Statistical Analysis Plan	
Study Code	D5680C00003
Edition Number	Final v3.0
Date	13SEP2023

A Randomised, Double-blind, Placebo-controlled, Dose-response Study of the Efficacy and Safety of MEDI7352 in Subjects with Painful Osteoarthritis of the Knee

A Randomised, Double-blind, Placebo-controlled, Dose-response Study of the Efficacy and Safety of MEDI7352 in Subjects with Painful Osteoarthritis of the Knee

Study Statistician (IQVIA)

PPD

Date

A Randomised, Double-blind, Placebo-controlled, Dose-response Study of the Efficacy and Safety of MEDI7352 in Subjects with Painful Osteoarthritis of the Knee

Study Statistician (AstraZeneca)

PPD

Date

A Randomised, Double-blind, Placebo-controlled, Dose-response Study of the Efficacy and Safety of MEDI7352 in Subjects with Painful Osteoarthritis of the Knee

Study Lead

PPD

Date

A Randomised, Double-blind, Placebo-controlled, Dose-response Study of the Efficacy and Safety of MEDI7352 in Subjects with Painful Osteoarthritis of the Knee

Study Physician

PPD

Date

TABLE OF CONTENTS

SIGNATURE OF STUDY STATISTICIAN 2
SIGNATURE OF STUDY STATISTICIAN 3
SIGNATURE OF STUDY LEAD 4
SIGNATURE OF STUDY PHYSICIAN 5
1 STUDY DETAILS 14
1.1 Study objectives 15
1.1.1 Primary objective 15
1.1.2 Secondary objectives 16
1.1.3 Safety objectives 17
1.1.4 CCI [redacted]
1.2 Study design 19
1.3 Number of participants 20
2 ANALYSIS SETS 20
2.1 Definition of analysis sets 20
2.2 Violations and deviations 22
3 PRIMARY AND SECONDARY VARIABLES 24
3.1 General principles 24
3.1.1 Missing data 24
3.1.2 Baseline 24
3.1.3 Change and percentage change from baseline 24
3.1.4 Early termination visit 25
3.1.5 Study periods 25
3.1.5.1 Pre-treatment period 25
3.1.5.2 Treatment period 25
3.1.5.3 Follow-up period 25
3.1.6 Exposure related periods 25
3.1.6.1 Prior treatment 25
3.1.6.2 On treatment 25
3.1.6.3 Post treatment 26
3.1.7 Reference start date 26
3.1.8 Analysis visits – Retests and Unscheduled Assessments 26
3.1.9 Concomitant medications and therapies 27
3.2 Primary efficacy variables 28
3.3 Secondary efficacy variables 30
3.3.1 WOMAC multiscale index 30
3.3.1.1 WOMAC pain subscale 31
3.3.1.2 WOMAC physical function subscale 31
3.3.1.3 WOMAC stiffness subscale 31
3.3.1.4 WOMAC overall score 31
3.3.2 Patient's Global Assessment (PGA) of Osteoarthritis (OA) 31

3.3.3 OMERACT-OARSI responder index 31
3.4 Pharmacokinetic and immunogenicity variables 32
3.4.1 Pharmacokinetics 32
3.4.2 Immunogenicity 32
3.5 Exploratory variables 33
3.5.1 Supportive efficacy and health related quality of life variables 33
3.5.1.1 Daily Sleep Interference Scale (DSIS) 33
3.5.1.2 European Quality of Life-5 Dimensions (EQ-5D-5L) 34
3.5.1.3 36-Item Short Form Health Survey Questionnaire 34
3.5.1.4 Work Productivity and Activity Impairment questionnaire for OA (WPAI:OA) 35
3.5.1.5 Healthcare Resource Utilisation (HCRU) 35
3.5.1.6 Lack of efficacy endpoints 35
3.5.1.7 Rescue medication use 36
3.5.2 Biomarker and Pharmacodynamic variables 36
3.5.3 Ambulatory blood pressure monitoring (ABPM) variables 37
3.6 Safety variables 38
3.6.1 Adverse events 38
3.6.1.1 Adverse events of special interest 39
3.6.1.2 Adverse events based on signs and symptoms 39
3.6.1.3 Adverse events based on examinations and tests 40
3.6.1.4 Pregnancy 40
3.6.2 Physical examination 40
3.6.2.1 Weight 40
3.6.3 Neurological examination 40
3.6.4 Total neuropathy score- nurse (TNSn) variables 41
3.6.5 Vital Signs 42
3.6.5.1 Orthostatic blood pressure 42
3.6.6 Survey of Autonomic Symptoms (SAS) 42
3.6.7 Electrocardiograms 43
3.6.8 Laboratory data 43
3.6.9 Injection site reactions 44
3.6.10 Radiographic (X-ray) and magnetic resonance imagining (MRI) assessments 44
3.6.10.1 X-ray assessments 44
3.6.10.2 MRI assessments 45
3.6.11 Study treatments 45
3.6.11.1 Exposure of study treatment 45
3.6.11.2 Compliance 45
4 ANALYSIS METHODS 46
4.1 General principles 47
4.1.1 Imputation Rules 47
4.1.1.1 Imputation of the primary and key secondary efficacy endpoints 47
4.1.1.2 Imputation of other study variables 50
4.1.1.3 Handling of missing/incomplete dates 50
4.2 Efficacy analysis 51
4.2.1.1 Multiple testing strategy 53

4.2.1.2 Analysis of the primary and key secondary endpoints 55
4.2.2 Other secondary efficacy endpoints 60
4.2.3 Exploratory analysis 60
4.2.3.1 ABPM 61
4.2.3.2 Pharmacodynamic analysis 61
4.2.3.3 Biomarker analysis 62
4.2.4 Pharmacokinetic analysis 62
4.2.5 Immunogenicity analysis 62
4.2.6 Safety analysis 63
4.2.6.1 Adverse Events 63
4.2.6.2 Joint replacement surgeries 65
4.2.6.3 Pregnancies 65
4.2.6.4 Laboratory results 66
4.2.6.5 Vital signs results 68
4.2.6.6 ECG results 69
4.2.6.7 Physical examination findings 69
4.2.6.8 Neurological examination findings 69
4.2.6.9 Total neuropathy score-nurse (TNSn) results 69
4.2.6.10 Injection site reaction 69
4.2.6.11 Survey of Autonomic Symptoms (SAS) 70
4.2.6.12 X-ray and MRI 70
4.2.7 Treatment exposure and compliance 70
4.2.8 Demographic and baseline characteristics 71
4.2.9 Accounting for the COVID-19 pandemic effects 73
5 INTERIM ANALYSES 74
6 CHANGES OF ANALYSIS FROM PROTOCOL 76
7 REFERENCES 78
8 APPENDICES 79
Appendix A Questionnaires derivations 79
Appendix A1 WPAI:OA 79
Appendix A2 Responses and scores of the HCRU questionnaire 80
Appendix B Multiple Imputation 82
Appendix B1 The pattern mixture framework 82
Appendix B2 The sampling from normal distribution approach 82
Appendix C MCP-Mod: Combining Estimates from Multiple Imputed Datasets 83
Appendix DMCP-Mod Code 85
Appendix D1 Deriving optimum contrast coefficients 85
Appendix D2 CCI [redacted]
Appendix D3 CCI
Appendix D4 Model estimates 93
Appendix D4.1 Calculation of Dose to achieve target effects 93
Appendix D4.2.1 Asymptomatic normal Distribution 94
Appendix D4.2.2 Delta method 94

LIST OF TABLES

Table 1 Analysis sets 20
Table 2 Intercurrent events and imputation strategy 29
Table 3 Threshold values for the SF-36v2 scale and summary measures 35
Table 4 Acceptable ABPM ranges 37
Table 5 Normal ABPM ranges 38
Table 6 Vital signs reference ranges 42
Table 7 Study treatments 45
Table 8 Pre-planned statistical analyses to be conducted 52
Table 9 Pre-defined hypothesis and hierarchy to be tested 53

LIST OF FIGURES

Figure 1 Study design 19
Figure 2 Formal statistical testing strategy 54

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
ABPM	Ambulatory blood pressure monitoring
ADA	Antidrug antibody
AE(s)	Adverse event(s)
AIC	Akaike information criterion
ANCOVA	Analysis of covariance
BMI	Body mass index
BOCF	Baseline-observation-carried-forward
BP	Blood pressure
CI	Confidence interval
COVID-19	Coronavirus disease 2019
CRP	C-reactive protein
CSP	Clinical study protocol
CTMS	Clinical Trial Management System
DCO	Data cut-off
DSIS	Daily Sleep Interference Scale
DSMB	Data and Safety Monitoring Board
DTG	Data transfer guidelines
ECG	Electrocardiogram
eCRF	Electronic case report form
ED50	Half of the Emax
ED90	90% of the Emax
E_{max}	Maximum response
EOT	End-of-treatment
ePRO	Electronic patient-reported outcome
CCI [redacted]	[redacted]
ET	Early termination
FAS	Full analysis set
CCI [redacted]	[redacted]
IA	Interim analyses
ICH	International Conference on Harmonization
IERC	Independent Efficacy Review Committee
IP	Investigational product
IPDs	Important protocol deviations

Abbreviation or special term	Explanation
IV	Intravenous
KL	Kellgren and Lawrence
LOCF	Last-observation-carried-forward
LLOQ	Lower limit of quantification
MAR	Missing at random
MCMC	Markov chain Monte Carlo
MCP-Mod	Multiple comparison procedure-modelling
MED	Minimum effective dose
MedDRA	Medical dictionary for regulatory activities
mg	Milligrams
MNAR	Missing not at random
MRI	Magnetic resonance imaging
NGF	Nerve growth factor
NRS	Numerical rating scale
OA	Osteoarthritis
OARSI	Osteoarthritis Research Society International
OMERACT-OARSI	Outcome Measures in Rheumatology-Osteoarthritis Research Society International
PDs	Protocol Deviations
PF	Physical function
PGA	Patient's Global Assessment
PK	Pharmacokinetic(s)
PT	Preferred term
Q2W	Every 2 weeks
RNA	Ribonucleic acid
RPOA	Rapidly Progressive Osteoarthritis
SAE(s)	Serious adverse event(s)
SAF	Safety analyses set
SAP	Statistical analysis plan
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SAS	Survey of Autonomic Symptoms
SAS®	Suite of analytics software
CCI	[redacted]
SD	Standard deviation
CCI	[redacted]
SIF	Subchondral Insufficiency Fracture

Abbreviation or special term	Explanation
SOC	System organ class
TNF	Tumour necrosis factor
WHO	World health organisation
WOMAC	Western Ontario and McMaster Universities Osteoarthritis
CCI [redacted]	[redacted]

AMENDMENT HISTORY

Category*: Change refers to	Date	Description of change	In line with the CSP?	Rationale
Other secondary endpoints	10AUG23	All dosing visits will be used in analysis and summary statistics.	No	Provided the data is available it was decided to analyse it despite the protocol pre-defined timepoints.
Data presentation	15AUG23	The document was updated throughout following the dry run prior to unblinding and final database lock.	Yes/ No	Data presentation was updated to allow more appropriate evaluation of the available data. In cases where such updates would deviate from the protocol, details are provided in section 6 of this document.
Other	15AUG23	Clarifications added throughout the document.	Yes	To support programming.
Other	12SEP23	Updated references to tables so that numbering aligns.	NA	Not applicable.

* Pre-specified categories are
Primary or secondary endpoints; Statistical analysis method for the primary or secondary endpoints; Derivation of primary or secondary endpoints; Multiple Testing Procedure; Data presentations; Other

1 STUDY DETAILS

This statistical analysis plan (SAP) contains a more detailed description of the final analyses as described in the clinical study protocol (CSP) version 4.0, dated 28th June 2022. The final analyses will be split into two sections:

Accelerated final efficacy analysis
This analysis will be performed once all participants complete or discontinue the treatment period by completing week 12 (EOT/ET) visit. This threshold will indicate a potential data cut-off (DCO) date that will be confirmed by the study team. This confirmed DCO date will be used for analysis purposes and data cleaning will be focused on critical data up to and including the DCO date. Therefore, any data with a date in-between the DCO and database export dates will not be used for accelerated analysis purpose.

The accelerated final efficacy analyses will consist primarily of the primary and secondary statistical analyses and formal statistical testing of the primary endpoint and one key secondary endpoint namely the Change in the WOMAC pain subscale from baseline to Week 12. In addition, the supportive and other statistical analyses of those two endpoints will be executed. Specific outputs that will be produced for this analysis will be summarised in the accelerated final efficacy shells file that will be approved prior to the accelerated final efficacy analysis database lock.

The accelerated final efficacy outputs will be prepared by the blinded IQVIA biostatistics team. The unblinding of the accelerated final efficacy analysis will be done by an independent unblinded IQVIA biostatistics team. The unblinded IQVIA biostatistics team will share the fully unblinded accelerated final efficacy analysis package with AstraZeneca via the restricted EntimICE unblinded folder. More details on the process and preserving the blind can be found in the biostatistical data transfer guidelines (DTG) and the unblinding plan.

Final analysis
This analysis will be performed at the end of the study once all participants either discontinue or complete the whole study including the follow-up period as defined in the clinical study protocol (CSP). This analysis will consist of the statistical analyses and formal statistical testing of the remaining key secondary endpoints and other statistical analyses that were not assessed nor performed during the accelerated final efficacy analysis. Outputs produced for this analysis will be summarised in the study specific shells file that will also include the specification of the outputs produced during the accelerated final efficacy analysis. Those outputs produced for the accelerated final efficacy analysis would be also rerun to ensure that only expected changes occurred (e.g., no ongoing participants in the study, changes associated with post treatment period). If any unexpected changes are observed, then the root cause analyses will be performed.

1.1 Study objectives

1.1.1 Primary objective

Primary Objective	Estimand description/endpoint
To assess the efficacy of MEDI7352 compared to placebo on chronic pain in participants with painful OA of the knee at Week 12	**Population:** Adults with moderate-to-severe chronic OA pain of the knee, persistent for 3 months or longer, not adequately controlled by standard-of-care treatments
	Primary endpoint: Change in weekly average of daily NRS pain scores from baseline to Week 12
	Intercurrent events: • Discontinuation due to lack of efficacy or an adverse event (AE) • Discontinuation due to other reasons such as loss to follow-up or external circumstances • Taking prohibited pain medication during the treatment period • Taking excessive permitted rescue medication
	Summary measures: • Difference in mean changes in weekly average daily NRS pain scores between MEDI7352 doses and placebo • An 'attributable' estimand strategy will be adopted for primary endpoint data missing or affected by any of the above intercurrent events: Missing or affected primary endpoint values will be imputed, but the method of imputation will differ depending upon the intercurrent event. Specifically, data missing due to discontinuation, due to lack of efficacy or an AE will be imputed assuming an unfavourable outcome. Similarly, data affected by prohibited or excessive rescue medication will be set to missing and imputed assuming an unfavourable outcome.

AE, adverse event; NRS, numerical rating scale; OA, osteoarthritis.

1.1.2 Secondary objectives

Secondary Objectives	Estimand descriptions/endpoints
To assess the efficacy of MEDI7352 compared to placebo on additional measures of efficacy in participants with painful OA of the knee	**Key secondary endpoints:** • Change in the WOMAC pain subscale from baseline to Week 12 • Change in the WOMAC physical function subscale from baseline to Week 12 • Change in the PGA of OA from baseline to Week 12
	Other secondary endpoints: • Change in the WOMAC pain subscale from baseline to Weeks 2, 4, 6, 8, 10, and 18 • Change in the WOMAC physical function subscale from baseline to Weeks 2, 4, 6, 8, 10, and 18 • Change in the WOMAC overall scores from baseline to Weeks 2, 4, 6, 8, 10, 12, and 18 • Change in the WOMAC stiffness scores from baseline to Weeks 2, 4, 6, 8, 10, 12, and 18 • Change in PGA of OA from baseline to Weeks 2, 4, 8, 10, and 18 • Percentage of responders as measured by the OARSI responder index using the OMERACT-OARSI definition (Pham et al 2004) at Weeks 2, 4, 8, 12, and 18 • Percentage of participants who have achieved an improvement of ≥ 2 points in PGA of OA at Weeks 2, 4, 8, 12, and 18 • Change in the weekly average of daily NRS pain scores from baseline to Weeks 2, 4, 6, 8, 10, and 18 • Percentage of participants who have achieved ≥ 30% and ≥ 50% reductions in the weekly average of daily NRS pain scores from baseline to Weeks 2, 4, 8, 12, and 18 • Percentage of participants who have achieved ≥ 30% and ≥ 50%, reductions in WOMAC pain subscale scores at Weeks 2, 4, 8, 12, and 18. • Percentage of participants who have achieved ≥ 30% and ≥ 50% reductions in WOMAC physical function subscale at Weeks 2, 4, 8, 12, and 18
To assess the PK of MEDI7352 in participants with painful OA of the knee	• Serum concentration of MEDI7352

Secondary Objectives	Estimand descriptions/endpoints
To assess immunogenicity of MEDI7352 in participants with painful OA of the knee	• Presence of ADA to MEDI7352 • ADA titre

ADA, anti-drug antibody; NRS, numerical rating scale; OA, osteoarthritis; OARSI, Osteoarthritis Research Society International; OMERACT-OARSI, Outcome Measures in Rheumatology-Osteoarthritis Research Society International; PGA, Patient's Global Assessment; PK, pharmacokinetics; WOMAC, Western Ontario and McMaster Universities Osteoarthritis.

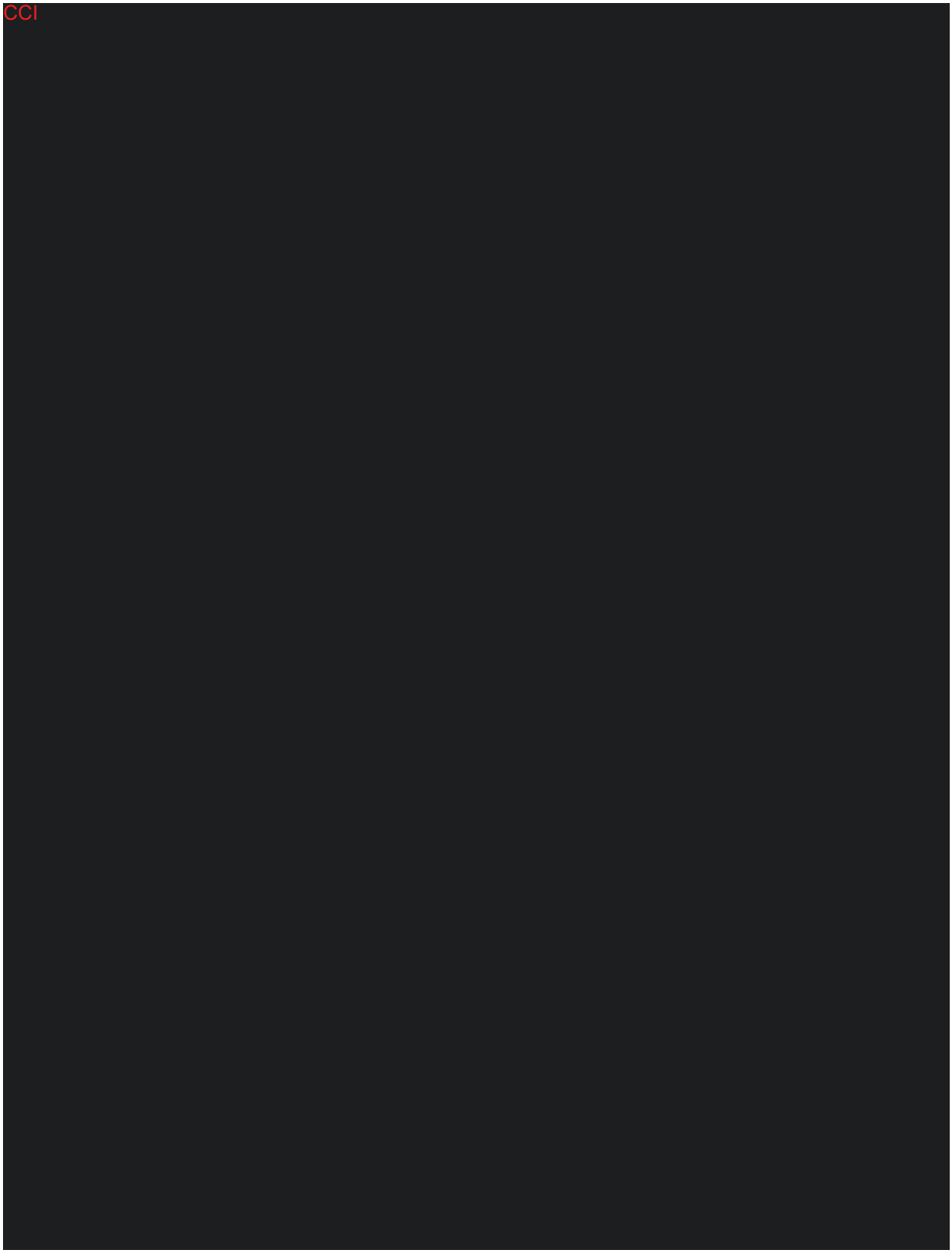
1.1.3 Safety objectives

Safety Objectives	Estimand descriptions/endpoints
To assess the safety and tolerability of MEDI7352 compared with placebo in participants with painful OA of the knee	Safety and tolerability will be evaluated based on AEs, vital signs, and clinical laboratory assessments, including but not limited to: • AEs and SAEs • Physical examinations • Neurological examinations • Total Neuropathy Score-nurse • Weight • Vital signs (supine and standing BP, pulse rate, temperature, respiratory rate) • Survey of Autonomic Symptoms • 12-lead ECGs • Clinical laboratory testing (haematology, chemistry, coagulation, and urinalysis) • CRP (inflammatory biomarker) • Concomitant medications and therapies • Injection site reactions • X-ray or/and MRI of large joints [a]

[a] Detailed information about the joints to be evaluated by X-ray and/or MRI as noted in the protocol section 8.2.6 Appendix I.
AE, adverse event; BP, blood pressure; CRP, C-reactive protein; ECG, electrocardiogram; MRI, magnetic resonance imaging; OA, osteoarthritis; SAE, serious adverse event.

1.1.4 CCI

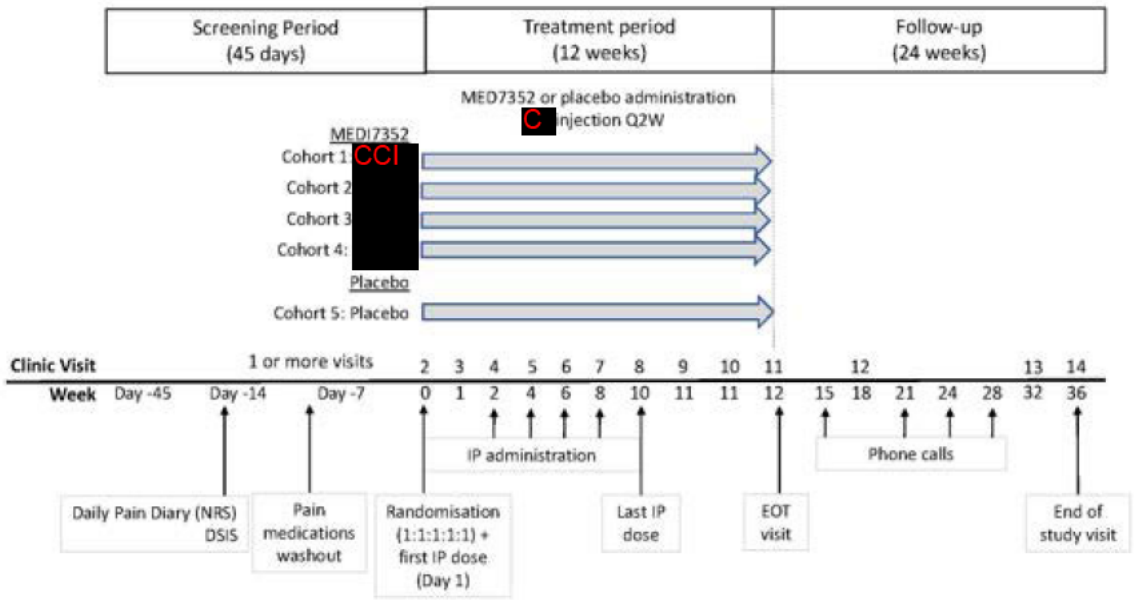
CCI



1.2 Study design

This is a Phase IIb, multicentre, multinational, randomised, double-blind, placebo-controlled, dose-response study of MEDI7352 in participants 18 to 80 years of age (inclusive) with moderate-to-severe chronic pain of the knee during the previous 3 months or longer, caused by osteoarthritis (OA), not adequately controlled by standard-of-care treatments. See Figure 1 for an overview of the study design. See CSP Table 1 for the schedule of activities planned, including details on the study period.

Figure 1 Study design



DSIS, daily sleep interference scores; EOT, end of treatment; IP, investigational product; NRS, numerical rating scale; Q2W, every 2 weeks; CCI

The screening period is up to 45 days; one rescreening attempt is allowed. Pain medication washout is at least 48 hours or 5 half-lives, whichever is longer, prior to the start of the numerical rating scale (NRS) pain score baseline period on Day -7. The baseline NRS pain score period comprises the scores from Day -7 to Day -1 (inclusive).

Visit 3 (Day 7), Visit 9 (Day 74) and Visit 10 (Day 77) may be performed at the participant's home by a qualified health care provider. Visits 9 and 10 cannot be on the same day.

Treatments and treatment duration:

The study consists of a screening period of up to 45 days (one or more visits), a 12-week double-blind treatment period (10 scheduled visits), and a 24-week follow-up period (3 scheduled visits and 4 phone calls).

After screening and confirmation of eligibility criteria, eligible participants will be randomised on Day 1 at a 1:1:1:1:1 ratio to receive MEDI7352 doses or placebo. Participants will receive a total of 6 CCI injections (every 2 weeks [Q2W] at Day 1, weeks 2, 4, 6, 8 and 10):

- Cohort 1: MEDI7352 CCI
- Cohort 2: MEDI7352 CCI
- Cohort 3: MEDI7352 CCI
- Cohort 4: MEDI7352 CCI
- Cohort 5: Placebo

Participants who complete the treatment period will come in for their end-of-treatment (EOT) visit at Week 12 and then enter the follow-up period. Participants who permanently discontinue investigational product (IP) will be asked to complete study assessments at the early termination (ET) visit (Week 12 assessments) as soon as possible and then enter the follow-up period. All participants are expected to complete the EOT/ET visit and the 24-week follow-up period.

The study incorporates two interim analyses, detailed in section 5.

1.3 Number of participants

Approximately 350 eligible participants will be randomly assigned to IPs with one of 4 dose levels of MEDI7352 or placebo. Recruitment will continue until the planned power of the study is achieved (i.e., when statistical information equivalent to 255 participants completing the treatment period is reached) or approximately 350 randomised participants, whichever is sooner (see CSP section 9.2 for more details).

2 ANALYSIS SETS

2.1 Definition of analysis sets

Details of the analysis sets as described in the CSP are presented in Table 1 below.

Table 1 Analysis sets

Analysis set	Definition
Screening set	All participants who provide informed consent and provide demographic and/or baseline screening assessments, regardless of the participant's randomisation and treatment status in the study. This set will be used for participants' disposition.

Analysis set	Definition
Full analysis set	All randomised participants. Participants will be analysed according to the intent-to-treat principle whereby their randomised study treatment will be analysed regardless of the study treatment actually received. This set will be used for demographics and other baseline characteristics, medication description and all efficacy analyses.
Safety analysis set	All participants who received at least one dose of double-blind study treatment. Participants will be analysed according to their actual study treatment received, regardless of the randomised study treatment. This set will be used for safety analyses.
PK analysis set	All participants who received at least one dose of double-blind study treatment per the protocol for whom any post-baseline PK data are available and who do not violate or deviate from the protocol in ways that would significantly affect the PK analyses. Participants will be analysed according to their actual study treatment received, regardless of the randomised study treatment.

PK, pharmacokinetics.

The assignment of actual treatment if an incorrect IP is administered will follow the below criteria:

- If a participant randomised to placebo receives at least one dose of MEDI7352 XX mg during the entire double-blind treatment period then the actual treatment arm = MEDI7352 XX, where XX corresponds to the lowest dose of MEDI7352 that participant received during the double-blind treatment period.
- If a participant took only MEDI7352 XX mg treatment during the entire double-blind treatment period, then the actual treatment arm = MEDI7352 XX, where XX corresponds to the dose of MEDI7352 that participant took during the entire double-blind treatment period regardless of what treatment participant was randomized to.
- If a participant took a mix of MEDI7352 XX mg treatments during the entire double-blind treatment period, then the actual treatment arm = MEDI7352 XX, where XX corresponds to the lowest dose of MEDI7352 that participant received during the double-blind treatment period regardless of what treatment participant was randomized to.
- If a participant randomised to an active treatment receives only placebo during the entire double-blind treatment period, then the actual treatment arm = Placebo.

2.2 Violations and deviations

AstraZeneca use International Conference on Harmonization (ICH) E3 terminology for protocol deviations (PDs), which are all important deviations related to study inclusion or exclusion criteria, conduct of the trial, participant management or participant assessment. All protocol deviations (PDs) identified during monitoring of the study are recorded in the clinical trial management system (CTMS). Important protocol deviations (IPDs) are a subset of protocol deviations that may significantly impact the completeness, accuracy, and/or reliability of the study data or that may significantly affect a participant's rights, safety, or well-being.

The reporting of PDs and hence the CTMS changed on 16MAY2022. The main changes were the addition of the flag identifying the IPDs as well as changes to the PD categories.

A detailed description of how PDs and IPDs are collected and identified is referenced in the Protocol Deviation Management Plan.

The final list of IPDs will be documented prior to unblinding the study data for the final DBL and re-visited if required once the study is unblinded (for example, incorrect IP administration). This list may include, but not limited to the following categories:

- Inclusion Criteria Deviations
- Exclusion Criteria Deviations
- Discontinuation Criteria for study product met but participants not withdrawn from study treatment
- Discontinuation Criteria for overall study withdrawal met but participant not withdrawn from study
- Deviations from IP management and administration
- Prohibited Medications Taken
- Deviations related to study procedure such as informed consent
- Other Important Deviations
    - Scheduled study assessments not done or incorrectly performed which could impact the primary or key secondary endpoints
    - Unblinding of treatment assignment for reasons unrelated to participant safety
    - Other severe non-compliance (such deviations will be clearly described in the CSR)

The IPDs will be listed including the full PD text as entered in the CTMS by site staff or CRAs and summarised by randomised treatment group including the old and new categories

depending on when the PD was reported (before or after 16MAY2022). IPDs will also be summarised separately by whether they are due to COVID-19 or not.

None of the deviations will lead to participants being directly excluded from the analysis sets described in section 2.1. The only exception being for the PK analysis set, if the deviation is considered to significantly impact upon PK assessments. That is:

- If a participant randomised to an active treatment receives only placebo, then such participant would be excluded from the PK analysis set.
- If a participant randomised to placebo receives only active treatment, then such participant would also be excluded from the PK analysis set (per protocol PK samples from participants randomised to placebo are not to be analysed).
- If a participant randomised to placebo receives only placebo, but has any post-baseline PK data available, such participant will be excluded from the PK analysis set.
- If a participant randomised to an active treatment receives an incorrect dose or placebo during the treatment period then such participant would be included in the PK analysis set, but PK assessment(s) after such dose would be excluded from summary outputs.
- Participants with missed IP dose will be included in the PK analysis set, but such participants will be discussed case by case prior to unblinding to agree on the PK assessments use. The generic rule would be to not include assessment(s) made after the missed dose in summary outputs. Such data would be set to missing on the ADaM level so that there is traceability (but will be included in the SDTMs).
- Participants that don't receive a full volume of the IP will be included in the PK analysis set if more than 80% of the volume was administered. If less than 80% of the volume was administered, then such cases will be reviewed prior to unblinding and handled case by case.

In addition, if a participant does not receive IP, this would lead to their exclusion from the safety set. A per protocol analysis excluding participants with specific important protocol deviations is not planned.

3 PRIMARY AND SECONDARY VARIABLES

3.1 General principles

3.1.1 Missing data

Unless specifically detailed otherwise, the attributable estimand approach will guide the analysis and summary of primary and key secondary variables, whereas other variables will be analysed and/or summarised using available data only.
For more information on handling of missing data and imputation refer to section 4.1.1.

3.1.2 Baseline

In general, baseline will be the last non-missing assessment value obtained prior to the first dose of randomised study treatment. If a participant is randomised and not treated, then baseline would be the last non-missing assessment prior or on the day of randomisation (Day 1).

However, for the primary endpoint of NRS pain score, CCI [redacted], the baseline is defined as the average (mean) of daily scores recorded during the 7-day period immediately prior to randomisation (Day -7 to Day -1, inclusive). At least four of the seven daily scores need to be recorded by the participant to obtain a valid baseline value for statistical analyses.

In addition, the baseline for rescue medication is defined as the sum of daily use recorded during the 7-day period immediately prior to randomisation (Day -7 to Day -1, inclusive).

3.1.3 Change and percentage change from baseline

In general, change from baseline outcome variables are computed as:

(post-baseline value - baseline value)

In general, percent change from baseline outcome variables are computed as:

((post-baseline value – baseline value) / (baseline value)) *100

However, for the primary endpoint, the change from baseline is defined as the weekly average of the daily NRS pain scores at Week x minus the baseline weekly average NRS pain score. The derivation of the weekly average of the daily NRS pain scores at Week x is described in section 3.2. In addition, for primary and key secondary endpoints the percent change from baseline is expressed in terms of improvement as follows:

((baseline value – Week 12 value) / (baseline value)) *100

3.1.4 Early termination visit

Assessments associated with an early termination of IP are the same as those for EOT (Week 12) visit. Consequently, summaries presented by visit will not differentiate whether the Week 12 assessments correspond to the EOT or the ET visit. The exceptions are the primary endpoint, secondary endpoints, CCI [redacted], TNSn and DSIS for which the handling of the data associated with ET visit can be found in sections 3.2, 3.3, 3.6.4 and 3.5.1.1 respectively.

3.1.5 Study periods

The below definitions of study periods are defined as per the protocol and study design Figure 1.

3.1.5.1 Pre-treatment period

The pre-treatment period is defined from signed informed consent until Day 1 prior to first administration of study treatment.

3.1.5.2 Treatment period

The treatment period is defined from the first administration of study treatment (Day 1) through to the Week 12 (EOT/ET) visit, regardless of when IP is discontinued.

3.1.5.3 Follow-up period

The follow-up period starts after Week12 (EOT/ET) visit and ends with end of study participation.

3.1.6 Exposure related periods

These periods are defined in terms of safety assessments and exposure to IP. If the assessment is not associated with the time and hence it cannot be determined whether the assessment on Day 1 is prior to or after the first dose of IP, then it will be assumed that assessment is after the IP unless otherwise stated.

3.1.6.1 Prior treatment

Prior treatment is defined from signed informed consent until Day 1 prior to first IP administration. In cases where the participant is randomised and IP is not administered then all assessments on Day 1 would be considered prior treatment.

3.1.6.2 On treatment

On treatment is defined from the day and time of first IP administration throughout to the last IP administration +14 days, based on MEDI7352 half-life of 3-4 days. Participants that are randomised and do not receive IP will not be part of ‘on treatment’ period.

3.1.6.3 Post treatment

Post treatment is defined from last IP+15 days through to the date of last study assessment.

3.1.7 Reference start date

The reference start date is defined as the date of the first dose of investigational product.

- If the date of the assessment/event is on or after the reference date, then:
  Relative study day = (date of event – reference date) + 1

- If the date of the assessment/event is prior to the reference date, then:
  Relative study day = (date of event – reference date)

In the situation where the assessment/event date is partial or missing, relative study day and corresponding durations will appear partial or missing in the data listings.

3.1.8 Analysis visits – Retests and Unscheduled Assessments

By-visit summaries will be presented by scheduled electronic case report form (eCRF) visit in form of a Day and a Week with a corresponding timepoint where applicable, unless otherwise specified. No analysis visit windowing will be applied unless otherwise specified.

If more than one assessment coincides with the date of the scheduled visit as per the eCRF visit form, then the following will apply:

- If assessment is associated with a scheduled timepoint as per the protocol, then such assessment will be used for analysis.
- If more than one assessment per specific visit is associated with a scheduled timepoint as per the protocol, then the latest assessment according to the assessment date/time with non-missing result will be used for analysis.
- If for such assessments a timepoint is not defined in the protocol, then the latest assessment according to the assessment date/time with non-missing result will be used for analysis.

In addition, due to the extension of the follow-up period as per the protocol amendment 2 dated 14OCT21, certain assessments were performed only during the Week 32 as supposed to Week 28 as done in the previous versions of the protocol. Provided that no windowing is applied some of such assessments will be mapped to Week 28 and some to Week 32 depending on at what visit the assessment was actually performed as per the eCRF visit form. Consequently, the by-visit summaries will include both visits.

NRS daily scores and **DSIS** will be mapped so that all assessments prior to Day 1 visit will be mapped to screening, the NRS daily score which date coincides with the Day 1 visit will be

mapped to that visit (Day 1) as well as all subsequently recorded scores (scheduled/unscheduled) up to but NOT including the NRS daily score that coincides with Week 1 visit date as recorded on the eCRF. This score will be mapped to Week 1 visit. The same logic will be applicable for the remaining visits and recorded daily scores. Only those averages associated with the protocol defined visits will be used for analysis and summaries.

All secondary endpoints, CCI [redacted] and **TNSn** will be mapped based on their date of assessment that coincides with the date of the visit (including unscheduled visits) as per the eCRF. The exception will be assessments prior to the first IP administration that will be automatically mapped to screening or rescreening visit depending on the assessment dates. Only those assessments associated with the protocol defined visits will be used for analysis and summarises.

ABPM assessments will be mapped only to two visits with labels: "Screening" and "Day 70, Week 10". If more than one assessment is available, then:

- Screening: baseline will be determined as per section 3.1.2 and remaining assessments will be mapped to screening;
- Post-baseline: the last recorded assessment will be used for analyses.

X-rays and MRI will be mapped only to two visits with labels: "Screening" and "Day 224, Week 32". If more than one assessment is available, then:

- Screening: baseline will be determined as per section 3.1.2 and remaining assessments will be mapped to screening;
- Post-baseline: the last recorded assessment will be used for analyses.

3.1.9 Concomitant medications and therapies

Medications will be classified as prior, or concomitant as follows:

- Prior medication: Defined as those medications that were taken and stopped prior to the date of first dose of study treatment.
- Concomitant medication: Defined as those medications taken at any time on or after the day of the first dose of study treatment.
- Prohibited medication: Concomitant medications assessed as disallowed by the medical review as specified in CSP section 6.5.2.

A medication started prior to the first dose of study treatment and ended after the first dose or is ongoing will be classified as both prior and concomitant. All records, including those

collected during the follow-up period as well as prior to the first IP administration will be considered and summarised.

3.2 Primary efficacy variables

The primary efficacy variable is the change in weekly average of daily NRS pain scores from baseline to Week 12.

The NRS is an 11-point Likert scale used to assess pain, where participants are asked to describe their average pain in the target knee by identifying a number from 0 = "no pain" to 10 = "most severe pain imaginable over the previous 24 hours". This will be recorded on a daily basis at approximately the same time every morning. The daily NRS pain scores will be recorded via electronic patient recorded outcome (ePRO) diary.

The relevant visit associated with the daily NRS scores will be derived based on the date of NRS pain scores that coincide with the relevant visit date as recorded on the electronic case report form (eCRF). Therefore, the Week 12 weekly average of daily NRS pain scores is the average of the daily NRS pain scores of 7 days up to and including the date on which the week 12 (EOT/ET) visit occurred.

For participants that discontinue the treatment period, the date of the ET visit as recorded on the eCRF in form of the week 12 (EOT/ET) visit will be associated with the protocol defined visit based on the study day, and the 7-day average for that specific visit will be derived using the same approach as described above for the Week 12 average. If the ET visit coincides with the target study day of the EOT visit, then this ET average will not be derived as it cannot represent a valid Week 12 average. In other words, only participants that complete the treatment will have Week 12 average derived. For those who prematurely withdraw from the treatment the Week 12 average will be imputed. At least four of the seven daily scores need to be recorded by the participant to obtain a weekly average value including baseline for the statistical analyses.

If there are less than four daily scores for the Week 12 average due to an inappropriate visiting schedule of Week 11 (Week 11 and Week 12 in close succession), the derivation of the Week 12 average will take priority over the Week 11 average. The weekly average will be derived in the same manner as described above but daily scores that may have been used for derivation of the Week 11 average would instead be used for derivation of the Week 12 average to ensure that at least 4 scores are available.

For the primary endpoint, an attributable estimand strategy will be adopted for missing or unevaluable data accounted for by any of the defined intercurrent events. Missing or unevaluable primary endpoint values will be imputed, but the method of imputation will differ depending upon the intercurrent event, as detailed in Table 2 below.

Primary endpoint data missing due to other reasons than those summarised in Table 2 will be imputed using missing-at-random (MAR) multiple imputation approach, as outlined in section 4.1.1.1.

In addition, only single LOCF imputation will be applied as specified in section 4.1.1.1.

Unevaluable data of the primary and key secondary variables are defined as cases where:

- Participant's cumulative daily dose of rescue medication of at least 4000 mg or prohibited pain medication were taken for a minimum of 4 days over the 7-day period that is being considered for deriving the weekly NRS average. In this instance the NRS average will be set to missing.

- Participant's cumulative daily dose of rescue medication of at least 4000 mg was taken within 24h prior of the WOMAC and PGA assessment or prohibited pain medication was taken a day before the WOMAC and PGA assessment. In this instance the WOMAC and PGA score is set to missing.

In addition, all daily NRS scores affected by use of prohibited pain medication or paracetamol intake of cumulative daily dose of at least 4000 mg will be set to missing and not used for derivation of the weekly averages.

Furthermore, there is potential for unevaluable primary and key secondary variable data in cases where prohibited pain medication was taken for less than 4 days over the 7-day period that is being considered for deriving the weekly NRS average or more than one day prior to WOMAC or PGA assessment.

Participants with potentially unevaluable data will be reviewed prior to study unblinding by the study team who will confirm whether the data is deemed unevaluable or not. Only evaluable data for such participants will be included in the analyses.

Table 2 Intercurrent events and imputation strategy

Intercurrent event	**Single imputation strategy (Used for tabulation purposes)**	**Multiple imputation strategy (Used for formal analyses)**
Discontinuation due to lack of efficacy or an AE	Missing data imputed assuming an unfavourable outcome, that is the BOCF imputation	MNAR: BOCF approach
Taking prohibited pain medication during the treatment period	Data affected will be set to missing and then imputed assuming an	

Taking excessive permitted rescue medication	unfavourable outcome, that is the BOCF imputation	
Discontinuation due to other reasons such as loss to follow-up or external circumstances	Missing data imputed assuming missing at random, that is the LOCF imputation	MAR approach

BOCF, baseline-observation-carried-forward; LOCF, last-observation-carried-forward; MAR, missing at random; MNAR, missing not at random.

More details concerning the imputation of missing or unevaluable primary endpoint data can be found in section 4.1.1.1.

3.3 Secondary efficacy variables

The following are the key secondary efficacy endpoints:

- Change in the WOMAC pain subscale from baseline to Week 12
- Change in the WOMAC physical function subscale from baseline to Week 12
- Change in the PGA of OA from baseline to Week 12

The same attributable estimand strategy as well as single LOCF imputation as for the primary efficacy endpoint will be applied for these key secondary efficacy endpoints. In addition, missing Week 12 WOMAC pain subscale scores will be imputed using a differing MI approach as described in section 4.1.1.1., as a supportive analysis to compare with previously published studies.

The other secondary efficacy endpoints are summarised in section 1.1.2. However, data of all available per protocol visits will be used for analysis and summarised (Weeks 2, 4, 6, 8, 10, 12 and 18) unless otherwise specified.

Finally, the ET assessments of all secondary endpoints will be handled in the same manner as for primary endpoint that is will be reassigned based on the study day.

3.3.1 WOMAC multiscale index

The Western Ontario and McMaster Universities Osteoarthritis (WOMAC) multiscale index is used to assess pain, stiffness, and joint functionality in the past 48 hours in participants with OA of the knee or hip. The WOMAC questionnaire will be completed for the target knee at Visit 2 (Week 0, baseline), Visit 4 (Week 2), Visit 5 (Week 4), Visit 6 (Week 6), Visit 7 (Week 8), Visit 8 (Week 10), Visit 11 (EOT or ET Week 12) and Visit 12 (Week 18) for only those participants who complete the treatment period. On dosing days, the WOMAC questionnaire will be completed prior to IP administration.

3.3.1.1 WOMAC pain subscale

The WOMAC pain subscale consists of 5 questions assessing the participant's pain due to OA in the target knee. Each question is scored on an NRS scale from 0 to 10, where higher scores represent higher pain. The WOMAC pain subscale score is calculated as the mean score from all 5 questions.

3.3.1.2 WOMAC physical function subscale

The WOMAC physical function (PF) subscale consists of 17 questions assessing the participant's difficulty in performing activities of daily living due to OA in the target knee. Each question is scored on an NRS scale from 0 to 10, and the WOMAC PF subscale score is calculated as the mean score from all 17 questions, where higher scores represent worse function.

3.3.1.3 WOMAC stiffness subscale

The WOMAC stiffness function subscale consists of 2 questions assessing stiffness due to OA in the target knee. Stiffness is defined as a sensation of decreased ease of movement in the target knee. Each question is scored on an NRS scale from 0 to 10, and the WOMAC stiffness function subscale score is calculated as the mean score from the 2 questions, where higher scores represent higher stiffness.

3.3.1.4 WOMAC overall score

The WOMAC overall score consists of all 24 questions reported in the WOMAC questionnaire to assess: i) pain subscale, ii) PF subscale and iii) stiffness subscale. WOMAC overall score is calculated as the mean score from all 24 questions where higher scores represent worse outcome.

3.3.2 Patient's Global Assessment (PGA) of Osteoarthritis (OA)

The PGA of OA is a 5-point Likert scale used to assess symptoms and activity impairment due to OA of the knee. Participants are asked to identify a number from 1 = "very good (asymptomatic and no limitation to normal activities)" to 5 = "very poor (very severe symptoms which are intolerable and inability to carry out all normal activities)" based on the question "Considering all the ways that OA of the knee affects you, how are you feeling today?". The PGA questionnaire will be administered at the time points specified in the CSP (Refer to CSP Table 1). Participants who discontinue IP early will have their last PGA of OA assessment at the ET visit. On dosing days, the PGA assessment will be completed prior to IP administration.

3.3.3 OMERACT-OARSI responder index

The OMERACT-OARSI responder index is calculated from the WOMAC Pain subscale, the WOMAC Physical Function Subscale and the PGA of OA. A participant is classified as a responder if either of the below occurs:

1. ≥2-point absolute change from Baseline to Week X or a ≥50% improvement is reported in the WOMAC Pain or the Physical Function subscales.
2. At least 2 of the following 3 conditions are true:
    - ≥1-point absolute change from Baseline to Week X or ≥20% improvement is reported in the WOMAC Pain subscale.
    - ≥1-point absolute change from Baseline to Week X or ≥20% improvement is reported in the WOMAC Physical Function subscale.
    - ≥1-point absolute change from Baseline to Week X is reported in the PGA of OA.

3.4 Pharmacokinetic and immunogenicity variables

Blood samples for PK and ADA assessments will be analysed by bioanalytical test sites. All serum samples will be collected prior to study treatment administration (on dosing days) at the time points specified (See CSP Table 1). On non-dosing days, PK and ADA samples will be collected in the morning and the time of sample collection must be recorded. Note that participants who discontinue IP early will have their last PK sampling at the ET visit.

3.4.1 Pharmacokinetics

The serum MEDI7352 concentrations will be determined only for participants randomised to an active treatment.

At a visit/time point for a specific cohort where less than or equal to 50% of the concentration values are not quantifiable (NQ) that is are below the lower limit of quantification (LLOQ), all such values will be set equal to the LLOQ value, and all descriptive statistics will be calculated accordingly (using the value of the LLOQ that will be included in the summary statistics).

At a visit/time point for a specific cohort where more than 50% (but not all) of the values are NQ, the gmean, gmean ± gSD and gCV% will be set to NC. The maximum value will be reported from the individual data, and the minimum and median will be set to NQ.

If all concentrations are NQ for a specific cohort at a visit/time point, no descriptive statistics will be calculated for that time point. The gmean, minimum, median and maximum will be reported as NQ and the gCV% and gmean ± gSD as NC.

3.4.2 Immunogenicity

For immunogenicity analysis, the presence of detectable (i.e., positive) anti-drug antibodies (ADAs) against MEDI7352 will be reported. ADA results from each sample are reported as either positive or negative. In addition, the ADA titer result will be reported for samples confirmed positive for the presence of ADAs. A participant is defined as being ADA-positive

if at least one positive ADA result is available at any time (including baseline and all post-baseline measurements); otherwise ADA negative.

The following ADA categories will be determined:

- **ADA positive** if a collected sample is tested positive at any time during the study, including baseline and/or post-baseline. (The percentage of these participants in a population is known as ADA prevalence)
- **Treatment-emergent ADA positive (TE-ADA+):** A positive post-baseline result and either of the following statements holds:
  - Baseline is ADA negative and at least one post-baseline assessment is ADA positive. This is called **treatment-induced ADA positive.**
  - Baseline is ADA positive, and the baseline titre is boosted by greater than the variability of the assay (i.e. ≥ X-fold increase, commonly 4-fold) at ≥1 post-baseline timepoint. This is called **treatment-boosted ADA positive.**
    (The percentage of these participants in a population is known as **ADA incidence)**
- **Non-Treatment-emergent ADA positive (non-TE-ADA+):** Participants who are ADA positive but not fulfilling the conditions for TE-ADA+
- **Treatment-emergent Persistently ADA positive**: ADA negative at baseline and having at least 2 post-baseline ADA positive measurements with ≥ 16 weeks (112 days) between first and last positive, or an ADA positive result at the last available post-baseline assessment
- **Treatment-emergent Transiently ADA positive**: ADA negative at baseline and at least one post-baseline ADA positive measurement and not fulfilling the conditions for persistently positive
- ADA positive post-baseline and positive at baseline.
- ADA not detected post-baseline and positive at baseline.

ADA evaluable participants are defined as the participants in the safety analysis set who have a non-missing baseline and at least one non-missing post-baseline ADA results.

3.5 Exploratory variables

3.5.1 Supportive efficacy and health related quality of life variables

The supportive efficacy and health related quality of life endpoints are summarised in section 1.1.4.

3.5.1.1 Daily Sleep Interference Scale (DSIS)

DSIS is an 11-point Likert scale used to assess how pain interferes with participants' sleep where 0= pain did not interfere with sleep and 10 = pain completely interfered with sleep. This will be recorded on a daily basis (upon awakening) via electronic patient recorded

outcome (ePRO) diary until Week 18 unless participants discontinue the IP early in which case last recording DSIS would be at ET visit.

The derivation of the averages will be the same as for the NRS pain score described in section 3.2 including the handling of the ET assessments. However, the average will be derived regardless of how many daily scores are available.

3.5.1.2 European Quality of Life-5 Dimensions (EQ-5D-5L)

The EQ-5D-5L questionnaire assesses 5 dimensions of health (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression). For each dimension, respondents will select which statement best describes their health on that day from a possible 5 options of increasing levels of severity (no problems, slight problems, moderate problems, severe problems, and extreme problems).

In addition to the descriptive system, the participants are asked to rate their health on the day of assessment on a visual analogue scale (VAS), ranging from 0 (worst imaginable health) to 100 (best imaginable health). This score is reported separately.

The European Quality of Life-5 Dimensions (EQ-5D-5L) may be analysed separately and not form part of this SAP.

3.5.1.3 36-Item Short Form Health Survey Questionnaire

The 36-Item Short Form Health Survey Questionnaire version 2 (SF-36v2) is a 36-item questionnaire on functional health and well-being, with a 1- to 4-week recall period. Responses to 35 of the 36 items are used to compute an 8-domain scores and two component summary measures. The remaining item (the 'health transition' item), asks participants to rate how their current state of health compares to their state of health one year ago. This item is not used to calculate domain scores.

The 8-domains are: Physical Functioning (PF), Role Limitations due to Physical Health (RP), Bodily Pain (BP), General Health Perceptions (GH), Vitality (VT), Social Functioning (SF), Role Limitations due to Emotional Problems (RE), and Mental Health (MH). The component summary measures (physical component summary [PCS] and mental health component summary [MCS]) are computed from domain scores to give a broader metric of physical and mental health-related quality of life. The transformed score range for each of the 8 domains and for PCS and MCS is 0-100; higher scores indicate better health state.

The change from baseline in SF-36v2 (PCS and MCS) scores will be calculated.

The threshold values for the SF-36v2 PCS, MCS, and domain scores listed in Table 3 are suitable for interpreting change at the participant level and are referred to as the responder thresholds or responder definitions (Maruish 2011).

Table 3 Threshold values for the SF-36v2 scale and summary measures

	SF-36v2 score									
Threshold	PCS	MCS	PF	RP	BP	GH	VT	SF	RE	MH
Group difference	2	3	3	3	3	2	2	3	4	3
Individual change	3.4	4.6	4.3	3.4	6.2	7.2	6.2	6.9	4.5	6.2

BP, bodily pain; GH, general health perceptions; MCS, mental health summary; MH, mental health; PCS, physical component summary; PF, physical functioning; RE, emotional problems; RP, role limitations due to physical health; SF, social functioning; VT vitality.

Categorical analyses of the change in each of the SF-36v2 scales and summary measures will be compared to its respective individual change threshold (see Table 3) to indicate whether each participant experienced a change deemed clinically significant and important. A participant will be classified as a responder if the change from baseline ≥ individual change threshold, or a non-responder if change from baseline < individual change threshold. If data are missing, then the participant will be classified as a non-responder.

3.5.1.4 Work Productivity and Activity Impairment questionnaire for OA (WPAI:OA)

The WPAI:OA is a self-administered tool comprised of 6 questions that measure absenteeism, presenteeism, work productivity, and activity impairment due to OA of knee. This tool captures data from the past 7 days. Each subscale score is expressed as a percentage impairment from 0 to 100, where higher numbers indicate greater impairment and less productivity.

The following WPAI:OA impairment scores are defined: i) Percent activity impairment due to Osteoarthritis, ii) Percent impairment while working due to osteoarthritis, iii) Percent overall work impairment due to osteoarthritis and iv) Percent work time missed due to Osteoarthritis. Further details about the questions and derivations of those endpoints are shown in Appendix A1.

3.5.1.5 Healthcare Resource Utilisation (HCRU)

Healthcare resource utilisation information (e.g., doctor office visits, hospitalisations, surgeries, or other procedures) used by each participant are collected by the investigator and study-site personnel for all participants at baseline (i.e., collecting data up to 3 months before) and various time points during the study as specified (See CSP Table 1). Protocol-mandated procedures, tests, and encounters are excluded. The responses and scores to the HRCU questionnaire are presented in Appendix A2 with questions recalling the period between visits reported in weeks for all visits except for baseline.

3.5.1.6 Lack of efficacy endpoints

The investigator is responsible to determine if a participant should discontinue treatment due to lack of efficacy and report the same on the DS_EOT page of eCRF.

For each participant, the time to permanent discontinuation of IP due to lack of efficacy will be determined as follows:

(Date of IP discontinuation as reported on the DS_EOT page of eCRF) – (date of first IP administration) +1

3.5.1.7 Rescue medication use

Rescue medication (paracetamol) use will be collected via the participant eDiary from Day -14 until Week 18. Participants who discontinue IP early, will stop recording rescue medication after the ET visit.

The following variables will be derived and/or summarised:

- Number of participants using the rescue medication
- Reason for using rescue medication
- Total number of days rescue medication was used
  - Each day on which rescue medication was used at least once is counted
- Cumulative consumption (mg) of rescue medication use
  - Calculated as the total dose of rescue medication (mg)
- Average daily dose (mg) of rescue medication use
  - Calculated as: cumulative consumption of rescue medication (mg) /total number of days rescue medication was used

In addition to the above and in order to evaluate the changes in rescue medication use (number of days and cumulative consumption) over time, the daily collected data will be assigned to a specific study week. This assignment will use the same principles as deriving the NRS pain scores. For details refer to the section 3.2.

3.5.2 Biomarker and Pharmacodynamic variables

The biomarker endpoints are defined in the section 1.1.4.

However, only the below will be considered for analyses and CSR reporting depending on data availability at the end of the study (final clinical database lock):

- Total NGF measurements in serum over time (also denominated as pharmacodynamic variable/endpoint)

- Baseline biomarkers/pharmacodynamics namely Total NGF, TNF alpha, IL-1 beta, IL-6 and high sensitivity C-reactive protein (CRP)

Additional endpoints may be analysed providing such data are available.

Samples for biomarker research are required and will be collected from all participants in this study as specified (see CSP Table 1). Note that participants who discontinue IP early will have their last pharmacodynamics/biomarkers sampling at the ET visit.

3.5.3 Ambulatory blood pressure monitoring (ABPM) variables

The ABPM endpoints are presented in section 1.1.4.

During ABPM, the systolic BP, diastolic BP, pulse pressure, pulse rate recorded as heart rate by the vendor, and mean arterial pressure readings will be recorded over a period of 24 hours with frequency of readings during the day (every 15 minutes) and overnight (every 30 minutes). Participants will wear the ABPM device for a period of 24 hours at 2 separate times during the study: one time during the screening period and one time during the double-blind treatment period (between Weeks 8 and 12 inclusive).

An average of the individual readings during the day as well as the night will be determined for each participant. The 24-hour assessments will be derived as an overall average of all readings of a specific parameter for each participant for that specific visit (screening or visit within Weeks 8 and 12).

Individual readings with missing or zero values or values out of the acceptable ranges as identified in Table 4 below as well as those set to 'not done' will not be included in the calculation of averages.

Table 4 Acceptable ABPM ranges

Parameter	ABPM acceptable ranges
Systolic blood pressure (SBP)	10 - 250 mmHg
Diastolic blood pressure (DBP)	5 - 200 mmHg
Pulse Pressure (PP)	5- 250 mmHg
Mean Arterial Pressure (MAP)	10 – 250 mmHg
Heart Rate (HR)	1 - 200 bpm

In addition, the entire set of the readings for a specific visit will be considered of poor quality and hence not used for analysis if any of the following is met:

- There is less than 20h of the total readings available for a specific parameter and visit for each participant

- There are less than 70% of valid measurements out of the total measurements recorded for a specific parameter and visit for each participant

Such readings that should be excluded will be identified by the vendor according to the above-mentioned criteria.

ABPM will be classified as normal if average is within normal reference range presented in Table 5 below and abnormal if the average is outside of this reference range.

Table 5 Normal ABPM ranges

Parameter	ABPM normal reference ranges
Day/Night/24-hours SBP	70 - 240 mmHg
Day/Night/24-hours DBP	40 - 150 mmHg
Pulse Pressure (PP)	20 - 150 mmHg
Mean Arterial Pressure (MAP)	40 - 200 mmHg
Day/Night/24-hours HR	20 - 200 bpm

3.6 Safety variables

Safety and tolerability of MEDI7352 will be evaluated based on the AEs, physical and neurological examination, total neuropathy scores, vital signs, survey of autonomic symptoms, electrocardiograms, injection site reactions, radiographic and magnetic resonance imagining assessment and clinical laboratory assessment.

3.6.1 Adverse events

AEs and serious AEs (SAEs) as defined in the protocol appendix C will be collected from the screening, throughout the study, and including the follow-up period. The Medical Dictionary for Regulatory Activities (MedDRA) will be used to code the AEs, using the latest version. AEs will be graded according to the Common Terminology Criteria for AEs (CTCAE version 5).

A treatment emergent adverse event (TEAE) is defined as:

- any AE with an onset at the time of or following the start of study treatment through the end of study participation as per DS page in eCRF;

- any AE starting prior to the start of study treatment administration but increasing in severity at the time of, or following, the start of study treatment through end of study participation as per DS page in eCRF.

Any AE occurring before the first dose of study treatment and without worsening after the first dose will be included in the data listings but will not be included in the AE summary tables.

The causal relationship between IP and each AE will be recorded. For SAEs, causal relationship will be assessed for other medication and study procedures.

3.6.1.1 Adverse events of special interest

An AE of special interest (AESI) is one of scientific and medical interest specific to understanding of the IP and may require close monitoring and collecting additional information by the investigator. An AESI may be serious or nonserious. The rapid reporting of AESIs allows ongoing surveillance of these events to characterise and understand them in association with the use of MEDI7352.

AESIs for MEDI7352 include, but are not limited to:

- Positively-adjudicated possible or probable Rapidly Progressive Osteoarthritis (RPOA), Subchondral Insufficiency Fracture (SIF), primary osteonecrosis, or pathological fracture
- Infections that meet SAE or severe AE criteria (see CSP Appendix C 2)
- Anaphylactic reactions; serious or severe hypersensitivity reactions; or injection site reactions that lead to permanent discontinuation of administration of IP
- Other – includes AESIs that do not meet any of the above criteria, but are deemed of interest by the study team

AESIs irrespective of their severity and seriousness should be reported immediately using the same procedure as for SAE reporting (See CSP section 8.3.9).

3.6.1.2 Adverse events based on signs and symptoms

All AEs spontaneously reported by the participant or care provider or reported in response to the open question from the study site staff: ‘Have you had any health problems since the previous visit/since you were last asked?’ or revealed by observation will be collected and recorded in the eCRF.

3.6.1.3 Adverse events based on examinations and tests

Deterioration as compared to baseline in protocol-mandated laboratory values and vital signs will only be reported as AEs if they fulfil any of the SAE criteria, are the reason for discontinuation of treatment with the IP or are considered to be clinically relevant as judged by the investigator (see CSP section 8.3.7).

In addition, occurrences in liver biochemistry of AST or ALT ≥ 3 × ULN together with total bilirubin ≥ 2 × ULN will be reported as SAEs as per Hy's Law (see CSP Appendix F).

3.6.1.4 Pregnancy

Abnormal pregnancy outcomes (e.g., spontaneous abortion, foetal death, stillbirth, congenital anomalies, and ectopic pregnancy) are considered SAEs. A woman who becomes pregnant during IP treatment or within 30 days of discontinuing the IP will be immediately withdrawn from study participation. Every pregnancy will be documented in EDC using the pregnancy report (PREGREP) page in eCRF.

3.6.2 Physical examination

The following assessments will be performed as part of the physical examination:

> General appearance, skin, head and neck, examination of the oral cavity for any lesions, lymph nodes, thyroid, abdomen (bowel sounds, liver, and spleen palpation), back (including costovertebral angle tenderness), musculoskeletal/extremities, cardiovascular, and respiratory systems.

The physical examination will be performed during screening, treatment period and follow-up period, as indicated in CSP (Table 1).

3.6.2.1 Weight

Weight will be measured at screening and at Week 12 as well as when change is suspected, or fluid retention occurs.

3.6.3 Neurological examination

The neurological examination performed ideally by a neurologist include:

> Assessment of mental status, cranial nerves, motor examination (muscle strength and tone), upper and lower extremity deep tendon reflexes, plantar responses, sensory system examination, coordination, and gait.

The neurological examination will be performed during screening, treatment period and follow-up period, as indicated in CSP (Table 1).

For participants with clinical manifestations consistent with worsening of pre-existing neuropathy, follow-up nerve conduction studies should be performed and documented at the time points deemed appropriate by the neurologist, but at least once during participation in the study.

3.6.4 Total neuropathy score- nurse (TNSn) variables

The total neuropathy score – nurse (TNSn) assessment is collected as scores as follows:

- Sensory symptom score
  - Sensory symptoms consist of six questions, each rated from 0 to 4.
  - Sensory symptom score is determined as the maximum of the six scores.
- Motor symptom score
  - Motor symptoms consist of four questions, each rated from 0 to 4.
  - Motor symptom score is determined as the maximum of the four scores.
- Autonomic symptom score
  - Autonomic symptoms consist of six symptoms rated as present of not (yes/no)
  - Autonomic symptom score is determined as the number of symptoms present, with a maximum score of 4 assigned for four or more symptoms present.
- Pin sensibility score
  - Pin sensibility is scored as 0 to 4 in upper and lower limbs both left and right, therefore four areas in total.
  - Pin sensibility score is determined as the maximum of the four scores.
- Vibration sensibility score
  - Vibration sensibility is scored as 0 to 4 in upper and lower limbs both left and right, therefore four areas in total.
  - Vibration sensibility score is determined as the maximum of the four scores.
- TNSn total
  - The total score is determined as the sum of the five aforementioned scores.

The TNSn assessment will be performed during screening, treatment period and follow-up period, as indicated in CSP (Table 1).

The ET assessments will be handled in the same manner as for primary endpoint that is will be reassigned based on the study day.

3.6.5 Vital Signs

Blood pressure (BP) and pulse rate will be measured at each clinic visit with details about the method, time and frequency described in CSP section 8.2.3.1. Body temperature and respiratory rates will be measured at different time points as indicated in CSP (Table 1).

Absolute values will be compared to the reference ranges in Table 6 and classified as low (below range), normal (within range or on limits) or high (above range). All values (absolute) falling outside the reference ranges will be flagged.

Table 6 Vital signs reference ranges

Parameter	Standard Unit	Lower Limit	Upper Limit
Diastolic Blood Pressure	mmHg	60	120
Systolic Blood Pressure	mmHg	100	160
Pulse Rate	Beats/min	40	120
Respiratory Rate	Breaths/Min	8	28
Body Temperature	Celsius	36.5	38

3.6.5.1 Orthostatic blood pressure

Orthostatic (BP) measurements will be obtained at all clinic visits and taken after scheduled supine measurements (1 and 3 minutes after the participant stands) and prior to any required blood draw or study treatment administration (and prior to randomisation on Day 1).

After stable BP is achieved in supine position (see CSP section 8.2.3.2), the participant will stand, and a BP measurement will be taken at 1 and 3 minutes after the participant stands. If the BP measurements also referred to as postural changes do not meet the criteria for orthostatic hypotension, no additional measurement is needed. If the BP measurement meets any of the criteria shown in CSP Table 6, investigators will repeat the supine and standing measurements up to 2 additional times. Blood pressure changes meeting the pre-specified criteria and confirmed will be designated as a confirmed orthostatic hypotension episode.

3.6.6 Survey of Autonomic Symptoms (SAS)

The SAS is an instrument that measures autonomic symptoms used for assessing autonomic neuropathies in clinical trials.

The SAS evaluates the presence of a symptom and the degree of its severity. It consists of 11 questions in women and 12 in men.

Each question has a “Yes” or “No” answer to symptoms occurring 6 months prior to IP administration. Questions answered with “Yes” are further rated from 1 to 5 by asking the

participant how much each symptom is bothering him or her. Each answer is scored on a scale from 1 to 5 where 1 = “not at all” and 5 = “a lot”, and a total symptom impact score is determined as the sum of the ratings provided for all reported symptoms, that is questions answered with ‘Yes”.

The following domains are assessed:

Orthostatic, sudomotor symptoms, vasomotor, gastrointestinal, urinary, and sexual dysfunction.

The assessment is done at Day 1 and Week 36 only.

3.6.7 Electrocardiograms

Electrocardiograms (ECGs) will be obtained during screening, during treatment period and follow-up period as indicated in CSP (Table 1).

The investigator will judge the overall interpretation as normal or abnormal clinically non-significant or abnormal clinically significant, and this evaluation will be reported in the eCRF.

3.6.8 Laboratory data

Blood and urine samples for determination of clinical chemistry, haematology, coagulation, serology, drug screening, pregnancy tests, and urinalysis will be taken at the visits indicated in CSP Table 1.

Blood and urine samples for haematology, serum chemistry, coagulation, urinalysis, and serology will be sent to a central laboratory for analysis. Urine drug screens will be conducted at the study sites. Serum pregnancy test will be conducted at the central laboratory; urine pregnancy test will be conducted at study sites.

Samples for laboratory tests will include the following:

- Haematology: haemoglobin, haematocrit, red blood cell (RBC) count, RBC indices, mean corpuscular haemoglobin, mean corpuscular haemoglobin concentration, platelet count (or estimate), white blood cell count, and white blood cell differential
- Serum chemistry: albumin, total and direct bilirubin, total protein, calcium, alkaline phosphatase (ALP), alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma glutamyl transferase (GGT), blood urea nitrogen (BUN), creatinine*, glucose, sodium, potassium, chloride, bicarbonate, calcium, phosphorus, lactate dehydrogenase (LDH), uric acid, C-reactive protein (CRP), cholesterol, triglycerides, creatine phosphokinase.
- Coagulation panel: prothrombin time, activated partial thromboplastin time, fibrinogen
- Urinalysis: pH, specific gravity, blood, leukocytes, glucose, protein, ketones, bilirubin

- Urine drug screen: including but not limited to amphetamine, barbiturate, cannabinoids, cocaine, methadone, methaqualone, opiate, phencyclidine, and propoxyphene.
- Urine pregnancy test (only in female participants not surgically sterile)
- Serology: human immunodeficiency virus (HIV), hepatitis B, and hepatitis C screen, tuberculosis (TB) (QuantiFERON), and anti-citrullinated protein antibodies (ACPA) (screening only)
- Endocrinology: follicle-stimulating hormone (FSH) (Female < 50 years of age only and at screening only, to confirm menopausal state), β-HCG beta human chorionic gonadotropin (female participants, not surgically sterile only, at screening only).
- Other: haemoglobin A1c

*Note: Serum chemistry reports should include calculation of creatinine clearance using the Cockcroft-Gault equation.

HIV, hepatitis B and C and TB QuantiFERON data is considered sensitive and as such, the data will remain at the sites and will not be transferred further hence will not be part of the analysis.

Change from baseline in hematology, chemistry, coagulation, urinalysis, and haemoglobin A1c variables will be calculated for each post-dose visit as per the protocol defined schedule.

3.6.9 Injection site reactions

Injection site reactions occurring during or immediately following study treatment administration will be evaluated. Local reaction to injectable product will be evaluated for pain, tenderness, erythema/redness, and induration/swelling, and classified as mild, moderate or severe as per CSP (Table 7).
In addition,

- For erythema/redness, the measurement associated to the local reaction at the greatest single diameter should be recorded as a continuous variable.
- Induration/swelling should be evaluated as the actual measurement.

3.6.10 Radiographic (X-ray) and magnetic resonance imagining (MRI) assessments

3.6.10.1 X-ray assessments

All participants should undergo diagnostic x-ray on the shoulder, knee, and hip joints during screening period while at Week 32 visit, x-ray imagining is performed on the knee and hip joints. Any potentially clinically significant findings detected on X-ray performed at week 32 will be further assessed and sent for adjudication. In addition, unscheduled X-ray should be performed “for cause” to follow up on any worsening symptoms. These assessments will aid

identification of joint safety events such as RPOA, SIF, primary osteonecrosis, or pathological fractures.

3.6.10.2 MRI assessments

MRI of the bilateral knee joints must be done during screening period unless participant undergone a total joint replacement and has an artificial nontarget joint or when a knee joint is no longer present due to leg amputation. In addition, unscheduled MRIs should be done "for cause" in the knee joints or the joints with a KL grade at baseline of ≥ 3.

3.6.11 Study treatments

Study treatment in this study refers to MEDI7352, and placebo. See Table 7 for further details.

Table 7 Study treatments

	Study treatment name	Route of administration	Dosing instructions
Active	MEDI7352	CCI [redacted]	CCI [redacted]
Placebo	Saline solution		0.9% (w/v) saline q2w.

CCI [redacted]; q2w Every 2 weeks; CCI [redacted]

3.6.11.1 Exposure of study treatment

Exposure (i.e. duration of treatment) will be defined as follows:

Total (or intended) exposure of study treatment in days

- Total (or intended) exposure = min (last dose date + 13 days or date of death) – (first dose date) +1

The total participant years of exposure is the sum of duration of exposure (days) of all participants in the respective treatment group divided by 365.25 (years).

3.6.11.2 Compliance

Study treatment compliance with IP administration is calculated as:

(Total doses administered / total doses expected) x 100.

The total number of doses expected includes all visits with protocol scheduled IP administration before a participant's Week 12 (EOT/ET) visit.

4 ANALYSIS METHODS

Formal statistical testing of the primary and key secondary efficacy endpoints with maintaining overall family-wise type-I error at 1-sided 2.5% level will be performed as follows:

Multiple comparison procedure-modelling (MCP-Mod) statistical tests

The main hypothesis tested:

- H_0**:** No dose-response effect, i.e., the linear contrasts of group means corresponding to each of the 6 pre-specified candidate dose-response models are all identically zero.

- H_1: Dose-response effect, i.e., at least one of the linear contrasts of group means corresponding to each of the 6 pre-specified candidate dose-response models differs to zero.

The individual test statistic and corresponding p-value for each candidate model are calculated by following steps 1-3 in section 4.2.1.2.1.1.

The minimum of the multiplicity adjusted p-value corresponding to each candidate model will be used as the basis for concluding statistical significance.

Pairwise statistical tests

There will be 4 treatment comparisons:

- MEDI7352 CCI vs placebo
- MEDI7352 CCI vs placebo
- MEDI7352 CCI vs placebo
- MEDI7352 CCI vs placebo

For each pairwise comparison, the following hypothesis will be tested:

- H_0: No difference between MEDI7352 and placebo
- H_1: Difference between MEDI7352 and placebo

However, pairwise testing of only two highest doses (MEDI7352 CCI and MEDI7352 CCI) versus placebo of the primary and key secondary endpoints will undergo formal statistical testing.

The testing process will commence with the primary endpoint followed by key secondary endpoints using a sequentially rejective multiple comparison approach detailed in section 4.2.1.1.

4.1 General principles

Efficacy outputs will be presented by cohorts also referred to as the treatment arms, while safety outputs will also include 'MEDI7352 Total' defined as the sum of all MEDI7352 cohorts. Study population outputs will be presented by cohorts, MEDI7352 Total and Total (defined as all MEDI7352 cohorts + placebo), except otherwise specified.

The below mentioned general principles will be followed throughout the study:

- Descriptive statistics will be used for all variables, as appropriate, and will be presented by cohort. Continuous variables will be summarised by the number of observations, mean, standard deviation (SD), median, minimum, and maximum. Categorical variables will be summarised by frequency counts and percentages for each category.
- Unless otherwise stated, percentages will be calculated out of the analysis set total for the corresponding cohort.
- For continuous data, the mean and median will be rounded to 1 additional decimal place compared to the original data. The SD will be rounded to 2 additional decimal places compared to the original data. Minimum and maximum will be displayed with the same accuracy as the original data unless otherwise stated.
- For categorical data, percentages will be rounded to 1 decimal place.
- Results of all statistical analysis will be presented using a 95% confidence interval (CI) and 1-sided p-value, unless otherwise stated.
- P-values will be rounded to 3 decimal places, except p-values less than 0.0005 (e.g. 0.0002) will be displayed as <0.001. P-values outputted as <0.0001 by statistical software will not be modified, hence displayed as '<0.0001'.
- SAS® version 9.4 or higher will be used for all analyses.

4.1.1 Imputation Rules

4.1.1.1 Imputation of the primary and key secondary efficacy endpoints

The imputation will be performed for participants': mean weekly NRS scores, mean scores of each WOMAC subscale and PGA score.

Single imputation

Single imputation will be used for tabulation of the primary and key secondary endpoints as well as ANCOVA analysis of primary and key secondary endpoints. In this case the missing or unevaluable data at Week 12 will be imputed using:

- A single imputation method LOCF or BOCF as specified in Table 2 for tabulation only.
- A single LOCF imputation method for ANCOVA supplementary analysis and tabulation.

Multiple imputation

Generally, it will be assumed that missing observations for the primary and key secondary endpoints are missing at random. The exception are observations that are missing due to lack of efficacy or an AE, also values that are invaluable or missing due to taking prohibited pain medication or excessive use of permitted rescue medication as presented in Table 2. Those observations are assumed to be missing not at random (MNAR). Consequently, the pattern mixture framework (Ratitch B. and O'Kelly M., 2011) will be implemented by multiple imputation using steps summarized below, where 100 imputed datasets will be created across which the Rubin's rule will be applied to combine the estimates. Execution details are provided in Appendix B1. However, prior to executing the below steps, missing baseline of the primary and key secondary endpoints will be imputed using the SAS PROC MI statement MONOTONE REG with seed number =568000003 with number of iterations set to 100. KL grade as reported by Bioclinica also referred to as Clario will be included in the imputation model.

Step 1: Intermediate missing data imputation
The non-monotone missing values will be imputed using the Markov chain Monte Carlo method assuming MAR and multivariate normality. This step will be executed using the SAS procedure PROC MI with the MCMC option with seed number=568000003 and number of iterations set to 1 and number of burn-in iterations fixed at 200. The treatment and Kellgren and Lawrence (KL) grade variables will be included in the imputation model.

Step 2: Imputation of missing patterns
The missing/unevaluable values of each pattern that are defined by the intercurrent events as specified in Table 2 will be imputed via the sequential regression method using the SAS PROC MI statement MONOTONE REG with seed number =568000003 as follows:

- MAR -The missing/unevaluable week 12 values of primary and key secondary endpoints will be estimated from the observed and imputed data from step 1

using the sequential modelling approach. That is, prior outcomes of the primary or key secondary endpoints at each visit will be considered for imputing missing week 12 values under the MAR assumption. In addition, the treatment and KL grade variables will be included in the model.

- MNAR – The missing/unevaluable week 12 values of primary and key secondary endpoints will be estimated from the observed and imputed data from step 1 using the baseline observation carried forward approach. Non-missing baseline records from participants from both the placebo and active treatment arms will be used to impute post-discontinuation records from both the placebo and active treatment arms with KL grade included in the model.

Step 3: Combining the imputed outcomes and Statistical analyses
The imputed data from first two steps along with observed data will be analysed using the MCP-Mod and analysis of covariance (ANCOVA) model as described in section 4.2

MCP-Mod
The estimated dose response parameters ($\hat{\mu}^{(q)}$) with their covariance matrices ($\hat{S}^{(q)}$) from each imputed dataset $q = 1, \ldots, Q$ obtained from ANCOVA model as specified in section 4.2 will be combined using Rubin's rule (as elaborated by Shomaker and Heumann, 2014). For more information refer to Appendix C.
Dose-response model will be selected using the estimates combined above.

ANCOVA
The estimates obtained from the ANCOVA models will be combined using the MIANALYZE SAS procedure with adjusted degrees of freedom (EDF = option) based on the Rubin's method (Barnard and Rubin 1999).

In addition to the above, the missing Week 12 WOMAC pain subscale scores will be multiply imputed using sampling from a normal distribution approach (Appendix B2) depending on the reason for missing data as follows:

- *Discontinuation due to lack of efficacy or an AE or death:*
  Sampling from normal distribution with mean of participant's baseline score and standard deviation of the observed WOMAC pain subscale scores at Week 12 over all participants in FAS regardless of treatment assignment.

- *Missing data due to reasons other than those stated above:*
  Sampling from normal distribution with mean of participant's last available score and standard deviation of the observed WOMAC pain subscale scores at

Week 12 over all participants in FAS regardless of treatment assignment. It is possible that participant's last available score would be baseline score, in which case this score would be used.

Missing baseline of WOMAC pain subscale score will be also imputed using the same approach of imputing missing baseline of primary and key secondary endpoints described in the first paragraph of the multiple imputation section above. In addition, this MI with seed number =568000003 will also generate 100 imputed datasets. Each imputed dataset will be used for ANCOVA analysis, and its estimates combined using the MIANALYZE SAS procedure with adjusted degrees of freedom (EDF = option) based on the Rubin's method (Barnard and Rubin 1999).

4.1.1.2 Imputation of other study variables

Age at randomization will be derived for analysis purpose from the date of birth (DM module) and the randomization date (IE1 module) on the eCRF at screening as: year (randomization date) – year (date of birth), -1 if "day and month" of the randomization date is before "day and month" of the date of birth.

Participants with a partial date of birth (i.e., for those countries where year of birth only is given due to local regulatory constraints) will have an assumed date of birth of 1st Jan ([given year]).

Unless otherwise specified, safety assessment values of the form of "< x" (i.e. below the lower limit of quantification) or "> x" (i.e. above the upper limit of quantification) will be imputed as "x" in the calculation of summary statistics but displayed as "< x" or "> x" in the listings.

Adverse events: all AEs will be considered as treatment-emergent unless the opposite can be clearly stated. Imputation will be done only in the context of identifying TEAEs.

Medications/therapies: all medications will be considered as concomitant unless the opposite can be clearly stated.

Additionally, AEs that have missing causality (after data querying) will be assumed to be related to study treatment.

4.1.1.3 Handling of missing/incomplete dates

The original incomplete or missing dates will be presented in the listings.

In practice, partial AE or medication start dates will be imputed as follows:

- Missing day: Impute the 1st of the month unless month is same as month of first dose of study drug then impute first dose date. If the onset is before the first dose of study drug then impute with date of informed consent

- Missing day and month: impute 1st January unless year is the same as first dose date then impute first dose date
- Completely missing – impute first dose date unless the end date suggests it could have started prior to this in which case impute the 1st January of the same year as the end date

When imputing a start date, ensure that the new imputed date is sensible i.e., is prior to the end date of the AE or medication.

Partial AE or medication end dates will be imputed as follows:

- Missing day: impute day as the earlier of either the end date of the study or the last day of the month
- Missing day and month: as the earlier of either the end date of the study or the last day of the year
- Completely missing: if the event is not ongoing then input as the date of the last visit unless the start date is prior to first dose date in which case this date would be used for imputation

The same logic presented above will be used to impute partial dates of the start of the OA in order to derive the duration of OA based on the available date parts.

If a participant is known to have died where only a partial death date is available, then the date of death will be imputed as the latest of the last date known to be alive +1 from the database and the death date using the available information provided:

- For missing day only – using the 1st of the month
- For missing day and month – using the 1st of January
- For missing year – the analysis date will not be imputed

No other imputation will be made.

Conversions

- 1 year=365.25 days
- 1 month=30.4573 days

4.2 Efficacy analysis

Results of all statistical analyses will be presented using a two-sided 95% confidence interval (CI) and one-sided nominal p-value, unless otherwise stated. In addition, for all primary and key secondary endpoints the asymmetric confidence intervals will be produced as per Frewer P. et. al., 2016, unless otherwise stated. Finally, details of all participants affected by intercurrent events will be summarized in the key participant information output and

participants with missing data for primary and key secondary endpoints will be listed specifying reasons for missingness. Also, the statistical testing process with alpha adjustments at each stage as well as final alpha assignments will be tabulated.

Table 8 below details which efficacy endpoints are to be subjected to formal statistical analysis, making it clear which analysis is regarded as primary for that endpoint.

Table 8 Pre-planned statistical analyses to be conducted

Endpoints analysed	Notes

Primary endpoint:
Change in weekly average of daily NRS pain scores from baseline to Week 12
	MCP-MOD for:
	• **Primary analysis**
	  ○ Presence of dose response effect for each of the 6 prespecified candidate dose model.
	• **Supportive analysis**
	  ○ Estimation of the dose-response parameters and target dose of MEDI17352 from the chosen model
	ANCOVA for:
	• **Secondary analysis**
	  ○ Pairwise comparisons of MEDI17352 CCI and CCI doses versus placebo
	• **Supportive analysis**
	  ○ Pairwise comparisons of MEDI17352 CCI and CCI doses versus placebo

Key secondary endpoints:
Change in the WOMAC pain subscale from baseline to week 12
	MCP-MOD for:
	• **Other analysis**
	  ○ Presence of dose response effect for each of the 6 prespecified candidate dose model.
	• **Supportive analysis**
	  ○ Estimation of the dose-response parameters and target dose of MEDI17352 from the chosen model
	ANCOVA for:
	• **Other secondary analysis**
	  ○ Pairwise comparisons of MEDI17352 CCI and CCI doses versus placebo
	• **Supportive analysis**
	  ○ Pairwise comparisons of MEDI17352 CCI and CCI doses versus placebo

Change in the WOMAC physical function subscale from baseline to week 12	ANCOVA for: • **Other secondary analysis** o Pairwise comparisons of MEDI17352 CCI and CCI doses versus placebo • **Supportive analysis** o Pairwise comparisons of MEDI17352 CCI and CCI doses versus placebo
Change in the PGA of OA from baseline to week 12	ANCOVA for: • **Other secondary analysis** o Pairwise comparisons of MEDI17352 CCI and CCI doses versus placebo • **Supportive analysis** o Pairwise comparisons of MEDI17352 CCI and CCI doses versus placebo

ANCOVA, Analysis of covariance; MCP-Mod, Multiple comparison procedure modelling; NRS, numerical rating scale; OA, osteoarthritis; PGA, Patient's Global Assessment; WOMAC, Western Ontario and McMaster Universities Osteoarthritis.

4.2.1.1 Multiple testing strategy

To control the overall family-wise type 1 error at 2.5% (1-sided) across primary and key secondary endpoints for pairwise testing, the sequentially rejective multiple comparison approach will be used as described by Bretz et al (2009).

Once the statistical significance is demonstrated for MCP-Mod statistical testing hypothesis for the primary endpoint as defined in section 4 then the testing process will commence for the remaining pre-defined hypotheses as summarised in Table 9 below.

Table 9 Pre-defined hypothesis and hierarchy to be tested

Hypothesis #	Endpoint	Test of:
H1	Change from baseline in weekly average NRS scores to week 12	Dose-response effect
H2	Change from baseline in weekly average NRS scores to week 12	Pairwise comparison: Difference between MEDI7352 CCI vs placebo
H3	Change from baseline in weekly average NRS scores to week 12	Pairwise comparison: Difference between MEDI7352 CCI vs placebo
H4	Change from baseline in WOMAC pain subscale scores to week 12	Dose-response effect
H5	Change from baseline in WOMAC pain subscale scores to week 12	Pairwise comparison: Difference between MEDI7352 CCI vs placebo
H6	Change from baseline in WOMAC pain subscale scores to week 12	Pairwise comparison: Difference between MEDI7352 CCI vs placebo
H7	Change from baseline in WOMAC physical function subscale scores to week 12	Pairwise comparison: Difference between MEDI7352 CCI vs placebo

H8	Change from baseline in WOMAC physical function subscale scores to week 12	Pairwise comparison: Difference between MEDI7352 CCI vs placebo
H9	Change from baseline in PGA of OA to week 12	Pairwise comparison: Difference between MEDI7352 CCI vs placebo
H10	Change from baseline in PGA of OA to week 12	Pairwise comparison: Difference between MEDI7352 CCI vs placebo

Hypotheses 2 to 10 will be tested using the sequential graphical weighted-Bonferroni approach described in Figure 2. The sequence isn't fixed as such but adapts to the result of testing of each hypothesis as specified in Figure 2 which also details the sequence of the formal statistical testing with the respective weights Bretz et al (2009).

The nominal p-values for all statistical hypotheses testing as per the section 4 will be presented in the outputs. However, if the null hypotheses are not rejected by the formal testing approach described above then no inference will be drawn from the nominal p-values.

Figure 2 Formal statistical testing strategy

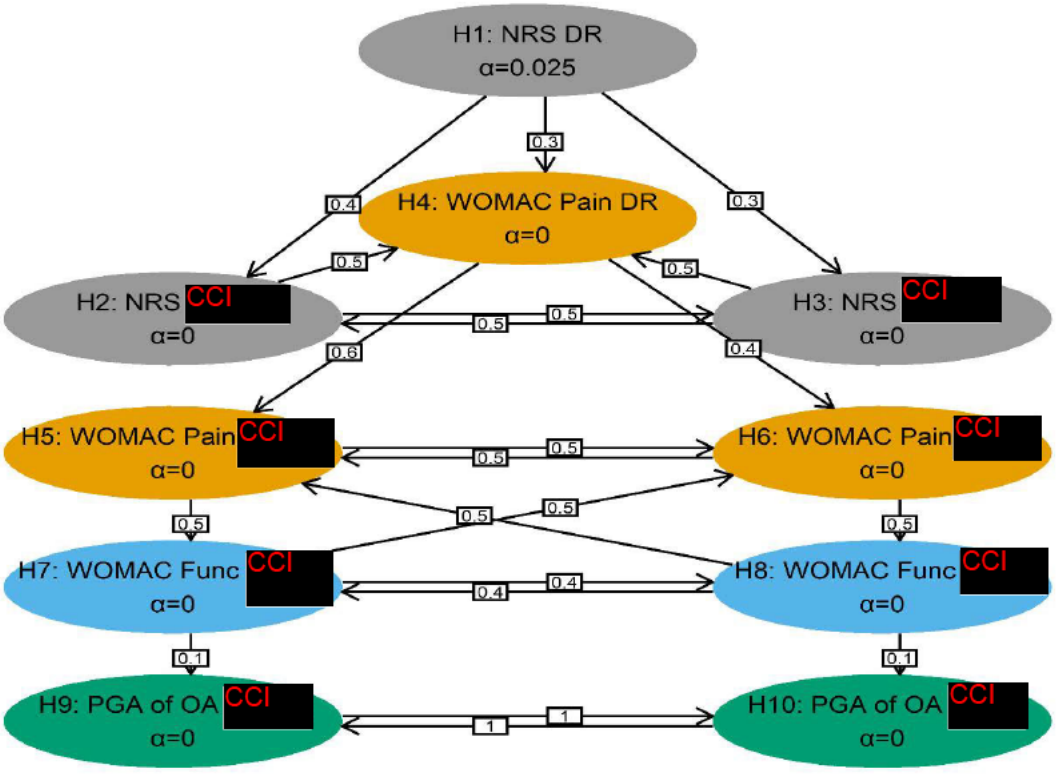


Figure 2 shows that full 1-sided alpha (p=0.025) is spent on the primary analysis (NRS Dose-Response test). Alpha is shared between doses for pairwise testing but weighted by power i.e., more allocated to CCI compared to CCI If NRS Dose-Response is significant, then test NRS Pain pairwise and WOMAC Dose-response simultaneously i.e. not all alpha is used for pairwise testing of NRS.

Generally, for any given endpoint, alpha is allowed to pass between low and high doses for pairwise testing.

Specifically, WOMAC pain is only tested if the WOMAC pain Dose-Response test is statistically significant. And WOMAC function is only tested if WOMAC pain is statistically significant on the same dose. Finally, if WOMAC function is statistically significant on a given dose then alpha is passed to WOMAC endpoints on other dose and PGA of OA on same dose, but with lower weight for the test of PGA of OA.

4.2.1.2 Analysis of the primary and key secondary endpoints

4.2.1.2.1 Primary endpoint

The primary endpoint is the change in weekly average of daily NRS pain scores from baseline to Week 12.

A summary table of the weekly average of daily NRS pain score will provide summary statistics on baseline and the change from baseline to Week 12 for:

- 'Observed cases' only (i.e. no imputation methods applied)
- According to single imputations (combination of LOCF/BOCF only)
- 'Observed cases' only (i.e. no imputation methods applied) by ADA positive and negative participants as defined in section 3.4.2.

In addition, a figure presenting change from baseline in NRS weekly average pain scores over time per treatment arm, based upon observed data and the mean of the multiply imputed datasets, will be produced as well as a figure of the cumulative distribution function of change from baseline for observed cases, with horizontal axis truncated to only show improvement. Finally, the NRS weekly averages will be listed.

4.2.1.2.1.1 Primary analysis of the primary endpoint

The primary efficacy analysis aims at evaluating the dose response relationship of the change in weekly average of daily NRS pain scores from baseline to Week 12. The main statistical analysis will use the MCP-Mod methodology to test the statistical hypothesis presented in section 4.

Multiple comparison procedure modelling (MCP-Mod)

There are two main steps to MCP-Mod:

- The 'MCP' step is a rigorous method used to establish the presence of a dose response while protecting the type I error and, if the dose-response relationship is statistically significant, then

- The 'Mod' step uses the selected candidate model(s) to produce inference on adequate doses, employing a model-based approach. That is producing estimates of the dose response function and associated model parameters, such as ED_{50} in the case of E_{max} relationship.

The underlying model will be an ANCOVA with dependent variable 'change from baseline to Week 12', and independent variables will include dose, 'baseline score' and KL grade at baseline as reported by Bioclinica. The random error is assumed to be normally and independently distributed with constant variance.

The KL grade covariate records the severity of knee OA using five grades (0-4) which are expected to be ≥3 at baseline, where 3 corresponds with moderate (moderate multiple osteophytes, definite narrowing of joint space and some sclerosis and possible deformity of bone ends) and 4 corresponds with severe (large osteophytes, marked narrowing of joint space, severe sclerosis, and definite deformity of bone ends).

According to Bornkamp B., Bezlyak V., and Bretz F.; 2015, these MCP and Mod steps can be further broken down as detailed below.

Step 1: Set of candidate dose-response shapes
This step has been established at the trial design stage, where 6 candidates from three dose-response model types were selected as follows:

Model Name	Model f (d, θ)	θ_2 (shape of the model)
E_{max}	$E_0 + E_{max}d/ (d + ED_{50})$	ED_{50} (CCI) ED_{50} (CCI) ED_{50} (CCI)
Linear	$E_0 + \delta d$	

Exponential	$E_0+E_1(\exp(d/\delta)-1)$	δ (CCI), corresponding to 10% of the maximum drug effect at CCI δ (CCI), corresponding to 25% of the maximum drug effect at CCI

Where:

- $\mu(d,\theta)=\theta_0+\theta_1 f^0(d,\theta_2)$ with q_0 and q_1 parameters that determine the location and scaling of the function f, and $f^0(d, q_2)$ is a nonlinear transformation of the dose levels depending on q_2
- E_0 is the basal effect at d=0
- E_1 is the scale parameter in the exponential model
- d is the dose
- ED_{50} is the dose that gives 50 percent of the maximum effect
- d is the rate of change for linear and exponential models

Step 2: Calculation of optimal contrasts and dose response parameters

Optimum contrast coefficients will be computed for each model to maximize the marginal power to detect a specific dose-response shape that is associated with the corresponding candidate model (Bornkamp B., Bezlyak V., and Bretz F.; 2015).

The computation of optimal covariate-adjusted contrasts is achieved by %optcont macro using the combined estimated covariance matrix $\hat{S}$ (Pinheiro et al.; (2004) and the candidate models. This matrix and estimated dose response parameters ($\hat{\mu}$) are generated using the estimated LSMEANS ($\hat{\mu}^{(q)}$) and covariance matrices of the LSMEANS $\hat{S}^{(q)}$ extracted from PROC GENMOD performed on multiply imputed datasets and then combined using the Rubin's rule (Shomaker and Heumann, 2014) as described in section 4.1.1.1. SAS example code to execute this step (excluding imputation) can be found in Appendix D1 .

Step 3: Dose response signal detection

The significance of the individual candidate dose-response model is tested in this step. This is done in terms of multiple linear contrast tests based on the optimal contrast coefficients calculated above. The significance of the individual candidate dose-response shapes will be evaluated based on the statistical hypotheses presented in section 4 where the individual test statistic and corresponding p-value for each candidate model are calculated using the estimated covariance matrix $\hat{S}$, the contrasts coefficients and the estimated dose response parameters calculated in step 1 using the methodology detailed in Appendix D2. Consequently, if the minimum adjusted p-value for the optimal contrasts for the candidate models is less than alpha (0.025), the null hypothesis will be rejected and the proof of concept of a non-null dose-response will be demonstrated.

PROC IML will be used to calculate the test statistics and corresponding p-values since it allows for the multiplication of vectors and matrices and sampling from a multivariate normal distribution. The seed 568000003 will be used when sampling from the multivariate normal distribution. Example SAS code is also included in Appendix D2.

If none of the candidate models are statistically significant, the procedure stops, suggesting that a dose-response relationship cannot be established from the observed data. However, in such circumstances the dose-response model corresponding to the MCP test contrast with the lowest p-value will still be estimated.

Step 4: Model selection
Out of the statistically significant candidate models the most appropriate one will be selected using the model selection criteria gAIC. Additional information alongside an example code can be found in Appendix D3.

The best/optimal model will be the one with the minimum gAIC criterion. The gAIC for each model will be displayed in the table and the optimal model with the minimum gAIC will be indicated.

Step 5: Dose-response and dose estimation (Target dose selection)
The estimates from the optimal dose-response model will be used to calculate the minimum effective dose (MED), ED50, 90% of the Emax (ED90) and dose to achieve selected target effects if/when applicable. The MED is defined as the minimum dose that the chosen model predicts will achieve 1.0, 1.2 and 1.5-point improvements over placebo. For details concerning the derivations of MED and other parameters refer to the Appendix D4.1.

To obtain the 95% CIs for the MED and estimates other than model estimates of the treatment effect at doses studied, the asymptotic normal distribution of the estimates from the optimal dose-response model will be determined as per Appendix D4.2.1. 10,000 samples from the asymptotic normal distribution will then be drawn (using seed 568000003) with each set of samples used to create a dose response curve and corresponding MED. The 0.025 and 0.975 percentiles of the MED values will be used to form the 95% confidence intervals. Same approach will be used to obtain the 95% CIs for estimates other than model estimates of the treatment effect at doses studied.

The 95% CIs for estimates associated with treatment comparisons will be constructed using the delta method described in Appendix D4.2.2.

Finally, the MCP-Mod process will be reported in a summary table including:

- Candidate dose-response models test statistics (including adjusted p-value)
- gAIC values for all candidate dose-response models (lowest value shows model with best fit)

- Model parameter estimates such as Emax, ED50, slope, delta, E1 with their corresponding 95% CIs
- Minimum effective doses: MED 1.0, 1.2 and 1.5 and ED90 with their corresponding percentile-based 95% CIs only for optimal model with the lowest gAIC. There will be two types of ED90 estimated. The first applies to all models and is defined as the dose needed to produce 90% of the effect of the highest dose and the "asymptotic ED90", which is only applicable to the Emax model. Therefore the "asymptotic ED90" and corresponding CI will be computed only if Emax model is selected as optimal model.
- Model estimates of the treatment effect at doses studied together with 80% and 90% 1-sided confidence intervals as well as 95% CIs only for optimal model with the lowest gAIC.

4.2.1.2.1.2 Secondary analysis of the primary endpoint

As a secondary analysis, pairwise testing will be performed as described in section 4.
The same ANCOVA model used for the MCP-Mod procedure will be used on singly and multiply imputed data as specified in section 4.1.1.1 and results including the 80% upper and 90% lower 1-sided confidence intervals as well as 95% CIs and nominal p-values tabulated.

4.2.1.2.1.3 CCI

CCI [redacted]

4.2.1.2.1.4 Sensitivity analyses of the primary endpoint

Sensitivity analyses of the attributable estimand will explore different 'missing not at random' imputation approaches other than return-to-baseline such as 'jump to' reference as described by Cro S. *et. al*., 2020. That is, the same imputation process as described in section 4.1.1.1 will be used except that in step 2 instead of using the BOCF approach for MNAR, the data of placebo participants would be used as a reference.

4.2.1.2.2 Key secondary endpoints

4.2.1.2.2.1 Change in WOMAC pain subscale scores

The same analysis as applied to the primary endpoint will be applied to the change from baseline to Week12 in WOMAC pain subscale scores.

The same summary outputs including listing will be produced as produced for the primary endpoint detailed in section 4.2.1.2.1.

4.2.1.2.2.2 Change in WOMAC physical function scores

The same analysis as applied to the primary endpoint, excluding the MCP-MOD analysis, will be applied to the change from baseline to Week12 in WOMAC physical function scores.

The same summary outputs including listing will be produced as produced for the primary endpoint detailed in section 4.2.1.2.1.

4.2.1.2.2.3 Change in PGA of OA scores

The same analysis as applied to the primary endpoint, excluding the MCP-MOD analysis, will be applied to the change from baseline to Week12 in PGA of OA scores.

The same summary outputs including listing will be produced as produced for the primary endpoint detailed in section 4.2.1.2.1.

4.2.2 Other secondary efficacy endpoints

For each endpoint reported under the other secondary efficacy variables detailed in section 3.3, a separate summary table for observed cases only will be produced, as described in section 4.1. That is, for other continuous secondary endpoints, the outputs will present the summary statistics for baseline and the change from baseline to corresponding Weeks.

Other continuous secondary endpoints will be analysed by observed cases only. The ANCOVA model with dose, baseline score and KL grade will be used, and results tabulated following the same presentation used for key secondary endpoints except for the asymmetric CI.

For binary secondary endpoints, a logistic regression analysis will be performed with the same covariates as for the respective continuous endpoints. That is, for example the logistic regression for the OMERACT-OARSI endpoint will include baseline WOMAC pain and physical function subscales scores, baseline PGA score and KL grade as covariates in the model. The results will be tabulated so that the number and percent of responders with odds ratios, 95% CIs and nominal p-values will be presented for each week of interest.

Finally, listings will be produced for all other secondary efficacy endpoints.

4.2.3 Exploratory analysis

A summary table will be produced for the change in DSIS score from baseline to Weeks 2, 4, 6, 8, 10, 12, and 18 as specified in section 4.1 following the same presentation used for summary statistics of primary endpoint.

A summary table will present the change from baseline to Weeks 6, 12, 18 and 32 in the overall health utility score from the EQ-VAS questionnaire, by treatment group. Data will be listed.

The number and percentage of participants above and below the threshold values for the SF-36v2 Scale domains, PCS and MCS will be presented as well as descriptive summary statistics

for scores and change from baseline will be produced by treatment group and visit. Data will be listed.

A separate summary table will be produced for each of the four WPAI:OA endpoints defined in section 3.5.1.4 for the change from baseline to Weeks 12,18 and 32. Data will be listed.

A summary table will be produced for the following HCRU endpoints due to OA at baseline and Weeks 12, 18, and 32: any type of medical service received, use of emergency room and frequency, hospitalisation, and length of stay in hospital, use of any aids or devices, any job loss and length of time. Data will be listed.

A table summarising the incidence and time to discontinuation due to lack of efficacy will be produce.

A summary table with use of rescue medication from baseline to Weeks 2, 4, 6, 8, 10, 12, and 18 will be produced. A separate listing will be produced also.

4.2.3.1 ABPM

Three ABPM parameters (systolic and diastolic blood pressure and pulse rate) will be evaluated separately for during daytime, night-time and over the 24-hour period as follows:

- Baseline and change from baseline will be summarised descriptively as per section 4.1
- ANCOVA for change from baseline with treatment and baseline ABPM as covariates will be fitted and results including 95% CIs and nominal p-values tabulated.

Absolute mean ABPM values and ABPM mean change from baseline for readings mentioned in section 3.5.3 will be listed for each participant. Abnormal values as defined in section 3.5.3 will be flagged within the listing.

4.2.3.2 Pharmacodynamic analysis

Only total NgF concentrations will be included and will be listed and summarised by treatment group and timepoint (further subset by ADA status), with the following descriptive statistics presented: n, N<LLOQ, Arithmetic mean and SD, Geometric mean (gSD) and CV% (gCV%), median, minimum and maximum.

Furthermore, the influence of the baseline total NgF on changes in NRS pain scores and any interaction with treatment will be assessed.

In addition, a figure of geometric mean (with and without gSD) total NGF serum concentration over time by treatment group will be generated on a linear and a semilog scales as well as spaghetti plot of individual participant profiles by treatment group (both further subset by ADA status). Available data will be listed.

Total NGF serum concentrations from the initial assay (Q2 assay) will be only listed using the latest raw dataset as provided by Q2. This data will not be mapped to SDTM nor ADaM and hence will not be included in any summary tables or figures nor included in any analyses.

4.2.3.3 Biomarker analysis

Biomarkers (i.e. soluble biomarkers (e.g., cytokines, chemokines), genomic biomarkers (e.g., RNA, microRNA) and urinary CTX-II concentration (cartilage degradation marker)) and their effects on PK, PD, and efficacy may be investigated. However, any such exploratory analyses will be performed and presented separately from the CSR.

Available biomarker data at time of DBL (e.g., baseline TNF alpha, IL-1 beta, IL-6 and high sensitivity CRP) will however be listed for each participant and summarised using descriptive statistics by cohort. Furthermore, the influence of the baseline biomarkers on changes in NRS pain scores and any interaction with treatment will be assessed.

4.2.4 Pharmacokinetic analysis

PK concentration data will be listed for each participant and each dosing day and timepoint, and descriptive summary statistics (same descriptive statistics as for total NGF concentrations) will be provided by treatment group and timepoint for all evaluable participants (PK analysis set) and further subset by ADA status.

For the calculation of the descriptive statistics refer to section 3.4.1.

In addition, a figure of geometric mean (with and without gSD) serum MEDI7352 concentration over time by treatment group will be generated on a linear and a semilog scales as well as spaghetti plot of individual participant profiles by treatment group (further subset by ADA status). Available data will be listed.

The population PKPD analysis of the study data will be described in a separate modelling analysis plan and will be reported separately from the main CSR.

4.2.5 Immunogenicity analysis

The number and percentage of participants who display detectable ADA within each ADA response category listed in section 3.4.2 will be summarized by treatment group and MEDI7352 Total, based on the safety set for ADA-evaluable participants. Median and range of maximum titer will be provided as appropriate (if a participant has multiple titers, the max post-baseline titer is used for summary statistics calculation).

A table providing ADA category definitions will be produced.

A summary will be provided of the number and percentage of participants who display detectable ADA by visit. Median, Q1, Q3, minimum and maximum of ADA titres by visit will also be included.

Impact of ADA on PK and total NGF concentrations will be explored by presenting serum MEDI7352 concentrations and total NGF concentrations descriptive statistics in ADA-positive and ADA-negative participants.

Summary of TEAEs by ADA categories table will include the number and percentage of participants who had:

- Any AE
- Any AE with an outcome of death
- Any SAE
- Any AE leading to discontinuation of study treatment
- Any AESI
- Any AE leading to withdrawal from study.

TEAEs in ADA positive participants will also be listed.

Immunogenicity results will be listed by participant regardless of ADA-evaluable status. ADA titer will be listed for samples confirmed positive for the presence of ADA.

4.2.6 Safety analysis

Safety summaries will be descriptive only. No formal statistical analyses are planned for safety data. All data will be reported using the safety population unless specified otherwise.

The following sections describe the planned safety summaries for safety variables as defined in section 3.6. However, additional safety summaries (not specified in this SAP) may need to be produced to aid interpretation of the safety data.

4.2.6.1 Adverse Events

Summary tables will include TEAEs as defined in section 3.6.1, unless otherwise specified.

An AE overview summary table will be produced showing the number (%) of participants with at least one AE in each of the following categories:

- Any AE
- Any AE with an outcome of death
- Any SAE

- Any AE leading to discontinuation of study treatment
- Any severe AE
- Any AESI
- Any AE leading to withdrawal from study

For all categories above, the number (%) of participants with at least one causally related AE in that category will also be provided.

An additional table will be produced showing the incidence of participants with the following type of adjudicated events:

- RPOA (type 1 and 2)
- SIF
- Primary osteonecrosis,
- Pathological fracture

The number of participants treated per treatment group will be used as the denominator to calculate percentages.

Summary tables will be produced by MedDRA SOC and PT for the following categories:

- AEs
  - By age group as defined in section 4.2.8
  - By sex
  - By periods as defined in section 3.1.6.
  - By ADA categories as defined in section 3.4.2
- Severe AEs
- By periods as defined in section 3.1.6. Causally related AEs and SAEs
  - By periods as defined in section 3.1.6.
- SAEs with outcome death
- SAEs
  - By periods as defined in section 3.1.6.
  - By ADA categories as defined in section 3.4.2
- AEs leading to discontinuation of study treatment
- AESIs
  - By periods as defined in section 3.1.6
  - Separate table to be provided for each AESI category as defined in section 3.6.1.1
    - By periods as defined in section 3.1.6
    - Joint related AESI table by study periods will also include the exposure adjusted rates, where AE event rate (per 100 participant years) is defined as the number of

participants with that AE divided by the total number of days at risk for AEs across all participants in given treatment arm and multiplied by 365.25 x 100 to present in terms of per 100 participant years.

- Most common non-serious AEs defined as AEs occurring in at least 5% of participants overall (in total column). For each SOC/PT, the number (%) of participants reporting at least one event at each level of summarisation will be presented. Participants with multiple occurrences of the same AE, the event will only be counted once per SOC/PT. Those will be sorted by international order for system organ class and in alphabetical order for preferred term.

In addition, a table with AEs and event counts sorted by decreasing frequency on preferred term will be produced as well as AEs by maximum intensity on preferred term by SOC and PT.

Additional table presenting mapping of AESIs will be produced.

A separate table with key participant information will be produced for:

- SAEs with outcome of death
- SAEs
- AEs leading to discontinuation of study treatment

A separate listing will be produced for the following AE categories:

- AEs
- SAEs
- AESIs
- AEs leading to discontinuation of study treatment

For any AE with change in intensity, listings will report all intensities as well as associated actions.

4.2.6.2 Joint replacement surgeries

Incidence of participants with the joint replacement surgeries occurred during the study participation will be summarised by treatment group depending on whether the surgeries were elective or due to an AE.

4.2.6.3 Pregnancies

Pregnancies entered on the pregnancy report (PREGREP) page in eCRF will be reconciled against safety database and the reconciled entries on (PREGREP) page in eCRF will be listed.

4.2.6.4 Laboratory results

Tables with descriptive summaries (mean, SD, median, minimum, and maximum) of actual (absolute) values and changes from baseline values will be produced for each treatment group at each visit. Tables will be produced for tests specified in section 3.6.8 except for urine drug screen, urine pregnancy test, serology, and endocrinology. Participants with abnormal results of these tests except for urine pregnancy test will be only listed including all available results.

Central laboratory reference ranges will be used for the identification of abnormalities. Shift tables will be produced for the same parameters for which summary tables will be produced with the exception of urinalysis to display low, normal, and high values.

The shift tables will present baseline and maximum/minimum on-treatment value, as applicable for each parameter and will include participants with both baseline and post-baseline data. The exception would be Haemoglobin A1c that is evaluated only at baseline and Week 32 hence this shift table will present baseline and maximum/minimum value during the study.

For urinalysis parameters collected as for example: Negative, Small, Moderate or Large, shift tables from baseline to maximum value during treatment will be provided. This will apply to bilirubin, blood, leucocytes, glucose, protein and ketones. Key participant information table listing only participants and assessments with abnormal values will be produced.

Laboratory values will also be presented as box plots, for absolute values and changes from baseline. All protocol defined visits will be summarised.

Each plot will show the following details:

- Mean (symbol within each box)
- Median (horizontal line within each box)
- Q1 and Q3 (bottom and top of each box)
- Whiskers extending to the most extreme observation within 1.5 times the interquartile range from the nearest quartile.
- Outliers outside the whiskers displayed individually.
- Reference ranges shown as dotted lines on the analysis value plots, not relevant to the change from baseline plot. If more than one reference range is available for an assessment (example: due to gender differences) then the narrowest range will be used in the graph.

The plots will be prepared for each test within the laboratory categories:

- Haematology: hemoglobin, hematocrit, RBC count, platelet count (or estimate), WBC count, WBC differential

- Serum chemistry: albumin, total and direct bilirubin, total protein, calcium, ALP, ALT, AST, GGT, BUN, creatinine, glucose, sodium, potassium, chloride, bicarbonate, phosphorus, LDH, uric acid, CRP, cholesterol, triglycerides, creatine phosphokinase.
- Coagulation panel: prothrombin time, activated partial thromboplastin time, fibrinogen
- Urinalysis: pH, specific gravity.

In addition, listings of abnormal safety laboratory results will be prepared. The listings will include only participants and tests with at least one abnormal value reported. All available results for that specific test will be presented in order to see the change over time.

4.2.6.4.1 Liver enzyme elevations and Hy's law

Liver function laboratory tests assessments (scheduled and unscheduled) up to 30 days after the last dose of study treatment will be considered.

The following summaries will include the number (%) of participants with at least one post-baseline liver function assessment who have:

- Elevated ALT, AST and total bilirubin as
  - ALT ≥3× to <5×, ≥5× to <10×, ≥10× to <20×, ≥20× upper limit of normal (ULN)
  - AST ≥3× to <5×, ≥5× to <10×, ≥10× to <20×, ≥20× ULN
  - Total bilirubin ≥2× ULN
  - ALT or AST ≥3× to <5×, ≥5× to <10×, ≥10× to <20×, ≥20× ULN
  - ALT or AST ≥3× ULN and total bilirubin ≥2× ULN (Potential Hy's law): The onset date of ALT or AST elevation should be prior to or on the date of total bilirubin elevation.

Percentages will be based on the number of participants who have at least one measurement available.

To identify potential Hy's Law cases, a listing of participants who have ALT or AST ≥3× ULN and total bilirubin ≥2× ULN (not necessarily at the same time) following first dose of study treatment will be provided.

A shift table will also present maximum ALT and AST observed values by maximum total bilirubin observed value, expressed as multiples of ULN. Categories will be included for:

- ALT <3, ≥3× to <5×, ≥5× to <10×, ≥10× to <20×, ≥20× ULN
- AST <3×, ≥3× to <5×, ≥5× to <10×, ≥10× to <20×, ≥20× ULN
- Total bilirubin <2×, ≥2× ULN

Key participant information table of participants with potential Hy's law will be produced.

4.2.6.5 Vital signs results

Table with descriptive summaries (n, mean, SD, median, minimum, maximum, Q1 and Q3) of observed values and changes from baseline values for each treatment group at each protocol scheduled time point will be produced for the following vital signs tests:

- Systolic BP
- Diastolic BP
- Pulse rate
- Respiratory rate
- Temperature

In addition, shift table from baseline to minimum/maximum value during treatment will be produced for all above mentioned parameters. The reference ranges provided in section 3.6.5 will be used to determine whether the minimum/maximum value during treatment is within the normal range or not.

A by-visit incidence summary of orthostatic hypotension will be prepared for the following result categories:

- Criteria met
- Confirmed
- Symptomatic

A table presenting the changes in postural BP at each timepoint/ visit will be produced. Consequently, the supine readings from VS (for non-dosing visits) and VS1 (for dosing visits) eCRF forms as well as postural readings from VS2 eCRF form will be summarised descriptively by cohorts at each visit and all available timepoints as per all three eCRF forms.

In addition, a listing of confirmed and symptomatic orthostatic hypotension cases will be prepared. The listing will include all participants with at least one assessment indicated as a confirmed or symptomatic case of orthostatic hypotension. All available results will be presented in order to see the change over time.

Vital signs data for participants/parameter with at least one abnormal result for the below mentioned parameters will be listed:

- Systolic BP (Including assessments from VS2 page in the eCRF)
- Diastolic BP (Including assessments from VS2 page in the eCRF)
- Pulse rate
- Respiratory rate
- Temperature

4.2.6.6 ECG results

Overall evaluation of electrocardiogram (ECG) will be summarised at each timepoint/visit reporting the number (%) of participants presenting a normal, abnormal not clinically significant or abnormal clinically significant evaluation.

In addition, a listing of abnormal ECG results will be prepared. The listing will include all participants with at least one abnormal value reported. All available results will be presented in order to see the change over time.

4.2.6.7 Physical examination findings

Findings, that is data of participants/examinations with at least one abnormal result will be listed for all available visits including unscheduled visits for all participants.

4.2.6.7.1 Weight

On top of listing the data as mentioned in section 4.2.6.7 above, the summary table presenting the actual (absolute) values and change from baseline to Week 12 will be prepared.

4.2.6.8 Neurological examination findings

A summary of the overall assessments across all neurological evaluations will be provided by treatment arms for each visit.

In addition, a key participant information table will be produced summarising data of all participants with at least one reported abnormality.

4.2.6.9 Total neuropathy score-nurse (TNSn) results

A summary of the total neuropathy score – nurse (TNSn) results will be prepared, including results for each of the following categories:

- Sensory symptom score
- Motor symptom score
- Autonomic symptom score
- Pin sensibility score
- Vibration sensibility score
- TNSn total

4.2.6.10 Injection site reaction

An incidence summary by severity (mild, moderate, severe) or presence (yes, no) and visit/timepoint will be prepared for the following injection site reactions:

- Pain (mild, moderate, severe)

- Tenderness (mild, moderate, severe)
- Erythema/redness (mild, moderate, severe)
- Induration/swelling (mild, moderate, severe)
- Necrosis (mild, moderate, severe)
- Ulceration (yes, no)

In addition, the quantitative result for erythema/redness and induration/swelling measurements (cm) will be summarised descriptively by visit/timepoint.

4.2.6.11 Survey of Autonomic Symptoms (SAS)

The SAS scores will be summarised by treatment group for the total number of symptoms reported and the total impact score. The summary will be presented by visit and will include the change from baseline.

4.2.6.12 X-ray and MRI

Summary table suggestive of abnormal findings for each treatment group at the protocol defined visits will be produced.

All data of assessable imaging will be listed for each participant. In addition, a separate listing of participants whose data was sent for adjudication will be produced presenting the outcome of the adjudication review.

4.2.7 Treatment exposure and compliance

The following summaries related to study treatment will be produced using the safety analysis set by actual treatment group and MEDI7352 Total:

- Participants who received an IP dose
- Maximum number of doses received
    - Participant will be included only in one category depending on the maximum number of doses received during the treatment period. That is if a participant receives only 4 doses during treatment, then this participant should be counted only once in the category '4'
- Duration of exposure as defined in section 3.6.11.1
- Number of participants with dose delay/interruption as reported on the eCRF
- Reasons for dose delays/interruptions as reported on the eCRF
- Cumulative exposure over time (in form of a graph) using the total treatment duration variable as defined in section 3.6.11.1.
- Compliance

Administration of study treatment will also be listed for all participants in the safety analysis set. In addition, overdose cases, as reported by the investigator, will be listed. Also, the kit details of the study treatment that were administered will be listed.

4.2.8 Demographic and baseline characteristics

Disposition, demographic data, and other baseline characteristics will be presented for participants in the FAS (unless otherwise specified below). No statistical testing will be carried out for demographic or other baseline characteristics. The following data will be summarized:

- Participant disposition (Screening set) namely:
    - Enrolled (number of participants only)
    - Rescreened (number of participants only)
    - Randomised/not randomised (including reason)
    - Received study treatment/did not receive study treatment (including reasons)
    - Completed study treatment/discontinued study treatment (including reasons)
    - Completed study/withdrawn from study (including reasons)

  Percentages will be based on the randomised participants.

  A listing will provide details on participants who were withdrawn from the study and participants who were enrolled and were not randomised. Also, a listing will report participants who completed the study as well as the randomisation codes of all randomised participants.

- Inclusion in analysis sets

  A listing will provide the reasons for exclusion from the analysis sets and the participant status for each of the analysis sets.

- Screen failure reasons
    - Number and percentage of participants that did not meet specific inclusion/exclusion criteria

  In addition, a listing will provide the reason(s) for screen failures (inclusion/exclusion criteria not met), including rescreened participants.

- Demographics namely:
    - Age (years) at the time of randomisation (derived)
    - Age groups (<65 and >=65)

  - Sex
    - Childbearing potential presented for females
    - Reasons for not of childbearing potential included
  - Race
  - Ethnic group
  - Country

- Participant characteristics at baseline namely:
  - Height (cm)
  - Weight (kg)
  - Body mass index (BMI) (kg/m2)
  - BMI group (kg/m2) (<18.5, >=18.5 - < 25, >=25 – < 30, >=30)

- Participant recruitment by region, country, and site

- Medical history and surgical history (as appropriate)
  - All medical/surgical history will be classified according to the latest version of the Medical Dictionary for Regulatory Activities (MedDRA).
  - The number (%) of participants with any medical/surgical history (records) will be tabulated by MedDRA system organ class (SOC) and PT. A participant will only be counted once within a particular SOC/PT, regardless of multiple conditions/diseases/surgeries in the same SOC/PT. SOC and PT will be presented in alphabetical order.

- All prior/concomitant medications will be coded using the latest World Health Organization (WHO) Drug Global B3-format dictionary.

  Summary tables of allowed and prohibited concomitant medications will present the number (%) of participants, using at least one medication by generic term within anatomical therapeutic chemical (ATC) classification. Participants will be counted once at each level of summarisation (ATC classification/generic term).

- OA related medical history, as collected in the OA medical history eCRF form, namely:
  - Duration of OA (years)
    - Calculated as: Randomization date - OA start date (collected on the OA medical history form), presented in years.
  - Duration of target knee pain (years)

        - Calculated as: Randomization date - target knee pain start date (collected on the OA medical history form), presented in years.
    - Target knee (left/right), n (%)
        - As collected on the OA medical history form.
    - Baseline KL grade
        - As reported by Bioclinica
    - Baseline pain intensity score (NRS)
        - As collected in the eDiary

In addition, disease-related treatment at screening will be summarised and will include:

- Previous/ongoing treatments
    - Nonsteroidal anti-inflammatory drugs (NSAIDs)/cyclooxygenase-2 (COX-2) inhibitors
    - Paracetamol
    - Opioids

Pain relief history evaluated in terms of:

- Not effective
- Intolerability
- Contraindication
- No access to opioids (only for opioids)
- Unwillingness to take opioids (only for opioids)

The number (%) of participants will be presented.

- Rescue medication variables defined in section 3.5.1.7 will be summarised by periods as defined in section 3.1.6.

Demographic and baseline characteristics data will also be listed.

4.2.9 Accounting for the COVID-19 pandemic effects

This study started during the severe acute respiratory syndrome coronavirus 2 (SARSCoV2)/Coronavirus disease 2019 (COVID19) pandemic, confirmed on 12th March 2020. The CSR will report how COVID-19 affected participants and/or study progression. COVID-19 assessments were carried out prior to the first IP administration and throughout the study and affected participants will be summarised.

A table will be produced to summarise any COVID-19 related study disruptions. The participant disposition table will incorporate COVID-19 as a further reason for discontinuing

study treatment. Participants affected by COVID-19 will be listed as well as those reported as having issues due to the COVID-19 pandemic in CTMS.

5 INTERIM ANALYSES

This section of the SAP is dedicated to the interim analyses (IA). Additional separate stand-alone SAPs were created for the Data and Safety Monitoring Board (DSMB) analyses as well as the Interim analyses.

The study incorporates two interim analyses:

- Futility analysis (IA 1):
  - When approximately 25% of participants are evaluable for the primary endpoint
  - The likely success of the study outcome will be assessed.
  - This analysis will enable the sponsor to make a go/no-go decision regarding continuation of the study as described by Frewer P. et. al., 2016. That is, the primary endpoint and WOMAC pain subscale endpoint will be analysed by most appropriate candidate model as specified in section 4.2.1.2.1.1 from which the pairwise model-based estimates versus placebo will be used for decision making purposes as follows:
    - Consider stopping the study if the lower limit of 1-sided 90% confidence interval for delta (MEDI7352 dose - Placebo) of change from baseline is below minus 1.5 (the target value).
    - Consider initiate planning for phase 3 if upper limit of 1-sided 80% confidence interval for delta (MEDI7352 dose - Placebo) of change from baseline is below minus 0.8 (the lower reference value).
  - Primary endpoint would be the main determinant of go/no go decision. If the primary endpoint is in the consider zone, then WOMAC pain subscale endpoint may help to determine the final decision.
  - This analysis will include disposition, baseline characteristics, primary efficacy, and key secondary efficacy (supportive) endpoint, which is change in the WOMAC pain subscale from baseline to Week 12.

- Interim analysis (IA 2):
  - When approximately 50% of participants are evaluable for the primary endpoint
  - This analysis will enable the sponsor to plan future project-related activities (without making any changes to the current study). At minimum, the same analyses as described for IA 1 will be performed. The decision making will proceed as follows:

- Consider initiate planning for phase 3 if upper limit of 1-sided 80% confidence interval for delta (MEDI7352 dose - Placebo) of change from baseline is below minus 0.8 (the lower reference value).

– This analysis will include the same evaluations as covered by the first interim analysis plus additional outputs as specified within this version of the IA SAP v2.0 and accompanying table shells. The interim analyses will be prepared by the blinded IQVIA biostatistics team. The unblinding of the interim analyses will be done by an independent unblinded IQVIA biostatistics team. The unblinded IQVIA biostatistics team will provide the unblinded interim analyses output to the Independent Efficacy Review Group (IERG).

6 CHANGES OF ANALYSIS FROM PROTOCOL

- The final analysis and its reporting will be split to two deliverables that would be supported by two database locks at timepoints specified in section 1. This is to evaluate the efficacy data once all participants complete the treatment period as it is not anticipated that this data will change rather than waiting once all participants compete the study especially after the follow-up period was extended from 12 to 24 weeks.
- Single imputation will be performed only for the primary and key secondary endpoints as it is not meaningful to perform the single imputation for endpoints that will only be summarised.
- The sequentially rejective multiple comparison approach includes not just pairwise comparisons, but also the dose response curve for the WOMAC pain subscale score. In addition, MCP-MOD analyses of the remaining key secondary endpoints will not be performed.
- Additional supplementary analyses not specified in the protocol were added for comparative purposes within the industry. Those are: ANCOVA analysis of the primary endpoint and WOMAC pain subscale score key secondary endpoint using singly imputed LOCF cases as well as ANCOVA analysis of the WOMAC pain subscale score key secondary endpoint using multiply imputed cases that were imputed using the sampling from normal distribution approach.
- The protocol presents the concomitant medication and therapies as a safety endpoint. However, this data does not necessarily evaluate the safety and hence was moved as part of participant disposition which is in alignment with AZ standard shells reporting.
- CCI [redacted].
- The definition of the treatment emergent AEs was clarified to account for instances where a participant withdraws from the study over a phone call and disposition (DS) page in eCRF is populated without the corresponding visit being recorded. In addition, the change in relationship was removed as pre-treatment AE that changes the relationship would need to be entered in the database as a new AE with a start date on or after the first IP administration date and hence would automatically become a treatment emergent AE as per the first part of the definition.
- Other secondary endpoints, DSIS and rescue medication use will be summarised and analysed using available data on protocol specified visits and not just those visits specified in the definition of the endpoint itself (Week 2, 4, 6, 8, 10, 12 and 18).
- Provided that the ECG evaluations are performed locally by sites, only the abnormal evaluations will be listed.
- Given that TNSn and Injection site reactions data are tabulated, the listings were not generated.

- An additional AESI category was defined in this SAP that does not feature in the protocol. This category was classed as ‘Other AESIs’ and has been defined as any other AEs that are deemed of an interest by the study team.

7 REFERENCES

Barnard, J., and Rubin, D. B.: Small-Sample Degrees of Freedom with Multiple Imputation. Biometrika. 1999; 86:948–955.

Billingsley P.: Probability and Measure, Second Edition. 1986.

Bornkamp B., Bezlyak V., and Bretz F.: Modern Approaches to Clinical Trials Using SAS Chapter 7 Implementing the MCP-Mod Procedure for Dose-Response Testing and Estimation 2015; 193-224.

Bretz F., Maurer W., Brannath W., and Posch M.: A graphical approach to sequentially rejective multiple test procedures. Statistics in Medicine. 2009; 28: 586-604.

Cro S., Morris T. P., Kenward M. G. and Carpenter J. R.: Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: A practical guide. Statistics in medicine. 2020; 21: 2815-2842.

Frewer P., Mitchell P., Watkins C. and Matcham J.: Decision-making in early clinical drug development. Pharmaceutical Statistics. 2016; 15: 255-263.

Little R.J.A. and Rubin D.B.: Statistical Analysis with Missing Data. 2002

CCI [redacted].

MEDI7352 - D5680C00003 Clinical study protocol version 1.0, dated 13AUG2020.

Pham T, van der Heijde D, Altman RD, Anderson JJ, Bellamy N, Hochberg M, et al. OMERACT-OARSI initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. Osteoarthritis Cartilage. 2004;12(5):389-99.

Pinheiro J *et al*: Model-Based dose finding under model uncertainty using general parametric models. Statistics in Medicine. 2014; 33: 1646-1661.

CCI [redacted].

Ratitch B. and O'Kelly M.: Implementation of Pattern-Mixture Models Using Standard SAS/STAT Procedures. PharmaSUG2021-Paper SP04. 2011.

Schomaker M. and Heumann Ch.: Model selection and model averaging after multiple imputation. Computational Statistics and Data Analysis 71 2014; 758-770

8 APPENDICES

Appendix A Questionnaires derivations

Appendix A1 WPAI:OA

The table below summarize the six questions of the WPAI:OA questionnaire, and the derivations of the four endpoints of the effect of impairment on activity and impairment.

Question	Question Wording	Scoring
1	Are you currently employed? [if No skip to question 6]	Yes, No
2	During the past seven days, how many hours did you miss from work due to problems associated with your OA of the knee or hip?	number of hours (free text)
3	During the past seven days, how many hours did you miss from work because of any other reason, such as vacation, holidays, time off to participate in this study?	number of hours (free text)
4	During the past seven days, how many hours did you actually work? (if '0' skip to Question 6)	number of hours (free text)
5	During the past seven days, how much did your OA of the knee or hip affect productivity while you were working?	0 to 10 scale with 0 being 'No effect on my work' and 10 being 'Completely prevented me from working'
6	During the past seven days, how much did your OA of the knee or hip affect your ability to do your regular daily activities, other than work at a job?	0 to 10 scale with 0 being 'No effect on my daily activities' and 10 being 'Completely prevented me from doing my daily activities'

WPAI endpoint	Calculation

Percent activity impairment due to Osteoarthritis	Q6*10
Percent impairment while working due to osteoarthritis	Q5*10
Percent overall work impairment due to osteoarthritis	$\{\frac{Q2}{Q2+Q4} + [1 - \left(\frac{Q2}{Q2+Q4}\right)](\frac{Q5}{10})\} * 100$
Percent work time missed due to Osteoarthritis	$\frac{Q2}{Q2+Q4} * 100$

Appendix A2 Responses and scores of the HCRU questionnaire

Question	Response	Scoring
During the last X weeks what services did you receive directly related to your osteoarthritis? • Primary Care Physician; • Neurologist; • Rheumatologist; • Physician Assistant or Nurse Practitioner; • Pain Specialist; • Orthopaedist; • Physical Therapist; • Chiropractor; • Alternative Medicine or Therapy (Concomitant Medications); • Podiatrist; • Nutritionist/Dietician; • Radiologist; • Home healthcare services; • Other;	Number of visits	Response not selected= 0 Number of visits= 1-999
During the past X weeks, have you visited the emergency room due to your osteoarthritis?	Yes, No	No = 0 Yes = 1
How many times?	Number of visits	0-999

During the past X weeks, have you been hospitalized due to your osteoarthritis?	Yes, No	No = 0 Yes = 1
How many nights in total did you stay in hospital due to your osteoarthritis in the last X weeks?	Number of Nights	0-999
Did you use these aids or devices to help you in doing things because of your osteoarthritis in the last X weeks? • Walking Aid • Wheelchair • Devices or utensils to help you dress, eat or bathe • Other	Did not use any aids or devices never, rarely, sometimes, often, always	Did not use any aids or devices= 0 Device not selected= 0 Never = 1 Rarely = 2 Sometimes = 3 Often = 4 Always = 5
Did you quit your job because of your osteoarthritis?	Yes, No	No = 0 Yes = 1 Not applicable = 2
How long ago did you quit your job because of your osteoarthritis?	Years and Months	0-99 Years and 0-99 Months (should be max of 11 months)

Appendix B Multiple Imputation

Appendix B1 The pattern mixture framework

The MAR imputation will be performed for all participants, regardless of the intercurrent event, on the observed and imputed data from step 1 (intermediate missing data imputation), including the imputed baseline from step 0 also. The same will be performed for MNAR approach. This would result in two sets of multiply imputed datasets: one set under MAR assumption and one set under MNAR assumption.

The final set that would then be used for further analysis will be compiled such that only imputed data from either MAR or MNAR sets would be selected based on the participants' intercurrent event. For example, if a participant discontinues treatment due to an adverse event, then only data from the MNAR set would be included in the final set, whilst data of a participant who discontinues treatment due to their decision would come from only MAR set.

Appendix B2 The sampling from normal distribution approach

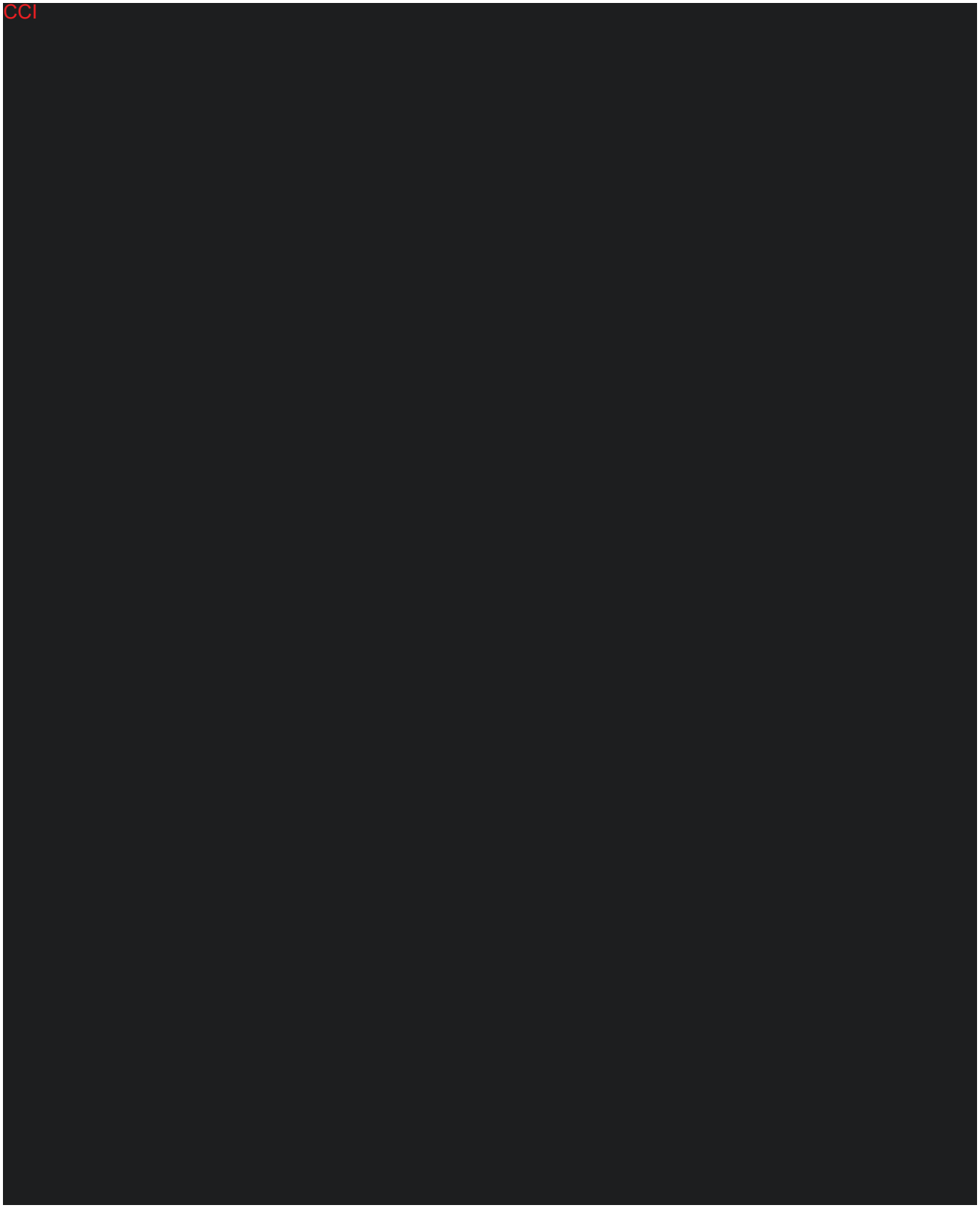
1. Obtain the standard deviation from observed WOMAC pain subscale scores at Week 12 for all participants in FAS
2. Sampling from normal distribution to be executed using the following data step: rand('NORMAL' BASE,stddev) or rand('NORMAL',LOCF,stddev)
    a. where the stddev is obtained in step 1
    b. BASE or LOCF are either participant's baseline score or participant's last observed WOMAC pain subscale score.
3. Step 2 to be performed 100 times
4. Data to be ordered by ParticipantID and imputation to have the same seed number and produce identical results on production and QC

NOTE:

- No intermediate missing data imputation step is needed.
- For LOCF, consider available WOMAC pain subscale scores of all post-baseline visits (not just protocol defined visits), except unscheduled visits. That is, if participant has Week 11 WOMAC score, this score can be used as last observation for LOCF imputation.

Appendix C MCP-Mod: Combining Estimates from Multiple Imputed Datasets

CCI



CCI

Appendix DMCP-Mod Code

Referenced in SAP section 4.2.1.2.1.1

Appendix D1 Deriving optimum contrast coefficients

The below code alongside the combined dose estimates and covariance matrix derived in appendix B can be used to obtain the optimum contrast coefficients.

```
/*Specify doses*/

data doses;
      input dose @@;
      datalines;
        0 7.5 25 75 150;
      ;
run;


/*Specify candidate models*/

data models;
      length modnam $25;
      input modnam par1-par3;
      datalines;
            emax 10 . .
            emax 40 . .
            emax 80 . .
            linear . . .
            exponential 10 . .
            exponential 25 . .
            ;
run;

/*Write optimal contrasts in data set optcontrasts and in global macro variable modcontr */

%optcont(models=models, dose=doses, sigma=scalemat, direction=incr, dosevarname=dose);

proc print data optcontrasts label noobs;
      title “Optimal contrasts for defined models”;
run;
```

%optcont macro code

```
%macro optcont(models=,doses=,sigma=,direction=,dosevarname=);

/*****************************************************************
```

```
Reading in models
*************************************************************************/
data &models;
    set &models;
    length modspec $ 50;
    if compress(modnam)="emax" then do;
        modelnum=1;
        modspec="emax "!!compress("(ed50="!!par1!!")");
    end;
    if compress(modnam)="exponential" then do;
        modelnum=2;
        modspec="exponential "!!compress("(delta="!!par1!!")");
    end;
    if compress(modnam)="linear" then do;
        modelnum=3;
        modspec=modnam;
    end;
    if compress(modnam)="linlog" then do;
        modelnum=4;
        modspec="linlog "!!compress("(off="!!par1!!")");
    end;
    if compress(modnam)="quadratic" then do;
        modelnum=5;
        modspec="quadratic "!!compress("(delta="!!par1!!")");
    end;
    if compress(modnam)="sigEmax" then do;
        modelnum=6;
        modspec="sigemax "!!compress("(ed50="!!par1!!", h="!!par2!!")");
    end;
    if compress(modnam)="betaMod" then do;
        modelnum=7;
        modspec="betamod "!!compress("(del1="!!par1!!", del2="!!par2!!",
        scal="!!par3!!")");
    end;
run;

proc sql noprint;
select count(distinct(dose))format=1.0 into :n_dose_ from &doses;
quit;

proc sql noprint;
select count(modelnum) into :n_mod_ from &models;
quit;

proc sql noprint;
select modnam into :mod1 - :mod%left(&n_mod_) from &models;
```

```
quit;

proc sql noprint;
select modspec into :modspec1 - :modspec%left(&n_mod_) from &models;
quit;

%if &direction=incr %then %do; %let dir=1; %end;
%if &direction=decr %then %do; %let dir=-1; %end;

/*****************************************************************
Calculation of contrasts
*****************************************************************/
proc iml ;
/* read in doses, covariance mat, models/guesstimates and model names*/
      use &doses var {dose };
      read all into dose;
      use &sigma;
      read all into sigma;
      use &models;
      read all var{modelnum par1 par2 par3} into par;
      use &models;
      read all var{modnam} into modlist;

      /* Unity matrix */
      unity=j(&n_dose_, 1, 1);

      /* Empty matrix for calculated mean vectors and contrasts */
      meanmat=j(&n_dose_, &n_mod_+1, -88);
      meanmat[, 1]=dose;

      /* row labels */
      contmat=j(&n_dose_, &n_mod_, -88);

      /* calculation of stand. mean-vector for each model*/
      do j=2 to &n_mod_+1;
            modnum=par[j-1, 1];
            do i=1 to &n_dose_; /* for each dose*/
                  if modnum=1 then meanmat[i, j]=dose[i]/(par[j-1, 2] + dose[i]);
                  if modnum=2 then meanmat[i, j]=exp(dose[i]/par[j-1, 2]) - 1;
                  if modnum=3 then meanmat[i, j]=dose[i];
                  if modnum=4 then meanmat[i, j]=log(dose[i] + par[j-1, 2]);
                  if modnum=5 then meanmat[i, j]=dose[i] + par[j-1, 2]*dose[i]**2;
                  if modnum=6 then meanmat[i, j]=dose[i]**par[j-1, 3]/(par[j-1,
                  2]**par[j-1,3] + dose[i]**par[j-1, 3]);
                  if modnum=7 then do;
                        maxDens=(par[j-1, 2]**par[j-1, 2]) * (par[j-1,3]**par[j-1,3])
```

```
                    /((par[j-1, 2] + par[j-1,3])**(par[j-1, 2] + par[j-1, 3]));
                    standdose=dose[i]/par[j-1, 4];
                    meanmat[i, j]=1/maxDens * (standdose**par[j-1, 2]) *
                    (1 - standdose)**par[j-1, 3];
                end;
            end;
        end;
        invsigma=inv(sigma);

        do j=2 to &n_mod_+1; /* calculation of contrasts*/
            mean_t=t(meanmat[, j]);
            _1=mean_t*invsigma*unity;
            _2=t(unity)*invsigma*unity;
            _3=_1/_2;
            _4=_3*unity;
            _5=meanmat[, j]-_4;
            contrast=invsigma*_5;
            c_norm=sqrt(ssq(contrast));
            contmat[, j-1]=&dir*contrast/c_norm;
        end;
        create optcontrasts from contmat;

        /* Export the final contrasts to dataset*/
        append from contmat;
        quit;
data &models;
        set &models;
        i=_n_;
run;

data _null_;
        set &models;
        do j=1 to &n_mod_;
            if j=i then do;
                call execute ("
                data optcontrasts; set optcontrasts;
                attrib col"!!compress(j)!!" label='"!!compress(modnam)!!"';
                run;");
            end;
        end;

data optcontrasts;
        merge &doses(keep=dose) optcontrasts;
run;
```

CCI

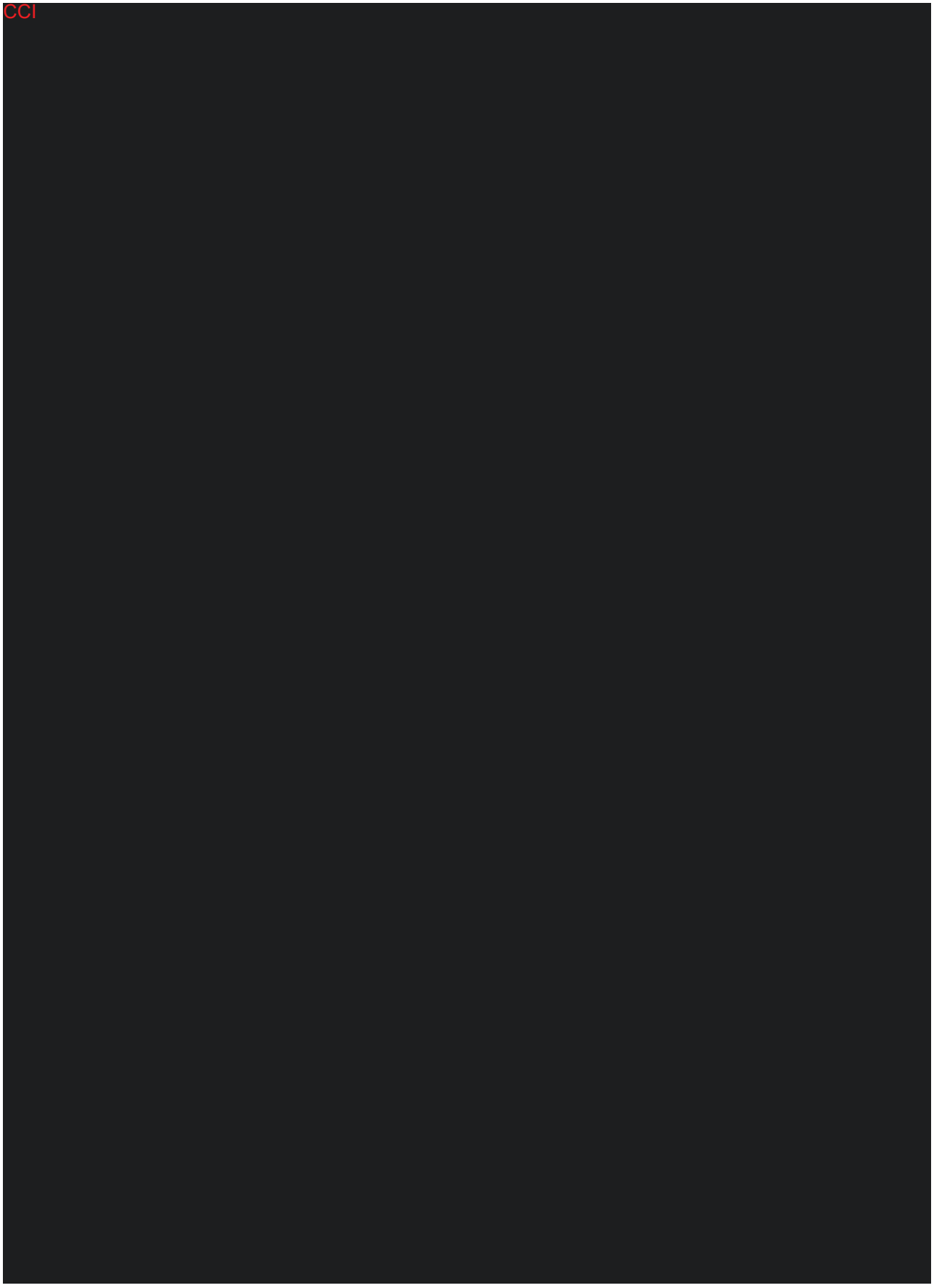
CCI

CCI

CCI

[redacted]

CCI



CCI

[redacted]

Appendix D4 Model estimates

Appendix D4.1 Calculation of Dose to achieve target effects

Target Dose TD_{Δ} to achieve a Delta (Δ) versus placebo (MED):

Linear: TDΔ= Δ/ δ, where δ= estimated linear slope

Emax: TDΔ= ED50/((Emax/Δ) - 1), where ED50 and Emax are estimated from the fitted Emax model

Exponential: TDΔ= δ*log((Δ/E1)+1), where δ and E1 are estimated from the fitted exp model

Dose (ED_P) to achieve a certain proportion P of the effect of the maximum dose studied versus placebo

Linear: ED_P= P*max(Doses) = P*150

Emax: ED_P= (P* ED50*max(doses))/(ED50+(1-P)*max(doses)) = (P* ED50*150)/(ED50+(1-P)*150)

Exponential: ED_P= δ *log((P*exp(max(doses)/ δ)-P+1)) = δ *log((P*exp(150/ δ)-P+1))

Note: Protocol defines EDp as the proportion P of the asymptotic maximum effect (Emax) for the Emax model that will be determined as follows:

Emax: asymptotic ED_P= (P*ED50) /(1-P)

Appendix D4.2.1 Asymptomatic normal Distribution

Sampling from the asymptotic normal distribution is used for constructing all 95% CIs except those for model estimates of the treatment effect at doses studied.
The parameters from the selected model are asymptotically normally distributed, following the distributed given below as described in Pinheiro et al (2014).

$$MNV\left(\hat{\theta}, \left[F(\hat{\theta})S^{-1}F(\hat{\theta})^{T^{-1}}\right]\right)$$

Where $F(\theta_0)$ is a $d \times k$ matrix of the partial derivatives $\frac{\partial f(x_i,\theta)}{\theta_h}$ $i = 1, \ldots, k$ $h = 1, \ldots d$ with each row being the partial derivative at each parameter and each column being the value of that derivative for a given dose.

As an example, $F(\theta_0)$ for the linear dose response candidate model with no additional covariates is given below where $\theta = \begin{pmatrix} E_0 \\ \delta \end{pmatrix}$; CCI [redacted]

Appendix D4.2.2 Delta method

95% CIs except for model estimates of the treatment effect at doses studied

The standard error for the response (difference from placebo) of any dose d can be calculated using the Delta method (Billingsley, P., 1986) as follows:

$$se(d) = \sqrt{v' \sum v}$$

Where Σ is the variance-covariance matrix of parameters (as given in the appendix D4.2.1), and v is the gradient vector evaluated at the parameter estimates and defined by:

For Emax model as:

$$v = \begin{pmatrix} 0 \\ \frac{d}{d + ED_{50}} \\ \frac{-E_{max} d}{(d + ED_{50})^2} \end{pmatrix}.$$

The corresponding SAS code:

CCI [redacted]

CCI