

Official Title: A Phase 2B, Randomized, Double Blind, Placebo and Active Comparator Controlled, Multicenter, Safety, and Efficacy Trial of ATX101 in Adults Undergoing Total Knee Arthroplasty (SPARK 2)

NCT Number: NCT06799845

Document Date: 28-Oct-2025

Statistical Analysis Plan
Sponsor Allay Therapeutics, Inc.
Protocol Number ATX-101-TKA-004
PCN Number ALLT238373

Sponsor	Allay Therapeutics, Inc.
Protocol Title:	A Phase 2B, Randomized, Double Blind, Placebo and Active Comparator Controlled, Multicenter, Safety, and Efficacy Trial of ATX101 in Adults Undergoing Total Knee Arthroplasty
Protocol Number:	ATX-101-TKA-004
Document Version:	Final Version 5.0
Document Date:	28-Oct-2025

Approvals

Role	Signatures
Lead Biostatistician	Print Name: [REDACTED]
	Sign Name: [REDACTED]
Senior Reviewer	Print Name: [REDACTED]
	Sign Name: [REDACTED]
Allay Therapeutics Representative	Print Name: [REDACTED]
	Sign Name: [REDACTED]



Statistical Analysis Plan
Sponsor Allay Therapeutics, Inc.
Protocol Number ATX-101-TKA-004
PCN Number ALLT238373

Final Version 1.0: 01-Aug-2024

Final Version 2.0: 01-Apr-2025

Final Version 3.0: 12-Jun-2025

Final Version 4.0: 9-Oct-2025

Final Version 5.0: 28-Oct-2025



Table of Contents

Approvals..... 1

Document History..... 1

Table of Contents..... 3

List of Tables 5

List of Figures..... 5

List of Abbreviations 6

1. Overview..... 9

2. Study Objectives and Endpoints 10

 2.1. Study Objectives..... 10

 2.2. Study Endpoints..... 10

 2.2.1. Efficacy Endpoints..... 10

 2.2.2. Safety Endpoints..... 12

3. Overall Study Design and Plan..... 13

 3.1. Overall Design..... 13

 3.2. Sample Size and Power..... 13

 3.3. Study Population..... 13

 3.4. Treatments Administered..... 13

 3.5. Method of Assigning Subjects to Treatment Groups..... 13

 3.6. Blinding and Unblinding..... 13

 3.7. Schedule of Events..... 14

4. Statistical Analysis and Reporting..... 15

 4.1. Introduction..... 15

 4.2. Interim Analysis and Data Monitoring 16

 4.2.1. Interim Analysis..... 16

 4.2.2. Safety Monitoring Committee 18

5. Analysis Populations..... 20

6. General Issues for Statistical Analysis..... 21

 6.1. Statistical Definitions and Algorithms..... 21

 6.1.1. Baseline..... 21

 6.1.2. Multiple Comparisons..... 21

 6.1.3. Handling of Dropouts or Missing Data..... 21

 6.1.4. Pooling of Sites 22

 6.1.5. Derived Variables 22

 6.1.6. Data Adjustments/Handling/Conventions 24

7. Study Participants and Demographics 27

 7.1. Disposition 27

7.2.	Protocol Deviations.....	27
7.3.	Demographics and Baseline Characteristics.....	27
7.4.	Medical History	28
7.5.	Prior and Concomitant Medication.....	28
7.6.	TKA Surgery Information.....	28
7.7.	Exposure and Compliance	28
8.	Efficacy Analysis.....	29
8.1.	Efficacy Assessments.....	29
8.1.1.	Range of Motion (ROM) Assessment.....	29
8.1.2.	Timed Up and Go (TUG) Test and NRS-A with Location of Pain	29
8.1.3.	Participant Self-Administered Questionnaires.....	29
8.2.	Primary Efficacy Endpoint	33
8.2.1.	Primary Estimand.....	33
8.2.2.	Main Estimation.....	35
8.2.3.	Multiple Imputation	35
8.2.4.	Sensitivity Analyses.....	37
8.2.5.	Subgroup Analyses	38
8.3.	Key Secondary Efficacy Endpoints	38
8.3.1.	Secondary Estimand #1.....	38
8.3.2.	Secondary Estimand #2.....	39
8.3.3.	Secondary Estimand #3.....	40
8.3.4.	Secondary Estimand #4.....	41
8.3.5.	Secondary Estimand #5.....	42
8.3.6.	Main Estimation.....	42
8.4.	Secondary Efficacy Endpoints.....	44
8.5.	Exploratory Efficacy Endpoints.....	47
9.	Safety and Tolerability Analysis.....	48
9.1.	Adverse Events	48
9.1.1.	Serious Adverse Events	50
9.1.2.	Adverse Events of Special Interest and LAST-Related Symptoms.....	50
9.1.3.	Adverse Events Leading to Study Discontinuation	52
9.1.4.	Deaths	52
9.2.	Clinical Laboratory Evaluations	53
9.2.1.	Liver Function Test Elevations.....	53
9.3.	Vital Signs.....	54
9.4.	12-Lead Electrocardiogram (ECG).....	54
9.5.	Physical Examination.....	54



9.6. Neurological Assessment..... 55

9.7. Wound Healing Assessment 55

10. Changes from Planned Analysis 56

11. Other Planned Analysis..... 57

11.1. Pharmacokinetic Analysis..... 57

12. References..... 58

13. Appendix..... 59

13.1. Tables, Listings, and Figures 59

13.1.1. Planned Table Descriptions 59

13.1.2. Planned Listing Descriptions 64

13.1.3. Planned Figure Descriptions 66

13.2. Tables, Listings, and Listing Shells 67

13.2.1. Standard Layout for all Tables, Listings, and Figures 67

13.3. Morphine Milligram Equivalent (MME) 69

List of Tables

Table 1: Study Participants and Demographics 59

Table 2: Efficacy Data 59

Table 3: Safety Data..... 63

Table 4: Planned Listings..... 64

Table 5: Planned Figures 66

List of Figures

Figure 1: Standardized Layout..... 68

List of Abbreviations

Abbreviation	Definition
AE	adverse event
AESI	adverse events special interest
ANCOVA	analysis of covariance
ANOVA	analysis of variance
ATC	anatomical therapeutic chemical
AUC	area under the curve
BMI	body mass index
CI	confidence interval
CP	conditional power
CRA	clinical research associate
CRF	case report form
CS	clinically significant
CSR	clinical study report
DBL	database lock
ECG	electrocardiogram
EDC	electronic data capture
EMA	European Medicines Agency
ET	early termination
FAS	full analysis set
FCS	fully conditional specification

Abbreviation	Definition
FDA	Food and Drug Administration
HR	heart rate
IA	interim analysis
ICH	International Conference on Harmonization
IRB	institutional review board
IRT	interactive response technology
KSS	Knee Society Score
LAST	Local Anesthetic Systemic Toxicity
LOE	Lack of efficacy
LS	least-square
MedDRA	Medical Dictionary for Regulatory Activities
MAR	missing at random
MCMC	Markov Chain Monte Carlo
MI	multiple imputation
MME	morphine milligram equivalents
MNAR	missing not at random
NCS	non-clinically significant
NRS-A	Numerical Rating Scale with Activity
NRS-R	Numerical Rating Scale at Rest
PD	protocol deviation
PHQ-9	Patient Health Questionnaire-9

Abbreviation	Definition
PK	pharmacokinetic
PP	per-protocol
PT	preferred term
REGPMM	regression and predictive mean matching
ROM	range of motion
RR	relative rate
SAE	serious adverse event
SAP	statistical analysis plan
SD	standard deviation
SDR	source data review
SE	standard error
SMC	Safety Monitoring Committee
SOC	system organ class
SRC	Statistical Review Committee
TEAE	treatment-emergent adverse event
TKA	total knee arthroplasty
TLF	table, listing, and figure
TUG	Timed Up and Go
WHO-DD	World Health Organization Drug Dictionary

1. Overview

This statistical analysis plan (SAP) describes the planned analysis and reporting for Allay Therapeutics, Inc. protocol number ATX-101-TKA-004 (A Phase 2B, Randomized, Double Blind, Placebo and Active Comparator Controlled, Multicenter, Safety, and Efficacy Trial of ATX101 in Adults Undergoing Total Knee Arthroplasty), amendment 5.0, dated 11-Jul-2025. Reference materials for this statistical plan include the protocol and the accompanying sample data collection documents. Operational aspects related to collection and timing of planned clinical assessments are not repeated in this SAP unless relevant to the planned analysis.

This SAP provides a comprehensive and detailed description of the strategy, rationale, and statistical techniques to evaluate the efficacy and safety endpoints. The purpose of the SAP is to ensure the credibility of the study findings by pre-specifying the statistical approaches to the analysis of study data prior to database lock (DBL) for the interim analysis and final analysis. The SAP will be finalized and signed off prior to DBL for the interim analysis and for the final analysis. The SAP may be amended, as necessary, to accommodate a protocol amendment(s) and the amended SAP will be finalized and signed off prior to DBL for the final analysis.

The structure and content of this SAP provides sufficient detail to meet the requirements identified by the Food and Drug Administration (FDA), European Medicines Agency (EMA), and International Conference on Harmonization (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use: Guidance on Statistical Principles in Clinical Trials (ICH, 1998). All work planned and reported for this SAP will follow internationally accepted guidelines, published by the American Statistical Association (ASA, 2018) and the Royal Statistical Society (RSS, 2014), for statistical practice.

The planned analyses identified in this SAP may be included in clinical study reports (CSRs), regulatory submissions, or future manuscripts. Also, post-hoc exploratory analyses not necessarily identified in this SAP may be performed to further examine study data. Any post-hoc or unplanned, exploratory analysis performed will be clearly identified as such in the final CSR.

The table, listing, and figure (TLF) shells are provided in a separate file based on this SAP. Upon approval of the SAP, some minor updates, including but not limited to titles, footnotes, headings, and re-numbering of tables on TLF shells are allowed without a SAP amendment as long as the updates within TLF shells do not conflict with the contents of the SAP.

The statistical plan described hereafter is an *a priori* plan. It will be approved before any unblinded inferential or descriptive analysis of data pertaining to the ATX-101-TKA-004 study.

2. Study Objectives and Endpoints

2.1. Study Objectives

The objectives of the trial are:

- To compare the efficacy of ATX101 1,500 mg with that of saline placebo and bupivacaine active comparator in participants undergoing primary unilateral total knee arthroplasty.
- To compare opioid consumption (morphine milligram equivalents (MME)) of participants administered ATX101 with that of participants administered saline placebo and bupivacaine active comparator.
- To evaluate the safety and tolerability of ATX101.

2.2. Study Endpoints

2.2.1. Efficacy Endpoints

2.2.1.1. Primary Efficacy Endpoint

The primary efficacy endpoint of this study is area under the curve (AUC) for the Numerical Rating Scale at Rest (NRS-R) of pain intensity from 30 minutes through 168 hours (Day 8) compared to placebo.

2.2.1.2. Key Secondary Efficacy Endpoints

The key secondary efficacy endpoints of this study include the following:

- AUC for the NRS-R of pain intensity from 30 minutes through 336 hours (Day 15) compared to placebo.
- AUC for the NRS-R of pain intensity from 30 minutes through 504 hours (Day 22) compared to placebo.
- Total (MME) post-surgical consumption of opioid medications from 30 minutes through 336 hours (Day 15) compared to placebo.
- AUC for the NRS-R of pain intensity from 30 minutes through 336 hours (Day 15) compared to active comparator.
- Difference in Numerical Rating Scale with Activity (NRS-A) pain intensity scores during the Timed Up and Go (TUG) Test at Day 15 compared to placebo.

2.2.1.3. Secondary Efficacy Endpoints

The secondary efficacy endpoints of this study include the following:

- Difference in NRS-A pain intensity scores during the TUG Test at Day 8 compared to placebo and active comparator.
- Difference in NRS-A pain intensity scores during the TUG Test at Day 15 compared to active comparator.
- Difference in NRS-A pain intensity scores during the TUG Test at Day 22 compared to placebo and active comparator.
- AUC for the NRS-R of pain intensity from 30 minutes through 168 hours (Day 8) compared to active comparator.
- AUC for the NRS-R of pain intensity from 30 minutes through 504 hours (Day 22) compared to active comparator.
- Total (MME) post-surgical consumption of opioid medications from 30 minutes through 168 hours (Day 8) compared to placebo and active comparator.
- Total (MME) post-surgical consumption of opioid medications from 30 minutes through 504 hours (Day 22) compared to placebo and active comparator.
- Total (MME) post-surgical consumption of opioid medications from 30 minutes to Day 8, Day 9 to Day 15, Day 16 to Day 22, and Day 23 to Day 30 compared to placebo.
- Time (days) to discontinuation of opioid analgesia compared to placebo and active comparator.
- Proportion of participants who are opioid free starting at Day 8, Day 15, Day 22, and Day 30 compared to placebo and active comparator.
- Proportion of participants who remain opioid free from 30 minutes through Day 8, Day 15, Day 22, and Day 30 compared to placebo and active comparator.
- AUC for the NRS-R of pain intensity from 168 hours (Day 8) through 336 hours (Day 15) compared to placebo and active comparator.
- AUC for the NRS-R of pain intensity from 336 hours (Day 15) through 504 hours (Day 22) compared to placebo and active comparator.
- AUC for the NRS-R of pain intensity from 30 minutes to 696 hours (Day 30) compared to placebo and active comparator.

- Time-weighted sum of pain intensity from 30 minutes to Day 8 (time interval in the primary endpoint), Day 9 to Day 15, Day 16 to Day 22, and Day 23 to Day 30 compared to placebo.
- Total (mg) post-surgical consumption of ibuprofen from 30 minutes through Day 30 compared to placebo and active comparator.
- AUC for the NRS of Worst Pain in the last 24 hours compared to placebo and active comparator.
- AUC of Sleep Quality Scale compared to placebo and active comparator.
- Difference in Knee Injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR) compared to placebo and active comparator.
- Difference in the Knee Society Score (KSS) compared to placebo and active comparator.
- Difference in percentage of participants reporting being pain free on the NRS-R of pain intensity compared to placebo and active comparator.
- Difference in NRS-A pain intensity scores during the TUG Test by visit compared to placebo and active comparator.
- Time (days) to 90 Degrees of Flexion compared to placebo and active comparator.
- Time (days) to -5 Degrees of Extension compared to placebo and active comparator.

2.2.1.4. Exploratory Efficacy Endpoints

The exploratory efficacy endpoint of this study is the difference in the duration of TUG Test compared to placebo and active comparator.

2.2.2. Safety Endpoints

The safety endpoints of this study include the following:

- Incidence of adverse events (AEs), adverse events of special interest (AESIs), and serious adverse events (SAEs) compared to placebo and active comparator.
- Reduction in the incidence of opioid-related AEs compared to placebo and active comparator.
- Wound healing assessment using the Southampton Wound Scoring System compared to placebo and active comparator.

3. Overall Study Design and Plan

3.1. Overall Design

This is a Phase 2B, randomized, double-blind, placebo and active comparator controlled, multicenter trial in participants undergoing primary unilateral total knee arthroplasty (TKA).

The duration of the clinical trial for participants is up to approximately 11 weeks, including a 30-day screening period. All participants will be screened within 30 days prior to surgery. Those who meet eligibility criteria will be randomized up to one business day prior to the day of TKA.

3.2. Sample Size and Power

Based on the Part A (ATX-101-TKA-003 study) results, an effect size in the range of 0.50 to 0.55 was deemed to be plausible with 0.52 chosen for an even number of 96 participants per group. This sample size yields greater than 90% power, for the primary comparison of ATX101 vs. placebo, with two-sided alpha of 5%. It is expected that the bupivacaine group will perform similarly to that of placebo for the first 24 hours, and superior from 24 hours onward. After the first 50% of participants have been treated and have either reached the Day 30 timepoint or have exited the trial, an interim analysis will be performed to re-estimate the sample size.

3.3. Study Population

This trial will enroll male and female participants aged 18 through 80, inclusive, undergoing primary unilateral TKA due to knee pain caused by osteoarthritis or post-traumatic arthritis.

3.4. Treatments Administered

Approximately 240 participants will be randomized and dosed to one of the three treatment groups in a 2:2:1 ratio to either ATX101 1,500 mg (three ATX101 500 mg implants; 96 participants), placebo (normal saline periarticular infiltration; 96 participants), or active comparator (bupivacaine hydrochloride 0.25% periarticular infiltration; 48 participants).

3.5. Method of Assigning Subjects to Treatment Groups

A randomization schedule will be computer generated using an interactive response technology (IRT) web-based system. Participants will be randomized in a 2:2:1 ratio to a one-time dose of ATX101 1,500 mg, saline placebo, or bupivacaine active comparator, administered during surgery. Participants who are deemed randomization failures (participants who are randomized but do not receive investigational product), will not be replaced.

3.6. Blinding and Unblinding

ATX-101-TKA-004 is a double-blind study. All participants, investigators, site personnel, and research staff will be blinded to treatment allocation, with the exception of those participating in the management of clinical supplies or participating in the surgical procedure.

The unblinded personnel will include an independent statistician who will generate and have access to the randomization code, as well as an IRT project manager overseeing the IRT system implementation.

The surgical staff will not be blinded to the trial treatment. Allay Therapeutics, Inc. will remain blinded during the treatment phase of trial, unless individual participants need to be unblinded due to safety concerns. Trained and pre-identified monitors will be unblinded throughout the trial to ensure compliance with trial protocols and investigational product accountability. Blinded monitors will be monitoring all other blinded data at the site.

Participants and blinded staff will need to refrain from reviewing any post-surgical x-ray or x-ray report to prevent unblinding.

The IRT will be programmed with blind-breaking instructions. In case of an emergency, the investigator has the sole responsibility for determining if unblinding of a participant's study product assignment is warranted. Participant safety must always be the first consideration in making such a determination. If the investigator decides that unblinding is warranted, the investigator should make every effort to contact the Medical Monitor prior to unblinding a participant's study product assignment, unless this could delay emergency treatment of the participant. If a participant's study product assignment is unblinded, the Medical Monitor must be notified immediately that the blind has been broken. The date and reason that the blind was broken must be recorded in the source documentation.

3.7. Schedule of Events

Please refer to the protocol for a detailed schedule of events.

4. Statistical Analysis and Reporting

This SAP is based on the approved clinical study protocol, amendment 5.0, dated 11-Jul-2025, and describes the statistical methods that are to be used in completing the final analyses of the study.

The reader of this SAP is encouraged to also read the clinical protocol, Safety Monitoring Committee (SMC) charter, and any other identified documents, for details on the planned conduct of the study. Operational aspects related to collection and timing of planned clinical assessments are not repeated in this SAP unless relevant to the planned analysis.

4.1. Introduction

All data recorded in the case report form (CRF) will be displayed in subject listings and sorted by treatment group, subject number, and visit (where applicable). Data from subjects excluded from an analysis set will be included in the applicable data listings but not included in the calculation of summary statistics for that analysis set.

Data processing, tabulation of descriptive statistics, calculation of inferential statistics, and graphical representations except for PK parameter estimation will primarily use SAS[®] (version 9.4 or higher). All PK parameter estimations will use WinNonlin[®] (version 8.3.5 or higher). If the use of other software is warranted, the final statistical methodology report will detail what software was used for what purposes.

Continuous (quantitative) variable summaries will include the number of subjects (n) with non-missing values, mean, standard deviation (SD), median, minimum, and maximum.

Categorical (qualitative) variable summaries will include the frequency and percentage of subjects who are in each category. In general, the denominator for the percentage calculation will be based upon the total number of subjects in the study population within each treatment group, unless otherwise specified. The denominator for by-visit displays will be the number of subjects in the relevant study population with non-missing data at each visit.

The minimum and maximum will be reported with the same degree of precision (i.e., the same number of decimal places) as the observed data. Measures of location (mean and median) will be reported to 1 degree of precision more than the observed data and measures of spread (SD) will be reported to 2 degrees of precision more than the observed data. Percentages will be presented to 1 decimal place, unless otherwise specified, though precision will be capped at a maximum of 4 decimal places due to readability.

All statistical hypothesis tests will be two-sided at the 5% significance level, with 95% corresponding confidence intervals (CIs). P-values will be rounded to four decimal places (e.g., 0.XXXX). If a p-value is less than 0.0001, it will be reported as “< 0.0001”.

This study is considered a confirmatory study. Statistical tests of the primary and key secondary will be regarded as formal tests of specific hypotheses.

Any p-values resulting from multiple comparisons of Treatment-Emergent Adverse Events (TEAEs) should be interpreted with caution. These p-values are not intended to infer statistical significance or clinical relevance and should be considered exploratory in nature.

4.2. Interim Analysis and Data Monitoring

4.2.1. Interim Analysis

An interim analysis (IA) will be performed when approximately 50% of the initially planned enrollment has completed Day 30. Enrollment will continue during the IA, with analyses to be performed by an unblinded Statistical Review Committee (SRC) separate from the team responsible for the conduct and analysis of the trial. The SRC will review the conditional power and recommend one of the following:

- Keep the current sample size and continue as planned, or
- Increase the sample size by up to 50% (100 subjects) using the methodology of Cui, Hung, and Wang as stated in Section 7.2.1.3 of the protocol

The conditional power will be calculated based on the first 50% of patients using the following formula:

$$CP(Z_1, \tilde{n}_2) = 1 - \Phi\left(\frac{1.96\sqrt{n_2} - Z_1\sqrt{n_1}}{\sqrt{\tilde{n}_2}} - \frac{Z_1\sqrt{\tilde{n}_2}}{\sqrt{n_1}}\right)$$

where:

Z_1 = the value of the Z score for the primary estimand at the interim; this will be calculated by taking the inverse normal of the one-sided p-value of the comparison, such that a large, positive Z corresponds to a finding in favor of active treatment.

n_1 = the actual sample size at the interim.

n_2 = the total planned sample size.

$\tilde{n}_2$ = the total planned sample size minus the actual sample size at the interim.

$\Phi()$ = the probability from a normal distribution.

The above conditional power will be calculated based on the primary and the key secondary endpoints, with efficacy analyses to be performed on the Full Analysis Set (FAS) population. These will be examined sequentially to determine the sample size recommendation.

Endpoint	Conditional Power (CP)	Recommendation
AUC _{30M-D8}	≥90%	Check AUC _{30M-D15} CP
AUC _{30M-D8}	<90% and ≥40%	Increase sample size in blocks of 10 participant by the smallest amount that raises the CP of AUC _{30M-D8} over 90%, capped at 100 participants total increase
AUC _{30M-D8}	<40%	No change to sample size
AUC _{30M-D15}	≥90%	Check AUC _{30M-D22} CP
AUC _{30M-D15}	<90% and ≥40%	Increase sample size in blocks of 10 participant by the smallest amount that raises the CP of AUC _{30M-D15} over 90%, capped at 100 participants total increase
AUC _{30M-D15}	<40%	No change to sample size
AUC _{30M-D22}	≥90%	No change to sample size
AUC _{30M-D22}	<90% and ≥40%	Increase sample size in blocks of 10 participant by the smallest amount that raises the CP of AUC _{30M-D22} over 90%, capped at 100 participants total increase
AUC _{30M-D22}	<40%	No change to sample size

AUC_{30M-D15} conditional power will only be considered in the case that the AUC_{30M-D8} has conditional power ≥90%; likewise, AUC_{30M-D22} conditional power will only be considered in the case that the AUC_{30M-D8} and AUC_{30M-D15} have conditional power ≥90%. The sequence will be stopped with the first endpoint that falls under 90% conditional power at the planned sample size and then either the sample size will be kept the same (conditional power for that endpoint <40%) or increased (conditional power <90% and ≥40%).

The trial team will only be informed of the non-binding recommendation to either maintain the original sample size or increase the sample size (by a fixed, pre-specified amount), as well as the endpoint that was the basis of the recommendation.

In the case of a sample size increase, the interim and post-interim cohorts will be combined using the methodology of Cui, Hung, and Wang:

$$Z_1 = \Phi^{-1}(1 - p_1) \quad (1)$$

$$\text{and } Z_2 = w_1 Z_1 + w_2 \Phi^{-1}(1 - p_2)$$

where:

Z_1 = the Z statistic for the first stage.

Z_2 = the combination test statistic at the end of the second stage.

$w_1, w_2 = \sqrt{0.5}$

p_1 = the first stage one-sided p-value.

p_2 = the second stage one-sided p-value based on second stage participants.

$\Phi^{-1}()$ = the inverse normal function.

The weights are defined prospectively according to the square root of the planned proportion of participants in the two stages, relative to the preplanned total enrollment of 160 participants for ATX101 and placebo groups, as $w_i = \sqrt{0.5}$. To control the type-I error, adaptive changes of the stagewise sample sizes will not lead to changes of the weights.

For each difference of interest, the combined point estimate of the difference and interim analysis adjusted two-sided 95% CI will be provided. The formulas for the inferential statistics are given below.

The combined point estimate of the treatment difference is provided in formula (2), and its adjusted confidence interval is provided in formula (3).

$$\hat{\delta} = \frac{r_1 \hat{\delta}^{(1)} + \sqrt{1-r_1} \sqrt{N^*-r_1} \hat{\delta}^{(2)}}{r_1 + \sqrt{1-r_1} \sqrt{N^*-r_1}} \quad (2)$$

$$\hat{\delta} \pm \left(\left| \frac{\hat{\delta}}{Z_2} \right| 1.96 \right) \quad (3)$$

where:

$\delta^{(1)}$ = LS Means difference from stage 1 analysis (interim cohort analysis).

$\delta^{(2)}$ = LS Means difference from stage 2 analysis (post-interim cohort analysis).

r_1 = total number of participants planned to be included at the interim analysis, divided by the total number of participants originally planned (planned sample size) (i.e., 0.5).

N^* = final total number of participants divided by the total number of participants originally planned.

Z_2 = combination test statistic at the end of the second stage obtained in equation (1).

The above weights and adjustments for the interim will apply to both the primary and key secondary endpoints listed in [Section 2.2.1](#).

The p-values from the Cui, Hung, and Wang weights will be used for the evaluation of the primary and key secondary endpoints in the formal statistical testing. Final analyses will also be conducted by pooling data from both stages, without usage of the methodology of Cui, Hung, and Wang.

4.2.2. Safety Monitoring Committee

A Safety Monitoring Committee (SMC) will review blinded safety data after approximately 25%, 50%, and 75% of the planned enrollment has completed Day 30. In the event that a study stopping criterion is met, the SMC will hold an ad-hoc meeting as soon as possible. The SMC will be composed of independent members, including, at a minimum, an anesthesiologist, cardiologist, and orthopedic surgeon. The SMC will operate under a written, detailed charter plan.

The SMC will evaluate trial data including AEs, AESIs, cardiac information, and safety laboratory results from participants. Should data arise where further information is needed to

address a safety concern, the SMC will escalate the data review to an unblinded, independent statistician and physician. The SMC can be unblinded for safety issues. The reviews will monitor for issues that may justify modifying, suspending, or terminating the study.

The trial will also be stopped should there be:

- One death where a clear alternative cause (other than ATX101) is not identified, or
- Two non-fatal SAEs where a clear alternate cause (other than ATX101) is not identified, or
- Five moderate cardiac or central nervous system AESIs with a PK concentration of more than 1,000 ng/mL and a clear alternate cause (other than ATX101) is not identified, or
- Three severe cardiac or central nervous system AESIs with a PK concentration of more than 1,000 ng/mL and a clear alternate cause (other than ATX101) is not identified, or
- Five wound healing AESIs for which a clear alternate cause (other than ATX101) is not identified.

In these instances, the trial blind may be broken for the SMC for the identified participants.

5. Analysis Populations

The following analysis populations are planned for this study:

- **Safety Population:** The Safety population includes all participants who are randomized and administered investigational product. All analyses using the Safety population will group participants according to actual treatment.
- **Full Analysis Set (FAS) Population:** The FAS population includes all participants who are randomized and administered investigational product. All analyses using the FAS population will group participants according to planned treatment.

Assignment of participants to populations will be confirmed at a blinded data review meeting to be held before the study database is locked.

6. General Issues for Statistical Analysis

6.1. Statistical Definitions and Algorithms

6.1.1. Baseline

Baseline is defined as the last observation recorded on or before the administration of study treatment.

6.1.2. Multiple Comparisons

All statistical tests will be two-sided tests, with the null hypothesis that there is no difference between the treatment groups.

To protect the family-wise type I error rate at the 5% significance level, p-values for the primary and key secondary endpoints will be analyzed using a fixed sequence hierarchical testing procedure.

If statistical significance is declared for the primary endpoint, comparisons between ATX101 and placebo or active comparator for the key secondary endpoints will be carried out in the order shown below until a statistically non-significant result is reached. Because the testing is sequential, the type I error rate of $\alpha = 0.05$ is maintained. Failure at any stage in the sequence implies automatic failure at all *subsequent* stages.

1. AUC for the NRS-R of pain intensity from 30 minutes through 336 hours (Day 15) compared to placebo.
2. AUC for the NRS-R of pain intensity from 30 minutes through 504 hours (Day 22) compared to placebo.
3. Total (MME) post-surgical consumption of opioid medications from 30 minutes through 336 hours (Day 15) compared to placebo.
4. AUC for the NRS-R of pain intensity from 30 minutes through 336 hours (Day 15) compared to active comparator.
5. Difference in Numerical Rating Scale with Activity (NRS-A) pain intensity scores during the Timed Up and Go (TUG) Test at Day 15 compared to placebo.

Nominal p-values will be reported for all endpoints but will be declared statistically non-significant once any proceeding comparison is not statistically significant. This also includes p-values for the analysis of secondary and exploratory efficacy, and safety endpoints, i.e., no formal statistical inferences will be drawn on the basis of these tests.

6.1.3. Handling of Dropouts or Missing Data

For handling of missing data, please refer to [Section 8](#) for full details.

6.1.4. Pooling of Sites

The clinical study will be conducted under a common protocol for each investigational site, with the intention of pooling the data for analysis. Every effort will be made to promote consistency in study execution at each investigational site. If an investigational site recruits ≤ 5 patients, it will be pooled into a common analysis site. All sites recruiting greater than this number of patients will not be pooled.

6.1.5. Derived Variables

- Day 1 = Date of administration of study treatment.
- Study day:
 - For assessments occurring before Day 1 = Date of assessment – date of Day 1.
 - For assessments occurring on or after Day 1 = (Date of assessment – date of Day 1) + 1.
- Duration:
 - For AE onset (start) dates occurring before Day 1 = End date – start date.
 - For AE onset (start) dates occurring on or after Day 1 = (End date – start date) + 1.
- Change from baseline = Value at current time point – value at baseline.
- Percent (%) change from baseline = $[(\text{Value at current time point} - \text{value at baseline}) / \text{Value at baseline}] * 100$.
- Prior medication = Any medication taken prior to administration of study treatment.
- Concomitant medication = Any medication taken on or after administration of study treatment.
- Pain-free = NRS-R for pain intensity of 0-1.
- Opioid free Starting at = No opioid use for a period of ≥ 24 hours and does not take for at least 30 days or remainder of the study.
- NSAID free Starting at = No NSAID use for a period of ≥ 24 hours and does not take for at least 30 days or remainder of the study.
- Rescue free Starting at = No rescue use for a period of ≥ 24 hours and does not take for at least 30 days or remainder of the study.

- Remain Opioid free = No opioid use during the noted time period.
- Remain NSAID free = No NSAID use during the noted time period.
- Remain Rescue free = No rescue use during the noted time period.
- Treatment duration = (Treatment end date – treatment start date) + 1.
- Treatment-emergent adverse event (TEAE) = Any AE that occurs on or after administration of study treatment.
- Treatment-related adverse event = Any TEAE with a relationship to study treatment of definitely, probably, or possibly related.
- AUC_{A-B} for the NRS-R of pain intensity = Pain intensity under the numeric rating scale (NRS)-time curve from timepoint A through timepoint B, calculated using the linear trapezoidal rule and the actual time points. The following formula will be used for AUC:

$$\sum_{i=1}^{i=n-1} \frac{(NRS_{i+1} + NRS_i)}{2} * (time_{i+1} - time_i)$$

where NRS_i values 1 to n are those that fall within the period of the AUC being calculated and an additional final calculated value placed exactly at the end of the nominal AUC period; time values are the actual times of NRS collection, and calculations will be done at the precision that the time values are collected (minutes) before being converted to hours.

For the last time point in the interval in the AUC calculation (boundary slice), we will interpolate using the collected time points on each side of the nominal time point unless there is an assessment at the exact nominal time point, in which case no interpolation is needed for the boundary slice. If there is no assessment after going through the steps above, including the imputations, at or after the last nominal time point, then we will use LOCF at the last nominal time point. Note that if this is during a rescue medication window, due to adding points at exactly the end of the window, it will be identical to imputing using the wLOCF approach.

- Time-Weighted sum of pain intensity_{A-B} = AUC_{A-B} for the NRS-R of pain intensity / (Time_B-Time_A)

That is the time-weighted sum of pain intensity will be the AUC divided by the interval of time. Thus, time-weighted sum of pain intensity values will range from a standardized scaling of 0 to 10.

- Time (days) to discontinuation of opioid analgesia = (Date of last reported use of opioid medication – date of administration of study treatment) + 1.
- Time (days) to discontinuation of NSAID analgesia = (Date of last reported use of NSAID medication – date of administration of study treatment) + 1.
- Time (days) to discontinuation of rescue analgesia = (Date of last reported use of rescue medication – date of administration of study treatment) + 1.
- Total (MME) post-surgical consumption of opioid medications = Sum (all doses of opioid medication during the period).
- Total (mg) post-surgical consumption of ibuprofen = Sum (all doses of ibuprofen during the period).

6.1.6. Data Adjustments/Handling/Conventions

Summary statistics and statistical analyses will generally only be performed for subjects included in the relevant analysis population. Data not subject to analysis according to this plan will not appear in any tables or figures, and such data will only be included in listings.

Summary tables and figures will be presented by planned or actual treatment received (depending on the analysis) and overall.

All data collected, including any re-screening data, will be presented in by-patient listings.

Adverse events will be coded using the most current version of Medical Dictionary for Regulatory Activities (MedDRA) at the time of study initiation and will be classified by system organ class (SOC) and preferred term (PT).

Prior and concomitant medications will be coded using the most current version of the World Health Organization Drug Dictionary (WHO-DD) at the time of study initiation and will be classified by anatomical therapeutic class (ATC) 3 and PT.

If partial dates occur in the recording of AEs or medications, the following conventions apply for the purposes of replacing missing dates for determination of TEAEs, duration of AEs, and prior or concomitant medications.

For AE and medication onset (start) dates:

- For medication start dates, no imputation will occur if the patient is not treated. For AE start dates, imputation will occur regardless of whether the patient is treated so that AE duration can be calculated for all patients.
- Imputation of end dates must occur before imputation of start dates.

- If the year is missing or the start date is completely missing, no replacement of the date will occur.
- If just day is missing, and the month and year are the same as the start of treatment, set the start day to the day of administration of study treatment.
- If just day is missing, and the month and year are not the same as the start of treatment or if the patient is not treated, set the start day to the first day of the month.
- If both month and day are missing, and the year is the same as the start of treatment, assign the month and day of administration of study treatment.
- If both month and day are missing, and the year is not the same as the start of treatment or if the patient is not treated, assign 'January 1st'.
- If end date (imputed or not) is not missing, and imputed start date is after end date (imputed or not), set the start date equal to the end date (imputed or not).

For AE and medication end dates:

- For medication end dates, no imputation will occur if the patient is not treated. For AE end dates, imputation will occur regardless of whether the patient is treated so that AE duration can be calculated for all patients.
- Imputation of end dates must occur before imputation of start dates.
- If the year is missing or the stop date is completely missing, no replacement of the date will occur, and the AE or medication will be classified as treatment-emergent or concomitant for patients who were treated.
- If just day is missing, set the end day to the last day of the month.
- If just month is missing, and the year is not the same as the end of study year, assign 'December'.
- If the day and month are missing, and the end year matches the end of study year, assign the end of study month and day.
- If the day and month are missing, and the end year is not the same as the end of study year, assign 'December 31st'.

If missing or partial times occur in the recording of AEs, the following conventions apply for the purposes of replacing missing times for determination of TEAEs and the duration of AEs.

- If the start time is completely missing (both hour and minute), and the start date is the same as the date of administration of study treatment, assign the time of administration of

study treatment. Otherwise, if the start date is not the date of administration of study treatment or if the patient is not treated, set the start time to '12:00 AM'.

- If just the start time minutes are missing, and the start date and hour is the same as the date and hour of administration of study treatment, assign the minutes of administration of study treatment. Otherwise, if the start date and hour is not the date and hour of administration of study treatment or if the patient is not treated, set the minutes to '00'.
- If the end time is completely missing, set the end time to '11:59 PM'.
- If just the end time minutes are missing, set the minutes to '59'.

7. Study Participants and Demographics

7.1. Disposition

Study disposition will be presented for all participants. Percentages will be calculated based on the FAS population. The following items will be summarized:

- Number of participants screened,
- Number of participants randomized,
- Number of participants enrolled (i.e., randomized and dosed),
- Number and percentage of participants in each analysis population, and
- Number and percentage of participants who completed the study or early terminated, along with the reasons for early termination.

Corresponding data listings will be provided separately for disposition, analysis populations and reasons for exclusion, and inclusion/exclusion criteria violated. Re-screened participants will be flagged.

7.2. Protocol Deviations

Protocol deviations will be tracked, recorded, and reviewed prior to DBL, following the Protocol Deviation Guidance Plan for this study.

Protocol deviations will be classified as important or non-important. An important deviation is one that may significantly impact the completeness, accuracy, or reliability of the study data or that may significantly affect the subject's rights, safety, or well-being. A non-important deviation is identified as any protocol deviation that does not meet the criteria for an important deviation.

The final decision regarding classification of protocol deviations will be based on a listing of protocol deviations as of the last participant completing the Day 45 (or ET) visit. This will be determined during a blinded review meeting before any unblinding occurs or DBL, with input from the Clinical and Biostatistics team members and approval from the sponsor.

The number and percentage of participants with important protocol deviations will be summarized by deviation type, treatment group, and overall, for the FAS population.

All important protocol deviations will be presented in a data listing.

7.3. Demographics and Baseline Characteristics

Descriptive statistics will be presented for age (years); sex; childbearing potential (yes/no); ethnicity; race; height (cm); weight (kg); body mass index (BMI) (kg/m²); TUG Test (seconds); NRS-A with location of pain; range of motion (degrees) for knee extension and knee flexion;

NRS-R of pain intensity with location of pain; Worst Pain in the last 24 hours; Sleep Quality Scale; Patient Health Questionnaire-9 (PHQ-9) Score; KOOS, JR Score and subscores; and KSS components and sub-scores.

The table summary will be presented by treatment group and overall, for the FAS population.

For the continuous variables, the number of non-missing values and the mean, SD, median, minimum, and maximum will be tabulated. For the categorical variables, the number and percentage of participants within each category will be tabulated.

All demographics and baseline characteristics will be presented in separate data listings.

7.4. Medical History

The number and percentage of participants reporting various medical histories will be summarized by SOC and PT, treatment group, and overall, for the Safety population. Medical history will be coded using the most current version of MedDRA at the time of study initiation.

All medical history will be presented in a data listing.

7.5. Prior and Concomitant Medication

Prior and concomitant medications will be summarized by ATC 3 and PT, treatment group, and overall. Prior medications, anesthesia regimen, and concomitant medications will all be presented separately.

Prior and concomitant medications will be coded using the most current version of the WHO-DD at the time of study initiation.

Any medications taken prior to administration of study treatment are considered prior medications. Any medications taken on or after administration of study treatment are considered concomitant. If a medication starts before administration of study treatment and continues after this date, it will be considered both prior and concomitant.

All prior and concomitant medications will be presented in a data listing.

7.6. TKA Surgery Information

Surgical information regarding TKA surgery for all participants will be summarized by treatment group and overall, for the Safety population. This includes surgical knee, type of anesthesia, duration of surgery, suprapatellar fat pad removal, administration of investigational product per protocol, tourniquet use, pressure, and duration.

Full details will be provided in a data listing.

7.7. Exposure and Compliance

This section is not applicable for single-dose treatments where participants' treatment is controlled by surgical staff in a hospital and compliance is not an issue.

8. Efficacy Analysis

Efficacy will be evaluated from NRS-R of pain intensity, rescue medication usage, NRS-A of pain intensity during TUG Test, Worst Pain in the last 24 hours, Sleep Quality Scale, KSS components, KOOS, JR Score, and post-surgical consumption of opioid medications.

All efficacy analyses will be performed on the FAS population.

8.1. Efficacy Assessments

8.1.1. Range of Motion (ROM) Assessment

Range of motion will be conducted using a goniometer to measure flexion and extension for the participant per Allay's provided ROM training.

8.1.2. Timed Up and Go (TUG) Test and NRS-A with Location of Pain

For the TUG Test, participants will be timed (stopwatch) on getting up from a chair, walking 10 feet, turn around, walk back to their chair, and sit down. The NRS-A with location will be completed immediately after the TUG Test in which they will then be asked about the amount of pain they had with this activity and the location of pain.

The NRS with activity (NRS-A) for pain intensity is an 11-item scale from 0 to 10 where participants rank the level of pain intensity after an activity. The NRS scale goes from No Pain (0) to Worst Imaginable Pain (10).

8.1.3. Participant Self-Administered Questionnaires

Participants will complete 5 days of assessments during the Screening period for training and compliance purposes. Participants will complete the e-diary post-surgical (as well as before spinal for NRS-R) and at multiple timepoints on Day 1 through Day 30, and in the clinic at Day 45 Visit. A participant may choose to use a provisioned device or their own device to complete the e-diary questionnaires/survey.

8.1.3.1. Knee Society Score (KSS)

The KSS has four components:

- Objective Knee Score (7 items: 100 points) – not collected
- Patient Expectations (3 items: 15 points)
- Patient Satisfaction Score (5 items: 40 points)
- Functional Knee Score (19 items: 100 points)

The KSS is only scored using the patient-reported responses and consists of three separate components: Function, Satisfaction, and Expectation, which should be reported as three separate scores (one for each component). The Objective component, which is a clinician's report, was not collected for this study and, therefore, will not be reported for this study.

8.1.3.2. Numeric Rating Scale at Rest (NRS-R) for Pain Intensity with Location of Pain

The NRS-R for pain intensity is an 11-item scale from 0 to 10 where participants rank the level of pain intensity. The NRS scale goes from No Pain (0) to Worst Imaginable Pain (10). If the participant records a pain intensity score of 1 or greater, they will be asked to indicate the single area of their knee with the most intense pain.

Participants should wait one hour from completion of ice/cold therapy to record their daily pain intensity scores. Prior to obtaining the NRS-R it is recommended participants be in a comfortable resting position (either seated or lying down) for at least 5 minutes prior to obtaining the pain intensity score.

Participants will complete the NRS-R with location of pain questionnaire in the e-diary three times daily for 5 days during the Screening period for training and compliance purposes.

For the first 24 hours after surgery/Time 0, pain assessments will be completed at 30 minutes and 1 hour (+10 min); 3 and 6 hours (+30 minutes); 9, 12, 18, 24 hours (+1 hour). After 24 hours, pain assessment will be completed every 6 hours through 72 hours (± 2 hours). On Day 4 through Day 30, participants will move to a morning, mid-day, and evening and as needed prior to taking any rescue medications (ibuprofen or opioid) for pain.

Prior to administration of rescue ibuprofen and opioid medications, NRS-R scores will be recorded. Participants will be allowed to rescue with either ibuprofen or opioid. The primary endpoint of AUC will be calculated with censoring scheduled pain scores after ibuprofen and opioid use and substituting the pre-rescue value.

NRS-R values will be censored/substituted for the duration of the effectiveness of rescue ibuprofen and opioid medications. Censoring for ibuprofen will be 6 hours and for opioids the actual duration of censoring will be dependent on each type of rescue opioid medication used.

When multiple rescues result in overlapping censoring periods, (e.g., in the immediate postoperative setting when multiple opioids might be used), then the pain intensity value prior to the last rescue medication will be used in place of the censored value. The censoring period will continue until the end of the duration effect of all overlapping analgesics, such that:

- If ibuprofen is followed by a short-acting opioid, and the opioid's censoring period expires while the ibuprofen is still active, the pre-rescue NRS-R score for the ibuprofen will be used until the start of the opioid's censoring period, upon which the pre-rescue NRS-R score for the opioid will be used until the end of the censoring period for all overlapping analgesics.

- If the censoring periods for ibuprofen and a short-acting opioid expire at the same time, the pre-rescue NRS-R score for the last administered rescue medication will be used. This ensures that the most recent pain experience prior to the last rescue medication is accurately reflected.
- If the ibuprofen censoring period expires while a short-acting opioid is still active, the pre-rescue NRS-R score for the opioid will be used until the end of the opioid's censoring period.

To ensure accurate reflection of the patient's pain experience and to avoid bias, the pain score prior to the administration of any rescue medication will be recorded and used for censoring. This approach will prevent the false decrease in pain experience in the comparator groups and maintain the assay sensitivity of the primary endpoint.

NRS censoring rules for rescue medication are as follows:

Rescue Medication	NRS Censoring Time (Hours)	Type of Rescue
The following are the censoring values for ibuprofen and the allowed oral opioid analgesics		
Ibuprofen	6	NSAID
Oxycodone	4	Opioid
Hydromorphone	4	Opioid
The following are the censoring values for all other rescue medications		
Hydromorphone (IM/IV)	3	Opioid
Hydrocodone with Acetaminophen	6	Opioid
Fentanyl (IV)	0.5	Opioid
Morphine (IV/IM)	3	Opioid
Tramadol	6	Opioid
Acetaminophen	6	Other
Codeine	4	Opioid
Acetaminophen with Codeine	6	Opioid

If other rescues are given outside of those enumerated here, then their censoring window will be based on their duration of action. If the name of the rescue medication can not be determined, then the censoring window will be considered 6 hours, and the type of rescue will be considered "Other".

8.1.3.3. Recording Rescue Pain Medication Consumption and Compliance

Participants will complete self-administered questionnaires/surveys in the e-diary to track ibuprofen and opioid consumptions and compliance. Prior to the administration of any rescue pain medication from surgery through the Day 30 Visit, participants will complete an NRS-R in the e-diary assessing the participant's pain intensity and location of pain. Participants should complete the NRS-R with location of pain prior to taking any rescue medications at any time. Medication selected as "Other" in the e-diary will be reconciled with the Concomitant Medication CRF in electronic data capture (EDC). If any discrepancies are identified, a query

will be issued to the site, and the clinical research associate (CRA) will perform the source data review (SDR).

8.1.3.4. Numeric Rating Scale for Worst Pain in the Last 24 hours

The NRS for Worst Pain in the last 24 hours is an 11-item scale from 0 to 10 where participants rank their most severe/worst level of pain intensity over the last 24 hours. The NRS scale goes from No Pain (0) to Worst Imaginable Pain (10).

8.1.3.5. Physical Therapy and Sleep Quality

Participants will complete questions related to physical therapy after TKA in the e-diary. Participants should be trained to record any type of exercise on their own or with a physical therapist as a form of physical therapy.

Participants will be asked in the morning upon awakening to rate their quality of sleep using the Sleep Quality Scale in the e-diary. Quality of sleep will be measured on a numerical scale from 0 (Best Possible Sleep) to 10 (Worst Possible Sleep).

8.1.3.6. Patient Health Questionnaire-9 (PHQ-9) Score

The PHQ-9 score is a measure used to assess the severity of depression. It is calculated by summing the scores of the nine questions, each of which is rated from 0 to 3 based on how often the symptoms have been experienced over the past two weeks. The total score ranges from 0 to 27, with higher scores indicating more severe depression.

Each of the nine questions on the PHQ-9 is scored based on the frequency of the symptoms:

- Not at all: 0 points
- Several days: 1 point
- More than half the days: 2 points
- Nearly every day: 3 points

Response points are totalled, providing a score, and the interpretation of results is as follows:

- 0-4: Minimal or no depression
- 5-9: Mild depression
- 10-14: Moderate depression
- 15-19: Moderately severe depression
- 20-27: Severe depression

8.1.3.7. Local Anesthetic Systemic Toxicity (LAST)-Related Symptoms Survey

Starting on Day 1 after surgery through Day 30, participants will complete a daily questionnaire asking if they have experienced any LAST-related symptoms. Prior to discharge, participants must be provided with detailed education regarding symptoms concerning LAST, and instructions to immediately contact the site staff should they develop any potential LAST-related symptoms. To ensure adequate monitoring for LAST, should any potential LAST-related symptom occur when a participant is completing the daily e-diary, the e-diary will alert the participant to call the site staff. If a participant does not complete the daily e-diary questionnaire for two consecutive days, the site staff will be alerted to call the participant.

8.1.3.8. Knee Injury and Osteoarthritis Outcome Scale for Joint Replacement (KOOS, JR)

The KOOS, JR questionnaire is used to assess the participant's opinion about their knee and associated problems. The assessments are taken directly from the parent KOOS questionnaire and focus on the three categories: joint pain, stiffness, and function in daily living. The questionnaire will be completed in the e-diary during the clinic visits.

The KOOS, JR contains 7 survey items, coded from 0 (none) to 4 (extreme). KOOS, JR is scored by summing the raw response (range 0-28) and then converting it to an interval score. The interval score ranges from 0 to 100, where 0 represents total knee disability and 100 represents perfect knee health.

8.2. Primary Efficacy Endpoint

8.2.1. Primary Estimand

The intercurrent events relevant to this study are as follows:

- IcEv1 (Rescue Medication): Consumption of rescue medication prior to scheduled assessments.
- IcEv2 (Related Treatment Discontinuation): Treatment discontinuation due to an AE or lack of efficacy (LOE).
- IcEv3 (Unrelated Study Discontinuation): Missing data resulting from study withdrawals due to any reason other than AE or LOE.
- IcEv4 (Initial Pain Score Missing): Initial pain score at 30 minutes post-surgical closure is missing.

The main analysis of the primary efficacy endpoint will be conducted using the Primary Estimand, which is described below.

Target Population

Patients requiring analgesia following primary unilateral TKA due to knee pain caused by osteoarthritis or post-traumatic arthritis.

Endpoint

AUC for the NRS-R of pain intensity from 30 minutes through 168 hours (Day 8) compared to placebo.

Population-Level Summary

The difference in least-square (LS) means (ATX101 minus placebo) of AUC_{30m-168h} for the NRS-R of pain intensity.

Intercurrent Event Strategy

A hypothetical strategy will be employed for IcEv1 (Rescue Medication), IcEv2 (Related Study Discontinuation), IcEv3 (Unrelated Study Discontinuation), and IcEv4 (Initial Pain Score Missing). This approach estimates the treatment effect under a scenario where the intercurrent events did not occur.

In relation to IcEv1, defined above, an NRS-R score for the current pain will be recorded just prior to the administration of rescue medication, i.e within 15 minutes of dosing; scheduled assessments collected while the rescue medication would be expected to influence the participant's NRS-R score (i.e., ≥ 15 minutes following rescue medication) will be censored and replaced with pre-rescue values.

As for IcEv2, missing data due to treatment discontinuation will be imputed via multiple imputation using worst observation carried forward (WOCF), under the assumption that participants that discontinue treatment due to an AE or LOE are missing not at random (MNAR) and will have poor outcomes for the unobserved values; these values are assumed to follow the distribution of the worst values through the time point being imputed. Using a regression model with treatment.

For IcEv3, missing data following study withdrawal due to any reason other than AE or LOE are assumed to be missing at random (MAR) and will be imputed using the distribution of the pain scores within treatment group for the time point being imputed. Using a regression model with treatment.

With regard to IcEv4, missing data are assumed to be MAR and are imputed using the distribution of pain scores within the treatment group for the 30-minute time point. Using a regression model with treatment.

Participant narratives will be produced for each participant who withdraws from the study and reviewed for possible indications that the withdrawal was related to AE or LOE; if these cannot be ruled out, the participant's imputation will be handled as described above for AEs and LOE (MNAR). All determinations of MNAR and MAR will be made prior to unblinding, and narratives will note the determination.

8.2.2. Main Estimation

The primary efficacy endpoint of this study is AUC for the NRS-R of pain intensity from 30 minutes through 168 hours (Day 8) compared to placebo.

The null and alternate hypotheses for the primary efficacy endpoint are as follows:

H₁₀: With respect to the AUC_{30m-168h} for the NRS-R of pain intensity, the difference between the ATX101 and placebo treatment groups is equal to zero.

H_{1A}: With respect to the AUC_{30m-168h} for the NRS-R of pain intensity, the difference between the ATX101 and placebo treatment groups is non-zero.

The primary efficacy endpoint will be based on the primary estimand and will be analyzed using a one-way Analysis of Variance (ANOVA) model with treatment group as a fixed effect. All statistical hypothesis tests will be two-sided at the 5% significance level (type I error rate of $\alpha = 0.05$).

A summary table will include LS means (standard error (SE)), 95% CIs, and p-values for each treatment group, with LS mean difference, 95% CI, and p-value to be presented for the comparison of ATX101 vs. placebo. In the case that MI is utilized, this model will be run within each MI repeat and combined using Rubin's method (see [Section 8.2.3](#)).

8.2.3. Multiple Imputation

A multiple imputation (MI) model will be used to impute missing NRS-R scores. Under both MAR- and MNAR-based MI approaches, missing data imputation will be performed using the SAS PROC MI procedure.

For the purposes of the current trial, the three patterns of missingness are:

- MNAR: missing NRS-R score from treatment discontinuation due to AE or LOE
- MAR: missing NRS-R score from study withdrawal due to any reason other than AE or LOE
- MAR: missing initial pain score at 30 minutes post-surgical closure

Fifty imputations will be made per missing NRS-R score using a seed of 523752. The seed may be changed after unblinding, in case of any issues with the imputation process, by adding 1 to the value of the seed until the issue is resolved or it is determined that the issue cannot be resolved by using a different seed. The main analysis will be repeated for each imputed data set and Rubin's method will be used, via SAS PROC MIANALYZE, to pool the results. Each imputation model will include data for each scheduled post-baseline NRS-R of pain intensity. The construction and analysis of these 50 datasets requires eight steps for the primary endpoint:

- **Step 1:** NRS-R values will be grouped with their nominal timepoints for the purposes of performing the multiple imputation.
 - Timepoints will be used per the schedule of assessments through 72 hours, and morning, mid-day, evening for day 4 through 30.
 - Data points may qualify for more than 1 timepoint, in which case it will be used for the imputation in both of the timepoints that it qualifies for.
 - If there is not enough data to converge imputation at a particular timepoint, then timepoints on that day will be averaged together.
- **Step 2:** Any substitution of values for use of rescue medication will then be performed using the raw data as described in [8.1.3.2](#).
 - Example, time points at exactly 4 hours after oxycodone receipt and 6 hours after ibuprofen receipt will be added for the AUC computation at the end of each rescue window.
- **Step 3:** If the 30 minute time point data is missing, then a fully conditional specification (FCS) with regression and predictive mean matching (REGPMM) will be used to impute this datapoint with treatment included in the model. This assumes MAR for the 30 minutes.

If the raw data has no missing data at the 30 minute time point, then no imputation will be performed and the process will start on step 3.

- **Step 4:** Using a standard MAR-based MI approach, the remaining monotone missing data will be imputed for subjects that do not discontinue treatment due to AE or LOE. The variables in the imputation model include each post-baseline visit, and treatment.
- **Step 5:** Using a MNAR-based MI approach, the remaining missing data will be imputed using a method for monotone missingness. The MI procedure used to complete the imputation will use a BY imputation number statement, and request one imputed dataset within each BY group. This step will use an iterative process to impute each time point based off the distribution of worst values through that time point:
 - The worst observation through that time point for each subject will be stored in a single variable. For each nominal time point following discontinuation from treatment due to AE or LOE, missing observations for that time point will be imputed using the regression equation with Worst Observation of NRS-R as the outcome and Treatment Group as the predictor. This will be iterated through all time points with missing data due to AE or LOE starting with the first one and ending with the last one. By doing this, only the treatment a subject is on will influence the worst observation generated via MI from the pool of worst observations through that time point. Moreover, a subject's non-missing responses, other than their worst observation in the pool of worst observations, will not influence their imputed WOCF values. The approach is analogous to the

strategy for pattern imputation based reasons for discontinuation described in Ratitch (2013).

Each imputation step will result in a fully imputed dataset through that time point and will be generated using a FCS REGPMM approach.

Once the process has iterated through all time points.

- **Step 6:** Steps 3 through 5 will be repeated until 50 imputed datasets are generated.
- **Step 7:** Calculation of AUC following formula in section 6.1.5 using actual times for collected data, and nominal times for imputed data. AUCs for subjects will be calculated for each of the 50 versions of the dataset.
- **Step 8:** An ANOVA model will be performed separately for each of the 50 complete analysis datasets, and the results will be combined into one multiple imputation inference (estimated treatment effect, standard error, p-value and associated 95% confidence interval) using Rubin's rule as implemented by the SAS MIANALYZE procedure. The treatment difference will be tested at the 2-sided 0.05 level and corresponding 95% CIs will be calculated.

In the case that there is no missing data, p-values and 95% confidence intervals will come from the ANOVA on the observed data.

8.2.4. Sensitivity Analyses

Three sensitivity analyses will be performed under alternative hypothetical estimands, using MI similar to that detailed in [Section 8.2.3](#):

- The first will impute all missing values due to dropout, regardless of reason for trial withdrawal, from the distribution of worst values observed through each timepoint, conditioned on the non-missing values under the assumption participants that discontinue receive no further benefit from their trial arm and receive no further rescue medication; the distribution of worst values is representative of what they would experience.
- The second will impute all missing values from the placebo distribution, conditioned on the non-missing values under the assumption that participants that dropout will continue to have access to prescribed medication similar to the rescue available to the placebo group, and these values are representative of what they would experience. All other estimand elements will be identical to the primary estimand. This will be achieved by following a control-based PMM MNAR approach.
- The third will impute all missing values from the placebo distribution, conditioned on the non-missing values under the assumption that participants that discontinue treatment due to AE or LOE will continue to have access to prescribed medication similar to the rescue available to the placebo group, and these values are representative of what they would experience. All other estimand elements will be identical to the primary estimand. This will be achieved by following a control-based PMM MNAR approach.

8.2.5. Subgroup Analyses

Subgroup analyses will be performed for the primary efficacy endpoint, using the same imputed data, model, and Rubin's rule as is being used for the primary analysis, and will be based on the primary estimand. Subgroups include, but are not limited to, the following:

- Age Group (years) (≥ 18 to ≤ 65 , > 65 to ≤ 80)
- Sex (Male, Female)
- Analysis Race (White, Black or African American, American Indian or Alaskan Native, Asian (Total), Chinese, Japanese, Korean, Native Hawaiian or Pacific Islander (Native Hawaiian, Guamanian or Chamorro, Samoan, Other Pacific Islander))
- Analysis Ethnicity (Hispanic or Latino, Not Hispanic or Latino)
- Analysis Site

If sample size is not at least 5 in each of the treatment groups for the primary analysis, then the subgroup analysis will not be performed for that particular subgroup.

8.3. Key Secondary Efficacy Endpoints

8.3.1. Secondary Estimand #1

Target Population

Patients requiring analgesia following primary unilateral TKA due to knee pain caused by osteoarthritis or post-traumatic arthritis.

Endpoint

AUC for the NRS-R of pain intensity from 30 minutes through 336 hours (Day 15) compared to placebo.

Population-Level Summary

The difference in least-square (LS) means (ATX101 minus placebo) of $AUC_{30m-336h}$ for the NRS-R of pain intensity.

Intercurrent Event Strategy

A hypothetical strategy will be employed for IcEv1 (Rescue Medication), IcEv2 (Related Study Discontinuation), IcEv3 (Unrelated Study Discontinuation), and IcEv4 (Initial Pain Score Missing). This approach estimates the treatment effect under a scenario where the intercurrent events did not occur.

In relation to IcEv1, an NRS-R score for the current pain will be recorded just prior to the administration of rescue medication, i.e within 15 minutes of dosing; scheduled assessments collected while the rescue medication would be expected to influence the participant's NRS-R

score (i.e., ≥ 15 minutes following rescue medication) will be censored and replaced with pre-rescue values.

As for IcEv2, missing data due to treatment discontinuation will be imputed via multiple imputation using WOCF, under the assumption that participants that discontinue treatment due to an AE or LOE are MNAR and will have poor outcomes for the unobserved values; these values are assumed to follow the distribution of the worst values through the time point being imputed.

For IcEv3, missing data following study withdrawal due to any reason other than AE or LOE are assumed to be MAR and will be imputed using the distribution of the pain scores within treatment group for the time point being imputed.

With regard to IcEv4, missing data are assumed to be MAR and are imputed using the distribution of pain scores within the treatment group for the 30-minute time point.

Participant narratives will be produced for each participant that withdraws from the study and reviewed for possible indications that the withdrawal was related to AE or LOE; if these cannot be ruled out, the participant's imputation will be handled as described above for AEs and LOE (MNAR). All determinations of MNAR and MAR will be made prior to unblinding, and narratives will note the determination.

8.3.2. Secondary Estimand #2

Target Population

Patients requiring analgesia following primary unilateral TKA due to knee pain caused by osteoarthritis or post-traumatic arthritis.

Endpoint

AUC for the NRS-R of pain intensity from 30 minutes through 504 hours (Day 22) compared to placebo.

Population-Level Summary

The difference in least-square (LS) means (ATX101 minus placebo) of $AUC_{30m-504h}$ for the NRS-R of pain intensity.

Intercurrent Event Strategy

A hypothetical strategy will be employed for IcEv1 (Rescue Medication), IcEv2 (Related Study Discontinuation), IcEv3 (Unrelated Study Discontinuation), and IcEv4 (Initial Pain Score Missing). This approach estimates the treatment effect under a scenario where the intercurrent events did not occur.

In relation to IcEv1, an NRS-R score for the current pain will be recorded just prior to the administration of rescue medication, i.e within 15 minutes of dosing; scheduled assessments collected while the rescue medication would be expected to influence the participant's NRS-R score (i.e., ≥ 15 minutes following rescue medication) will be censored and replaced with pre-rescue values.

As for IcEv2, missing data due to treatment discontinuation will be imputed via multiple imputation using WOCF, under the assumption that participants that discontinue treatment due to an AE or LOE are MNAR and will have poor outcomes for the unobserved values; these values are assumed to follow the distribution of the worst values through the time point being imputed.

For IcEv3, missing data following study withdrawal due to any reason other than AE or LOE are assumed to be MAR and will be imputed using the distribution of the pain scores within treatment group for the time point being imputed.

With regard to IcEv4, missing data are assumed to be MAR and are imputed using the distribution of pain scores within the treatment group for the 30-minute time point.

Participant narratives will be produced for each participant that withdraws from the study and reviewed for possible indications that the withdrawal was related to AE or LOE; if these cannot be ruled out, the participant's imputation will be handled as described above for AEs and LOE (MNAR). All determinations of MNAR and MAR will be made prior to unblinding, and narratives will note the determination.

8.3.3. Secondary Estimand #3

Target Population

Patients requiring analgesia following primary unilateral TKA due to knee pain caused by osteoarthritis or post-traumatic arthritis.

Endpoint

Total (MME) post-surgical consumption of opioid medications from 30 minutes through 336 hours (Day 15) compared to placebo.

Population-Level Summary

The difference in LS means (ATX101 minus placebo) of total post-surgical consumption of opioid medications from 30 minutes through 336 hours (Day 15).

Intercurrent Event Strategy

A hypothetical strategy will be employed for IcEv2 (Related Treatment Discontinuation), and IcEv3 (Unrelated Study Discontinuation). This approach estimates the treatment effect under a scenario where the intercurrent events did not occur.

For IcEv2, missing data due to treatment discontinuation will be imputed via multiple imputation using WOCF, under the assumption that participants that discontinue treatment due to an AE or LOE are MNAR and will have poor outcomes for the unobserved values; these values are assumed to follow the distribution of the worst values through the day being imputed.

For IcEv3, missing data following study withdrawal due to any reason other than AE or LOE are assumed to be MAR and will be imputed using the distribution of the daily MME within treatment group for the day being imputed.

8.3.4. Secondary Estimand #4

Target Population

Patients requiring analgesia following primary unilateral TKA due to knee pain caused by osteoarthritis or post-traumatic arthritis.

Endpoint

AUC for the NRS-R of pain intensity from 30 minutes through 336 hours (Day 15) compared to active comparator.

Population-Level Summary

The difference in least-square (LS) means (ATX101 minus active comparator) of AUC_{30m-336h} for the NRS-R of pain intensity.

Intercurrent Event Strategy

A hypothetical strategy will be employed for IcEv1 (Rescue Medication), IcEv2 (Related Study Discontinuation), IcEv3 (Unrelated Study Discontinuation), and IcEv4 (Initial Pain Score Missing). This approach estimates the treatment effect under a scenario where the intercurrent events did not occur.

In relation to IcEv1, an NRS-R score for the current pain will be recorded just prior to the administration of rescue medication, i.e within 15 minutes of dosing; scheduled assessments collected while the rescue medication would be expected to influence the participant's NRS-R score (i.e., ≥ 15 minutes following rescue medication) will be censored and replaced with pre-rescue values.

As for IcEv2, missing data due to treatment discontinuation will be imputed via multiple imputation using WOCF, under the assumption that participants that discontinue treatment due to an AE or LOE are MNAR and will have poor outcomes for the unobserved values; these values are assumed to follow the distribution of the worst values through the time point being imputed.

For IcEv3, missing data following study withdrawal due to any reason other than AE or LOE are assumed to be MAR and will be imputed using the distribution of the pain scores within treatment group for the time point being imputed.

With regard to IcEv4, missing data are assumed to be MAR and are imputed using the distribution of pain scores within the treatment group for the 30-minute time point.

Participant narratives will be produced for each participant that withdraws from the study and reviewed for possible indications that the withdrawal was related to AE or LOE; if these cannot be ruled out, the participant's imputation will be handled as described above for AEs and LOE (MNAR). All determinations of MNAR and MAR will be made prior to unblinding, and narratives will note the determination.

8.3.5. Secondary Estimand #5

Target Population

Patients requiring analgesia following primary unilateral TKA due to knee pain caused by osteoarthritis or post-traumatic arthritis.

Endpoint

Difference in NRS-A pain intensity scores during the TUG Test at Day 15 compared to placebo.

Population-Level Summary

The difference in LS means (ATX101 minus placebo) of NRS-A pain intensity scores during the TUG Test at Day 15.

Intercurrent Event Strategy

A treatment policy strategy will be employed for IcEv1 (Rescue Medication). This approach is used to estimate the effect irrespective of whether the intercurrent event occurred.

A hypothetical strategy will be employed for IcEv2 (Related Treatment Discontinuation), IcEv3 (Unrelated Study Discontinuation), and IcEv4 (Initial Pain Score Missing). This approach estimates the treatment effect under a scenario where the intercurrent event did not occur.

For IcEv2, missing data due to treatment discontinuation will be imputed via multiple imputation using WOCF, under the assumption that participants that discontinue treatment due to an AE or LOE are MNAR and will have poor outcomes for the unobserved values; these values are assumed to follow the distribution of the worst values through the time point being imputed.

For IcEv3, missing data following study withdrawal due to any reason other than AE or LOE are assumed to be MAR and will be imputed using the distribution of the pain scores within treatment group for the time point being imputed.

8.3.6. Main Estimation

The following key secondary efficacy endpoints will be summarized separately:

- AUC for the NRS-R of pain intensity from 30 minutes through 336 hours (Day 15) compared to placebo.
- AUC for the NRS-R of pain intensity from 30 minutes through 504 hours (Day 22) compared to placebo.
- Total (MME) post-surgical consumption of opioid medications from 30 minutes through 336 hours (Day 15) compared to placebo.
- AUC for the NRS-R of pain intensity from 30 minutes through 336 hours (Day 15) compared to active comparator.
- Difference in NRS-A pain intensity scores during the TUG Test at Day 15 compared to placebo.

For analysis purposes, each of the key secondary efficacy endpoints will be evaluated in a fixed sequence, specified below, with the endpoints tested hierarchically. To protect the family-wise error rate, key secondary efficacy endpoints will only be tested if the primary efficacy endpoint is significant at the 5% level.

The null and alternate hypotheses for the key secondary efficacy endpoints are as follows:

H2₀: With respect to the AUC_{30m-336h} for the NRS-R of pain intensity, the difference between the ATX101 and placebo treatment groups is equal to zero.

H2_A: With respect to the AUC_{30m-336h} for the NRS-R of pain intensity, the difference between the ATX101 and placebo treatment groups is non-zero.

H3₀: With respect to the AUC_{30m-504h} for the NRS-R of pain intensity, the difference between the ATX101 and placebo treatment groups is equal to zero.

H3_A: With respect to the AUC_{30m-504h} for the NRS-R of pain intensity, the difference between the ATX101 and placebo treatment groups is non-zero.

H4₀: With respect to the total post-surgical consumption of opioid medications from 30 minutes through 336 hours (Day 15), the difference between the ATX101 and placebo treatment groups is equal to zero.

H4_A: With respect to the total post-surgical consumption of opioid medications from 30 minutes through 336 hours (Day 15), the difference between the ATX101 and placebo treatment groups is non-zero.

H5₀: With respect to the AUC_{30m-336h} for the NRS-R of pain intensity, the difference between the ATX101 and active comparator treatment groups is equal to zero.

H5_A: With respect to the AUC_{30m-336h} for the NRS-R of pain intensity, the difference between the ATX101 and active comparator treatment groups is non-zero.

H6₀: With respect to the NRS-A pain intensity scores during the TUG Test at Day 15, the difference between the ATX101 and placebo treatment groups is equal to zero.

H6_A: With respect to the NRS-A pain intensity scores during the TUG Test at Day 15, the difference between the ATX101 and placebo treatment groups is non-zero.

The key secondary efficacy endpoints will be tested hierarchically, with each pair of hypotheses for each endpoint to be tested under the fixed sequence procedure, i.e., H_{20} and $H_{2A} \rightarrow H_{30}$ and $H_{3A} \rightarrow H_{40}$ and $H_{4A} \rightarrow H_{50}$ and $H_{5A} \rightarrow H_{60}$ and H_{6A} . In each test of the hierarchy, point estimates will also be used to confirm that ATX101 was found to be significantly better than placebo or active comparator.

The endpoints related to AUC for the NRS-R of pain intensity compared to placebo will be based on the primary estimand ([Section 8.2.1](#)). Analysis will be performed using a one-way ANOVA model—similar to that of the primary efficacy endpoint described in [Section 8.2.2](#)—and will be summarized in separate tables.

The endpoint, total post-surgical consumption of opioid medications from 30 minutes through 336 hours (Day 15) compared to placebo, will be based on the secondary estimand. Analysis will be performed using a one-way ANOVA model—similar to that of the primary efficacy endpoint described in [Section 8.2.2](#)—and will be summarized in a separate table.

The endpoint, AUC for the NRS-R of pain intensity from 30 minutes through 168 hours (Day 8) compared to active comparator, will be based on the secondary estimand #4. Analysis will be performed using a one-way ANOVA model—similar to that of the primary efficacy endpoint described in [Section 8.2.2](#)—and will be summarized in a separate table.

The endpoint, difference in NRS-A pain intensity scores during the TUG Test at Day 15 compared to placebo, will be based on the secondary estimand #5. Analysis will be performed using a one-way ANOVA model of observed values at Day 15—similar to that of the primary efficacy endpoint described in [Section 8.2.2](#)—and will be summarized in a separate table.

MI of key secondary efficacy endpoints will be performed to that detailed in [Section 8.2.3](#), with specific details of intercurrent events differing as described under the estimands sections.

All statistical hypothesis tests will be two-sided at the 5% significance level (type I error rate of $\alpha = 0.05$).

8.4. Secondary Efficacy Endpoints

The following secondary efficacy endpoints will be summarized separately:

- Difference in NRS-A pain intensity scores during the TUG Test at Day 8 compared to placebo and active comparator.
- Difference in NRS-A pain intensity scores during the TUG Test at Day 15 compared to active comparator.

- Difference in NRS-A pain intensity scores during the TUG Test at Day 22 compared to placebo and active comparator.
- AUC for the NRS-R of pain intensity from 30 minutes through 168 hours (Day 8) compared to active comparator.
- AUC for the NRS-R of pain intensity from 30 minutes through 504 hours (Day 22) compared to active comparator.
- Total (MME) post-surgical consumption of opioid medications from 30 minutes through 168 hours (Day 8) compared to placebo and active comparator.
- Total (MME) post-surgical consumption of opioid medications from 30 minutes through 504 hours (Day 22) compared to placebo and active comparator.
- Total (MME) post-surgical consumption of opioid medications from 30 minutes to Day 8, Day 9 to Day 15, Day 16 to Day 22, and Day 23 to Day 30 compared to placebo.
- Time (days) to discontinuation of opioid analgesia compared to placebo and active comparator.
- Proportion of participants who are opioid free starting at Day 8, Day 15, Day 22, and Day 30 compared to placebo and active comparator.
- Proportion of participants who remain opioid free from 30 minutes through Day 8, Day 15, Day 22, and Day 30 compared to placebo and active comparator.
- AUC for the NRS-R of pain intensity from 168 hours (Day 8) through 336 hours (Day 15) compared to placebo and active comparator.
- AUC for the NRS-R of pain intensity from 336 hours (Day 15) through 504 hours (Day 22) compared to placebo and active comparator.
- AUC for the NRS-R of pain intensity from 30 minutes to 696 hours (Day 30) compared to placebo and active comparator.
- Time-weighted sum of pain intensity from 30 minutes to Day 8 (time interval in the primary endpoint), Day 9 to Day 15, Day 16 to Day 22, and Day 23 to Day 30 compared to placebo.
- Total (mg) post-surgical consumption of ibuprofen from 30 minutes through Day 30 compared to placebo and active comparator.
- AUC for the NRS of Worst Pain in the last 24 hours compared to placebo and active comparator from 30 minutes to Day 8, Day 15, Day 22, and Day 30.

- AUC of Sleep Quality Scale compared to placebo and active comparator.
- Difference in KOOS, JR score compared to placebo and active comparator.
- Difference in the KSS compared to placebo and active comparator.
- Difference in percentage of participants reporting being pain free on the NRS-R of pain intensity compared to placebo and active comparator.
- Difference in NRS-A pain intensity scores during the TUG Test by visit compared to placebo and active comparator.
- Time (days) to 90 Degrees of Flexion compared to placebo and active comparator.
- Time (days) to -5 Degrees of Extension compared to placebo and active comparator.

All endpoints related to AUC for the NRS-R of pain intensity (including time-weighted sum of pain intensity) will be analyzed using a one-way ANOVA model—similar to that of the primary efficacy endpoint described in [Section 8.2.2](#)—and will be summarized in separate tables. Additionally, a box plot will be produced of NRS-R of pain intensity over time and will be presented by treatment group and visit as well as a mean plot over time by treatment group. AUC for NRS-R of pain intensity analysis will be repeated for intervals starting at 2 hours through Days 8, 15, 22, and 30 respectively.

All endpoints related to the difference in NRS-A pain intensity scores during the TUG Test will be analyzed using a one-way ANOVA model—similar to that of the primary efficacy endpoint described in [Section 8.2.2](#)—and will be summarized in separate tables.

The endpoint, total (MME) post-surgical consumption of opioid medications from 30 minutes to Day 8, Day 9 to Day 15, Day 16 to Day 22, and Day 23 to Day 30 compared to placebo, will be analyzed using a one-way ANOVA model—similar to that of the primary efficacy endpoint described in [Section 8.2.2](#)—and will be summarized by visit.

All other endpoints related to the total post-surgical consumption of opioid medications or ibuprofen will be analyzed using a one-way ANOVA model—similar to that of the primary efficacy endpoint described in [Section 8.2.2](#)—and will be summarized in separate tables.

A Kaplan-Meier plots will be produced for each of time to discontinuation of opioid, NSAID, and any rescue analgesia and will be presented by treatment group. Accompanying tables will also be produced, summarizing the number of participants with the event and the time to event (in days), as well as associated Kaplan-Meier estimates (including 95% CI) for the 25th percentile, median, and 75th percentile in addition to Hazard Ratios (including 95% CI) and p-value from a Cox Proportional Hazards model.

All endpoints related to the proportion of participants: opioid free, NSAID free, and rescue free will each be analyzed using Chi-Square test. Note: if the expected cell count is less than 5 for any cell, then Fisher's exact test will be used as an alternative for that test. A summary table will

be presented by visit and treatment group and will include the proportion of participants for each treatment group, with the difference in proportion of participants, 95% CI, and p-value to be presented for contrast comparisons against ATX101. Additional timepoints will be also be considered for analysis on these endpoints.

The endpoint, AUC for the NRS of worst pain in the last 24 hours, will be analyzed over time using a one-way ANOVA model—similar to that of the primary efficacy endpoint described in [Section 8.2.2](#)—and will be summarized by visit. Additionally, a box plot will be produced of the NRS of worst pain in the last 24 hours over time and will be presented by treatment group and visit.

The endpoint, AUC of sleep quality scale, will be analyzed over time using a one-way ANOVA model—similar to that of the primary efficacy endpoint described in [Section 8.2.2](#)—and will be summarized by visit. Additionally, a box plot will be produced of sleep quality scale scores over time and will be presented by treatment group and visit.

The endpoints, difference in KOOS, JR score, subscales and difference in the KSS, will be analyzed separately using a one-way ANOVA model—similar to that of the primary efficacy endpoint described in [Section 8.2.2](#)—and will be summarized by visit. KSS will also be summarized by component type (Function, Satisfaction, Expectation) within the same table.

The endpoint, difference in percentage of participants reporting being pain free on the NRS-R of pain intensity, will be analyzed using Chi-Square test. Note: if the expected cell count is less than 5 for any cell, then Fisher's exact test will be used as an alternative for that test. A summary table will be presented by visit and treatment group and will include the proportion of participants for each treatment group, with the difference in proportion of participants, 95% CI, and p-value to be presented for contrast comparisons against ATX101.

For the endpoints, time to 90 degrees of flexion and time to -5 degrees of extension, separate Kaplan-Meier plots will be presented by treatment group. Accompanying tables will also be produced, summarizing the number of participants with the event and the time to event (in days), as well as associated Kaplan-Meier estimates (including 95% CI) for the 25th percentile, median, and 75th percentile in addition to Hazard Ratios (including 95% CI) and p-value from a Cox Proportional Hazards model.

The endpoint, difference in NRS-A pain intensity scores during the TUG Test by visit, will be analyzed over time using a one-way ANOVA model—similar to that of the primary efficacy endpoint described in [Section 8.2.2](#)—and will be summarized by visit.

All statistical hypothesis tests will be two-sided at the 5% significance level (type I error rate of $\alpha = 0.05$).

8.5. Exploratory Efficacy Endpoints

The exploratory efficacy endpoint, difference in the duration of TUG Test compared to placebo and active comparator, will be analyzed using a one-way ANOVA model—similar to that of the primary efficacy endpoint described in [Section 8.2.2](#)—and will be summarized by visit.

9. Safety and Tolerability Analysis

Safety will be evaluated from reported AEs, changes in clinical laboratory values (hematology and chemistry), vital signs, 12-lead electrocardiogram (ECG) results, physical examination findings, neurological assessment findings, and wound healing assessment results.

All safety analyses will be performed on the Safety population.

9.1. Adverse Events

Adverse events will be collected from informed consent through Day 45 (or ET).

Treatment-emergent adverse events (TEAEs) are defined as AEs that are new or worsening after the patient has received the investigational product.

All AEs will be coded using the most current MedDRA version at the time of study initiation.

For partially missing dates and times, the imputation rules described in [Section 6.1.6](#) will be followed, and imputed dates will be used to determine whether an AE is treatment-emergent and the duration of the AE. If the start date is completely missing, then the AE is considered treatment-emergent if the end date and time (imputed or not) is on or after the date and time of administration of study treatment. If both the start and end dates are completely missing, then the AE is considered treatment-emergent. If the start or end date of the AE is missing, and the AE is resolved, the AE duration will be set as missing.

All AEs will be classified by severity (mild, moderate, severe, life-threatening), seriousness (yes, no), relationship to investigational product (unrelated, unlikely related, probably related, possibly related, definitely related), relationship to opioid use (unrelated, related), and relationship to TKA (unrelated, unlikely related, probably related, possibly related, definitely related). The AE outcome will be classified as recovered/resolved, recovered/resolved with sequelae, recovering/resolving, not recovered/not resolved, fatal, or unknown. Only AEs with an outcome of death will be reported as fatal.

The following summary tables for TEAEs will be presented by treatment group and overall:

- Overall summary of TEAEs, including number of patients with
 - any TEAE,
 - any TEAE by maximum severity,
 - any treatment-emergent SAE,
 - any treatment-emergent AESI,
 - any TEAE with Local Anesthetic Systemic Toxicity (LAST)-related symptom,
 - any TEAE requiring medical treatment,

- any TEAE related to investigational product,
 - any TEAE related to opioid use,
 - any TEAE related to TKA,
 - any TEAE leading to study discontinuation, and
 - any TEAE leading to death.
- Number and percentage of patients reporting TEAEs by SOC and PT.
- Number and percentage of patients reporting TEAEs by SOC, PT, and maximum severity.
- Number and percentage of patients reporting TEAEs related to investigational product by SOC and PT.
- Number and percentage of patients reporting opioid-related TEAEs by SOC and PT (as specified in [Section 9.1.2.1](#)).
- Number and percentage of patients reporting TEAEs related to TKA by SOC and PT.
- Number and percentage of patients reporting treatment-emergent SAEs by SOC and PT.
- Number and percentage of patients reporting treatment-emergent AESIs by SOC and PT.
- Number and percentage of patients reporting TEAEs leading to study discontinuation by SOC and PT.
- Number and percentage of patients with TEAEs leading to death by SOC and PT.

All summary tables by SOC and PT will be sorted by descending frequency of SOC and then by descending frequency of the PT within each SOC (based on the total frequency across all treatment groups). SOC terms and PTs with the same frequency will be presented alphabetically. A participant will be counted only once by the highest severity grade within a SOC and preferred term, even if the participant experienced more than one TEAE within a specific SOC and preferred term.

For the summary tables by severity, the maximum severity (life-threatening > severe > moderate > mild) will be reported. TEAEs with an unknown or missing severity will be classified as ‘Severe’.

For the summary table of opioid-related TEAEs, comparisons against ATX101 will be analyzed—at the SOC level—using a risk difference with associated pvalue and 95% confidence interval of ATX101 against placebo and active comparator groups. These differences will be

estimated using an exact confidence interval calculated per Chan and Zhang 1999, and a pvalue from a Fisher's exact test.

For the summary tables by relationship to investigational product or TKA, AEs reported as definitely related, probably related, or possibly related will be classified as 'Related', while AEs reported as unlikely related or unrelated will be classified as 'Not Related'. TEAEs with an unknown or missing relationship will be classified as 'Related'.

All AEs will be presented in a data listing. The AE listing will include treatment group, patient identifier, demographic information (age, race, sex), SOC, PT, and reported term, start and end dates of AE (including start and end days), AE duration, severity grade, seriousness, relationship to investigational product, opioid use, and TKA, action taken with respect to AE, and AE outcome. AEs that are treatment-emergent will be flagged.

9.1.1. Serious Adverse Events

An SAE is any untoward medical occurrence that has any of the following consequences:

- Results in death,
- Is life-threatening,
- Requires in-patient hospitalization or prolongation of existing hospitalization,
- Results in persistent or significant disability/incapacity, or
- Is a congenital anomaly/birth defect.

Note: The term "life-threatening" in the definition of "serious" refers to an event in which the participant was at risk of death at the time of the event; it does not refer to an event that hypothetically might have caused death if it were more severe.

Medical and scientific judgment should be exercised in deciding whether expedited reporting is appropriate in other situations, such as important medical events that may not be immediately life-threatening, or result in death or hospitalization, but may jeopardize the participant or may require intervention to prevent one of the other outcomes listed in the definition above. These should also usually be considered serious. Examples of such events are intensive treatment in an emergency room or at home for allergic bronchospasm, blood dyscrasia, or convulsions that do not result in hospitalization; or development of drug dependency or drug abuse. A new diagnosis of cancer should be considered serious and should be reported on the SAE form.

All SAEs will be presented in a data listing.

9.1.2. Adverse Events of Special Interest and LAST-Related Symptoms

Adverse Events of Special Interest (AESI), including LAST-related symptoms, will also be collected in this trial and include the following:

Central Nervous System Symptoms/Events

- Tongue and perioral numbness
- Metallic taste
- Dysarthria
- Tinnitus
- Muscle fasciculations
- Tremors
- Seizures
- Global CNS depression
- Apnea (not sleep apnea)
- Severe dizziness
- Severe drowsiness
- Severe change in mental status

Cardiovascular Symptoms/Events

- Sinus bradycardia
- Atrioventricular blocks
- Conduction defects (prolonged PR or prolonged QRS)
- Ventricular dysrhythmias
- Cardiac arrest
- Asystole

Investigations

- Elevated liver enzymes (Grade 3)

Renal and Urinary Disorders

- Moderate chronic urinary retention

Wound Healing Symptoms/Events

- Wound dehiscence
- Wound pus/purulent discharge (Grade 4 from Southampton Wound Scoring System)
- Deep wound infection (Grade 5 from Southampton Wound Scoring System)
- Cellulitis near the incision site (Grade 5 from Southampton Wound Scoring System)

All AESIs and LAST-related symptoms will be presented in a data listing.

9.1.2.1. Opioid-Related Adverse Events

For reporting purposes, the following AEs are deemed to be opioid-related:

- Constipation
- Nausea
- Vomiting
- Urinary retention
- Pruritis
- Delirium
- Hallucinations
- Sedation
- Respiratory depression
- Paralytic Ileus

9.1.3. Adverse Events Leading to Study Discontinuation

All adverse events leading to study discontinuation will be presented in a data listing.

9.1.4. Deaths

All adverse events leading to death will be presented in a data listing.

9.2. Clinical Laboratory Evaluations

Laboratory data will be summarized by type of laboratory test (hematology, chemistry), visit, treatment group, and overall, and descriptive statistics for continuous laboratory parameters at each assessment visit will be presented for the observed and change from baseline values.

The following laboratory parameters will be measured:

- Hematology: hemoglobin (gm/dL), white blood cell ($10^3/\text{uL}$), lymphocytes ($10^3/\text{uL}$), neutrophils ($10^3/\text{uL}$), eosinophils ($10^3/\text{uL}$), and platelets ($10^3/\text{uL}$).
- Chemistry: sodium (mmol/L), potassium (mmol/L), blood urea nitrogen (mg/dL), creatinine (mg/dL), phosphorous (mg/dL), albumin (g/dL), alkaline phosphate (U/L), aspartate aminotransferase (U/L), alanine aminotransferase (U/L), total bilirubin, and gamma-glutamyl transferase.

Laboratory results outside the normal range will be flagged. The abnormal values will be flagged with 'L' (low) for values below the lower limit of the laboratory's normal range or 'H' (high) for values above the upper limit of the laboratory's normal range.

The number of subjects with clinical laboratory values below, within, or above the normal range by visit and in relation to baseline will be summarized in a shift table for each clinical laboratory analyte by treatment group and overall.

Separate listings will be provided for hematology and chemistry laboratory data. Drug screen and pregnancy test results will not be tabulated but will be presented in data listings.

9.2.1. Liver Function Test Elevations

Frequency of liver function test elevations (i.e., using the maximum post-baseline value) will be provided in a summary table for the following events:

- $\geq 1x, \geq 3x, \geq 5x$ ULN elevation in aspartate aminotransferase (AST),
- $\geq 1x, \geq 3x, \geq 5x$ ULN elevation in alanine aminotransferase (ALT),
- $\geq 1x, \geq 3x, \geq 5x$ ULN elevation in AST or ALT at the same visit,
- $\geq 1x, \geq 2x, \geq 3x$ ULN elevation in alkaline phosphatase (ALP), and
- $\geq 1x, \geq 2x, \geq 3x$ ULN elevation in total bilirubin.

Additionally, potential Hy's Law cases will be identified based on the following criteria:

- $\text{AST or ALT} \geq 3x \text{ ULN}$
- $\text{Total bilirubin} > 2x \text{ ULN}$

- ALP < 2x ULN (to rule out cholestasis)

Note: ULN = Upper Limit of Normal.

Maximum of post-baseline ALT and Bilirubin values will be plotted against each other with Bilirubin on the Y-axis, and ALT on the X-axis with all values adjusted for upper limit of normal.

9.3. Vital Signs

Descriptive summaries of observed and changes from baseline values will be calculated for weight (kg), systolic blood pressure (mmHg), diastolic blood pressure (mmHg), pulse rate (beats/min), respiration rate (breaths/min), body temperature (°C), and oxygen saturation (%). Results will be summarized by visit, treatment group, and overall.

All vital signs will be presented in a data listing.

9.4. 12-Lead Electrocardiogram (ECG)

Triplicate ECGs will be taken 2 minutes apart (± 1 minute). Participants will be at rest in a supine position for 10 minutes before the ECGs are obtained.

The corrected QT interval (QTc) will be calculated using the Fridericia correction method (QTcF) as follows:

$$QTcF \text{ (msec)} = \frac{QT \text{ (ms)}}{\sqrt[3]{RR(\text{ms})/1000}}$$

Descriptive summaries will be presented for the average triplicate ECG measures of HR (beats/min), RR interval (msec), PR interval (msec), QRS interval (msec), QT interval (msec), and QTcF interval (Fridericia method). Summaries of observed values will be presented by treatment group.

Additionally, counts and percentages for overall investigator interpretation (Normal, Abnormal—NCS, Abnormal—CS) will be summarized by treatment group.

All ECG measurements will be presented in a data listing.

9.5. Physical Examination

A complete physical exam will be performed at the Screening Visit. An abbreviated/directed physical exam will be performed to include an assessment of participant's symptoms as well as auscultation of heart and lungs on Day 1 and on Day 45 (or ET).

All physical examination findings will be presented in a data listing.

9.6. Neurological Assessment

Neurological assessments are once daily (around evening mealtime) during the hospitalization/inpatient period from Day 1 through Day 4. On Day 4, neurological assessments should be conducted prior to discharge which may be earlier than the evening mealtime.

The participant will be queried on abnormal senses (ears/hearing, sight, touch, smell, and taste).

All neurological assessment findings will be presented in a data listing.

9.7. Wound Healing Assessment

The wound healing assessment, based on the Southampton Wound Scoring System, will be completed from Day 4 through Day 45 (or ET).

Descriptive summaries of observed values will be presented by visit, treatment group, and overall.

All wound healing assessment results will be presented in a data listing.

10. Changes from Planned Analysis

1. The Safety population definition is inclusive of all treatment groups, and not just ATX101 patients.
2. Further sensitivity analyses were also added, per FDA request (see [Section 8.2.4](#)).
3. Estimand #3 MME incurrent event strategies were updated to use a MNAR approach for discontinuation due to AE or LOE.

11. Other Planned Analysis

11.1. Pharmacokinetic Analysis

A PK sample is required for bupivacaine monitoring for an AESI or LAST-related symptom.

If LAST-related symptom is reported before discharge from inpatient setting, a PK sample will be collected every day until discharge and at all regularly scheduled visits.

If LAST-related symptom is reported after discharge, a PK sample will be promptly collected (unscheduled or scheduled visit) and additional PK samples collected at all regularly scheduled visits following the event.

Collection and handling of PK samples is described in the Laboratory Manual.

All PK data will be provided in a data listing.

12. References

1. ASA. (2018) Ethical Guidelines for Statistical Practice. Prepared by the Committee on Professional Ethics, April 2018. <http://www.amstat.org/about/ethicalguidelines.cfm>
2. Chan, I. S., & Zhang, Z. (1999). Test-based exact confidence intervals for the difference of two binomial proportions. *Biometrics*, 55(4), 1202-1209.
3. Cui L, Hung HM, Wang SJ. Modification of sample size in group sequential clinical trials. *Biometrics*. 1999;55(3):853-7.
4. ICH (1998). ICH Harmonised Tripartite Guideline. Statistical Principles for Clinical Trials E9; 1998. https://database.ich.org/sites/default/files/E9_Guideline.pdf
5. Lehman W, Wassmer G. Adaptive sample size calculations in group sequential trials. *Biometrics*. 1999;55:1286-90.
6. Lawrence J, Hung HMJ. Estimation and confidence intervals after adjusting the maximum information. *Biometrical Journal*. 2003;45(2):143-52.
7. Mehta C, Pocock S. Adaptive increase in sample size when interim results are promising: A practical guide with examples. *Statistics in Medicine*. 2010;30:3267-84.
8. Ratitch B, O'Kelly M, Tosiello R. Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. *Pharmaceut. Statist.* 2013, 12 337–347.
9. RSS. (2014) The Royal Statistical Society: Code of Conduct, 2014. <https://rss.org.uk/about/policy-and-guidelines/code-of-conduct/>.
10. Rubin, DB. Inference and missing data. *Biometrika*. 1976;63:581-592.

13. Appendix

13.1. Tables, Listings, and Figures

All tables, listings, and figures will have a header showing the sponsor company name and protocol, and a footer showing the version of SAS, the file name and path, and the source of the data (CRF page or listing number), as illustrated in [Section 13.2.1](#).

13.1.1. Planned Table Descriptions

The following are planned summary tables for protocol number ATX-101-TKA-004. The table numbers are placeholders only and will be determined when the tables are produced.

13.1.1.1. Study Participants and Demographics

Table 1: Study Participants and Demographics

Table Number	Population	Table Title / Summary
Table 14.1.1	All Participants	Study Disposition and Analysis Populations
Table 14.1.2	FAS	Demographics and Baseline Characteristics
Table 14.1.3	Safety	Medical History
Table 14.1.4.1	Safety	Prior Medications
Table 14.1.4.2	Safety	Concomitant Medications
Table 14.1.5	FAS	Important Protocol Deviations
Table 14.1.6	Safety	Total Knee Arthroplasty (TKA) Surgery Information

13.1.1.2. Efficacy Data

Table 2: Efficacy Data

Table Number	Population	Table Title / Summary
Table 14.2.1	FAS	Summary of ANOVA difference in AUC for the NRS-R of pain intensity from 30 minutes through 168 hours (Day 8) compared to placebo
Table 14.2.1.1	FAS	Sensitivity Analysis #1: Summary of ANOVA difference in AUC for the NRS-R of pain intensity from 30 minutes through 168 hours (Day 8) compared to placebo – Discontinuation following Worst Value Distribution Imputation
Table 14.2.1.2	FAS	Sensitivity Analysis #2: Summary of ANOVA difference in AUC for the NRS-R of pain intensity from 30 minutes through 168 hours (Day 8) compared to placebo – Discontinuation following Placebo Distribution Imputation
Table 14.2.1.3	FAS	Sensitivity Analysis #3: Summary of ANOVA difference in AUC for the NRS-R of pain intensity from 30 minutes through 168 hours (Day 8) compared to placebo – AE and LOE Discontinuation following Placebo Distribution Imputation

Table 14.2.1.4	FAS	Subgroup Analysis #1: Summary of ANOVA difference in AUC for the NRS-R of pain intensity from 30 minutes through 168 hours (Day 8) compared to placebo by age group
Table 14.2.1.5	FAS	Subgroup Analysis #2: Summary of ANOVA difference in AUC for the NRS-R of pain intensity from 30 minutes through 168 hours (Day 8) compared to placebo by sex
Table 14.2.1.6	FAS	Subgroup Analysis #3: Summary of ANOVA difference in AUC for the NRS-R of pain intensity from 30 minutes through 168 hours (Day 8) compared to placebo by race
Table 14.2.1.7	FAS	Subgroup Analysis #4: Summary of ANOVA difference in AUC for the NRS-R of pain intensity from 30 minutes through 168 hours (Day 8) compared to placebo by ethnicity
Table 14.2.1.8	FAS	Subgroup Analysis #5: Summary of ANOVA difference in AUC for the NRS-R of pain intensity from 30 minutes through 168 hours (Day 8) compared to placebo by site
Table 14.2.2.1	FAS	Summary of ANOVA difference in AUC for the NRS-R of pain intensity from 30 minutes through 336 hours (Day 15) compared to placebo
Table 14.2.2.2	FAS	Summary of ANOVA difference in AUC for the NRS-R of pain intensity from 30 minutes through 504 hours (Day 22) compared to placebo
Table 14.2.2.3	FAS	Summary of ANOVA difference in total (MME) post-surgical consumption of opioid medications from 30 minutes through 336 hours (Day 15) compared to placebo
Table 14.2.2.4	FAS	Summary of ANOVA difference in AUC for the NRS-R of pain intensity from 30 minutes through 336 hours (Day 15) compared to placebo and active comparator
Table 14.2.2.5	FAS	Summary of ANOVA difference in NRS-A pain intensity scores during the TUG Test at Day 15 compared to placebo
Table 14.2.3.1	FAS	Summary of ANOVA difference in NRS-A pain intensity scores during the TUG Test at Day 8 compared to placebo and active comparator
Table 14.2.3.2	FAS	Summary of ANOVA difference in NRS-A pain intensity scores during the TUG Test at Day 15 compared to active comparator
Table 14.2.3.3	FAS	Summary of ANOVA difference in NRS-A pain intensity scores during the TUG Test at Day 22 compared to placebo and active comparator
Table 14.2.3.4	FAS	Summary of ANOVA difference in AUC for the NRS-R of pain intensity from 30 minutes through 168 hours (Day 8) compared to active comparator
Table 14.2.3.5	FAS	Summary of ANOVA difference in AUC for the NRS-R of pain intensity from 30 minutes through 504 hours (Day 22) compared to active comparator
Table 14.2.3.6	FAS	Summary of ANOVA difference in total (MME) post-surgical consumption of opioid medications from 30 minutes through 168 hours (Day 8) compared to placebo and active comparator
Table 14.2.3.7	FAS	Summary of ANOVA difference in total (MME) post-surgical consumption of opioid medications from 30 minutes through 504 hours (Day 22) compared to placebo and active comparator

Table 14.2.3.8	FAS	Summary of ANOVA difference in total (MME) post-surgical consumption of opioid medications from 30 minutes to Day 8, Day 9 to Day 15, Day 16 to Day 22, and Day 23 to Day 30 compared to placebo
Table 14.2.3.9	FAS	Summary of Time (days) to discontinuation of opioid analgesia
Table 14.2.3.10	FAS	Summary of Time (days) to discontinuation of NSAID analgesia
Table 14.2.3.11	FAS	Summary of Time (days) to discontinuation of rescue analgesia
Table 14.2.3.12.1	FAS	Proportion of participants who are opioid free starting at 30 minutes, Day 2, Day 3, Day 4, Day 8, Day 15, Day 22, and Day 30 by visit
Table 14.2.3.12.2	FAS	Proportion of participants who remain opioid free from 30 minutes, Day 2, Day 3, Day 4 through Day 8, Day 15, Day 22, and Day 30 by visit
Table 14.2.3.13.1	FAS	Proportion of participants who are NSAID-free starting at 30 minutes, Day 2, Day 3, Day 4, Day 8, Day 15, Day 22, and Day 30 by visit
Table 14.2.3.13.2	FAS	Proportion of participants who remain NSAID-free from 30 minutes, Day 2, Day 3, Day 4 through Day 8, Day 15, Day 22, and Day 30 by visit
Table 14.2.3.14.1	FAS	Proportion of participants who are rescue-free starting at 30 minutes, Day 2, Day 3, Day 4, Day 8, Day 15, Day 22, and Day 30 by visit
Table 14.2.3.14.2	FAS	Proportion of participants who remain rescue-free from 30 minutes, Day 2, Day 3, Day 4 through Day 8, Day 15, Day 22, and Day 30 by visit
Table 14.2.3.15	FAS	Summary of ANOVA difference in AUC for the NRS-R of pain intensity from 168 hours (Day 8) through 336 hours (Day 15) compared to placebo and active comparator
Table 14.2.3.16	FAS	Summary of ANOVA difference in AUC for the NRS-R of pain intensity from 336 hours (Day 15) through 504 hours (Day 22) compared to placebo and active comparator
Table 14.2.3.17	FAS	Summary of ANOVA difference in AUC for the NRS-R of pain intensity from 30 minutes to 696 hours (Day 30) compared to placebo and active comparator
Table 14.2.3.18	FAS	Summary of ANOVA difference in time-weighted sum of pain intensity from 30 minutes to Day 8 (time interval in the primary endpoint), Day 9 to Day 15, Day 16 to Day 22, and Day 23 to Day 30 compared to placebo
Table 14.2.3.19	FAS	Summary of ANOVA difference in total (mg) post-surgical consumption of ibuprofen from 30 minutes through 696 hours (Day 30) compared to placebo and active comparator
Table 14.2.3.20	FAS	Summary of ANOVA difference in AUC for the NRS of Worst Pain in the last 24 hours from 30 minutes through Day 8, Day 15, Day 22, and Day 30 compared to placebo and active comparator
Table 14.2.3.21	FAS	Summary of ANOVA difference in AUC of Sleep Quality Scale from 30 minutes through Day 8, Day 15, Day 22, and Day 30 compared to placebo and active comparator by visit
Table 14.2.3.22	FAS	Summary of ANOVA differences in Knee Injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR) compared to placebo and active comparator by visit
Table 14.2.3.23	FAS	Summary of ANOVA differences in Knee Society Score (KSS) compared to placebo and active comparator by visit and component type

Table 14.2.3.24	FAS	Proportion of participants reporting being pain free on the NRS-R of pain intensity by visit
Table 14.2.3.25	FAS	Summary of ANOVA difference in NRS-A pain intensity scores during the TUG Test compared to placebo and active comparator by visit
Table 14.2.3.26	FAS	Summary of Time (days) to 90 Degrees of Flexion
Table 14.2.3.27	FAS	Summary of Time (days) to -5 Degrees of Extension
Table 14.2.3.28	FAS	Summary of ANOVA difference in AUC for NRS-R of pain intensity from 2 hours through Day 8, Day 15, Day 22, and Day 30 compared to placebo and active comparator
Table 14.2.4.1	FAS	Summary of paired t-test differences in the duration of TUG Test compared to placebo and active comparator by visit

13.1.1.3. Safety Data

Table 3: Safety Data

Table Number	Population	Table Title / Summary
14.3.1 Displays of Adverse Events		
Table 14.3.1.1	Safety	Overall Summary of Treatment-Emergent Adverse Events
Table 14.3.1.2	Safety	Summary of Treatment-Emergent Adverse Events by SOC and PT
Table 14.3.1.3	Safety	Summary of Treatment-Emergent Adverse Events by SOC, PT, and Maximum Severity
Table 14.3.1.4	Safety	Summary of Treatment-Emergent Adverse Events Related to Investigational Product by SOC and PT
Table 14.3.1.5	Safety	Summary of Opioid-Related Treatment-Emergent Adverse Events by SOC and PT
Table 14.3.1.6	Safety	Summary of Treatment-Emergent Adverse Events Related to TKA by SOC and PT
14.3.2 Summary of Deaths, Other Serious and Significant Adverse Events		
Table 14.3.2.1	Safety	Summary of Treatment-Emergent Serious Adverse Events by SOC and PT
Table 14.3.2.2	Safety	Summary of Treatment-Emergent AESIs by SOC and PT
Table 14.3.2.3	Safety	Summary of Treatment-Emergent Adverse Events Leading to Study Discontinuation by SOC and PT
Table 14.3.2.4	Safety	Summary of Treatment-Emergent Adverse Events Leading to Death by SOC and PT
14.3.3 Narratives of Deaths, Other Serious and Certain Other Significant Adverse Events		
14.3.4 Abnormal Laboratory Values		
14.3.5 Laboratory Data		
Table 14.3.5.1.1	Safety	Summary of Hematology Results by visit
Table 14.3.5.1.2	Safety	Shift from Baseline in Hematology Results by visit
Table 14.3.5.2.1	Safety	Summary of Chemistry Results by visit
Table 14.3.5.2.2	Safety	Shift from Baseline in Chemistry Results by visit
Table 14.3.5.3	Safety	Summary of Liver Function Test Elevations
14.3.6 Other Safety Data		
Table 14.3.6.1	Safety	Summary of Vital Signs by visit
Table 14.3.6.2.1	Safety	Summary of ECG Measurements at Baseline
Table 14.3.6.2.2	Safety	Summary of ECG Overall Investigator Interpretation at Baseline
Table 14.3.6.3	Safety	Summary of Wound Healing Assessment by visit
Table 14.3.6.4	Safety	Summary of Investigational Product Administration for ATX101 Participants

13.1.2. Planned Listing Descriptions

The following are planned data listings for protocol number ATX-101-TKA-004.

In general, one listing will be produced per CRF domain. All listings will be sorted by treatment, site, and patient number. All calculated variables will be included in the listings.

In all listings, a blank line will be placed between each patient. Within a data listing, if an item appears line after line (eg, repetition of patient number), then only the first occurrence will be displayed.

In data listings, the information for one patient will be kept on one page, if at all possible, rather than splitting a patient's information across pages.

Table 4: Planned Listings

Listing Number	Population	Data Listing Title / Summary
16.2.1 Patient Discontinuations/Completions		
Listing 16.2.1.1	All Participants	Study Disposition
16.2.2 Protocol Deviations		
Listing 16.2.2.1	All Participants	Inclusion/Exclusion Criteria Violations
Listing 16.2.2.2	FAS	Important Protocol Deviations
16.2.3 Patients Excluded from the Efficacy Analyses		
Listing 16.2.3	All Participants	Analysis Populations and Reasons for Exclusion
16.2.4 Demographic Data and Other Baseline Characteristics		
Listing 16.2.4.1	FAS	Demographics
Listing 16.2.4.2	FAS	Baseline Characteristics
Listing 16.2.4.3	Safety	Medical History
Listing 16.2.4.4	Safety	Prior and Concomitant Medications
Listing 16.2.4.5	Safety	Total Knee Arthroplasty (TKA) Surgery Information
16.2.5 Compliance and/or Drug Concentration Data		
16.2.6 Individual Efficacy Response Data		
Listing 16.2.6.1	FAS	Numerical Rating Scale at Rest (NRS-R) with Location of Pain
Listing 16.2.6.2	FAS	Timed Up and Go (TUG) Test and Numerical Rating Scale with Activity (NRS-A) with Location of Pain
Listing 16.2.6.3	FAS	Range of Motion (ROM)
Listing 16.2.6.4	FAS	Rescue Medication Consumption
Listing 16.2.6.5	FAS	Numeric Rating Scale of Worst Pain in the Last 24 Hours
Listing 16.2.6.6	FAS	Physical Therapy
Listing 16.2.6.7	FAS	Sleep Quality Scale

Listing 16.2.6.8	FAS	Knee Society Score (KSS)
Listing 16.2.6.9	FAS	Knee Injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR) Score
16.2.7 Adverse Events		
Listing 16.2.7.1	Safety	Adverse Events
Listing 16.2.7.2	Safety	Serious Adverse Events
Listing 16.2.7.3	Safety	Adverse Events of Special Interest and LAST-Related Symptoms
Listing 16.2.7.4	Safety	Adverse Events Leading to Study Discontinuation
Listing 16.2.7.5	Safety	Adverse Events Leading to Death
16.2.8 Laboratory Values		
Listing 16.2.8.1	Safety	Hematology Laboratory Results
Listing 16.2.8.2	Safety	Chemistry Laboratory Results
Listing 16.2.8.3	Safety	Drug Screen Results
Listing 16.2.8.4	Safety	Pregnancy Test Results
16.2.9 Other Clinical Observations and Measurements		
Listing 16.2.9.1	Safety	Vital Signs
Listing 16.2.9.2	Safety	ECG Results
Listing 16.2.9.3	Safety	Physical Examination
Listing 16.2.9.4	Safety	Neurological Assessment
Listing 16.2.9.5	Safety	Wound Healing Assessment
16.2.10 Other Study Measurements or Assessments		
Listing 16.2.10.1	Safety	Pharmacokinetic Results

13.1.3. Planned Figure Descriptions

The following are planned summary figures for protocol number ATX-101-TKA-004. The figure numbers are placeholders only and will be determined when the figures are produced.

Table 5: Planned Figures

Figure Number	Population	Figure Title/Summary
Figure 14.2.1.1	FAS	Box Plot for the NRS-R of Pain Intensity over time by treatment group and visit – Imputed Data
Figure 14.2.1.2	FAS	Box Plot for the NRS-R of Pain Intensity over time by treatment group and visit – Observed Data
Figure 14.2.1.3	FAS	Mean NRS-R of Pain Intensity over time by treatment group – Imputed Data
Figure 14.2.1.4	FAS	Mean NRS-R of Pain Intensity over time by treatment group – Observed Data
Figure 14.2.2	FAS	Kaplan-Meier Plot of Time (days) to Discontinuation of Opioid Analgesia by treatment group
Figure 14.2.3	FAS	Kaplan-Meier Plot of Time to Discontinuation of NSAID by treatment group
Figure 14.2.4	FAS	Kaplan-Meier Plot of Time to Discontinuation of Rescue by treatment group
Figure 14.2.5	FAS	Kaplan-Meier Plot of Time (days) to 90 Degrees of Flexion by treatment group
Figure 14.2.6	FAS	Kaplan-Meier Plot of Time (days) to -5 Degrees of Extension by treatment group
Figure 14.2.7	FAS	Box Plot for the NRS of Worst Pain in the Last 24 Hours over time by treatment group and visit
Figure 14.2.8	FAS	Box Plot of Sleep Quality Scale over time by treatment group and visit
Figure 14.2.9	Safety	Peak on-treatment ALT/ULN vs BILI/ULN

13.2. Tables, Listings, and Listing Shells

13.2.1. Standard Layout for all Tables, Listings, and Figures

The following standard layout will be applied to all tables, listings, and figures in support of this study. Note that programming notes may be added if appropriate after each TLF shell.



Figure 1: Standardized Layout

Allay Therapeutics, Inc.	Page xx of xx
Protocol: ATX-101-TKA-004	
<Table, Listing, or Figure> xx.x.x <Title of Table, Listing, or Figure> <Study Population and (if applicable) Subgroup Description>	
Body of Table, Listing, or Figure	
<Note: If directly applicable> Footnote 1 <if applicable> Footnote 2 <if applicable> Footnote n <if applicable> Footnote n+1 <pgm path and name>, <date>	

13.3. Morphine Milligram Equivalent (MME)

Opioid Oral Morphine Milligram Equivalent (MME) Conversion Factors^{i,ii}

Type of Opioid (strength units)	MME Conversion Factor
Buprenorphine film/tablet ⁱⁱⁱ (mg)	
Buprenorphine patch ⁱⁱⁱ (mcg/hr)	
Buprenorphine film ⁱⁱⁱ (mcg)	
Butorphanol (mg)	7
Codeine (mg)	0.15
Dihydrocodeine (mg)	0.25
Fentanyl buccal or SL tablets, or lozenge/troche ^{iv} (mcg)	0.13
Fentanyl film or oral spray ^v (mcg)	0.18
Fentanyl nasal spray ^{vi} (mcg)	0.16
Fentanyl patch ^{vii} (mcg)	7.2
Hydrocodone (mg)	1
Hydromorphone (mg)	4
Levorphanol tartate (mg)	11
Meperidine hydrochloride (mg)	0.1
Methadone ^{viii} (mg)	3
>0, <= 20	4
>20, <= 40	8
>40, <= 60	10
>60	12
Morphine (mg)	1
Opium (mg)	1
Oxycodone (mg)	1.5
Oxymorphone (mg)	3
Pentazocine (mg)	0.37
Tapentadol ^{ix} (mg)	0.4

Tramadol (mg)	0.1
---------------	-----

ⁱThe MME conversion factor is intended only for analytic purposes where prescription data are used to calculate daily MME. Use the formula: Strength per Unit X (Number of Units/ Days Supply) X MME conversion factor = MME/Day. This value does not constitute clinical guidance or recommendations for converting patients from one form of opioid analgesic to another. Please consult the manufacturer's full prescribing information for such guidance. Use of this file for the purposes of any clinical decision-making warrants caution. This is particularly true with regard to methadone (see viii below).

ⁱⁱNational Center for Injury Prevention and Control. CDC compilation of benzodiazepines, muscle relaxants, stimulants, zolpidem, and opioid analgesics with oral morphine milligram equivalent conversion factors, 2017 version. Atlanta, GA: Centers for Disease Control and Prevention; Available at <https://www.cdc.gov/drugoverdose/resources/data.html>. For more information, send an email to Mbohm@cdc.gov.

ⁱⁱⁱBuprenorphine products are listed but do not have an associated MME conversion factor. These buprenorphine products, as partial opioid agonists, are not expected to be associated with overdose risk in the same dose-dependent manner as doses for full agonist opioids. The conversion factors for drugs prescribed or provided as part of medication-assisted treatment for opioid use disorder should not be used to benchmark against dosage thresholds meant for opioids prescribed for pain.

^{iv}The MME conversion factor for fentanyl buccal tablets, sublingual tablets, and lozenges/troche is 0.13. This conversion factor should be multiplied by the number of micrograms in a given tablet or lozenge/troche.

^vThe MME conversion factor for fentanyl film and oral spray is 0.18. This reflects a 40% greater bioavailability for films compared to lozenges/tablets and 38% greater bioavailability for oral sprays compared to lozenges/tablets.

^{vi}The MME conversion factor for fentanyl nasal spray is 0.16, which reflects a 20% greater bioavailability for sprays compared to lozenges/tablets.

^{vii}The MME conversion factor for fentanyl patches is based on the assumption that one milligram of parenteral fentanyl is equivalent to 100 milligrams of oral morphine and that one patch delivers the dispensed micrograms per hour over a 24 hour day. Example: 25 ug/hr fentanyl patch X 24 hrs = 600 ug/day fentanyl = 60 mg/day oral morphine milligram equivalent. In other words, the conversion factor not accounting for days of use would be 60/25 or 2.4. However, since the fentanyl patch remains in place for 3 days, we have multiplied the conversion factor by 3 (2.4 X 3 = 7.2). In this example, MME/day for ten 25 ug/hr fentanyl patches dispensed for use over 30 days would work out as follows: Example: 25 ug/hr fentanyl patch X (10 patches/30 days) X 7.2 = 60 MME/day. Please note that because this allowance has been made based on the typical dosage of one fentanyl patch per 3 days, you should first change all Days Supply in your prescription data to follow this standard, i.e., Days Supply for fentanyl patches= # of patches X 3.

^{viii}The CDC MME conversion factor to calculate morphine milligram equivalents of methadone is 3. Calculating MME for methadone in clinical practice often involves a sliding-scale approach whereby the conversion factor increases with increasing dose since the conversion factor of 3 for methadone could underestimate MME for a given patient. CMS uses this conversion factor when analyzing Medicare population opioid use. CMS uses the graduated methadone MME conversion factors to calculate MME within the Overutilization Monitoring System (OMS) for identifying and reporting potential opioid overutilizers.
https://www.cdc.gov/drugoverdose/pdf/calculating_total_daily_dose-a.pdf.