STATISTICAL ANALYSIS PLAN for the NORD-STAR trial

Administrative information:

<table>
<thead>
<tr>
<th>Sponsor name</th>
<th>Ronald van Vollenhoven, The Karolinska Institute</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sponsor address</td>
<td>Stockholm, Sweden</td>
</tr>
<tr>
<td>EudraCT number / REC no</td>
<td>2011-004720-35</td>
</tr>
<tr>
<td>Trial title</td>
<td>A multicenter, randomized, open-label, blinded-assessor, phase 4 study in patients with early rheumatoid arthritis to compare active conventional therapy versus three biologic treatments, and two de-escalation strategies in patients who respond to treatment.</td>
</tr>
<tr>
<td>Trial ID</td>
<td>NORD-STAR</td>
</tr>
<tr>
<td>Trial registration number</td>
<td>NCT01491815</td>
</tr>
</tbody>
</table>

SAP and protocol version:

<table>
<thead>
<tr>
<th>SAP version and date:</th>
<th>This SAP is version 1.0, dated 18 May, 2021</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protocol version</td>
<td>This document has been written based on information contained in the study protocol version 7.0 December 2013</td>
</tr>
</tbody>
</table>

SAP revision history:

<table>
<thead>
<tr>
<th>Protocol version</th>
<th>SAP version</th>
<th>Section number changed</th>
<th>Description and reason for change</th>
<th>Date changed</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.0</td>
<td>1.0</td>
<td>NA</td>
<td>Final version relating to the first study period from screening to week 24</td>
<td>02 Sep 2019</td>
</tr>
<tr>
<td>7.0</td>
<td>1.1</td>
<td>NA</td>
<td>First updated version relating to the data from screening to week 48</td>
<td>26 June 2020</td>
</tr>
</tbody>
</table>
## STATISTICAL ANALYSIS PLAN for the NORD-STAR trial

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>[No changes required]</th>
<th>[SAP reviewed against protocol amendments]</th>
<th>7 May 2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.0</td>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Valid from October 2018

Oslo University Hospital
STATISTICAL ANALYSIS PLAN for the NORD-STAR trial

SIGNATURE PAGE

SPONSOR/COORDINATING INVESTIGATOR:
Ronald van Vollenhoven, Professor,
The Karolinska Institute, Stockholm; Sweden

Signature

Date (dd/mmm/yyyy)

19 MAY 2021

TRIAL STATISTICIAN:
Inge Christoffer Olsen, PhD
Dept. of Research Support for Clinical Trials
Oslo University Hospital, Norway

Signature

Date (dd/mmm/yyyy)

18/May/2021

PRINCIPAL INVESTIGATOR
Mikkel Østergaard, Professor
Copenhagen Center for Arthritis Research,
Rigshospitalet, University of Copenhagen, Denmark

Signature

Date (dd/mmm/yyyy)

18/May/2021

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)
## STATISTICAL ANALYSIS PLAN for the NORD-STAR trial

### ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AE</td>
<td>Adverse Event</td>
</tr>
<tr>
<td>ATC</td>
<td>Anatomical/Therapeutic/Chemical</td>
</tr>
<tr>
<td>BMI</td>
<td>Body Mass Index</td>
</tr>
<tr>
<td>BPM</td>
<td>Beats Per Minute</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>DMC</td>
<td>Data Monitoring Committee</td>
</tr>
<tr>
<td>ECG</td>
<td>Electrocardiogram</td>
</tr>
<tr>
<td>eCRF</td>
<td>electronic Case Report Form</td>
</tr>
<tr>
<td>EDC</td>
<td>Electronic Data Capture</td>
</tr>
<tr>
<td>MedDRA</td>
<td>Medical Dictionary for Regulatory Activities</td>
</tr>
<tr>
<td>SAE</td>
<td>Serious Adverse Event</td>
</tr>
<tr>
<td>SAS</td>
<td>Statistical Analysis System</td>
</tr>
<tr>
<td>SD</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>SOC</td>
<td>System Organ Class</td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS

1 INTRODUCTION ........................................................................................................................ 6  
1.1 Background and rationale.................................................................................................... 6  
1.2 Trial Objectives .................................................................................................................... 6  
2 TRIAL METHODS ...................................................................................................................... 7  
2.1 Trial Design.......................................................................................................................... 7  
2.2 Randomisation ..................................................................................................................... 8  
2.3 Sample size .......................................................................................................................... 8  
2.4 Statistical Framework .......................................................................................................... 8  
2.5 Statistical Interim Analyses and Stopping Guidance ........................................................... 9  
2.6 Timing of Final Analysis....................................................................................................... 9  
2.7 Timing of Outcome Assessments ........................................................................................ 10  
3 STATISTICAL PRINCIPLES .................................................................................................... 11  
3.1 Confidence Intervals and p-values .................................................................................... 11  
3.2 Adherence and Protocol Deviations .................................................................................. 11  
3.3 Analysis Populations ......................................................................................................... 13  
4 TRIAL POPULATION ............................................................................................................... 13  
4.1 Screening Data, Eligibility and Recruitment ...................................................................... 13  
4.2 Withdrawal/Follow-up ........................................................................................................ 14  
4.3 Baseline Patient Characteristics ........................................................................................ 14  
5 ANALYSIS ................................................................................................................................ 15  
5.1 Outcome Definitions .......................................................................................................... 15  
5.2 Analysis Methods .............................................................................................................. 27  
6 SAFETY ANALYSES ............................................................................................................... 31  
6.1 Adverse Events ................................................................................................................... 31  
6.2 Clinical Laboratory Parameters ......................................................................................... 31  
6.3 Vital Signs ........................................................................................................................... 31  
7 STATISTICAL SOFTWARE ..................................................................................................... 31  
8 REFERENCES ......................................................................................................................... 32  
8.1 Literature References ......................................................................................................... 32  
8.2 Reference to Data Management Plan ................................................................................ 32  
8.3 Reference to the Trial Master File and Statistical Master File .......................................... 32  
8.4 Reference to other Standard Operating Procedures or Documents .................................... 32
1 Introduction

1.1 Background and rationale

This is the Statistical Analysis Plan (SAP) describing the statistical analyses of the NORD-STAR trial until 48 weeks of the first treatment period (TP1). This analysis plan should be seen as a part of the trial protocol, detailing all analyses to be undertaken for this time period. If there are inconsistencies between this SAP and the protocol, the SAP will be the governing document.

Further background and rationale is described below and in the latest version of the protocol, version 8.0 dated 10 February 2020.

Rheumatoid arthritis (RA) is a common inflammatory disorder with a reported prevalence of approximately 1% of the population. It is characterized by a progressive inflammatory synovitis with joint swelling and tenderness. Over time, structural joint damage evidenced by radiographic progression occurs and joint function diminishes. The most significant factor that has a great impact on prognosis of RA is pharmacologic intervention. A delay in starting a disease-modifying anti-rheumatic drug (DMARD) therapy has a significant negative impact. Patients treated early have a significant reduction of radiographic progression. Patients with more aggressive diseases seemed to benefit most from early DMARD initiation. MTX, given at 15 mg or more/weekly and folic acid; 5-15 mg/weekly, has remained the anchor drug for treatment of RA. In some countries initial combination therapy is preferred. Glucocorticoids are part of the early standard treatment in RA patients and widely used as add-on therapy to conventional DMARDs because of their effectiveness in controlling inflammation as well as reducing radiographic damages for a low cost. However, they also have adverse effects when used long-time, among others: negative effect on glucose and bone metabolism, increased risk for infection, osteoporosis, skin atrophy, osteonecrosis, cataract, hypertension, etc. Consequently, they are not recommended for long time use, but short time use of moderate-to-high dose glucocorticoids provides rapid initial control of active synovitis in early RA. Most rheumatologists recommend the use of biological DMARD after a trial of at least one or two conventional DMARDs in patients who continue having active synovitis. A recent recommendation suggests the use of biologics even in patients who have not yet failed non-biological DMARDs in case the patient has high disease activity and negative prognostic markers. Clinical data from the initial 24 weeks of the trial is covered in a previous SAP and has recently been analysed (Hetland et al, ACR abstract 2019).

1.2 Trial Objectives

1.2.1 Primary Objective

The objective of this part or the trial is to assess and compare radiographic and clinical outcomes after 48 weeks of active conventional therapy (ACT) versus each of three biological therapies (1) certolizumab-pegol, (2) abatacept and (3) tocilicumab.

The overall objective of the part of the NORDSTAR study covered by this SAP is, in RA patients treated with active conventional therapy (ACT) versus each of three different biologic therapies (certolizumab-pegol, abatacept and tocilicumab), to assess and compare the radiographic progression from baseline to the 48-week follow-up as well to assess and compare the proportion of subjects who achieve remission at the 48 week follow-up.

1.2.1 Secondary and Exploratory Objectives
STATISTICAL ANALYSIS PLAN for the NORD-STAR trial

Secondary and exploratory objectives are to assess radiographic progression and achievement of remission at other available time points, as well as differences in outcomes related to treatment response, disease activity, patient reported outcomes including quality of life, and safety ((serious) adverse events).

2 Trial Methods

2.1 Trial Design

This SAP describes the 48 week clinical and radiographic outcome in a 160 week, multicenter, randomized, open-label, blinded-assessor, double-treatment period study designed to compare the safety and efficacy of active conventional therapy (ACT; 1A: MTX+Prednisolone; 1B: MTX + sulphasalazine (SSZ) + hydroxochloroquine (HCQ) intraarticular glucocorticoids) and three biologic treatments (MTX+certolizumab; MTX+abatacept; MTX+tocilizumab) in subjects with early RA. (Figure 1).

After 8th of May 2014 tocilizumab was no longer available in Finland, and patients randomised to tocilizumab were treated with ACT.
2.2 Randomisation

All screening laboratory results were reviewed by the investigator prior to randomization. Subjects who met the selection criteria proceeded to randomization.

Randomization has been done through the trial center at the Karolinska Institute. There, at the outset of the study, randomization lists were generated with separate lists based on the three protocol-specified stratifications: by country, sex, and ACPA positivity. Thus, for each of the participating countries there were four randomization lists, one each for ACPA positive female, ACPA positive male, ACPA negative female and ACPA negative male participants. Each randomization list was generated by running an open-access internet-based random number generator set for four levels (1-2-3-4) in equal proportions (blocks of four). For practical reasons lists were generated of length 200 for Sweden and 100 for all other countries. Participating sites were provided with a dedicated telephone number in Stockholm, manned by site personnel at all business hours. After a patient had consented to participation and fulfilled inclusion criteria, and without exclusions, personnel at the local site dialed the randomization phone line, and informed trial center personnel of country, sex and ACPA status of the patient. Trial center personnel read off from the top of the appropriate list to what arm the patient was randomized, and noted the patient trial number (site number and patient number) on that list. They also sent a confirmatory e-mail to the site.

2.3 Sample size

The sample size calculation as described in the protocol was done for the whole trial.

Under the assumption that the true remission rates in the treatment arms are 0.12 in the ACT arm, 0.22 in the certulizumab-pegol arm, 0.22 in the abatacept arm and 0.26 in the tocilizumab, 724 patients must be randomised to reach 85% power to reject the null hypothesis that the remission rates are equal in the four treatment arms using the standard chi-square test. With 90% power, 832 patients have to be randomised.

2.4 Statistical Framework

2.4.1 Hypothesis Tests

This part of the trial is designed to establish the superiority of at least one of the biologic treatments compared to active conventional treatment on (1) avoiding progression in the radiographic Sharp-van der Heijde Score from randomisation to 48 weeks and (2) achieving CDAI remission at 48 weeks. Thus, there are six separate null hypotheses to be tested in this trial.

\( H_0^{RX,C} \): The progression in the radiographic Sharp van der Heijde score from randomisation to 48 weeks is equal when receiving certulizumab-pegol compared to active conventional therapy

\( H_0^{RX,A} \): The progression in the radiographic Sharp van der Heijde score from randomisation to 48 weeks is equal when receiving abatacept compared to active conventional therapy
STATISTICAL ANALYSIS PLAN for the NORD-STAR trial

$H_{0}^{x vdH}$: The progression in the radiographic Sharp van der Heijde score from randomisation to 48 weeks is equal when receiving tocilizumab compared to active conventional therapy.

$H_{0}^{CDAI}$: The probability of achieving CDAI remission at 48 weeks is equal when receiving certulizumab-pegol compared to active conventional therapy.

$H_{0}^{CDAI a}$: The probability of achieving CDAI remission at 48 weeks is equal when receiving abatacept compared to active conventional therapy.

$H_{0}^{CDAI t}$: The probability of achieving CDAI remission at 48 weeks is equal when receiving tocilizumab compared to active conventional therapy.

These hypotheses will be tested adjusting for multiplicity by testing the null hypotheses regarding CDAI remission and radiographic outcome separately and adjusting for multiplicity by Dunnett’s procedure within each outcome family.

There will be no other hypotheses tested, and all other efficacy and safety analyses will be regarded as supportive or exploratory.

2.4.2 Decision Rule

Each of the two outcome families (CDAI remission and Sharp-van der Heijde) will be tested to an overall significance level of 0.025. If any of the six hypotheses are rejected on the 0.025 level using adjusted p-values according to Dunnet’s method when you have a common comparator (according to expression (5) in Hothorn, Bretz and Westfall (2008)), a difference is claimed either in favour of active conventional therapy or the corresponding biological therapy dependent on the direction of the contrast measure.

2.5 Statistical Interim Analyses and Stopping Guidance

There have been no interim analyses in this trial.

2.6 Timing of Final Analysis

The main analysis for this part of the trial is planned when all patients have concluded 48 ± 1 week of treatment, all data up to 48 weeks have been entered, verified and validated and the primary database has been locked.

The last patient entered the study in December 2018 and the last patient’s week 48 visit (LPLV) was performed in December 2019. A soft lock of the database for the first 48 weeks of treatment is planned to be performed in Q2 20201. The hard lock target date is Q4 2020.

Prior to DBL a final determination is made that the data have achieved minimal quality standards. Each country develops a procedure for the national level of data cleansing and data management. It has been established that Study sites have been regularly monitored and that the three Data Centers have well-established internal routines for QC.
STATISTICAL ANALYSIS PLAN for the NORD-STAR trial

After solving of queries the national databases are exported to Zitelab for additional QC in accordance with a written procedure. Test analysis will be performed by the statistician with random allocation of patients to the different treatment arms to detect any outliers and errors in the database. Continuous variables will be visualized in histograms for the same purpose. Mock datasets (i.e. all variables, soft-locked) will be sent to the statistician to allow him to initiate the writing of code.

This SAP has been reviewed by an independent statistician for quality control before final approval. Unresolved issues regarding the softlocked dataset will be discussed and decided upon by the Steering Committee. Thereafter, the final, locked and merged dataset will be delivered to the statistician from Zitelab.

2.7 Timing of Outcome Assessments

For all clinically planned measures, visits should occur within a window of the scheduled visit. Visits outside visit window is regarded a protocol deviation. The target day and visits window is defined in the protocol as:

<table>
<thead>
<tr>
<th>Visit Label</th>
<th>Target Day</th>
<th>Definition (Day window)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screening</td>
<td>-1</td>
<td>-28 to -1 day</td>
</tr>
<tr>
<td>V1. Baseline</td>
<td>Day 1 (Randomization)</td>
<td>Day 1</td>
</tr>
<tr>
<td>V2 Week 2</td>
<td>15</td>
<td>Target day ± 3 days</td>
</tr>
<tr>
<td>V3 Week 4</td>
<td>29</td>
<td>Target day ± 3 days</td>
</tr>
<tr>
<td>V4 Week 6</td>
<td>43</td>
<td>Target day ± 3 days</td>
</tr>
<tr>
<td>V5 Week 8</td>
<td>57</td>
<td>Target day ± 3 days</td>
</tr>
<tr>
<td>V6 Week 12</td>
<td>85</td>
<td>Target day ± 14 days</td>
</tr>
<tr>
<td>V7 Week 16</td>
<td>113</td>
<td>Target day ± 14 days</td>
</tr>
<tr>
<td>V8 Week 20</td>
<td>141</td>
<td>Target day ± 14 days</td>
</tr>
<tr>
<td>V9 Week 24</td>
<td>169</td>
<td>Target day ± 7 days</td>
</tr>
<tr>
<td>V10 Week 32</td>
<td>225</td>
<td>Target day ± 30 days</td>
</tr>
<tr>
<td>V11 Week 40</td>
<td>281</td>
<td>Target day ± 30 days</td>
</tr>
<tr>
<td>V12 Last study visit* Week 48</td>
<td>337</td>
<td>Target day ± 7 days</td>
</tr>
</tbody>
</table>

*The last study visit is defined as the visit following the last visit with randomised treatment, and where there is a study end statement.
STATISTICAL ANALYSIS PLAN for the NORD-STAR trial

For analysis and tabulation purposes, we define study time points as

<table>
<thead>
<tr>
<th>Time Point Label</th>
<th>Target Day</th>
<th>Definition (Day window)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP1. Baseline</td>
<td>Day 1</td>
<td>Information up to randomisation</td>
</tr>
<tr>
<td>TP2 Week 2</td>
<td>15</td>
<td>Day 2 to 21</td>
</tr>
<tr>
<td>TP3 Week 4</td>
<td>29</td>
<td>Day 22 to 35</td>
</tr>
<tr>
<td>TP4 Week 6</td>
<td>43</td>
<td>Day 36 to 49</td>
</tr>
<tr>
<td>TP5 Week 8</td>
<td>57</td>
<td>50 to 70</td>
</tr>
<tr>
<td>TP6 Week 12</td>
<td>85</td>
<td>71 to 98</td>
</tr>
<tr>
<td>TP7 Week 16</td>
<td>113</td>
<td>99 to 126</td>
</tr>
<tr>
<td>TP8 Week 20</td>
<td>141</td>
<td>127 to 154</td>
</tr>
<tr>
<td>TP9 Week 24</td>
<td>169</td>
<td>155 to 196</td>
</tr>
<tr>
<td>TP10 Week 32</td>
<td>225</td>
<td>197 to 252</td>
</tr>
<tr>
<td>TP11 Week 40</td>
<td>281</td>
<td>253 to 308</td>
</tr>
<tr>
<td>TP12 Last study visit*</td>
<td>337</td>
<td>309 to 364</td>
</tr>
</tbody>
</table>

If more than one visit fall into the same time point interval, information on all visits will be used in the analyses.

3 Statistical Principles

3.1 Confidence Intervals and p-values

All calculated p-values will be two-sided and compared to a 5% family wise error rate. All efficacy estimates will be presented with two-sided confidence intervals. When the efficacy estimates relates to one of the pre-specified null hypotheses as defined in section 2.4.1, adjusted confidence intervals will be used when applicable. Otherwise, unadjusted 95% confidence intervals will be presented.

3.2 Adherence and Protocol Deviations

3.2.1 Adherence to Allocated Treatment

3.2.2 Protocol Deviations
STATISTICAL ANALYSIS PLAN for the NORD-STAR trial

The following are pre-defined major protocol deviations regarded to affect the efficacy of the intervention:

1. Patient withdrew from study (Early termination visit before TP1-w48)
2. Randomized study medication (RSM) was permanently stopped before TP1-w48
3. RSM was interrupted for 12 weeks or more between TP1-Day1 and TP1-w24 or between TP1-w24 and TP1-w48
4. DMARD therapy was added – except SSZ/HCQ added in arm 1 (Sweden, Norway, The Netherlands or Iceland).
5. For Sweden, Norway, The Netherlands and Iceland:
   5.1: Prednisolone (or equivalent) >10 mg/day and/or greater than baseline dose is given for >8 weeks
   5.2: Prednisolone (or equivalent) > 10 mg/day and/or greater than baseline dose is given from week 20 to 24 or from week 44 to 48
   5.3: Prednisolone (or equivalent) > 10 mg/day and/or greater than baseline dose is given on any day during the week prior to the week 24 visit or on any day during the week prior to the week 48 visit
6. For Denmark and Finland:
   6.1: Any oral prednisolone
7. For all countries:
   7.1: IA injection of triamcinolonehexacetonid (Lederspan) or equivalent >4 mL or > 4 joints at one visit after week 4
   7.2: IA injection of triamcinolonehexacetonid (Lederspan) or equivalent >2 mL from week 20 to 24 or from week 44 to 48
   7.3: Any IA injection of triamcinolonehexacetonid (Lederspan) or equivalent during the week prior to the week 24 visit or during the week prior to the week 48 visit.

The following deviations are defined as minor; deviations that exceed those mentioned here are regarded as major protocol deviations:

1. Methotrexate: up to four missed doses between TP1-Day1 and TP1-w24 and up to four missed doses between TP1-w24 and TP1-w48. Deviations from the scheduled dosing regimen are allowed when medically indicated
2. Certolizumab-pegol: up to two missed doses between TP1-Day1 and TP1-w24 and up to two missed doses between TP1-w24 and TP1-w48. A single mistakenly administered extra dose is allowed, but NOT later than the week 20 visit or 4 weeks prior to the week 48 visit;
3. Abatacept SC: up to four missed doses between TP1-Day1 and TP1-w24 and up to four missed doses between TP1-w24 and TP1-w48. A single mistakenly administered extra dose (but NOT later than the week 20 visit or 4 weeks prior to the week 48 visit)
4. Abatacept IV: up to one missed dose between TP1-Day1 and TP1-w24 and up to one missed dose between TP1-w24 and TP1-w48
5. Tocilizumab SC: up to four missed doses between TP1-Day1 and TP1-w24 and up to four missed doses between TP1-w24 and TP1-w48. A single mistakenly administered extra dose (but NOT later than the week 20 visit or 4 weeks prior to the week 48 visit)
STATISTICAL ANALYSIS PLAN for the NORD-STAR trial

6. Tocilizumab IV: up to one missed dose between TP1-Day1 and TP1-w24 and up to one missed dose between TP1-w24 and TP1-w482.

The number (and percentage) of patients with major and minor protocol deviations will be summarised by treatment group with details of type of deviation provided. The patients that are included in the ITT analysis data set will be used as the denominator to calculate the percentages. No formal statistical testing will be undertaken.

3.3 Analysis Populations

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

The Intention to Treat (ITT) population will be defined as all patients randomly assigned to a treatment group except 17 patients from Finland for whom the allocated treatment (tocilizumab) was not available.

The as-treated population will be defined as all randomized patients who received at least one dose of randomized study medication (i.e, at least one dose of MTX in combination with at least one dose of prednisolone or at least one intra-articular injections of corticosteroid (arm 1A and 1B, respectively)), at least one dose of certolizumab-pegol (arm 2), at least one dose of abatacept (arm 3), or at least one dose of tocilizumab (arm 4). Thus, patients who received only MTX (and no glucocorticoid) are not included in the as-treated population).

The Safety Analysis Set will be identical to the as-treated population.

The Per Protocol Analysis Set (PPS) will include all randomised patients meeting the study eligibility criteria and with no major protocol deviations affecting the treatment efficacy as defined in section 3.2.2.

The ITT population will be regarded as the primary analysis population. All results based on the as-treated and PP populations are considered sensitivity or robustness analyses.

4 Trial Population

4.1 Screening Data, Eligibility and Recruitment

The total number of screened patients and reasons for not entering the trial will be summarised and tabulated.

A CONSORT flow diagram (appendix A) will be used to summarise the number of patients who were:

- assessed for eligibility at screening
- eligible at screening
- ineligible at screening*
- eligible and randomised
- eligible but not randomised*

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance Supplemental material placed on this supplemental material which has been supplied by the author(s) Ann Rheum Dis

doi: 10.1136/ard-2023-224116
**STATISTICAL ANALYSIS PLAN for the NORD-STAR trial**

- received the randomised allocation
- did not receive the randomised allocation*
- lost to follow-up*
- discontinued the intervention*
- randomised and included in the primary analysis
- randomised and excluded from the primary analysis*

*reasons will be provided.

### 4.2 Withdrawal/Follow-up

The status of eligible and randomised patients at trial end will be tabulated by treatment group according to

- completed intervention and assessments
- completed assessments but not intervention
- withdrew consent
- lost to follow-up

Time from randomisation to treatment discontinuation and time from randomisation to withdrawal/lost to follow-up will be presented graphically using the Kaplan-Meier estimator.

### 4.3 Baseline Patient Characteristics

The patient demographics and baseline characteristics to be summarised include age in years, gender, Body Mass Index (BMI), smoking status, symptom duration, disease duration, RF/ACPA status, Tender Joint Count (TJC), Swollen Joint Count, Clinical Disease Activity Index (CDAI), Disease activity Score in 28 joints (DAS28), Patient’s Global Assessment of disease activity (PGA), Physician’s Global assessment of Disease Activity (PhGA), pain on a Visual Analogue Scale (VAS), questionnaires and radiographic damage score.

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. There will be no statistical analysis of treatment difference. Any clinically important imbalance between the treatment groups will be noted.
STATISTICAL ANALYSIS PLAN for the NORD-STAR trial

5 Analysis

5.1 Outcome Definitions

5.1.1 General Definitions and Derived Variables

5.1.1.1 Body Mass Index
Body Mass Index (BMI) = Body weight in kilograms divided by the square of the height in meters.

5.1.1.2 Change from baseline
Change from baseline (Δ) = time-point value - baseline value.

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

5.1.1.3 Joint Counts
The Tender Joint Count of 68 joints (TJC68) is the number of joints with pain in 68 joints (temporomandibular (TMJs), sternoclavicular (SCs), acromioclavicular (ACs), shoulder, elbows, wrists, metacarpophalangeal (MCPs), finger proximal interphalangeal (finger PIPs), distal interphalangeal (DIPs), hips, knees, ankles, tarsus, metatarsophalangeal (MTPs) and toe proximal interphalangeal (toe PIPs)).

The Swollen Joint Count of 66 joints (SJC66) is the number of joints with swelling in 66 joints. The joints are the same as for TJC68 except hips.

The Tender and Swollen Joint Counts of 28 joints (TJC28/SJC28) are based on the following joints: shoulders, elbows, wrists, MCP, finger PIPs and knees.

5.1.1.4 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 erosion ranges from 0 to 160 in the hands and from 0 to 120 in the feet (erosion total 280). The score for JSN ranges from 0 to 120 in the hands and from 0 to 48 in the feet (JSN total 168). The total vdBSS score is the sum of scores of erosion and JSN, the maximum score is 448.

Each radiograph has been read twice by two independent readers. Consensus reads was elicited by a difference in change in Total Sharp van der Heijde score (Total SvdH score) from 0-48 weeks >= 2 between readers. Cases selected for consensus reads was read by the two readers together, and a final assessment was made (joint by joint level).

The final joint score used for analysis will be the consensus read if available or the mean of the two joint scores if no consensus read is needed.

Table 5.1 Overview of vdBSS

<table>
<thead>
<tr>
<th>Area</th>
<th>Joints</th>
<th>Erosion left</th>
<th>Erosion right</th>
<th>JSN left</th>
<th>JSN right</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Valid from October 2018

Oslo University Hospital
## STATISTICAL ANALYSIS PLAN for the NORD-STAR trial

<table>
<thead>
<tr>
<th></th>
<th>Metacarpophalangeal (MCP)</th>
<th>Proximal interphalangeal (IP/PIP)</th>
<th>Carpometacarpal (CMC)</th>
<th>Wrist</th>
<th>Foot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hand</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- First (MCP1)</td>
<td>0-5</td>
<td>0-5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Second (MCP2)</td>
<td>0-5</td>
<td>0-5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Third (MCP3)</td>
<td>0-5</td>
<td>0-5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Fourth (MCP4)</td>
<td>0-5</td>
<td>0-5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Fifth (MCP5)</td>
<td>0-5</td>
<td>0-5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- First (IP1)</td>
<td>0-5</td>
<td>0-5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Second (PIP2)</td>
<td>0-5</td>
<td>0-5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Third (PIP3)</td>
<td>0-5</td>
<td>0-5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Fourth (PIP4)</td>
<td>0-5</td>
<td>0-5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Fifth (PIP5)</td>
<td>0-5</td>
<td>0-5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Third (CMC3)</td>
<td>NA</td>
<td>NA</td>
<td>0-4</td>
<td>0-4</td>
<td></td>
</tr>
<tr>
<td>- Fourth (CMC4)</td>
<td>NA</td>
<td>NA</td>
<td>0-4</td>
<td>0-4</td>
<td></td>
</tr>
<tr>
<td>- Fifth (CMC5)</td>
<td>NA</td>
<td>NA</td>
<td>0-4</td>
<td>0-4</td>
<td></td>
</tr>
<tr>
<td>- First metacarpal base (MCB)</td>
<td>0-5</td>
<td>0-5</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Radius bone</td>
<td>0-5</td>
<td>0-5</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Ulna bone</td>
<td>0-5</td>
<td>0-5</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Trapezium/trapezoid (multangular)</td>
<td>0-5</td>
<td>0-5</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Navicula</td>
<td>0-5</td>
<td>0-5</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Capitatum naviculare</td>
<td>0-5</td>
<td>0-5</td>
<td>0-4</td>
<td>0-4</td>
<td></td>
</tr>
<tr>
<td>Multangular navicular</td>
<td>NA</td>
<td>NA</td>
<td>0-4</td>
<td>0-4</td>
<td></td>
</tr>
<tr>
<td>Radiocarpal</td>
<td>NA</td>
<td>NA</td>
<td>0-4</td>
<td>0-4</td>
<td></td>
</tr>
</tbody>
</table>

### 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
- Capitatum naviculare 0-5 0-5 0-4 0-4
- Multangular navicular NA NA 0-4 0-4
- Radiocarpal NA NA 0-4 0-4

### Foot
- Metatarsophalangeal (MTP) 0-10 0-10 0-4 0-4
STATISTICAL ANALYSIS PLAN for the NORD-STAR trial

| - 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 |

Rules for handling of missing data / data imputation

Hierarchical rules for imputation:

1. If some or all data from one of three timepoints are missing: Impute by Linear interpolation/extrapolation on the lowest possible level (i.e. per joint)
2. If some or all data from two of three timepoints are missing only one time point is available: Use last observation carried forwards (LOCF) or First observation carried backwards (FOCB)
3. If data from some joints within a hand or foot are missing from all time points: Use the mean of neighboring joints.
4. If data from one entire foot or one entire hand are missing at all time points: Use the score from the contralateral foot or hand for the missing foot/hand.
5. If data from two feet are missing at all timepoints: Use scores of the two hands and normalize to the total score range for both hands and feet of the Total SvdH score (0-448)/JSN score (0-168) and erosion score (0-280). (Range erosion score both hands 160, range JSN score two hands 120).
6. If data from two hands are missing at all timepoints: Use scores from the two feet and normalize to the total score range for both hands and feet of the Total SvdH score (448)/JSN score (168) and erosion score (280)

5.1.1.5 CDAI

The Clinical Disease Activity Index (CDAI) includes TCJ28, SJC28, Patient’s Global Assessment of disease activity on a VAS 0-100 mm (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:

\[
\text{CDAI} = \text{TCJ28} + \text{SJC28} + \frac{\text{PGA}}{10} + \frac{\text{PhGA}}{10}
\]

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
STATISTICAL ANALYSIS PLAN for the NORD-STAR trial

5.1.1.6 SDAI
The Simplified Disease Activity Index (SDAI) includes TCJ28, SJC28, PGA, PhGA and C-reactive protein (CRP) in mg/L.

The SDAI is calculated as follows:

\[ \text{SDAI} = \text{TCJ28} + \text{SJC28} + \frac{\text{PGA}}{10} + \frac{\text{PhGA}}{10} + \frac{\text{CRP}}{10} \]

According to SDAI, the following cut-points are used:

- High disease activity: \( \text{SDAI} > 26.0 \)
- Moderate disease activity: \( 26.0 \geq \text{SDAI} > 11.0 \)
- Low disease activity: \( 11.0 \geq \text{SDAI} > 3.3 \)
- In remission: \( \text{SDAI} \leq 3.3 \)

5.1.1.7 DAS28-CRP
The 28-joint Disease Activity Score (DAS28) includes TJC28, SJC28, CRP and PGA.

The DAS28-CRP is calculated as follows:

\[ \text{DAS28} = 0.56\times\sqrt{\text{TJC28}} + 0.28\times\sqrt{\text{SJC28}} + 0.36\times\ln(\text{CRP}+1) + 0.014\times\text{PGA} + 0.96 \]

If any of the components are missing, then the DAS28 is missing.

According to DAS28, the following cut-points are used:

- High disease activity: \( \text{DAS28} > 5.1 \)
- Moderate disease activity: \( 5.1 \geq \text{DAS28} > 3.2 \)
- Low disease activity: \( 3.2 \geq \text{DAS28} > 2.6 \)
- In remission: \( \text{DAS28} < 2.6 \)

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

- \( \text{TJC68} \leq 1 \)
- \( \text{SJC66} \leq 1 \)
- \( \text{CRP} \leq 1 \)
5.1.1.9 HAQ-DI

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).

According to Maska, L., Anderson, J. and Michaud, K. (2011), Measures of functional status and quality of life in rheumatoid arthritis: Health Assessment Questionnaire Disability Index (HAQ), Modified Health Assessment Questionnaire (MHAQ), Multidimensional Health Assessment Questionnaire (MDHAQ), Health Assessment Questionnaire II (HAQ-II), Improved Health Assessment Questionnaire (Improved HAQ), and Rheumatoid Arthritis Quality of Life (RAQoL). Arthritis Care Res, 63: S4-S13. doi:10.1002/acr.20620:

“Eight categories, reviewing a total of 20 specific functions evaluate patient difficulty with activities of daily living over the past week. Categories include dressing and grooming, arising, eating, walking, hygiene, reaching, gripping, and errands and chores. Also identified are specific aids or devices utilized for assistance, as well as help needed from another person (aids/help).

There are 41 total items: 20 4-point Likert-scale questions assessing specific activities of daily living, 13 additional questions assessing use of assistive devices, and 8 additional questions assessing help received from another.

Twenty specific activities are assessed on a 4-point Likert scale where 0 = without difficulty, 1 = with some difficulty, 2 = with much difficulty, and 3 = unable to do. The 20 activities are grouped into 8 functional categories with each category given a single score equal to the maximum value of their component activities (0, 1, 2, or 3).

There are 3 steps to scoring the HAQ (with aids/help): 1) identify the highest subcategory score from each of the 8 categories. Adjust for use of aids/help by increasing the category score from 0 or 1 to a 2 if use of aids/help for that category (utilize table of companion aids/help for HAQ categories). If the category score is already a 2 or 3, no adjustment is made; 2) sum the category scores; and 3) divide the final sum by the number of categories answered to obtain the final HAQ score rounded to the nearest value evenly divisible by 0.125. Requires a minimum of 6 categories answered; if less, do not score.”

5.1.1.10 ACR response

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

- 20% improvement in TJC68, AND
- 20% improvement in SJC66, AND
- 20% improvement in at least 3 of 5 other core set items

The other core set items consist of:

- PhGA
- PGA
- Patient’s Global Assessment of Pain on a VAS 0-100 mm
STATISTICAL ANALYSIS PLAN for the NORD-STAR trial

- HAQ-DI
- ESR/hsCRP

ACR50, ACR70 and ACR90 are defined in a similar manner with 50%, 70% and 90% improvement, respectively. High sensitivity CRP (hsCRP) 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.

5.1.1.11 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>
<thead>
<tr>
<th>Change from baseline in DAS28</th>
<th>Good</th>
<th>Moderate</th>
<th>None</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAS28 at time-point</td>
<td>ΔDAS28 ≤ -1.2</td>
<td>-1.2 &lt; DAS28 ≤ -0.6</td>
<td>DAS28 ≥ 0.6</td>
</tr>
<tr>
<td>DAS28 ≤ 3.2</td>
<td>Good</td>
<td>Moderate</td>
<td>None</td>
</tr>
<tr>
<td>3.2 &lt; DAS28 ≤ 5.1</td>
<td>Moderate</td>
<td>Moderate</td>
<td>None</td>
</tr>
<tr>
<td>DAS28 &gt; 5.1</td>
<td>Moderate</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>

5.1.1.12 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](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.

5.1.1.13 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.

5.1.1.14 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):

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

Percent activity impairment due to problem: \( \frac{Q6}{10} \)
5.1.2 Primary Outcome Definition
The primary radiographic outcome is the change in total van der Heijde-modified Sharp Score from baseline to week 48 (ΔvdHSS). The primary radiographic outcome is continuous.

The primary clinical outcome is the occurrence of CDAI remission at week 48. The primary clinical outcome is dichotomous.

5.1.3 Secondary Outcomes Definitions

<table>
<thead>
<tr>
<th>Group</th>
<th>Endpoint</th>
<th>Assessment time</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Key Radiology</strong></td>
<td>No radiographic progression (ΔvdHSS at 48 weeks ≤ 0.5)</td>
<td>At 48 weeks</td>
<td>Dichotomous</td>
</tr>
<tr>
<td></td>
<td>ΔErosion score</td>
<td>0 to 48 weeks</td>
<td>Continuous Δ</td>
</tr>
<tr>
<td></td>
<td>ΔJSN</td>
<td>0 to 48 weeks</td>
<td>Continuous Δ</td>
</tr>
<tr>
<td></td>
<td>ΔvdHSS</td>
<td>0 to 24 weeks</td>
<td>Continuous Δ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24 to 48 weeks</td>
<td></td>
</tr>
<tr>
<td><strong>Key Clinical</strong></td>
<td>ACR/EULAR Boolean remission</td>
<td>At 48 weeks</td>
<td>Dichotomous</td>
</tr>
<tr>
<td></td>
<td>DAS28 remission</td>
<td>At 48 weeks</td>
<td>Dichotomous</td>
</tr>
<tr>
<td></td>
<td>SDAI remission</td>
<td>At 48 weeks</td>
<td>Dichotomous</td>
</tr>
<tr>
<td></td>
<td>EULAR good response</td>
<td>At 48 weeks</td>
<td>Dichotomous</td>
</tr>
<tr>
<td><strong>Other Radiology</strong></td>
<td>Radiographic progression above smallest detectable change (ΔvdHSS at 48 weeks &gt; SDC*)</td>
<td>At 48 weeks</td>
<td>Dichotomous</td>
</tr>
<tr>
<td></td>
<td>Rapid radiographic progression (ΔvdHSS at 48 weeks &gt; 5)</td>
<td>At 48 weeks</td>
<td>Dichotomous</td>
</tr>
<tr>
<td></td>
<td>ΔErosion score</td>
<td>0 to 24 weeks</td>
<td>Continuous Δ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24 to 48 weeks</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ΔJSN</td>
<td>0 to 24 weeks</td>
<td>Continuous Δ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24 to 48 weeks</td>
<td></td>
</tr>
</tbody>
</table>
## STATISTICAL ANALYSIS PLAN for the NORD-STAR trial

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Timepoint</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Radiographic progression</td>
<td>At 24 weeks</td>
<td>Dichotomous</td>
</tr>
<tr>
<td>(ΔvdHSS at 24 weeks ≤ 0.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radiographic progression</td>
<td>At 48 weeks</td>
<td>Dichotomous</td>
</tr>
<tr>
<td>(ΔvdHSS between 24 and 48 weeks &gt; 0.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Erosion progression</td>
<td>At 48 weeks</td>
<td>Dichotomous</td>
</tr>
<tr>
<td>(ΔErosion at 48 weeks &gt; 0.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Erosion progression</td>
<td>At 24 weeks</td>
<td>Dichotomous</td>
</tr>
<tr>
<td>(ΔErosion at 24 weeks &gt; 0.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No erosion progression</td>
<td>At 48 weeks</td>
<td>Dichotomous</td>
</tr>
<tr>
<td>(ΔErosion between 24 and 48 weeks ≤ 0.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Joint space narrowing progression</td>
<td>At 48 weeks</td>
<td>Dichotomous</td>
</tr>
<tr>
<td>(ΔJSN at 48 weeks &gt; 0.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No joint space narrowing progression</td>
<td>At 24 weeks</td>
<td>Dichotomous</td>
</tr>
<tr>
<td>(ΔJSN at 24 weeks ≤ 0.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No joint space narrowing progression</td>
<td>At 48 weeks</td>
<td>Dichotomous</td>
</tr>
<tr>
<td>(ΔJSN between 24 and 48 weeks ≤ 0.5)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## STATISTICAL ANALYSIS PLAN for the NORD-STAR trial

<table>
<thead>
<tr>
<th>Other Clinical</th>
<th>CDAI remission</th>
<th>All post-baseline visits, except 48 weeks</th>
<th>Dichotomous</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACR/EULAR Boolean remission</td>
<td>All post-baseline visits, except 48 weeks</td>
<td>Dichotomous</td>
<td></td>
</tr>
<tr>
<td>DAS28 remission</td>
<td>All post-baseline visits, except 48 weeks</td>
<td>Dichotomous</td>
<td></td>
</tr>
<tr>
<td>SDAI remission</td>
<td>All post-baseline visits, except 48 weeks</td>
<td>Dichotomous</td>
<td></td>
</tr>
<tr>
<td>EULAR good response</td>
<td>All post-baseline visits, except 48 weeks</td>
<td>Dichotomous</td>
<td></td>
</tr>
<tr>
<td>CDAI Low disease activity</td>
<td>All post-baseline visits, except 48 weeks</td>
<td>Dichotomous</td>
<td></td>
</tr>
<tr>
<td>DAS28 Low disease activity</td>
<td>All post-baseline visits</td>
<td>Dichotomous</td>
<td></td>
</tr>
<tr>
<td>ACR20 response</td>
<td>All post-baseline visits</td>
<td>Dichotomous</td>
<td></td>
</tr>
<tr>
<td>ACR50 response</td>
<td>All post-baseline visits</td>
<td>Dichotomous</td>
<td></td>
</tr>
<tr>
<td>ACR70 response</td>
<td>All post-baseline visits</td>
<td>Dichotomous</td>
<td></td>
</tr>
<tr>
<td>EULAR Good or moderate response</td>
<td>All post-baseline visits</td>
<td>Dichotomous</td>
<td></td>
</tr>
<tr>
<td>ΔDAS28</td>
<td>All post-baseline visits</td>
<td>Continuous Δ</td>
<td></td>
</tr>
<tr>
<td>ΔSDAI</td>
<td>All post-baseline visits</td>
<td>Continuous Δ</td>
<td></td>
</tr>
<tr>
<td>ΔCDAI</td>
<td>All post-baseline visits</td>
<td>Continuous Δ</td>
<td></td>
</tr>
<tr>
<td>ΔSJC66</td>
<td>All post-baseline visits</td>
<td>Continuous Δ</td>
<td></td>
</tr>
</tbody>
</table>
### STATISTICAL ANALYSIS PLAN for the NORD-STAR trial

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Assessment time</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔSJC28</td>
<td>Continuous Δ</td>
<td>All post-baseline visits</td>
<td></td>
</tr>
<tr>
<td>ΔTJC68</td>
<td>Continuous Δ</td>
<td>All post-baseline visits</td>
<td></td>
</tr>
<tr>
<td>ΔHAQ-DI</td>
<td>Continuous Δ</td>
<td>All post-baseline visits</td>
<td></td>
</tr>
<tr>
<td>Morning stiffness</td>
<td>Dichotomous</td>
<td>All post-baseline visits</td>
<td></td>
</tr>
<tr>
<td>ΔPGA</td>
<td>Continuous Δ</td>
<td>All post-baseline visits</td>
<td></td>
</tr>
<tr>
<td>ΔPhGA</td>
<td>Continuous Δ</td>
<td>All post-baseline visits</td>
<td></td>
</tr>
<tr>
<td>ΔJointPain</td>
<td>Continuous Δ</td>
<td>All post-baseline visits</td>
<td></td>
</tr>
<tr>
<td>ΔESR</td>
<td>Continuous Δ</td>
<td>All post-baseline visits</td>
<td></td>
</tr>
<tr>
<td>ΔCRP</td>
<td>Continuous Δ</td>
<td>All post-baseline visits</td>
<td></td>
</tr>
</tbody>
</table>

Smallest detectable change (SDC): \( SDC = 1.966 \times SD_{\text{between readers}} \times \sqrt{\frac{1}{k}} \), where \( k \) is the number of readers (\( k=2 \)). For this method one firstly calculates the differences between the change-scores obtained in the repeated reading session (i.e. the change in score between baseline and follow-up for reader 1 is subtracted from the change in score between baseline and follow-up for reader 2). Secondly, the SD of these differences is calculated. This SDDIFF(CHANGE-SCORES) reflects the measurement error of the difference between two change-scores that is, the measurement error when discriminating between two change-scores. (Bruynesteen et al, Ann Rheum Dis 2005; 64: 179–182).

#### 5.1.4 Patient reported outcome measures

<table>
<thead>
<tr>
<th>Group</th>
<th>Endpoint</th>
<th>Assessment time</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>SF-36</td>
<td>ΔPhysical functioning</td>
<td>All post-baseline visits</td>
<td>Continuous Δ</td>
</tr>
<tr>
<td></td>
<td>ΔBodily pain</td>
<td>All post-baseline visits</td>
<td>Continuous Δ</td>
</tr>
<tr>
<td></td>
<td>ΔRole limitations due to physical health problems</td>
<td>All post-baseline visits</td>
<td>Continuous Δ</td>
</tr>
<tr>
<td></td>
<td>ΔRole limitations due to personal or emotional problems</td>
<td>All post-baseline visits</td>
<td>Continuous Δ</td>
</tr>
</tbody>
</table>
### STATISTICAL ANALYSIS PLAN for the NORD-STAR trial

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Measurement Period</th>
<th>Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔEmotional well-being</td>
<td>All post-baseline visits</td>
<td>Continuous Δ</td>
</tr>
<tr>
<td>ΔSocial functioning</td>
<td>All post-baseline visits</td>
<td>Continuous Δ</td>
</tr>
<tr>
<td>ΔEnergy/fatigue</td>
<td>All post-baseline visits</td>
<td>Continuous Δ</td>
</tr>
<tr>
<td>ΔGeneral health perception</td>
<td>All post-baseline visits</td>
<td>Continuous Δ</td>
</tr>
<tr>
<td>ΔPhysical health composite score</td>
<td>All post-baseline visits</td>
<td>Continuous Δ</td>
</tr>
<tr>
<td>ΔMental health composite score</td>
<td>All post-baseline visits</td>
<td>Continuous Δ</td>
</tr>
<tr>
<td>RAID total score</td>
<td>All post-baseline visits</td>
<td>Continuous Δ</td>
</tr>
<tr>
<td>ΔRAID index value</td>
<td>All post-baseline visits</td>
<td>Continuous Δ</td>
</tr>
<tr>
<td>ΔAbsenteeism</td>
<td>All post-baseline visits</td>
<td>Continuous Δ</td>
</tr>
<tr>
<td>ΔPresenteeism</td>
<td>All post-baseline visits</td>
<td>Continuous Δ</td>
</tr>
<tr>
<td>ΔWork productivity loss</td>
<td>All post-baseline visits</td>
<td>Continuous Δ</td>
</tr>
<tr>
<td>ΔActivity impairment</td>
<td>All post-baseline visits</td>
<td>Continuous Δ</td>
</tr>
<tr>
<td>Patient Acceptable Symptom State</td>
<td>All post-baseline visits</td>
<td>Dichotomous</td>
</tr>
<tr>
<td>FACIT-Fatigue subscale score</td>
<td>All post-baseline visits</td>
<td>Continuous Δ</td>
</tr>
</tbody>
</table>

### 5.1.5 Safety outcome definitions

#### 5.1.5.1 Treatment emerging adverse events

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

#### 5.1.5.2 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.
STATISTICAL ANALYSIS PLAN for the NORD-STAR trial

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.

5.2 Analysis Methods

5.2.1 Primary Outcome

Primary Analysis
The change in total van der Heijde-modified Sharp score from baseline to week 48 will be analysed using analysis of covariance (ANCOVA), adjusted for baseline score and the stratification factor in the randomisation (gender, anti-CCP status and country).

The occurrence of CDAI remission at 48 weeks will be analysed using logistic regression, adjusted for the stratification factors in the randomisation (gender, anti-CCP status and country).

The primary analyses will be performed on the ITT population as defined in section 3.3.

Summary Measures
Descriptive statistics will include cumulative probability plots, and number and percentage by treatment group. Descriptive statistics will be based on non-imputed data, thus the number of evaluable outcome measurements at the time of primary interest (48 weeks) will also be presented.

The primary radiographic effect estimates will be the adjusted difference in mean ΔvdHSS between ACT and each of the three biologic treatments. Each effect estimate together with the 95% confidence interval and p-value of the null hypothesis test will be presented, adjusted for multiplicity by the Dunnett’s procedure.

The primary clinical effect estimate will be the adjusted risk difference in CDAI remission between ACT and each of the three biologic treatments, computed from the logistic regression effect estimate using the delta method. The adjusted relative risk will also be reported together with the p-value of the null-hypothesis test of no treatment difference from the logistic regression. P-values and confidence intervals will be adjusted for multiplicity by the Dunnett’s procedure.
STATISTICAL ANALYSIS PLAN for the NORD-STAR trial

5.2.1.3 Assumption Checks and Alternative Analyses
The ANCOVA model will be checked for deviations from the normal assumption by plotting the residuals both against the fitted values and the quantiles against quantiles. Blinded analyses have revealed an expected deviation from the normality assumption. However, we assume the sample size is sufficient for the model to provide unbiased estimates according to the central limit theorem.

A robustness analysis will be undertaken where the adjusted p-values will be calculated based on normal scores i.e. that the values of ΔvdHSS will be transformed to the corresponding quantile according to the normal distribution. The same ANCOVA model will be used, and p-values will be adjusted according to the Dunnett’s method.

The logistic regression relies on few assumptions, and there will be no assumption checks for the primary analysis of CDAI remission.

5.2.1.4 Missing Data
For the primary radiographic outcome missing observations will be imputed as described in section 5.1.1.4

For the primary clinical outcome, missing data will be imputed with worst outcome which is no response. From the blinded review, it is known that the majority of missing data for this outcome is lack of response to the allocated treatment. Thus, a non-response imputation is the only clinically relevant method.

5.2.1.5 Sensitivity Analyses
For both primary endpoints, the following robustness analyses will be performed:

- unadjusted analyses on the ITT population
- adjusted on the PP population
- longitudinal analyses based on mixed models

Details will be given in the corresponding section for dichotomous and radiographic outcomes.

5.2.1.6 Subgroup Analyses
There will be no subgroup analyses in this part of the trial.

5.2.2 Dichotomous Secondary Outcomes at 48 weeks
Dichotomous secondary outcomes at 48 weeks will be analysed similar to the primary clinical CDAI remission outcome. There will be no adjustments for multiplicity in the secondary outcome analyses.

5.2.2.1 Main Analysis
Dichotomous secondary outcomes will be analysed using logistic regression, adjusted for the stratification factors in the randomisation (gender, anti-CCP status and country).

The main analysis will be performed on the ITT population.

5.2.2.2 Summary Measures
Summary measures will include descriptive numbers and percentages in addition to adjusted risk differences. Effect measures will be presented with unadjusted 95% confidence limits, but without...
5.2.2.3 Assumption Checks
There will be no assumptions checks.

5.2.2.4 Missing Data
Missing data will be imputed by worst outcome.

5.2.2.5 Sensitivity Analyses
There will be no sensitivity analyses.

5.2.2.6 Subgroup Analyses
There will be no subgroup analyses.

5.2.1 Longitudinal Analyses of Dichotomous Secondary Outcomes
Dichotomous secondary outcomes for all visits post baseline will be analysed using a repeated measures methodology. There will be no adjustments for multiplicity.

5.2.1.1 Main Analysis
Dichotomous secondary outcomes will be analysed using repeated measures logistic regression, adjusted for the stratification factors in the randomisation (gender, anti-CCP status and country). Within-patient dependencies will be handled by estimating the variance-covariance matrix using generalized estimating equations.

This analysis will be performed on the ITT population.

5.2.1.2 Summary Measures
Summary measures will include descriptive numbers and percentages in addition to adjusted risks and risk differences. Effect measures will be presented with unadjusted 95% confidence limits, but without p-values. Plots with the adjusted risks and confidence limits by treatment will be presented. The adjusted risk differences and corresponding 95% confidence intervals will be based on the averaged predictive margins and the delta method.

5.2.1.3 Assumption Checks
There will be no assumptions checks.

5.2.1.4 Missing Data
Missing data will be imputed by worst outcome.

5.2.1.5 Sensitivity Analyses
The following sensitivity analysis will be performed:

- unadjusted analyses on the ITT population
- adjusted on the PP population

5.2.1.6 Subgroup Analyses
There will be no subgroup analyses.
5.2.2 Continuous Secondary Radiographic Outcomes

Continuous secondary radiographic outcomes will be analysed similar to the primary radiographic outcome. There will be no adjustments for multiplicity in the secondary outcome analyses.

5.2.2.1 Main Analysis

Secondary continuous radiographic outcomes will be analysed using ANCOVA, adjusted for the stratification factors in the randomisation (gender, anti-CCP status and country) and baseline value. This analysis will be performed on the ITT population.

5.2.2.2 Summary Measures

Summary measures will include cumulative probability plots and adjusted difference in adjusted mean between ACT and each of the three biologic treatments. Effect measures will be presented with unadjusted 95% confidence limits, but without p-values. The adjusted estimates and corresponding 95% confidence intervals will be based on the averaged predictive margins and the delta method.

5.2.2.3 Assumption Checks

There will be no assumptions checks.

5.2.2.4 Missing Data

All missing observations will be imputed as described in section 5.1.1.4. A sensitivity analysis on un-imputed data will be performed using mixed models with random intercept and random slope as described in the next section.

5.2.2.5 Sensitivity Analyses

The following sensitivity analysis will be performed:

- unadjusted analyses on the ITT population
- adjusted on the PP population

In addition, sensitivity analyses will be performed using mixed models with time by treatment as fixed covariates, and random intercept and random slope as random covariates. This analysis will be done on non-imputed data. The structure of the covariance matrix will be set to unrestricted. The summary measure from this model will be the difference in average yearly radiographic slope (deterioriation) between the treatment groups.

5.2.2.6 Subgroup Analyses

There will be no subgroup analyses.

5.2.3 Continuous Secondary Outcomes

5.2.3.1 Main Analysis

Secondary continuous longitudinal outcomes other than radiographic outcomes will be analysed using generalized linear mixed model with random intercept, adjusted for the stratification factors in the randomisation (gender, anti-CCP status and country). For skewed variables (CRP and ESR) we will
STATISTICAL ANALYSIS PLAN for the NORD-STAR trial

use a gamma model, for count variables (joint counts) we will use a negative binomial model. For all other variables we will use the normal model.

5.2.3.2 Summary Measures
Summary measures will include estimated marginal means and corresponding unadjusted 95% confidence limits, both by treatment group and time, and by treatment difference and time. The adjusted estimates and corresponding 95% confidence intervals will be based on the averaged predictive margins and the delta method.

5.2.3.3 Assumption Checks
The residuals will be checked in models without treatment term

5.2.3.4 Missing Data
Missing data will be handled within the longitudinal mixed model framework.

5.2.3.5 Sensitivity Analyses
There will be no sensitivity analyses.

5.2.3.6 Subgroup Analyses
There will be no subgroup analyses.

5.2.4 Additional Analyses
None.

6 Safety Analyses
General safety evaluations will be based on the incidence, intensity, and type of adverse events (AEs) Safety variables will be tabulated and presented for all patients in the safety set.

6.1 Adverse Events
Adverse events will be coded using MedDRA, version 22.0. The investigator records the maximum intensity of each AE using the levels mild, moderate and severe. For tabulations, only treatment emerging adverse events (TEAEs) will be presented. TEAEs are defined as AEs with a start date on or after date of first randomised treatment. Any AEs prior to treatment will be listed but not tabulated.

The number (%) of subjects with any TEAEs, with 1, 2 or > 3 TEAEs, with treatment related TEAEs, with treatment emerging serious AEs (TESAE) and TEAEs of special interest will be summarised by treatment group. TEAEs of special interest are infections, cardiovascular diseases, cataracts, demyelinating diseases, diabetes mellitus, herpes zosters, malignancies, osteoporosis, and weight gain. 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 serious AEs and for AEs of special interest. In addition, a summary table of AEs reported by >= 2% of all patients will be presented by SOC and PT.

7 Statistical Software
STATISTICAL ANALYSIS PLAN for the NORD-STAR trial

All statistical analyses will be done in Stata v16 (StataCorp. 2020. Stata Statistical Software: Release 16. College Station, TX, USA), and R version 4.03 (R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/)

8 References

8.1 Literature References
8.2 Reference to Data Handling Plan
8.3 Reference to the Trial Master File and Statistical Documentation
8.4 Reference to other Standard Operating Procedures or Documents