STATISTICAL ANALYSIS PLAN

Direct comparison of intra-articular saline injections with an education plus exercise program for treatment of knee osteoarthritis symptoms: SAP for the randomised, open label, controlled, evidence-based DISCO trial

Trial Registration

Clinicaltrials.gov Trial registration identifier: NCT03843931
European Clinical Trials Database (EudraCT number): 2019-000809-71
Internal protocol number: APPI2-PT-2019-01

Protocol Version and Date

This document has been written based on information contained in the study protocol version 1.5, 28 June 2019.

Statistical Analysis Plan Version and Date

Version 1.0
Date: September 23, 2020

Statistical Analysis Plan Authors

Marius Henriksen
Elisabeth Bandak
Robin Christensen

CHANGE HISTORY

Protocol version	Updated SAP version	Section Number Changed	Description of and reason for change	Date changed

1 SIGNATURES

Approved by Study Director Marius Henriksen

Marius Henriksen, Prof, PhD
The Parker Institute

Signature

Date: sept 25/2020

Approved by Trial Manager Elisabeth Ann Bandak

Elisabeth Ann Bandak, PhD
The Parker Institute

Signature

Date: Sept. 25/2020

Approved by Principal Investigator Henning Bliddal

Henning Bliddal, Prof., DMSc.
The Parker Institute

Signature

Date: 25/9 20

Approved by Senior Biostatistician Robin Christensen

Robin Christensen, Prof., PhD
The Parker Institute

Signature

Date: Sept. 25/2020

2 TABLE OF CONTENT

3 PURPOSE

This statistical analysis plan (SAP) describes detailed aspects of data preparation and analysis and was set up before starting the final analysis. The SAP is based on the final trial protocol (Version 1.5, 28 June 2019).

4 STUDY SYNOPSIS

Background and rationale:	Knee osteoarthritis (OA) is a highly prevalent musculoskeletal condition mainly affecting older people, causing pain, physical disability, and reduced quality of life. Exercise and patient education is recommended as treatment, but placebo controlled studies do not exist. This trial has been designed to compare the effects of a widely used exercise and education program (the Good Life with osteoArthritis in Denmark; GLAD concept) with a widely used placebo-comparator; IA saline injections, on improving knee pain in individuals with knee OA.
Objectives:	*Primary objective:* To assess efficacy equivalence between GLAD vs. 4 intra-articular saline injections, on changes in knee pain in individuals with knee OA. *Key secondary objectives:* To compare GLAD vs intra-articular saline injections on changes from baseline at week 9 in patient-reported: Physical function, knee-related quality of life, and the patients' global assessment of impact of knee OA *Other secondary objectives:* To compare GLAD vs intra-articular saline injections on changes from baseline in: Patient reported knee OA symptoms and physical function in sports and recreational activities, physical performance tests, and clinical assessment of presence of swelling in the target knee joint and the number of OMERACT-OARSI responders at week 9 as well as comparisons of changes from baseline in all outcomes at week 12. *Exploratory objectives:* To assess efficacy equivalence of GLAD vs 4 intra-articular saline injections on: Changes from baseline in in morning knee pain; Changes from baseline in constant and intermittent pain (ICOAP) score; Analgesics use.
Outcomes:	*Primary outcome:* Change from baseline in the Knee injury and Osteoarthritis Outcome Score (KOOS) pain subscale at 9 weeks. *Key secondary outcomes:* Change from baseline to week 9 the Function and Quality of life subscales of the KOOS as well as the patient's global assessment of impact of OA. *Other secondary outcomes:* Changes from baseline to week 9 in the Symptoms and Sports and recreational subscales of the KOOS questionnaire, 4x10 meter fast walk test, 30 seconds chair stand test, stair climbing test. Also, the number of treatment responders as per the OMERACT-OARSI response criteria. *Safety outcomes:* Swollen study knee joint count, joint effusion and aspiration volume at week 9 and 12. *Exploratory outcomes:* Change from baseline in Morning pain and the ICOAP score at week 9 and 12. Analgesics use at week 9.
Study design:	The trial is a randomized, controlled, open-label, equivalence trial with two parallel groups comparing the GLAD concept with intra-articular saline injections.
Statistical considerations:	Primary analyses will be based on an intention-to-treat principle. Continuous scores will be analyzed using mixed linear models adjusted for baseline values of the scores, taking randomization stratification factors into account. Missing data will not be imputed. Dichotomous scores will be analyzed using logistic regression models. Sensitivity analyses will be done on the per-protocol population and using data sets with missing data replaced using multiple imputation, and baseline observation carried forward. Adverse events will be presented in a descriptive way for both groups.

For further details regarding the trial design, please see the protocol version 1.5, 28 June 2019.

5 INTRODUCTION

5.1 Background and rationale

In brief, knee osteoarthritis (OA) is a highly prevalent musculoskeletal condition mainly affecting older people, causing pain, physical disability, and reduced quality of life. Exercise and patient education is recommended as a primary treatment strategy, and based on these recommendation the GLAD concept has been developed and widely adopted. However, no placebo-controlled trials of exercise and education interventions exist. This trial has been designed to compare the effects of the GLAD concept (8 weeks exercise and education) with a well-established placebo (intra-articular saline injections) in patients with knee OA.

For more details about the background, rationale and evidence base of the trial, please see the protocol version 1.5, 28 June 2019.

6 STUDY METHODS

6.1 Trial Design

The trial is a single centre, randomised, parallel-group, open-label, equivalence, 12 weeks trial comparing an 8-week Exercise plus Education program (the GLAD concept) and 4 intra-articular saline injections (IA Saline) separated by two weeks with a primary endpoint at week 9 (9 weeks after initiation of treatment) and a further assessment 12 weeks after initiation of treatment.

The trial is conducted among individuals with painful knee OA. A planned total of 200 patients will be randomly assigned to one of the two treatments, GLAD concept or IA Saline injections.

6.2 Study Objectives

The primary objective of this trial is to assess efficacy equivalence between the GLAD concept vs. intra-articular saline injections, on change from baseline to week 9 in knee pain in individuals with knee OA.

The key secondary objectives are to compare the GLAD concept vs intra-articular saline injections on changes from baseline at week 9 in patient-reported: Physical function, knee-related quality of life, and the patients' global assessment of impact of knee OA

Our other secondary objectives include comparisons of the GLAD concept vs intra-articular saline injections on changes from baseline in patient reported knee OA symptoms and physical function in sports and recreational activities, changes from baseline in physical performance tests, and clinical

assessment of presence of swelling in the target knee joint and the number of OMERACT-OARSI responders at week 9 as well as comparisons of changes from baseline in all outcomes at week 12.

6.3 Randomization

Randomization is equal, i.e. on a 1:1 basis. Each randomisation is via minimisation incorporating a random element stratified for the following baseline conditions:

Stratification factors	Criterion
BMI	≥30
Swollen study knee	Present
Bilateral OA	K/L grade ≥ 2 in both knees from bilateral radiographs
Physically active as young adult	Positive answer ('yes') to the question "*When you were in your 20'ies did you participate in sports activities for at least 1 hour 2 times or more per week*"?

6.4 Blinding

This is an open-label trial. Participants and clinicians who deliver the interventions are not blinded.

Outcome assessors, and study personnel performing data queries and management, and the statistician will be blinded until all primary and secondary analyses are completed.

6.5 Sample Size and Power

This trial was designed as an equivalence trial. The sample size was calculated to test the equivalence of the GLAD concept versus IA saline injections in the assessment of change in KOOS pain from randomisation to the end of treatment, 9 weeks after first treatment.

In a two one-sided tests analysis for additive equivalence of two-sample normal means with equivalence bounds of -8 and 8 KOOS pain subscale points (0-100 scale) for the mean difference and a statistical significance level of 0.05, assuming a mean difference of 0 KOOS pain subscale points (0-100 scale) and a common standard deviation of 15 KOOS pain subscale points (0-100 scale), a total sample size of 154 assuming a balanced design is required to obtain a statistical power of 90.1%. To account for a possible dropout rate (of no more than 23%) 200 patients in total are required.

6.5.1 Statistical power calculation for potential superiority claim

A sample size of 200 in total will provide strong statistical power to detect group differences in favour of either of the two investigational treatments.

For a two-sample pooled t test of a normal mean difference with a two-sided significance level of 0.05 (P<0.05), assuming a common standard deviation of 15 KOOS-pain points, a total sample size of 200 assuming a balanced design has a power of 80.4% to detect a mean difference of 6 KOOS-Points (corresponding to a small effect size of 0.4).

6.6 Framework

This is an equivalence trial.

6.7 Statistical Interim Analyses and Stopping Guidance

No statistical interim analysis has been planned and there is no guidance for stopping the trial.

6.8 Timing of Final Analysis

Final analysis will take place in one stage: The first (and main) report/publication of the trial will be prepared for the GLAD/IA Saline comparison when every trial participant has completed the week 12 assessment and data for the primary and secondary outcomes have been received and cleaned (anticipated to be February 2021).

6.9 Timing of Outcome Assessments

The schedule of study procedures and visit windows are given in the Table 1. The start time for each calculation is the scheduled day of the participant's first treatment. Then, 9 and 12 weeks are added to determine the expected date for the week 9 and 12 assessment visits.

Table 1. Schedule of study procedures and visit windows

	Pre-screening	Screening	Baseline	Week 1	2	3	4	5	6	7	8	9	10 -11	12
Day		-42 to -14	-14 to 0	1-7	8-14	15-21	22-28	29-35	36-42	43-49	50-56	57-63	64-77	78-84
Written information	●													
Oral information	x	●												
Procedure														
Eligibility criteria	x	●												
Informed consent		●												
Radiograph		●												
Randomization			●											
Interventions														
GLA:D: Education†				●	●									
GLA:D: Exercise†						●●	●●	●●	●●	●●	●●			
IA Saline: Injection*				●		●		●		●				
Outcomes														
KOOS			●	●	●	●	●	●	●	●	●	●		●
Patient Global			●									●		●
OMERACT/OARSI												●		●
Morning pain			●	●	●	●	●	●	●	●	●	●		●
ICOAP			●									●		●
Analgesics use			●									●		
Performance tests			●									●		●
Swollen joint count			●									●		●
Joint Aspiration				(●*)		(●*)		(●*)		(●*)		(●)		(●)

x indicates that the procedure is only partly completed
Bullets in parentheses ( ) indicate 'if possible' and pertain to joint fluid aspiration that can only be done if excess joint fluid is present.
* IA Saline group only.
†GLA:D group only. The education and exercise sessions may be scheduled differently than illustrated.

7 OUTCOMES

7.1 Study knee

At inclusion a study knee must be selected, which will be subject to all subsequent assessment:

- The study knee will be defined as the symptomatic knee with a diagnosis of OA
- If both knees are eligible, the more symptomatic knee will be selected (selected by participant)
- If both knees have equivalent pain scores, the knee with the greater radiographic K/L grade will be chosen
- If the both the pain scores and K/L grades are equivalent, the study knee will be randomly assigned.

7.2 Primary outcome

The primary outcome is assessed at week 9 as change from baseline in the Knee injury and Osteoarthritis Outcome Score (KOOS) *pain* subscale – a widely used and well-validated survey instrument evaluating pain in OA. We will measure the difference in the changes from baseline in the KOOS pain subscale in the study knee between GLAD vs IA Saline after 9 weeks.

7.3 Key Secondary outcomes

The following outcomes are assessed as key secondary outcomes:

- Change from baseline in the *function in activities of daily living* subscale of the KOOS questionnaire at week 9
- Change from baseline in the *knee-related quality of life* subscale of the KOOS questionnaire at week 9
- Change from baseline in patient's global assessment (PGA) of impact of osteoarthritis at week 9

7.4 Other secondary outcomes

The following outcomes are assessed as other secondary outcomes:

- Change from baseline in the *symptoms* subscale of the KOOS questionnaire at week 9
- Change from baseline in the *function in sports and recreational activity* subscale of the KOOS questionnaire at week 9
- Number of treatment responders as per the OMERACT-OARSI response criteria at week 9
- Change from baseline in 4x10 meter fast walk test at week 9
- Change from baseline in the 30 seconds chair stand test at week 9
- Change from baseline in Stair climbing test at week 9
- All outcomes assessed at the week 12 assessment

7.4.1 Safety outcomes

- Presence of study knee effusion at week 9 and 12
- Study knee joint aspiration volume (ml) at week 9 and 12
- Number of swollen study knee joints count at week 9 and 12

7.4.2 Exploratory outcomes

- Change from baseline in average morning pain at week 9 and 12

- Change from baseline in the intermittent and constant osteoarthritis pain (ICOAP) score at week 9 and 12
- Number of participants who has discontinued paracetamol and/or ibuprofen use at week 9

7.5 Definition of outcome variables

7.5.1 Knee injury and Osteoarthritis Outcome Score (KOOS)

The Knee injury and Osteoarthritis Outcome Score (KOOS), a disease-specific instrument designed to assess health related quality of life (QoL) in patients with knee OA (26). The KOOS consists of 42 items covering five domains, namely, *Pain* (9 items), *Function* (in Activities of Daily Living) (17 items), *Knee-related QoL* (4 items), *Symptoms* (7 items), and *Sports and Recreation* (5 items). The KOOS uses a five-point Likert scale scoring system (ranging from 0 (least severe) to 4 (most severe)).

We will calculate the primary outcome KOOS pain, as well as the other KOOS domains, from the questionnaire values as outlined in the user guide (www.koos.nu[1]). Normalized scores are calculated for each domain with 100 indicating no symptoms and functional impairment and 0 indicating extreme symptoms and functional impairment. If the number of missing items is less than or equal to 2 in a subscale, they will be substituted by the average item value for that subscale. If more than two items of the subscale are omitted the response will be considered invalid and no subscale score calculated.

7.5.2 Patient's global assessment of impact of osteoarthritis (PGA)

The PGA is assessed using a 100 mm analogue scale (VAS) relating to the degree of the participants perceived impact of their knee OA on their overall life with anchors: 0= "*No impact*" and 100 = "*Worst imaginable impact*".

7.5.3 OMERACT-OARSI Response

OMERACT–OARSI have developed a set of standardized criteria to identify responders and non-responders in clinical trials in patients with OA of the knee. It is based on changes after treatment in three symptomatic domains (pain, function, and patient's global assessment) and reported as a single dichotomous variable.

[1] www.koos.nu accessed on September 23, 2020

7.5.4 4x10 meter fast walk test

The 4x10 meter fast walk test (40mFWT) is a physical performance test that quantifies short distance walking performance. Time of one trial, with turn time excluded, is recorded and expressed as speed in m/s.

7.5.5 30 seconds chair stand test

The 30 seconds chair stand test (30sCST) is a physical performance test that quantifies how many sit-to-stand movements an individual is able to perform within 30 seconds. The total number of complete chair stands is counted.

7.5.6 Stair climbing test

A stair climbing test (SCT) is a physical performance test that quantifies how fast an individual is able to ascend and descend a flight of stairs in a usual manner. Total time to ascend and descend is recorded in seconds.

7.5.7 Study knee effusion

During an ultrasound examination of the study knee, presence of excess joint fluid will be recorded binarily as present/absent.

7.5.8 Study knee joint aspiration volume (ml)

If possible, any excess joint fluid detected in the study knee during the ultrasound examination will be aspirated by inserting a needle into the joint cavity (under ultrasound guidance). The volume (in ml) of the aspirated fluid will recorded. If fluid is detected but not aspirated this value will be set to 0 ml.

7.5.9 Study knee swollen knee joint count

An investigator (medical doctor) will examine the study knee and record if it is swollen or not based on the presence of palpable effusion. The outcome of the examination will be recorded binarily (present/absent).

7.5.10 Morning pain

This is measured using a 100 mm analogue scale (VAS) relating to the degree of the patient's perceived averaged morning knee pain during the last week with anchors: 0 = "No pain" and 100 = "Worst imaginable pain".

7.5.11 Intermittent and constant osteoarthritis pain (ICOAP) score

The ICOAP is an OA-specific 11-item questionnaire designed to assess the pain experience within the last week among people suffering from knee and hip OA (21). The questionnaire is divided into two domains, a 5-item scale for constant pain and a 6-item scale for intermittent pain. Each domain captures pain intensity as well as related distress and the impact of OA pain on quality of life. All items are scored on anchored rating scales with five levels of response (0–4). A score is separately produced for the constant pain subscale (0–20) and the intermittent pain subscale (0–24), and for total pain (0–44). Normalized scores for the two subscales and for the total pain score, from 0 (no pain) to 100 (extreme pain), are calculated.

7.5.12 Use of paracetamol and ibuprofen

The participants' use of analgesics (paracetamol and NSAIDs) is recorded as during an interview with an investigator at baseline and at the primary endpoint; week 9. The analgesics use is dichotomised to either *'use analgesics'* (covering both regular use and PRN usage of paracetamol and/or NSAIDs) or *'do not use analgesics'* (meaning that no paracetamol and/or NSAIDs are used neither regularly nor PRN).

7.6 Adverse and serious adverse events

The investigators and clinical staff monitor each participant for evidence of adverse events (AEs) and serious adverse events (SAEs) throughout the study. The investigator will assess and record any AE and SAE in detail including the date of onset, description, severity, duration and outcome, relationship to study treatment, and any action(s) taken.

An investigator will adjudicate all reported AEs and SAEs based on available and relevant medical records.

8 DATA MANAGEMENT

8.1 Data validation

All variables used in the database, including derived variables, will be checked for missing values, outliers and inconsistencies. We do not expect many faulty data points because error checks and warnings were implemented into the eCRF and database system, and because the trial is monitored by an external monitoring committee (The Good Clinical Practice Unit at Copenhagen University Hospitals). The monitor has visited the trial site before trial commencement and regularly during the. The monitor checks trial procedures, including safety assessments, drug handling, data recording and complete source data verification procedures, and participant confidentiality.

8.2 Data preparation

8.2.1 Changes from baseline

The primary outcome is change from baseline in KOOS pain at week 9. This will be calculated for each individual as the baseline value subtracted from the week 9 value:

$$KOOSPAIN_{change\_week9} = KOOSPAIN_{week9} - KOOSPAIN_{baseline}$$

Thus, a positive change value indicate that the week 9 value is greater than the baseline value, which suggest an improvement in the KOOS pain score (= less pain).

The same calculation will be applied for all other outcomes defined as change from baseline at various time points in the trial:

$$VARIABLE_{change\_week\mathit{i}} = VARIABLE_{week\_\mathit{i}} - VARIABLE_{baseline}$$

The interpretation of calculated change values are as follows:

OUTCOME	INTERPRETATION OF POSITIVE CHANGE
KOOS Pain (primary outcome)	Improvement
KOOS other subscales	Improvement
PGA	Worsening
40m Fast Walk Test	Worsening
30s Chair Stand Test	Improvement
Stair climbing test	Worsening
Morning knee pain	Worsening
ICOAP, total and subscales	Worsening

For assessment of change from baseline in analgesics usage, we assign each participant as either '*user'* or '*non-user'* at the baseline and week 9 visits. To assess change we will assign each participant with one of the following values:

-1 = Have discontinued analgesics at week 9 relative to baseline
0 = No change in analgesics use at week 9 relative to baseline
+1 = Have started using analgesics at week 9 relative to baseline

8.2.2 Composite and calculated outcomes

The secondary outcome, number of OMERACT-OARSI responders, is based on a calculation of positive response for each individual at the 9- and 12-week assessments. A positive OMERACT-OARSI response is calculated as follows:

- In either pain (KOOS pain subscale) or function (KOOS function subscale) a > 50% improvement from Baseline and an absolute change from Baseline of > 20 normalized units (0-100 scale)

OR

- Improvement in at least two of the following three:
    1. Improvement in the KOOS pain subscale defined as > 20% improvement from Baseline and an absolute change from Baseline of > 10 normalised KOOS pain points (0-100 scale)
    2. Improvement in the KOOS function subscale defined as > 20% improvement from Baseline and an absolute change from Baseline of > 10 normalised KOOS function points (0-100 scale)
    3. Improvement in PGA defined as > 20% improvement from Baseline and an absolute change from Baseline of > 10 mm (0-100 scale)

9 TRIAL POPULATIONS

9.1 Participant flow

A CONSORT participant flow diagram will be drawn following the CONSORT standards (see Shell Figure 1).

The flow diagram will be used to summarise the number of patients who were:

- assessed for eligibility at screening
- ineligible at screening*
- eligible and randomised
- eligible but not randomised*
- allocated to each intervention
- received the randomly allocated intervention
- did not receive the randomised allocation*
- discontinued the intervention*
- lost to follow-up at week 9 and 12*
- randomised and included in the primary analysis (intention-to-treat population)
- randomised and included in secondary analysis (per-protocol population)

*reasons will be provided.

9.2 Intention-To-Treat population

The Intention-To-Treat (ITT) population consist of all randomised patients irrespective of whether the patient actually received study intervention or the patient's compliance with the study protocol, in the treatment group to which the participant was assigned at randomisation (i.e. referring to the ITT principle). A patient will be considered randomised as soon as a treatment is assigned according to the allocation sequence (i.e. breaking the allocation concealment for an enrolled individual).

The participant demographics and baseline data for the ITT population will be summarised in a table (shell Table 1). Participants will be described with respect to baseline age, sex, height, body mass, body mass index, radiographic disease severity (K/L grade), stratification factors and baseline values of primary, secondary, safety and exploratory outcomes, separately for the two groups.

Continuous data will be summarised by means and SDs. Categorical data will be summarised by numbers and percentages. Tests of statistical significance will not be undertaken for baseline characteristics; rather the clinical importance of any imbalance will be noted.

9.3 Per Protocol Population

The per protocol (PP) population consists of all participants in the ITT population who did not have any major protocol deviations that could make the interpretation of analyses on the ITT population difficult.

The following are pre-defined major protocol violations with a direct bearing on the primary outcome:

- Not adherent to the allocated intervention (see below for definition of adherence) unless non-adherence is due to adverse reactions⁂ to study treatments
- Initiation of opioids during trial participation, except for treatment of adverse reactions⁂ to study treatments
- Initiation of oral glucocorticoids during trial participation, except for treatment of adverse reactions⁂ to study treatments
- Intra-articular injections in the lower extremity (other than described in the protocol) during trial participation, except for treatment of adverse reactions⁂ to study treatments
- Non-pharmacological treatments for lower extremity OA (ankle, hip, non-study knee), except those described in this protocol and for treatment of adverse reactions⁂ to study treatments
- Surgery to the lower extremity during trial participation, except for treatment of adverse reactions⁂ to study treatments
- Failure to perform primary endpoint assessment, i.e. KOOS pain not assessed at week 9, except for failure due to adverse reactions⁂ to study treatments
- Early discontinuation of trial participation (before week 3), except for discontinuation due to adverse reactions⁂ to study treatments
- Week 9 visit not completed within +/- 7 days of the specified time window
- Non-compliance with any of the eligibility criteria

⁂ an adverse reaction is defined as an adverse event that is deemed related to the study treatment by an investigator.

The number (and percentage) of patients with major protocol deviations will be summarised by treatment group with details of type of deviation provided. The number of randomised participants in each group will be used as the denominator to calculate the percentages. No formal statistical testing will be undertaken.

9.4 Adherence

For each trial participant, compliance is assessed based on the percent of the scheduled number of treatments that was received. The number of scheduled treatments is predefined in the trial protocol.

- For the GLAD group, the scheduled number of treatments are 2 educational sessions and 12 exercise sessions; 14 sessions in total.
- For the IA saline group, the scheduled number of treatments are 1 injection every two weeks; 4 injections in total.

The following pre-defined (see protocol version 1.5 page 32) criteria for treatment adherence have been set:

- For the GLAD group: Have attended at least 1 of the educational sessions AND have attended at least 8/12 of the scheduled exercise appointments; 9 attendances in total (75%).
- For the IA saline group: Have received at least 3 of the 4 scheduled injections (75%).

Descriptive statistics on the percent compliance (Mean, SD) will be summarized by randomisation group. Also, the number and % of participants receiving at least 75% of the prescribed treatment will be presented by treatment group.

9.5 Safety population

The safety population consists of all participants in the ITT population who has received at least 1 IA Saline injection (IA Saline group) or attended at least 1 exercise session (GLAD group).

10 STATISTICAL ANALYSES

10.1 General considerations

In the primary analysis, all participants will be analysed using the ITT population according to the intention-to-treat principle. All 95% confidence intervals will be two sided. We will not apply explicit adjustments for multiplicity, rather we will analyse the key secondary outcomes in a prioritized order (i.e. “inverse gatekeeping procedure”): The analyses of the secondary and exploratory outcomes will be performed in sequence until one of the analyses fails to show equivalence.

The hierarchy of the key secondary outcomes including equivalence margins are as follows:

1. KOOS function in activities of daily living
2. KOOS knee related quality of life
3. Patient’s global assessment (PGA)

All other secondary outcome will be analysed, i.e. no hierarchy applied.

All safety outcomes will be analysed, i.e. no hierarchy applied.

All exploratory outcomes will be analysed, i.e. no hierarchy applied.

The statistician will be blinded to the treatment allocation at the time of the primary analysis of primary and secondary outcomes. Once the primary analysis is accomplished, the statistician may be unblinded.

10.2 Equivalence margins

As this is an equivalence trial, the following equivalence margins has been set prior to the analyses. *Equivalence* will be claimed if the computed 95% confidence interval of the estimated group difference in an outcome does not include the below equivalence margins.

OUTCOME	EQUIVALENCE MARGINS
Primary outcome	
KOOS pain	± 8 points (www.koos.nu[1])
Key Secondary outcomes	
KOOS function	± 8 points (www.koos.nu[1])
KOOS quality of life	± 8 points (www.koos.nu[1])
Patient's global assessment (PGA)	± 15 mm (1)
Other secondary outcomes	
OMERACT-OARSI response	± 15%
4x10 meter fast walk test	± 0.2 m/s (2)
30 seconds chair stand test	± 2.0 stands (2)
Stair climbing test	± 5.2 s* (2)
KOOS function in sports & recreation	± 8 points (www.koos.nu[1])
KOOS symptoms	± 8 points (www.koos.nu[1])
Safety outcomes	
Knee effusion	No equivalence margins set
Aspiration volume	No equivalence margins set
Swollen knee joint count	No equivalence margins set
Pain at treatment visits	No equivalence margins set
Exploratory outcomes	
Average morning pain (VAS)	± 15 mm (1)
ICOAP Total	± 18.5 points (3)
ICOAP Constant	± 18.7 points (3)
ICOAP Intermittent	± 18.4 points (3)

[1] www.koos.nu accessed on September 23 2020.
* No validated margins (minimal clinical important difference) exist for knee OA, so for this study we set the margins at 2×minimal detectable change.

10.3 Missing Data and Robustness

Our primary (efficacy) analyses will be based on the ITT population, including all randomised participants with available data at baseline. For continuous outcomes (incl. both the primary and key secondary outcome measures), missing data will be handled indirectly by statistically using Mixed Linear models based on the repeated-measurements framework. These models are generally considered valid if data are '*Missing at Random*' (MAR); i.e. where "*Any systematic difference between the missing values and the observed values can be explained by differences in observed data*" (4). Contrasts between groups will be estimated based on repeated-measures mixed linear models across all timepoints (i.e., with explicit estimates derived at 9 and 12 weeks from baseline, respectively).

Robustness is a concept that refers to the sensitivity of the overall conclusions to various limitations of the data, assumptions, and analytic approaches to data analysis. Robustness implies that the treatment effect and primary conclusions of the trial are not substantially affected when analyses are carried out based on alternative assumptions or analytic approaches.

Loss to follow-up and missing data for various reasons is difficult to avoid in randomized trials and in particular in pragmatic trials. We will apply the analysis framework suggested by White et al (2011) in which missing data related to the ITT approach depend on making plausible assumptions about the missingness of the data and including all participants in subsequent sensitivity analyses (5); we will:

1. Attempt to follow up all randomized participants, even if they withdraw from allocated treatment (i.e., contact all individuals unless they explicitly stated that they had withdrawn their consent)
2. Perform a main analysis of all observed data that are valid under a plausible assumption about the missingness of the data (i.e., Model-based: data as observed; using linear mixed models, assuming that data are 'Missing at Random' [MAR])
3. Perform sensitivity analyses to explore the effect of departures from the assumption made in the main (#2) analysis (i.e., conservative imputation methods of missing data; these models will potentially be informative even if data are 'Missing Not At Random' [MNAR])
4. Account for all randomized participants, at least in the main and sensitivity analyses (covered by #2 and #3 above, plus the corresponding analyses based on the per protocol population).

For sensitivity analyses (see section 10.6 below) we will use multiple imputation techniques for repeated replacement of missing data at week 9. Each outcome variable will be imputed separately.

We will use baseline variables, outcome measures at week 9, and group allocation as predictors in the imputation models.

The primary analysis will be repeated based on participants in the ITT population, but by imputing missing postbaseline observations of the primary and key secondary outcomes at the week 9 visit (where the outcome was scheduled to be measured) using a multiple imputation procedure. The imputation will be performed according to the following steps:

1. Missing values are imputed based on observed data using a Markov Chain Monte Carlo method where 5 copies of the dataset will be generated;
2. For each of the 5 copies, missing values at the week 9 visit will be analysed using an ANCOVA model including treatment group, and stratification factors (BMI, Swollen study knee, Bilateral tibiofemoral OA, and Active as young adult) as fixed factors and the baseline level as covariates;
3. From this repeated standard ANCOVA model, estimated differences between groups in each of the imputed datasets will differ (at least slightly) because of the variation introduced in the imputation of the missing values, and they are only useful when averaged together to give overall estimated associations. The corresponding standard errors will be calculated using Rubin's rules (6), which take account of the variability in results between the imputed datasets, reflecting the uncertainty associated with the missing values. Valid inferences are obtained because we will be averaging over the distribution of the missing data given the observed data.

10.4 Missing data due to COVID-19 pandemic trial suspension

From March12 to April 20, 2020 all DISCO trial activities (screening visits, baseline assessments, intervention delivery, and all clinical outcome assessments) were suspended as an urgent safety measure due to the lockdown of all non-critical activities in the public sector in Denmark was decreed by the Danish government because of the COVID-19 pandemic.

For the participants who had begun the investigational treatments in the DISCO trial at the time of suspension (GLAD n=13; IA Saline n=9) patient reported outcome measures at week 9 and 12 (including primary outcome) were collected via telephone interviews at the scheduled time points. The outcomes that require a clinical visit have not been collected and are defined as *missing completely at random* ('MAR') for these participants.

10.5 Primary analysis

Our primary analysis population will be all participants with available data at baseline, statistically modelled using repeated-measures linear mixed models (see above). These models will be valid if data are 'MAR'.

The primary analyses will be conducted according to the ITT principle. The ITT principle asserts the effect of a treatment policy (that is, the planned treatment regimen), rather than the actual treatment given (i.e., it is independent of treatment adherence). Accordingly, participants allocated to a treatment group will be followed up, assessed and analysed as members of that group, irrespective of their adherence to the planned course of treatment (i.e., independent of withdrawals and cross-over phenomena). Primary and secondary outcomes will be assessed using mixed linear models adjusted for baseline values and stratification factors.

10.5.1 Primary analysis of primary outcome

The primary outcome analysis will be an *equivalence analysis* based on the ITT population, asking whether the GLAD and IA Saline treatments are equivalent regarding change from baseline in KOOS Pain scores at the end of the treatment period (week 9). We will use a repeated measures linear mixed model regression analysis model adjusted for stratification factors and the baseline score of the KOOS Pain. An interaction for time and group will be included.

$$\text{KOOSPAIN}_{\text{change}} \approx \text{GROUP} + \text{WEEK} + \text{GROUP} \times \text{WEEK} + \text{Stratification}_i + \text{KOOSPAIN}_{\text{baseline}}$$

Analyses will include all collected data, and effects will be estimated at week 9 and 12; missing data will be handled implicitly via the mixed methods (maximum likelihood) approach. From this model the observed differences in the change from baseline in KOOS pain between GLAD and IA Saline at week 9 will be estimated together with the associated 95% confidence interval (and the p-value) corresponding to the test of the hypothesis of no difference between treatments. The result of the primary analysis of the primary outcome will be presented in a table (shell Table 2) and in a figure (shell Figure 2).

Equivalence (i.e. a significant similarity between groups) will be claimed if the 95% confidence interval exclude differences greater than what is typically being interpreted as being a clinically meaningful difference; e.g., if the computed 95% confidence interval of the estimated difference between groups in the change from baseline in the KOOS pain at week 9 exclude differences greater than (±) 8 units between groups will be interpreted as indicating the absence of a clinically meaningful difference.

Superiority on the other hand, will potentially be claimed if the computed 95% confidence interval of the estimated group difference in the change from baseline in the KOOS pain at week 9 exclude the null.

10.5.2 Primary analysis of secondary, safety and exploratory outcomes

The primary analyses of the secondary, safety, and exploratory outcomes, will be *equivalence analyses* using the ITT population. Missing data will not be imputed but handled in mixed linear models. Continuous secondary, safety and exploratory outcomes will be analysed identically to the primary outcome and adjusted for the stratification factors and the respective baseline value if available. We will compute differences with unadjusted two-sided 95% confidence intervals interpreted based on the equivalence paradigm.

Dichotomous outcomes and handling of missing data: Categorical changes for dichotomous outcomes (i.e. the proportion of OMERACT-OARSI responders, Presence of study knee effusion, and number of swollen study knee joints count, Analgesics use) will be analysed with the use of logistic regression with the same fixed effects as the respective mixed linear models explained above; however, the models will only include data from the before-and-after setting (i.e. without a repeated measures analysis element) at week 9 and week 12, respectively. For ease of interpretation the corresponding Odds Ratios will be converted back into Adjusted Risk Ratios and/or Risk Differences based on the number of responders in the IA Saline group (7).

Handling of missing data for the dichotomous outcomes: We will use a fixed-set multiple imputation framework for our missing dichotomous "efficacy outcomes" (8). For this binary outcome, we will a use a methodology that imputes the extreme displays that reveal the effects of all outcomes for each randomised individual, using combinations of the values of missing data in the first arm (GLAD group) and the second arm (IA saline group); the imputation technique is based on the idea of 'tipping-point' (TP) analysis (9). We enhance this idea by formalizing the process of robust estimation using a more detailed display in conjunction with multiple imputation (MI) of missing data. We will semi-automatically generate **5** individual data sets based on the following approach:

- Data as observed (including missing data)
- Worst ($Y_{Mis} = 0$) *AND* Worst ($Y_{Mis} = 0$) case imputation in each group, respectively
- Best ($Y_{Mis} = 1$) *AND* Best ($Y_{Mis} = 1$) case imputation in each group, respectively
- Worst ($Y_{Mis} = 0$) *AND* Best ($Y_{Mis} = 1$) case imputation in each group, respectively
- Best ($Y_{Mis} = 1$) *AND* Worst ($Y_{Mis} = 0$) case imputation in each group, respectively.

From these 4 simulated (complete) data sets as well as the original data set (with missing data), 5 different sets of the point and variance estimates will be computed. Using Rubin's rules (6), which take account of the variability in results between the imputed datasets, reflecting the uncertainty associated with the missing values, we will combine these results and generate valid and robust statistical inference about the multiply imputed Odds Ratio, as well as the proportions responding in each group.

The results of the primary analysis of the secondary, safety and exploratory outcomes will be presented in tables (shell Table 2 for week 9 and Appendix shell Table x1 for week 12).

10.6 Secondary analyses

The interpretation of the corresponding statistical measures of uncertainty of the treatment effect and treatment comparisons will involve consideration of the potential contribution of bias to the 95% confidence intervals and of the inference in general. Therefore, we will perform sensitivity analyses for the primary and key secondary outcomes.

First, we will repeat the primary analyses on the Per Protocol population that includes only participants who adhered to the allocated treatment without major protocol violations as defined above (section 9.3). This analysis will be conducted without imputation of missing data.

Secondly, we will perform an analysis of covariance of the primary and key secondary outcomes at week 9 on the ITT population (defined in section 9.2), with multiple imputation of missing data at week 9 (see section 10.3) adjusted for stratification factors and the baseline values

$$\text{VARIABLE}_{\text{change\_week9}} \approx \text{GROUP} + \text{Stratification}_i + \text{VARIABLE}_{\text{baseline}}$$

Thirdly, we will repeat the analysis of covariance of the primary and key secondary outcomes at week 9 on the ITT population (defined in section 9.2), with a baseline observation carried forward imputation of missing data at week 9 (see section 10.3) adjusted for stratification factors and the baseline values

$$\text{VARIABLE}_{\text{change\_week9}} \approx \text{GROUP} + \text{Stratification}_i + \text{VARIABLE}_{\text{baseline}}$$

If the sensitivity analyses are in agreement, and the sensitivity analyses and the primary analysis lead to essentially the same conclusions, confidence in the results is increased.

The result of the secondary analyses will be presented in supplementary tables (shell Tables x2-4).

10.7 Assessment of statistical assumptions

For the linear models of the primary, secondary, safety and exploratory outcomes, we will check for the normality of residuals by visual inspection of residual plots.

10.8 Statistical Software

The analysis will be carried out using the statistical software SAS version 9.4 (SAS Institute Inc., Cary, NC, USA). Linear mixed-effect models will be fitted using the MIXED procedure (proc mixed).

10.9 Harms

Analyses of AEs and SAEs will be performed on the Safety Population (see section 9.5).

The number (and percentage) of patients experiencing AEs and SAEs will be presented for each treatment arm categorised by severity. For each patient, only the maximum severity experienced of each type of AE/SAE will be displayed.

AEs and SAEs will be assessed for relationship with the trial treatment and the number (and percentage) of related AE/SAE will be presented for each treatment arm.

Deaths and AEs/SAEs leading to discontinuation of study treatment will be listed.

No formal statistical testing will be undertaken.

The AEs/SAEs will be presented in a table (Shell Table 3)

10.10 Timing of analyses

When this statistical analysis plan was signed, recruitment to the DISCO trial had not been completed. We expect recruitment to be completed by the end of September 2020. We will close the database 2 months after the last participant's last visit at the latest. Statistical analyses are expected to be completed after additionally 2 months.

11 DEVIATIONS FROM THE PROTOCOL

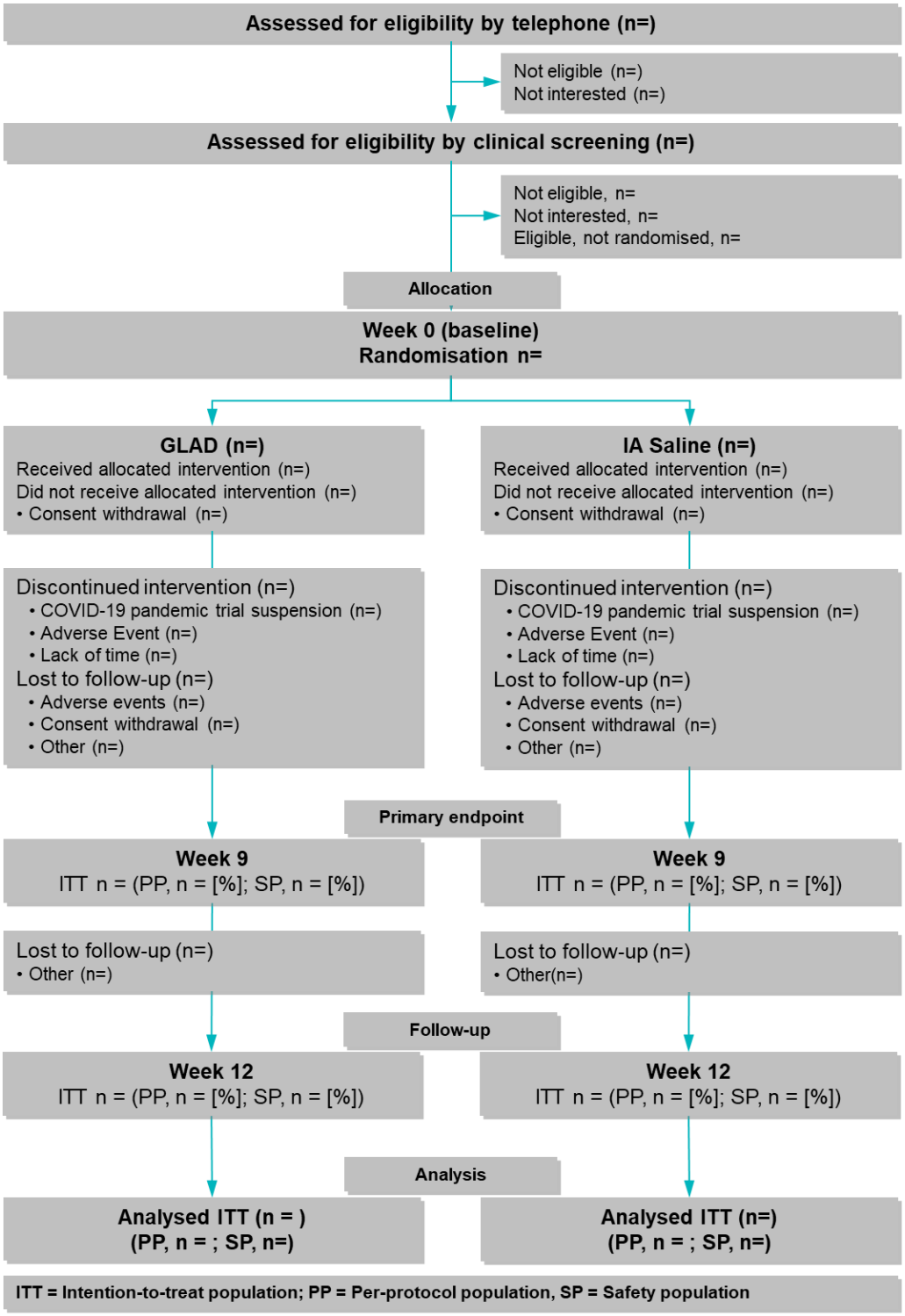
The following details in this SAP represents deviations from trial protocol version 1.5

Header in protocol	Deviation	Reason
9.2 Secondary outcomes	Secondary outcomes divided into 'Key Secondary Outcomes' and 'Other Secondary Outcomes'	Together, the Primary and Key Secondary Outcomes represent the OMERACT-OARSI Core Domain Set for Measurement in Clinical Trials of Hip and/or Knee Osteoarthritis and thus considered most important.
10.7 Swollen joint count	Changed from 'both knees' to 'study knee'	Non-study knee not assessed in the DISCO trial.
	Changed from 'Change in swollen knee joint count' to 'Study knee swollen knee joint count'	
10.12 Patient reported paracetamol and ibuprofen use	Use is recorded via interview with investigator at baseline and week 9 rather than patient-report.	Part of routine medical review interview in the outpatient clinic.
	Use of analgesics (paracetamol and NSAIDs) is dichotomized at baseline and week 9 rather than actual dosages.	The investigators reported to the steering committee that for many participants there were no dosages recorded as the participants use analgesics on a PRN basis. This precludes meaningful analyses of changes in dosages in this trial.
	Participant analgesics diaries are not used for analyses	A blinded review of a sample of diaries revealed inconsistent data not useful for statistical analyses.
14.4.1 Per protocol population	Defined in more detail compared to the protocol	Necessary to be operational.
14.5 General statistical approach	Changed from a repeated measures ANCOVA model to repeated measures mixed effects linear model	It will yield the same results, but mixed model handles missing data implicitly in a repeated measures trial design.
	Changed analysis population in equivalence claim from PP population to ITT	We suspect the PP population to be a biased population

12 MANUSCRIPT OUTLINE

12.1 Shell Figure 1 (Flow Diagram)

Figure X: CONSORT flow diagram



12.2 Shell Table 1 (Baseline)

Table X: Demographics and Baseline Characteristics

	GLAD	IA Saline
	n=	n=
Demographics		
Age, years	xx.x (xx.x)	xx.x (xx.x)
Female sex, n[%]	xx (xx.x%)	xx (xx.x%)
Body mass, kg	xx.x (xx.x)	xx.x (xx.x)
Height, m	xx.x (xx.x)	xx.x (xx.x)
Body Mass Index, kg/m2	xx.x (xx.x)	xx.x (xx.x)
Radiographic disease severity (K/L grade) §		
2, n[%]	xx (xx.x%)	xx (xx.x%)
3, n[%]	xx (xx.x%)	xx (xx.x%)
4, n[%]	xx (xx.x%)	xx (xx.x%)
Stratification factors		
Body Mass Index ≥30, n[%]	xx (xx.x%)	xx (xx.x%)
Swollen study knee, n[%]	xx (xx.x%)	xx (xx.x%)
Bilateral tibiofemoral OA (K/L≥2), n[%]	xx (xx.x%)	xx (xx.x%)
Active as young adult, n[%]	xx (xx.x%)	xx (xx.x%)
KOOS scores, 0-100 ∫		
Pain*	xx.x (xx.x)	xx.x (xx.x)
Physical Function in activities of daily living†	xx.x (xx.x)	xx.x (xx.x)
Quality of Life†	xx.x (xx.x)	xx.x (xx.x)
Symptoms	xx.x (xx.x)	xx.x (xx.x)
Physical Function in Sports & Recreation	xx.x (xx.x)	xx.x (xx.x)
Patient Global Assessment, VAS 0-100†	xx.x (xx.x)	xx.x (xx.x)
Morning pain, VAS 0-100	xx.x (xx.x)	xx.x (xx.x)
ICOAP scores, 0-100		
Constant	xx.x (xx.x)	xx.x (xx.x)
Intermittent	xx.x (xx.x)	xx.x (xx.x)
Total	xx.x (xx.x)	xx.x (xx.x)
Performance tests		
4x10 meter fast walk test, m/s	xx.x (xx.x)	xx.x (xx.x)
30 seconds chair stand test, reps	xx.x (xx.x)	xx.x (xx.x)
Stair climbing test, seconds	xx.x (xx.x)	xx.x (xx.x)
Clinical assessment		
Study knee effusion – ultrasound, n[%]	xx (xx.x%)	xx (xx.x%)
Analgesics use		
Paracetamol or Ibuprofen user, n[%]	xx (xx.x%)	xx (xx.x%)

VAS: Visual Analogue Scale.
§ Scores on the Kellgren–Lawrence scale range from 0 to 4, with a score of 2, 3, or 4 indicating definite osteoarthritis and higher scores indicating more severe disease.
∫ Scores on the Knee Injury and Osteoarthritis Outcome Score (KOOS) subscales range from 0 (worst) to 100 (best).
* Primary outcome measure; † Key secondary outcome measures

12.3 Shell Table 2 (Primary analysis week 9)

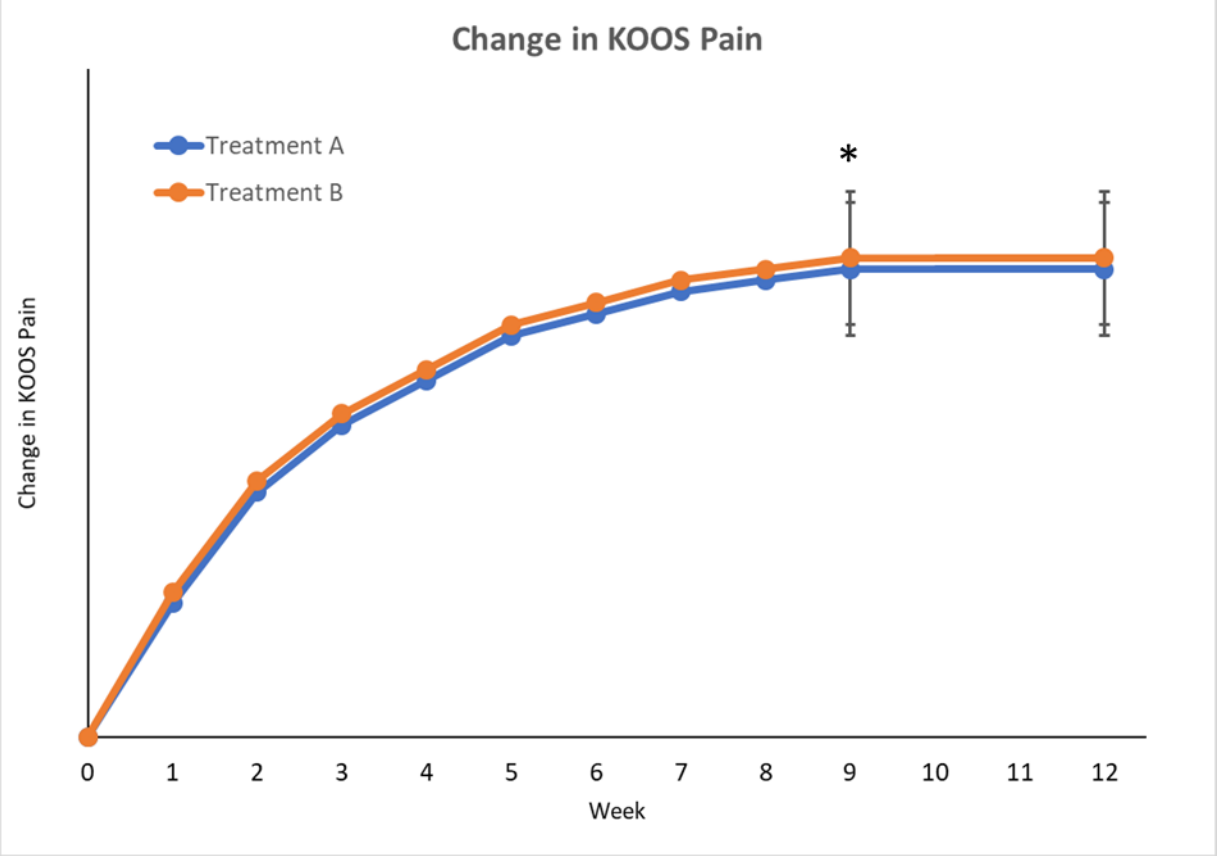
Table X: Change from baseline in Primary and Secondary Outcomes at week 9 in the ITT population. CI denotes 95% confidence interval. Based on repeated measures mixed linear models, where missing data is modelled implicitly.

	GLAD (N=)	IA Saline (N=)	Estimated treatment difference	P-value
	LSMean (SE)	LSMean (SE)	ΔLSMean (95% CI)	
Primary outcome:				
Change in KOOS Pain score; equivalence test*	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	0.xxx*
Change in KOOS Pain score; superiority test*				0.xxx*
Key Secondary outcomes:				
Change in KOOS Function score	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in KOOS Quality of life score	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in PGA – VAS (mm)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Other Secondary Outcomes:				
Change in KOOS Sports and recreation– score	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in KOOS Symptoms – score	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
OMERACT-OARSI responders - no. (%)†	xx (xx.x %)	xx (xx.x%)	xx.x (xx.x to xx.x)	
Change in 4x10 meter fast walk test (m/s)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in 30 seconds chair stand test (number)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in Stair climbing test (seconds)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Safety outcomes:				
Swollen study knee, clinical - no. (%)†	xx (xx.x %)	xx (xx.x%)	xx.x (xx.x to xx.x)	
Study knee effusion, ultrasound - no. (%)†	xx (xx.x %)	xx (xx.x%)	xx.x (xx.x to xx.x)	
Study knee aspiration volume (ml)⁂	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Exploratory outcomes:				
Change in average morning pain – VAS (mm)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in ICOAP Total score	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in ICOAP Constant Pain subscore	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in ICOAP Intermittent Pain subscore	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Paracetamol and Ibuprofen discontinued - no. (%)†	xx (xx.x %)	xx (xx.x%)	xx.x (xx.x to xx.x)	
Treatment adherence:				
Treatment adherence (%)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Treatment adherence ≥75% - no. (%)†	xx (xx.x %)	xx (xx.x%)	xx.x (xx.x to xx.x)	

Values are least squares means ± standard error unless otherwise stated.
* Primary outcome will be analysed using both a test for equivalence and a test for superiority.
† Missing data in binary outcomes (after 9 weeks) will be handed using an extreme-set multiple imputation technique followed by applying Rubin's rule to both the observed and 4 extreme case scenarios (*i*: Data as observed; *ii*; Worst-Worst case; *iii*: Worst-Best case; *iv*: Best-Worst case, and *v*: Best-Best case scenario).
⁂ Aspiration only performed in case of effusion detected on ultrasound
KOOS: Knee injury and osteoarthritis outcome score.
PGA: Patient Global Assessment
VAS: Visual Analog Scale

12.4 Shell Figure 2 (KOOS Pain trajectories)

Figure X: Exemplar (hypothetical) trajectories for our primary efficacy outcome measure (i.e. primary endpoint, change from baseline in KOOS pain) in the ITT population. Based on repeated measures mixed linear models, where missing data is modelled implicitly. Error bars indicates standard error of the estimate. The asterisk indicates the time of primary outcome assessment (week 9).



12.5 Shell Table 3 (Adverse events)

Table X. Adverse events in the safety population defined as participants in the ITT population who has received at least one IA Saline injection (IA Saline group) or attended at least 1 exercise session (GLAD group).

	GLAD (n=)	IA Saline (n=)
Exposure time – patient weeks	xx.x	xx.x
AE - no. of patients (%)	xx (xx%)	xx (xx%)
AE - no. of events	xx (xx%)	xx (xx%)
AEs leading to discontinuation - no. of patients (%)	xx (xx%)	xx (xx%)
Max. reported severity of AEs*		
Mild - no. of patients (%)	xx (xx%)	xx (xx%)
Moderate - no. of patients (%)	xx (xx%)	xx (xx%)
Severe - no. of patients (%)	xx (xx%)	xx (xx%)
AEs, relationship to trial treatment		
Not related - no. of events (%)	xx (xx%)	xx (xx%)
Probably not related - no. of events (%)	xx (xx%)	xx (xx%)
Probably related - no. of events (%)	xx (xx%)	xx (xx%)
AEs, classification		
Infections & infestations- no. of events (%)	xx (xx%)	xx (xx%)
General and administrative site conditions- no. of events (%)	xx (xx%)	xx (xx%)
Musculoskeletal and connective tissue disorders- no. of events (%)	xx (xx%)	xx (xx%)
Skin and subcutaneous tissue disorders- no. of events (%)	xx (xx%)	xx (xx%)
Injury, poisoning, and procedural complications- no. of events (%)	xx (xx%)	xx (xx%)
SAE - no. of patients (%)	xx (xx%)	xx (xx%)
SAE - no. of events (%)	xx (xx%)	xx (xx%)
SAEs leading to discontinuation - no. of patients (%)	xx (xx%)	xx (xx%)
SAEs, relationship to trial treatment		
Not related - no. of events (%)	xx (xx%)	xx (xx%)
Probably not related - no. of events (%)	xx (xx%)	xx (xx%)
Probably related - no. of events (%)	xx (xx%)	xx (xx%)
Deaths - no. of events (%)	xx (xx%)	xx (xx%)

12.6 Appendix Shell Table x1 (Primary analysis week 12)

Table X: Changes from baseline in Primary and Secondary Outcomes at week 12 in the ITT population. CI denotes 95% confidence interval. Based on repeated measures mixed linear models, where missing data is modelled implicitly.

	GLAD (N=)	IA Saline (N=)	Estimated treatment difference	P-value
	LSMean (SE)	LSMean (SE)	ΔLSMean (95% CI)	
Primary outcome:				
Change in KOOS Pain score; equivalence test*	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	0.xxx*
Change in KOOS Pain score; superiority test*				0.xxx*
Key Secondary outcomes:				
Change in KOOS Function score	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in KOOS Quality of life score	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in PGA – VAS (mm)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Other Secondary Outcomes:				
Change in KOOS Sports and recreation– score	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in KOOS Symptoms – score	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
OMERACT-OARSI responders - no. (%)†	xx (xx.x %)	xx (xx.x%)	xx.x (xx.x to xx.x)	
Change in 4x10 meter fast walk test (m/s)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in 30 seconds chair stand test (number)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in Stair climbing test (seconds)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Safety outcomes:				
Swollen study knee, clinical - no. (%)†	xx (xx.x %)	xx (xx.x%)	xx.x (xx.x to xx.x)	
Study knee effusion, ultrasound - no. (%)†	xx (xx.x %)	xx (xx.x%)	xx.x (xx.x to xx.x)	
Study knee aspiration volume (ml)⁂	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Exploratory outcomes:				
Change in average morning pain – VAS(mm)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in ICOAP Total score	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in ICOAP Constant Pain subscore	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in ICOAP Intermittent Pain subscore	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	

Values are least squares means ± standard error unless otherwise stated.
* Primary outcome will be analysed using both a test for equivalence and a test for superiority.
† Missing data in binary outcomes (after 9 weeks) will be handed using an extreme-set multiple imputation technique followed by applying Rubin's rule to both the observed and 4 extreme case scenarios (*i*: Data as observed; *ii*; Worst-Worst case; *iii*: Worst-Best case; *iv*: Best-Worst case, and *v*: Best-Best case scenario).
⁂ Aspiration only performed in case of effusion detected on ultrasound
KOOS: Knee injury and osteoarthritis outcome score.
PGA: Patient Global Assessment
VAS: Visual Analog Scale

12.7 Appendix Shell Table x2 (week 9 PP)

Table X: Sensitivity analyses of the changes from baseline in primary, secondary, safety and exploratory outcomes at week 9 in the PP population. CI denotes 95% confidence interval. Based on repeated measures mixed linear models, where missing data is modelled implicitly.

	GLAD (N=)	IA Saline (N=)	Estimated treatment difference	P-value
	LSMean (SE)	LSMean (SE)	ΔLSMean (95% CI)	
Primary outcome:				
Change in KOOS Pain score; equivalence test*	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	0.xxx*
Change in KOOS Pain score; superiority test*				0.xxx*
Key Secondary outcomes:				
Change in KOOS Function score	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in KOOS Quality of life score	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in PGA – VAS (mm)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Other Secondary Outcomes:				
Change in KOOS Sports and recreation– score	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in KOOS Symptoms – score	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
OMERACT-OARSI responders - no. (%)†	xx (xx.x %)	xx (xx.x%)	xx.x (xx.x to xx.x)	
Change in 4x10 meter fast walk test (m/s)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in 30 seconds chair stand test (number)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in Stair climbing test (seconds)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Safety outcomes:				
Swollen study knee, clinical - no. (%)†	xx (xx.x %)	xx (xx.x%)	xx.x (xx.x to xx.x)	
Study knee effusion, ultrasound - no. (%)†	xx (xx.x %)	xx (xx.x%)	xx.x (xx.x to xx.x)	
Study knee aspiration volume (ml)⁂	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Exploratory outcomes:				
Change in average morning pain – VAS(mm)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in ICOAP Total score	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in ICOAP Constant Pain subscore	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in ICOAP Intermittent Pain subscore	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Paracetamol and Ibuprofen discontinued - no. (%)†	xx (xx.x %)	xx (xx.x%)	xx.x (xx.x to xx.x)	
Treatment adherence:				
Treatment adherence (%)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Treatment adherence ≥75% - no. (%)†	xx (xx.x %)	xx (xx.x%)	xx.x (xx.x to xx.x)	

Values are least squares means ± standard error unless otherwise stated.
* Primary outcome will be analysed using both a test for equivalence and a test for superiority.
† Missing data in binary outcomes (after 9 weeks) will be handed using an extreme-set multiple imputation technique followed by applying Rubin's rule to both the observed and 4 extreme case scenarios (*i*: Data as observed; *ii*; Worst-Worst case; *iii*: Worst-Best case; *iv*: Best-Worst case, and *v*: Best-Best case scenario).
⁂ Aspiration only performed in case of effusion detected on ultrasound
KOOS: Knee injury and osteoarthritis outcome score.
PGA: Patient Global Assessment
VAS: Visual Analog Scale

12.8 Appendix Shell Table x3 (Week 9 MI)

Table X: Sensitivity analyses of changes from baseline in the primary and key secondary outcomes at week 9 in the ITT population. CI denotes 95% confidence interval. Based on analysis of covariance, where missing data is replaced using multiple imputation.

	GLAD (N=)	IA Saline (N=)	Estimated treatment difference	P-value
	Mean (SE)	Mean (SE)	Mean (95% CI)	
Primary outcome:				
Change in KOOS Pain – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	0.xxx*
Key Secondary outcomes:				
Change in KOOS Function – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in KOOS Quality of life – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in PGA – VAS (0 to 100 mm)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	

Values are least squares means ± standard error unless otherwise stated.
*Primary outcome will be analyzed using both a test for equivalence and a test for superiority.
KOOS: Knee injury and osteoarthritis outcome score.
PGA: Patient Global Assessment
VAS: Visual Analog Scale

12.9 Appendix Shell Table x4 (Week 9 BOCF)

Table X: Sensitivity analyses of changes from baseline in the primary and key secondary outcomes at week 9 in the ITT population. CI denotes 95% confidence interval. Based on repeated measures mixed linear models, where missing data is conservatively imputed using baseline observation carried forward.

	GLAD (N=)	IA Saline (N=)	Estimated treatment difference	P-value
	Mean (SE)	Mean (SE)	Mean (95% CI)	
Primary outcome:				
Change in KOOS Pain – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	0.xxx*
Key Secondary outcomes:				
Change in KOOS Function – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in KOOS Quality of life – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in PGA – VAS (0 to 100 mm)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	

Values are least squares means ± standard error unless otherwise stated.
*Primary outcome will be analyzed using both a test for equivalence and a test for superiority.
KOOS: Knee injury and osteoarthritis outcome score.
PGA: Patient Global Assessment
VAS: Visual Analog Scale

13 REFERENCES

1. Tubach F, Ravaud P, Martin-Mola E, Awada H, Bellamy N, Bombardier C, et al. Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: Results from a prospective multinational study. Arthritis Care Res (Hoboken). 2012;64(11):1699-707.
2. Bennell K, Dobson F, Hinman R. Measures of physical performance assessments: Self-Paced Walk Test (SPWT), Stair Climb Test (SCT), Six-Minute Walk Test (6MWT), Chair Stand Test (CST), Timed Up & Go (TUG), Sock Test, Lift and Carry Test (LCT), and Car Task. Arthritis Care Res (Hoboken). 2011;63 Suppl 11:S350-S70.
3. Singh JA, Luo R, Landon GC, Suarez-Almazor M. Reliability and clinically important improvement thresholds for osteoarthritis pain and function scales: a multicenter study. J Rheumatol. 2014;41(3):509-15.
4. Detry MA, Ma Y. Analyzing Repeated Measurements Using Mixed Models. JAMA. 2016;315(4):407-8.
5. White IR, Horton NJ, Carpenter J, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. BMJ. 2011;342:d40.
6. Rubin DB. Multiple Imputation for Nonresponse in Surveys: Wiley; 1987. 258 p.
7. Zhang J, Yu KF. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. JAMA. 1998;280(19):1690-1.
8. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. BMC Med Res Methodol. 2017;17(1):162.
9. Liublinska V, Rubin DB. Sensitivity analysis for a partially missing binary outcome in a two-arm randomized clinical trial. Stat Med. 2014;33(24):4170-85.