Safety clinical laboratory parameters were collected and assessed, but only used to identify adverse events. No analyses of clinical laboratory parameters will be performed.

9.1.3 Vital Signs
Changes in vital signs (including systolic and diastolic blood pressure [mmHg], heart rate [beats per minute] and weight [kg]) will be summarised by assessment time and treatment arm.

9.1.4 Other Safety Analyses
Changes in BMD, T-score and Z-score for available DEXA measurements, as well as proportions of patients with osteopenia and osteoporosis, will be summarized by assessment time and treatment arm.

9.2 Interim Analyses
There are no planned interim analyses for efficacy.

9.2.1 Independent Data Monitoring Committee (IDMC)
No analyses have been made for the IDMC, and will not be reported in the clinical study report.

10. REFERENCES

Hammer, H. B., Bolton-King, P., Bakkeheim, V., Berg, T. H., Sundt, E., Kongtorp, A. K., & Haavardsholm, E. A. (2011). Examination of intra and interrater reliability with a new ultrasonographic reference atlas for scoring of synovitis in patients with rheumatoid arthritis. *Annals of the Rheumatic Diseases*, *70*(11), 1995–1998. http://doi.org/10.1136/ard.2011.152926

Reeve, R., Pang, L., Ferguson, B., O'Kelly, M., Berry, S., & Xiao, W. (2013). Rheumatoid Arthritis Disease Progression Modeling. *Therapeutic Innovation & Regulatory Science*, *47*(6), 641–650. http://doi.org/10.1177/2168479013499571

Ware, J. E., Kosinski, M., & Keller, S. D. (1994). SF-36 physical and mental health summary scales. Boston, MA.

11. DATA ANALYSES FOLDER PLAN
All programs and datasets will be organized according to the following plan.

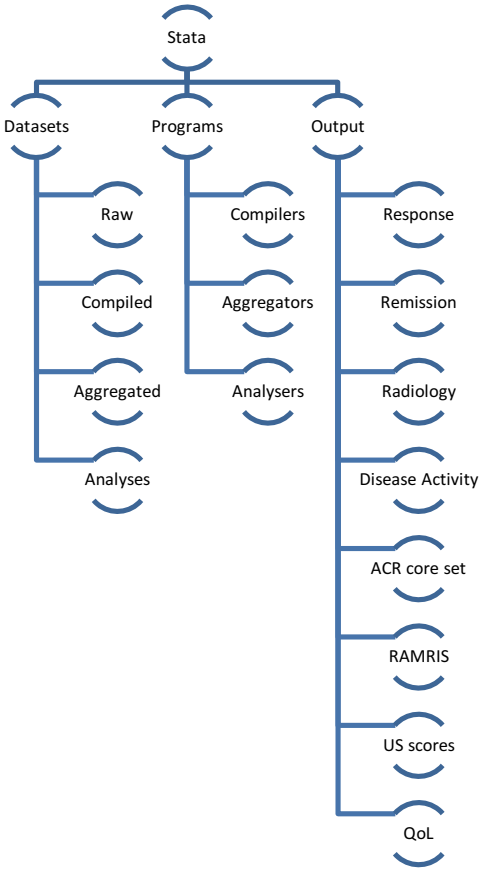


Figure 11.1 Data analysis folder plan

- Datasets
  - Raw; all raw datasets, exported from the study database in a flat file or excel format
  - Compiled; Compiled datasets using Compiler programs. All compiled datasets correspond to raw datasets
  - Aggregated; Aggregated datasets are combinations of compiled datasets, including calculations. The Disease Activity dataset is an example, where information is combined to compute the DAS and DAS28
  - Analyses; Analyses datasets are further compiled datasets made ready for analyses
- Programs
  - Compilers: Programs to import, format and prettify the raw datasets. Results in datasets stored in the Compiled Datasets folder
  - Aggregators: Programs to combine compiled datasets and make calculated variables. Results in datasets stored in the Aggregated and Analyses Datasets folders
  - Analyzers: Programs performing analyses and producing tables and figures from the Analyses Datasets. Results stored in Output folder.
- Output
  - Output from the Analyzers according to SAP.

12. LIST OF PLANNED TABLES, FIGURES AND LISTINGS

This section contains lists of all the summary tables, figures and patient data listings for this study.

12.1 Data Tables

Data tables will be configured according to publication requirements.

12.2 Data Listings

Data listings will be provided as needed.

12.3 Data Figures

Data figures will be configured according to publication requirements.