STATISTICAL ANALYSIS PLAN (SAP) for MRI studies in BackToBasic

Administrative information:

Sponsor name	Prof. Dr. Med. John-Anker Zwart, Forskning og Utvikling, Nevroklinikken, Oslo University Hospital
Sponsor address	Nydalen, 0424, Oslo, Norway
EudraCT number / REC no	2017-004861-29 / 2017/2450/REK sør-øst D
Trial title	BackToBasic: The effect of Infliximab in patients with chronic low back pain and Modic changes. A randomized, double blind, placebo-controlled, multicenter trial.
Trial ID	BackToBasic
Trial registration number	EudraCT number: 2017-004861-29 ClinicalTrials.gov ID: NCT03704363

SAP and protocol version:

SAP version and date:	This supplemental SAP is version 1.0, dated 4 March 2024
Protocol version	This document has been written based on information contained in the study protocol version 2.1, dated 19 March 2019

This SAP supplements the original SAP for BackToBasic describing the analyses of clinical outcomes and clinical subgroups, available at ClinicalTrials.gov (NCT06169488). The statisticians who signed the original SAP are Erica Ponzi PhD (Trial Statistician) and Morten Valberg PhD (QC Statistician).

The current SAP is authored by Ansgar Espeland PhD, and has been reviewed by Lars Christian H Braathen PhD, Magnhild Dagestad MD, Elisabeth Gjefsen MD, Kristina Gervin PhD, Kjersti Storheim PhD, John-Anker Zwart PhD, and Jörg Aßmus PhD (Statistician).

SIGNATURE PAGE

PRINCIPAL/COORDINATING INVESTIGATOR:

Kjersti Storheim, PT, PhD
Forsknings- og formidlingsenheten for muskelskjelettlidelser (FORMI), Oslo University Hospital

Kjersti Storheim — Signature

04.03.2024 — Date (dd/mmm/yyyy)

MRI RESPONSIBLE INVESTIGATOR:

Ansgar Espeland, PhD
Department of Radiology, Haukeland University Hospital

Ansgar Espeland — Signature

04.03.2024 — Date (dd/mmm/yyyy)

STATISTICIAN:

Jörg Allmus, PhD
Centre for Clinical Research, Haukeland University Hospital

Signature

04/03-2024 — Date (dd/mmm/yyyy)

STATISTICAL ANALYSIS PLAN (SAP) for MRI studies in BackToBasic

Introduction

This supplemental SAP was prepared after main results for treatment efficacy in the trial and in clinical subgroups were available, based on analyses described in the original SAP for BackToBasic. However, the **Secondary Objectives B** and **E** described below in the present SAP were defined in the study protocol version 2.1., dated 19 March 2019, prior to any results being available.

Background and rationale

BackToBasic included patients with chronic low back pain (LBP) and Modic changes type 1 (MC1) on magnetic resonance imaging (MRI), suggesting vertebral bone marrow edema/inflammation. Patients were randomized to anti-inflammatory treatment with a tumor necrosis factor (TNF)-alpha inhibitor (infliximab) or placebo. It is not clear if this treatment reduces MC related bone marrow edema (MC edema) or if subgroups with more MC edema modify the treatment effect. Short tau inversion recovery (STIR) MRI sequences are sensitive to MC edema, which increases the signal intensity.

The present Study Objectives

The first objective (**Secondary Objective B** in the study protocol) is to assess if infliximab is superior to placebo for reducing MC edema from baseline to 6 months. The corresponding secondary endpoint is the change in STIR signal (intensity and extent) of MCs from baseline to 6 months MRI (reduced, unchanged, increased), dichotomized into reduced or not reduced (i.e., unchanged or increased).

The next objective (**Secondary Objective E** in the study protocol) is to assess if the baseline STIR signal (intensity, extent) of MCs **predicts** clinical outcome. The corresponding secondary endpoints (clinical outcomes) are the Oswestry Disability Index (ODI) - and the LBP intensity score – at the primary 5 months clinical follow-up. ODI scores can range from 0 to 100 and LBP intensity scores from 0 to 10. We will also **explore** if the baseline apparent diffusion coefficient (ADC) of MCs (representing diffusion of water molecules) predicts these endpoints. This implies to assess whether the STIR signal or ADC of MCs predicts any effect of infliximab vs placebo on ODI or LBP intensity score at 5 months adjusted for the baseline score, i.e., whether the STIR signal or the ADC are **effect modifiers**.

We also aim to **explore** whether the STIR signal or ADC of MCs predicts any effect of infliximab vs placebo on ODI or LBP intensity score at the final clinical 9 months safety follow-up.

The last objective is to **explore** if the baseline STIR signal (intensity and extent) and ADC of MCs are **prognostic factors** for ODI and LBP intensity scores at 5 months follow-up.

MRI assessment

Volume and intensity of the STIR signal, i.e., of the MC oedema on STIR (= STIR oedema), are assessed per endplate level, including all oedema regions at that level. ADC is also recorded per endplate level.

Change in STIR oedema (intensity and extent) from baseline to 6 months MRI is visually categorized by two raters, and when they disagree, by a third rater. The follow-up MRI was planned to be performed at 6 months, not at 5 months, since it was to be requested at the 5 months clinical follow-up, and organizing the MRI visit required some time. The categories used both per endplate level and per patient are reduced, unchanged, or increased oedema. At each endplate level, a change in STIR

oedema (e.g., 'reduced') is noted if it is present on ≥2 slices and on ≥2 more slices than any opposite change (e.g., 'increased'). In each patient, STIR oedema is 'unchanged' if unchanged at all endplate levels, 'reduced' if reduced at ≥1 endplate and increased at 0 endplates, and 'increased' if increased at ≥1 endplate and reduced at 0 endplates. If there are opposite changes in STIR oedema at different endplate levels, a change in the patient is reported if it outweighs any opposite change on ≥2 slices.

STIR oedema volume is independently and visually categorized by three raters according to which percentage of the vertebral body marrow (of the vertebra containing the oedema) it is visually judged to occupy. The volume of the oedema is evaluated considering all slices and all oedema regions at the given endplate level. The categories used are (1) <10%, (2) 10% to <25%, (3) 25-50%, and (4) >50%.

STIR oedema volume is also measured in cm^3 by one rater and reported as a percentage of the measured volume of the vertebral body marrow containing the oedema.

At each endplate level, oedema volume is measured by measuring the oedema area on each slice, summing these areas across all slices, and multiplying that sum by slice thickness plus interslice gap. The volume of the vertebral body marrow is derived similarly, based on manual annotation (of S1) or automated annotation (of Th12-L5) of the vertebral body marrow on each slice. The automated annotation is based on artificial intelligence (deep learning) and is corrected manually when needed.

Mean intensity of the STIR oedema across the whole oedema volume is measured by one rater. It is reported as STIRint%, which describes where the mean intensity of the STIR oedema is placed on the percentage scale from normal vertebral body marrow intensity (0%) to cerebrospinal fluid (CSF) intensity (100%).

In a subsample of 20 patients, volumes and intensities of STIR oedema are independently measured by a second rater and re-measured by the first rater for analysis of inter- and intra-rater reliability.

ADC values are measured independently by two raters. MC-ADC is the highest mean ADC value in a circular region with diameter 7 mm placed in the most intense MC related area on the ADC map. MC-ADC% describes where this MC-ADC value is located on the percentage scale from ADC measured in normal vertebral body marrow (0%) to ADC measured in CSF (100%).

Potential effect modifiers and prognostic factors

The following baseline STIR/ADC variables will be analyzed as effect modifiers and prognostic factors:

1. **STIRmax**: maximum volume of STIR oedema at any endplate level visually assessed as ≥25% of vertebral body marrow volume (yes or no, categorical variable; **primary effect-modifier variable**)

2. **STIRsum**: sum of the two* largest measured STIR oedema volumes at two different endplate levels in percent of measured vertebral body marrow volume (continuous variable)

3. **STIRvolint**: sum of the two* largest measured values for the product "STIR volume (in percent) × STIR intensity (in percent)/100%" at two different endplate levels (continuous variable)

4. **MC-ADC%**: maximum MC related ADC value in percentage on the scale from ADC in normal vertebral body marrow (0%) to ADC in CSF (100%) (continuous variable)

***** The two largest values are used since few patients had STIR oedema at a third endplate level, and the sum of the three largest values from three endplate levels and the sum of the two largest values from two endplate levels correlate strongly (Spearman's rho for STIRsum 0.98 and for STIRvolint 0.95).

STATISTICAL ANALYSIS PLAN (SAP) for MRI studies in BackToBasic

Clinically important treatment effect

In the original trial SAP, a clinically important treatment effect for ODI (primary outcome) was defined as a 10 points difference in improvement between the treatment groups at 5 months follow-up.

Predefined hypotheses

Our research hypotheses are that:

A. Infliximab is superior to placebo for reducing MC edema on STIR from baseline to 6 months.
B. STIRmax ≥25%, high STIRsum, high STIRvolint, and high MC-ADC% at baseline each predicts lower 5 months ODI and LBP intensity scores adjusted for baseline scores in the infliximab vs placebo group.
C. STIRmax ≥25%, high STIRsum, high STIRvolint, and high MC-ADC% at baseline are prognostic factors for higher 5 months ODI and LBP intensity scores adjusted for baseline scores.

The statistical null hypotheses are that:

A. Infliximab does not differ from placebo in reducing MC edema on STIR.
B. Each of the STIR/ADC variables does not modify the effect of infliximab on ODI or LBP intensity.
C. Each of the STIR/ADC variables is not a prognostic factor for ODI or LBP intensity.

The statistical two-sided alternative hypotheses are that:

A. Infliximab differs (any direction) from placebo in reducing MC edema on STIR.
B. Each of the STIR/ADC variables modifies the effect of infliximab on ODI and LBP (any direction).
C. Each of the STIR/ADC variables is a prognostic factor for ODI/ LBP intensity (any direction).

Analyses

All analyses in this SAP are regarded a priori analyses as they are defined prior to being conducted. Any post hoc analyses will be reported as such. All statistical tests will be 2-sided. All CIs will be 95% and 2-sided. The significance level will be 0.05 (2-sided) to minimize type 2 error.

The assumption of normal distribution will be checked via residual plots / Q-Q plots. If the assumption is violated, the variable may be transformed.

Samples for analysis

Primary analyses for Secondary objectives B and E will be based on the per protocol analysis set (PP). The PP set (n=78) includes all randomized patients meeting the study eligibility criteria and with no major protocol deviations affecting the treatment efficacy, as defined in the original SAP for the trial. This means that we focus primarily on the effect of infliximab and not on the effect of the randomization. The effect of infliximab on MC oedema, and whether MRI findings predict effect of infliximab, appear most relevant to study in PP patients treated in agreement with the protocol.

Sensitivity analyses for Secondary objectives B and E will be based on the full analysis set (FAS) (n=128), defined as all randomized patients having received at least one study treatment infusion.

We expect any treatment effect on MC oedema, and any ability of a STIR/ADC variable to predict treatment effect, to be weaker in the FAS vs the PP set. Yet, the larger FAS provides better power to detect a relevant result as statistically significant, why both the FAS and the PP set are analyzed.

The STIR/ADC variables will be analyzed as prognostic factors in the FAS at 5 months.

Missing data

We will exclude a few patients from the relevant analyses due to lack of MRI data: (a) patients lacking a 6-month MRI (5 of 128 patients, 3.9 %; 2 of 78 PP patients, 2.6%) will be excluded from the analyses of treatment effect on MC oedema (secondary objective B in the study protocol) and (b) patients lacking a baseline ADC map (also 5 of 128 patients, 3.9 %) will be excluded from analyses involving an ADC variable. We assume that the data in (a) and (b) are lacking at random and we will not perform data imputation as a sensitivity analysis, since the data are lacking in very few patients (< 5%).

A few patients will be excluded from relevant analyses due to missing data for ODI at 5 months (6 of 128 patients, 4.7%) or 9 months (14/128 patients, 11 %) or LBP intensity at 5 months (9/128 patients, 7 %) or 9 months (17/128 patients, 13 %). In the PP set (n=78) at the primary follow-up visit at 5 months, no patient lacks data on the primary outcome ODI, and only two lack data on LBP intensity.

We will exclude patients with missing data, since very few lack relevant data in the PP set used in the primary analyses of the predefined Secondary objectives B and E, and the FAS is used only for sensitivity analyses of these objectives. We may, in a further sensitivity analysis, impute missing values of ODI and LBP intensity based on the multiple imputation models used in the AIM study.

Analyses of whether infliximab reduces STIR oedema

All patients have STIR oedema at baseline in this study of patients with MC1 (oedema type MC).

Logistic regression will be conducted with change in STIR oedema dichotomized into reduced vs. not reduced (= unchanged or increased) as dependent variable. The independent variable will be treatment group, adjusted for study center (a stratification factor in the randomization) and baseline STIR oedema (STIRmax).

We will report the odds ratio (OR) with a 95% CI and the p value from Wald's test. The model-predicted marginal probabilities of reduced STIR signal may also be estimated for both treatment groups, and the difference between the two probabilities reported with a 95% CI. We may also report number needed to treat (NNT) with 95% CIs.

We will also report unadjusted descriptive raw data for all categories of change in STIR signal of MCs (reduced, unchanged, increased) from baseline to 6 months follow-up MRI for each treatment group.

The interaction term STIRmax × treatment group will be added in a sensitivity analysis to see if the baseline STIR oedema modifies any effect of the treatment in reducing the STIR oedema.

The main result for effect of infliximab on STIR oedema (Secondary objective B, Hypothesis A) will be the result for effect visible on 6 months MRI in the PP analysis set.

Analyses of effect modifiers

Baseline STIRmax, STIRsum, STIRvolint, and MC-ADC% will be analyzed each at the time as potential modifiers of effect of infliximab vs placebo on ODI and LBP intensity at 5 and 9 months, using ANCOVA. ODI (or LBP intensity) at 5 (or 9) months will be dependent variable. Independent variables will be treatment group, the potential effect modifier, and their interaction, adjusted for baseline ODI (or LBP intensity respectively) and study center (stratification factor in the randomization).

Previous participation in the Antibiotics In Modic changes (AIM) study was also a stratification factor in the randomization. However, since < 5% of the patients had participated in the AIM study, this variable will not be adjusted for to avoid problems with the convergence and model fitting.

We will report the effect of infliximab vs placebo in each subgroup as the predicted mean marginal effect (95% CI) from the ANCOVA, report p for interaction ($p < 0.05$ implies statistically significant effect modification), report B values with 95% CIs, and assess the fit of the models by eta squared.

We will check that assumptions for ANCOVA are fulfilled. This will include inspection of residual plots.

We may plot ODI/LBP intensity at 5 and 9 months against STIR/ADC variables in each treatment group and estimate the correlation between the variables.

If a continuous potential predictor shows significant interaction with treatment group ($p < 0.05$), we may perform analyses to estimate which cutoff of that variable provides the best prediction.

The main result for effect modifiers (Secondary objective E, Hypothesis B) will be the result for the ability of STIRmax to modify the effect of infliximab vs placebo on ODI at 5 months in the PP set.

Analyses of prognostic factors

The relevant above ANCOVA models will be used to analyze each STIR/ADC variable as a prognostic factor for ODI and LBP intensity at 5 months in the FAS. These models include each STIR/ADC variable as an independent variable, i.e., as a prognostic factor. If the interaction between the variable and treatment group is non-significant and does not improve model fit, it will be omitted from the model used to assess the prognostic factor.

We may plot ODI/LBP intensity at 5 and 9 months against STIR/ADC variables for both treatment groups combined and estimate the correlation between the variables.