STATISTICAL ANALYSIS PLAN

Protocol Number:	ABSK021-301
Protocol Title:	A Phase 3, Randomized, Double-blind, Placebo-Controlled, Multicenter Study of ABSK021 to Assess the Efficacy and Safety in Patients with Tenosynovial Giant Cell Tumor
Protocol Version and Effective Date:	V2.2 2024-04-06
Investigational Product:	ABSK021
Study Phase:	Phase 3
SAP Version:	V1.0
SAP Effective Date:	2024-10-30

SAP SIGNATURE PAGE

[redacted]

Biostatistician, SAP Author Date (yyyy-mm-dd)

[redacted]

[redacted]

Primary Medical Reviewer Date (yyyy-mm-dd))

[redacted]

[redacted]

Clinical Pharmacologist Reviewer Date (yyyy-mm-dd)

[redacted]

[redacted]

Head of Biometrics or Designee Date (yyyy-mm-dd)

[redacted]

Table of Content

LIST OF ABBREVIATIONS

Abbreviation	Term
ADI	Actual Dose Intensity
AE	Adverse Event
AECI	Adverse Events of Clinical Interest
ALT	Alanine Aminotransferases
ANC	Absolute Neutrophil Count
AST	Aspartate Aminotransferases
ATC	Anatomical Therapeutic Classification
BIRC	Blinded Independent Review Committee
BOR	Best Overall Response
BPI	Brief Pain Inventory
BQL	Below Quantification Limit
CI	Confidence Interval
CK	Creatine Phosphokinase
COA	Clinical Outcome Assessments
CR	Complete Response
CRF	Case Report Form
CSR	Clinical Study Report
CTCAE	Common Terminology Criteria for Adverse Events
DBP	Diastolic Blood Pressure
DOR	Duration of Response
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EOT	End of Treatment
EQ-5D-5L	EuroQol 5-Dimension, 5-Level Questionnaire
ITT	Intent to Treatment
J2R	Jump to Reference
LSMEAN	Least Squares Mean
MAR	Missing at Random
MCAR	Missing Completely at Random
MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed Models for Repeated Measurements
MNAR	Missing not at Random
NRS	Numeric Rating Scale
ORR	Objective Response Rate
PDI	Planned Dose Intensity

PDs	Protocol Deviations
PID	Percentage Intended Dose
PKS	Pharmacokinetic Analysis Set
PMM	Pattern Mixture Models
PN	Preferred Name
PPS	Per Protocol Set
PR	Partial Response
PRO	Patient-Reported Outcome
PROMIS-PF	Patient-Reported Outcomes Measurement Information System – Physical Function
PT	Preferred Term
QTc	Corrected QT Interval
RDI	Relative Dose Intensity
RECIST	Response Evaluation Criteria in Solid Tumors
ROM	Range of Motion
SAE	Serious Adverse Events
SAP	Statistical Analysis Plan
SBP	Systolic Blood Pressure
SD	Stable Disease
SOC	System Organ Class
SST	Safety Analysis Set
TEAE	Treatment-emergent Adverse Event
TFL	Tables, Figures and Listings
TGCT	Tenosynovial Giant Cell Tumor
TPR	Time Point Response
TVS	Tumor Volume Score
ULN	Upper Limit of the Normal Range
VAS	Visual Analogue Scale
WBC	White Blood Cells
WHODD	World Health Organization Drug Dictionary

AMENDMENT HISTORY

Version	Date [yyyy-mm-dd]	Brief description of changes
1.0	2024-10-29	Abbisko Initialization

1. INTRODUCTION

This Statistical Analysis Plan (SAP) provides detailed statistical methods and data handling approaches for Study ABSK021-301 based on Protocol Version 2.2. It is prepared for the global clinical study report (CSR) and will be finalized before the clinical database locks at the end of study Part 1. Additional analyses required for regional submissions after the initial SAP approval will be prespecified in either SAP amendments or addendums and will be submitted to the appropriate authorities.

Any deviations from this analysis plan will be substantiated by statistical rationale and will be documented in the final clinical study report (CSR).

1.1. Study Objectives, Endpoints and Estimands

Primary Objective(s)

Objective(s)	Endpoint(s)
• To compare the Objective Response Rate (ORR) by Blinded Independent Review Committee (BIRC) within 25 weeks after treatment with ABSK021 or placebo in TGCT patients based on RECIST v1.1.	• 25-Week ORR by BIRC based on RECIST v1.1 (BIRC-25-Week ORR)

Secondary Objective(s)

Key Secondary Efficacy	
Objectives	**Endpoints**
• To compare the Objective Response Rate (ORR) by BIRC within 25 weeks after treatment with ABSK021 or placebo in TGCT patients based on Tumor Volume Score (TVS); • To compare the effects of ABSK021 and placebo on the Range of Motion (ROM) in TGCT patients at Week 25; • To compare the effects of ABSK021 and placebo on Patient-Reported Outcome	• 25-Week ORR by BIRC based on TVS (BIRC-25-Week ORR-TVS); • Mean change from baseline in Range of Motion (presented as relative ROM) of the affected joint at Week 25, according to the reference criteria for the same joint; • Mean change from baseline in the Worst Stiffness Numeric Rating Scale (NRS) score at Week 25; • Mean change from baseline in the BPI Worst Pain Numeric Rating Scale (NRS) score at Week 25; • Mean change from baseline in the Patient-reported Outcomes Measurement Information System (PROMIS) Physical Functioning score at Week 25.

(PRO) in TGCT patients at Week 25.	
Other Secondary Efficacy	
• To compare the Objective Response Rate (ORR) by investigator within 25 weeks after treatment with ABSK021 or placebo in TGCT patients based on RECIST v1.1; • To compare the Duration of Response (DOR) by BIRC after treatment with ABSK021 or placebo in TGCT patients based on RECIST v1.1 and TVS, respectively; • To compare the DOR by investigator within 25 weeks after treatment with ABSK021 or placebo in TGCT patients based on RECIST v1.1; • To compare the effects of ABSK021 and placebo on Patient-Reported Outcome (PRO) for EQ-5D-5L in TGCT patients at Week 25.	• 25-week ORR by Investigator per RECIST v1.1: proportion of patients with a BOR of CR or PR assessed by the Investigator within 25 weeks according to RECIST v1.1; • DOR by BIRC based on RECIST v1.1: the time (months) from the first documentation of objective response (CR or PR as assessed by BIRC per RECIST v1.1 criteria) to the first documentation of radiographic disease progression (PD) or death due to any cause, whichever occurs first; • DOR by BIRC based on TVS: the time (months) from the first documentation of objective response (CR or PR as assessed by BIRC per TVS) to the first documentation of radiographic disease progression (PD) or death due to any cause, whichever occurs first; • DOR by Investigator per RECIST v1.1: the time (months) from the first documentation of objective response (CR or PR as assessed by the Investigator per RECIST v1.1 criteria) to the first documentation of radiographic PD or death due to any cause, whichever occurs first; • Mean change from baseline in EuroQol 5-dimension, 5-level questionnaire (EQ-5D-5L) visual analogue scale (VAS) score at Week 25.
Safety	
• To compare the safety of ABSK021 and placebo in TGCT patients;	• Safety endpoints, including but not limited to treatment-emergent adverse events (TEAEs), dose modifications, laboratory tests, vital signs, electrocardiograms (ECGs), echocardiography, and related items in NCI-PRO-CTCAE.
Pharmacokinetic	
• To evaluate the pharmacokinetic (PK) profile of oral ABSK021.	• PK profile of ABSK021.

Primary Estimand

Primary Estimand: The BIRC-25-Week ORR based on RECIST v1.1 will be compared between ABSK021 and matching placebo using the mean difference (ABSK021 – Placebo) as

a population-level summary in patients with TGCT who meet the eligibility criteria, regardless of (1) dose modifications due to toxicity or (2) early discontinuation of treatment not due to PD or (3) before receiving an alternative anti-tumor therapy (including surgery).	
Estimand Attributes	
Treatment	ABSK021 50 mg QD vs. matching placebo QD for 24 weeks
Population	Eligible TGCT patients
Variable/Endpoint	The BIRC-25-Week ORR is defined as the proportion of patients achieving a best overall response of CR or PR within 25 Weeks based on RECIST v1.1 as assessed by BIRC in Part 1.
Intercurrent events	**Handling Strategy**
Dose modifications (including dose interruption and dose reduction) due to toxicity	Treatment policy strategy: continue to collect and use the data even if the patient experiences the intercurrent event
Early discontinuation of treatment not due to PD	Treatment policy strategy: continue to collect and use the data even if the patient experiences the intercurrent event
Receiving an alternative anti-tumor therapy (including surgery)	While on treatment policy strategy: outcome after the event is considered irrelevant to the treatment effect. Data before the intercurrent event will be used.
Population-level summary	The mean difference (ABSK021 – Placebo) in BIRC-25-Week ORR based on RECIST v1.1.

Key Secondary Estimands

Tumor Volume Score (TVS)

Secondary Estimand 1: The BIRC-25-Week ORR based on TVS will be compared between ABSK021 and matching placebo using the mean difference (ABSK021 – Placebo) as a population-level summary in patients with TGCT who meet the eligibility criteria, regardless of (1) dose modifications due to toxicity or (2) early discontinuation of treatment not due to PD or (3) before receiving an alternative anti-tumor therapy (including surgery).

Estimand Attributes	
Treatment	The same as the primary estimand
Population	The same as the primary estimand
Variable/Endpoint	The BIRC-25-Week ORR is defined as the proportion of patients achieving a best overall response of CR or PR within 25 Weeks based on TVS as assessed by BIRC in Part 1.
Intercurrent events	**Handling Strategy**
The same as the primary estimand.	
Population-level summary	The mean difference (ABSK021 – Placebo) in BIRC-25-Week ORR based on TVS.

Range of Motion (ROM)

Secondary Estimand 2: The change from baseline in the relative ROM of the affected joint at Week 25 will be compared between ABSK021 and matching placebo using the mean difference (ABSK021 – Placebo) of change from baseline at Week 25 as a population-level summary in patients with TGCT who meet the eligibility criteria, regardless of (1) dose modifications due to toxicity or (2) early discontinuation of treatment or (3) before receiving an alternative anti-tumor therapy (including surgery).	
Estimand Attributes	
Treatment	The same as the primary estimand.
Population	The same as the primary estimand.
Variable/Endpoint	For each affected joint, the endpoint of change from baseline in relative ROM at Week 25 will be derived based on the last non-missing assessment average prior to the first Part 1 dose as the baseline.
Intercurrent events	**Handling Strategy**

Dose modifications (including dose interruption and dose reduction) due to toxicity	Treatment policy strategy: continue to collect and use the data even if the patient experiences the intercurrent event
Early discontinuation of treatment 1) not due to PD; 2) due to PD	Treatment policy strategy: continue to collect and use the data even if the patient experiences the intercurrent event
Receiving an alternative anti-tumor therapy (including surgery)	Treatment policy strategy: continue to collect and use the data even if the patient experiences the intercurrent event
Population-level summary	The Least Squares Means difference (ABSK021 – Placebo) of change from baseline in relative ROM at Week 25

Worst Stiffness Numeric Rating Scale

Secondary Estimand 3: The change from baseline in Worst Stiffness NRS at Week 25 will be compared between ABSK021 and matching placebo using the mean difference (ABSK021 – Placebo) of change from baseline at Week 25 as a population-level summary in patients with TGCT who meet the eligibility criteria, regardless of (1) dose modifications due to toxicity or (2) early discontinuation of treatment or (3) before receiving an alternative anti-tumor therapy (including surgery).	
Estimand Attributes	
Treatment	The same as the primary estimand.
Population	The same as the primary estimand.
Variable/Endpoint	The change from baseline in Worst Stiffness NRS at Week 25 will be derived based on the average of evaluations completed prior to randomization and evaluation on C1D1 as the baseline
Intercurrent events	**Handling Strategy**
The same as the estimand for ROM.	
Population-level summary	The Least Squares Means difference (ABSK021 – Placebo) of change from baseline in Worst Stiffness NRS at Week 25

Brief Pain Inventory (BPI) Worst Pain Numeric Rating Scale (NRS)

Secondary Estimand 4: The change from baseline in BPI Worst Pain NRS at Week 25 will be compared between ABSK021 and matching placebo using the mean difference (ABSK021 – Placebo) of change from baseline at Week 25 as a population-level summary in patients with TGCT who meet the eligibility criteria, regardless of (1) dose modifications due to toxicity or (2) early discontinuation of treatment or (3) before receiving an alternative anti-tumor therapy (including surgery) or (4) change in stable dose of background analgesic treatment.	
Estimand Attributes	
Treatment	The same as the primary estimand.
Population	The same as the primary estimand.
Variable/Endpoint	The change from baseline in BPI Worst Pain NRS at Week 25 will be derived based on the average of evaluations completed prior to randomization and evaluation on C1D1 as the baseline
Intercurrent events	**Handling Strategy**
The same as the estimand for ROM for intercurrent events 1-3; Additionally, if patients had change in stable dose of background analgesic treatment, the NSR will be collected up to Week 25 and included in the analysis.	
Population-level summary	The Least Squares Means difference (ABSK021 – Placebo) of mean change from baseline in BPI Worst Pain NRS at Week 25

PROMIS Physical Function (PROMIS-PF) Scale

Secondary Estimand 5: The change from baseline in PROMIS-PF T-score at Week 25 will be compared between ABSK021 and matching placebo using the mean difference (ABSK021 – Placebo) of change from baseline at Week 25 as a population-level summary in patients with TGCT who meet the eligibility criteria, regardless of (1) dose modifications due to toxicity or (2) early discontinuation of treatment or (3) before receiving an alternative anti-tumor therapy (including surgery).	
Estimand Attributes	

Treatment	The same as the primary estimand.
Population	The same as the primary estimand.
Variable/Endpoint	The change from baseline in PROMIS-PF T-score at Week 25 will be derived based on average of evaluation prior to randomization and evaluation on C1D1 as the baseline.
Intercurrent events	**Handling Strategy**
The same as the estimand for ROM.	
Population-level summary	The Least Squares Means difference (ABSK021 – Placebo) of change from baseline in PROMIS-PF T-score at Week 25

1.2. Study Design

Overall design:

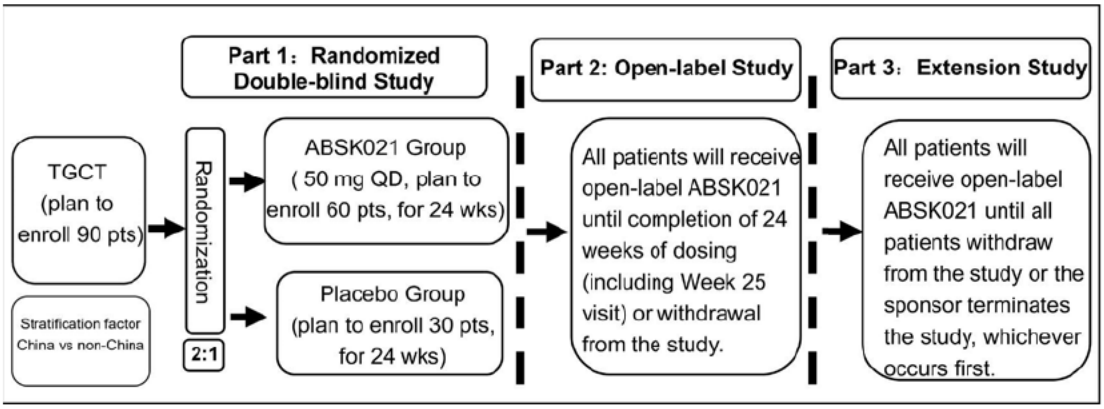
This is a randomized, double-blind, placebo-controlled Phase 3 clinical study to evaluate the efficacy and safety of ABSK021 at the dose of 50 mg QD in patients with TGCT. This study consists of Part 1, Part 2 and a long-term extension treatment phase (i.e., Part 3). Part 1 is a double-blind phase, eligible patients will be randomized in a 2:1 ratio to ABSK021 treatment group or matching placebo group and will receive 50 mg QD of ABSK021 or matching placebo (28 days/cycle) until completion of treatment and follow-up in Part 1 (i.e., completion of ABSK021 administration at Week 24 and completion of Week 25 follow-up visit including MRI) or withdraw from the study. Randomization will be stratified by China and non-China sites. All patients who complete Part 1 treatment and meet eligibility criteria will be eligible to continue in Part 2 of the study.

Part 2 is an open-label treatment phase, and all patients entering this phase will receive 50 mg QD of open-label ABSK021 until completion of 24 weeks of dosing or withdrawal from the study (If a patient has dose modification in Part 1, the patient will continue to be administered at the modified dose in Part 2). Patients must continue to meet eligibility criteria to continue on to Part 2. After the patient has achieved a sustained tumor response (defined as continuous radiological PR or CR for more than 6 months by BIRC or Investigator based on RECIST v1.1), the Investigator and the sponsor will discuss and decide whether the current recommended starting dose can be reduced to 25 mg QD during the subsequent maintenance treatment. If the Investigator assesses that the patient will not benefit from subsequent treatment, the patient may be discontinued from the study treatment. Part 2 will end after all patients have completed 24 weeks of dosing (including follow-up at Week 25) or have withdrawn from the study.

All patients who complete 24 weeks of dosing in Part 2 will be eligible to enter the open-label extension treatment phase (i.e., Part 3) for a longer period treatment and safety follow-up. Part 3

will end when all patients withdraw from the study, or the sponsor decides to terminate the study, whichever occurs first. The study design is specified in the following figure:

Figure 1.1 Study Schema



An independent Data Monitoring Committee (IDMC) will be established in Part 1 and Part 2 of this study to continuously monitor the safety profile and oversee the overall conduct of the study. In order to improve the integrity of the study, the IDMC may also make recommendations on patient selection, recruitment, and management in order to improve patient compliance and improve data management and quality control procedures. A separate IDMC charter will define IDMC members, roles, and responsibilities, as well as the process of providing advice to the sponsor.

1.3. Sample Size Consideration

The sample size of this study is calculated based on the primary objective of the study, which is to compare the objective response rate within 25 weeks after treatment with ABSK021 or placebo in TGCT patients.

The target ORR of 40% for ABSK021-301 was anticipated from this study (i.e., 85% of the lower limit of 95% CI for interim ORR in ABSK021-101) based on the following considerations: a) Limited sample size in ABSK021-101 interim analysis; b) The inclusion of TGCT patients in the Phase 3 study will be more stringent than ABSK021-101 study which could lead to the enrolment of patients with a more severe condition. This may have a detrimental influence on treatment effectiveness; c) A single region preliminary ORR results from ABSK021-101 may not completely reflect the true response from the multi-region trial in a placebo control study.

Patients will be randomized in a 2:1 ratio to ABSK021 treatment group or placebo group. Assuming a 25-week ORR of 6% (P_0) (the upper 95% CI of the placebo group in the ENLIVEN[1] study) in the placebo group and 40% (P_1) for ABSK021, by using Fisher's exact test with a two-sided significance level α of 0.05, a sample size of 75 evaluable patients provides a 90% power to detect the difference in the ORR between the two groups (H0: $P_1 = P_0$ vs. Ha $P_1 \neq P_0$). The power is calculated using the binomial distribution enumeration method (PASS 2021，v21.0.3). Considering a dropout rate of about 15%, approximately 90 TGCT patients need to be enrolled, including 60 patients in the ABSK021 treatment group and 30 patients in the placebo group.

1.4. Randomization and Blinding

In study Part 1, patients are randomized in a 2:1 ratio using a stratified randomization method to ABSK021 treatment group or matching Placebo treatment group. Randomization will be performed across all sites using a central interactive web response system (IWRS) and will be stratified by China sites and non-China sites. Treatment assignment will remain unknown to the patients, investigators, study site personnel, safety laboratory personnel, central imaging readers, reviewers, and the sponsor. The appearance of ABSK021 and placebo capsules will be identical.

2. PLANNED ANALYSIS

No formal interim analysis is planned for this study. Data analyses will be performed according to the following analysis plan to support the study objectives.

2.1. First Planned Analysis – Primary Analysis

The first planned analysis is the primary analysis of this study, which will be performed after all randomized patients have either completed 24 weeks of treatment in Part 1 and a follow-up MRI scan at Week 25 visit or have withdrawn from informed consent or died. Specifically, the date of data cutoff (DCO) for the primary analysis will be the date of the last patient's date of Week 25 visit (C7D1).

The effect of ABSK021 on ORR and key secondary efficacy endpoints will be primarily evaluated in this analysis. All data from Part 1 to Part 3 up to the data cut-off date for the primary analysis will be cleaned before database locks. Part 1 data will be locked, Part 2 and Part 3 data will be snapshot data. Treatment codes will be unblinded for the analysis, but this unblinding will only apply to the sponsor and contract research organization (CRO), who will be involved in data analysis (i.e., the study is still single-blinded).

Efficacy analyses will be primarily based on patients' data up to the Week 25 visit in Part 1.

Safety analyses will include all patients' data from Part 1 to Part 3 up to the DCO. In addition, patients, who don't enter the study Part 2, should complete safety follow-up visit and the data will be included in the safety analysis.

2.2. Second Planned Analysis – Final analysis

The second planned analysis (i.e., the final analysis) will be performed after the completion of study Part 2 to supplement a long-term efficacy, safety, and tolerability profile. All data from Part 1 to Part 3 before the data cut-off date will be cleaned and analyzed before the database locks for the analysis. The DCO will be the date of the last patient's Week 49 visit.

The treatment comparison of effectiveness between ABSK021 and placebo will be mainly based on data from the double-blinded treatment period, along with supportive analyses for efficacy durability based on data from Part 2 and Part 3 up to DCO

Safety analyses will include all patients' data from Part 1 to Part 3 up to the DCO. In addition, patients, who don't enter the study during Part 3, should complete safety follow-up visit and the data will be included in the safety analysis.

2.3. End of Study Analysis

The end of study analysis will be performed at the end of study Part 3. All collected data will be cleaned before the full database locks. Data analyses will be provided for supplemental purposes to evaluate the long-term efficacy and safety of ABSK021.

3. ANALYSIS POPULATIONS

The following analysis populations will be included in this study.

3.1. Screening Set

The Screening Set consists of all patients who signed the informed consent form. It will only be used for patient disposition summary and listing.

3.2. Intent-to-treat Analysis Set (ITT)

The **ITT** includes all randomized patients. The analyses for the ITT analysis set will be conducted according to the randomly assigned treatment regardless of whether the patient received study treatment or was compliant with the protocol. The ITT analysis set is the primary analysis set for the primary and key secondary efficacy analyses unless otherwise specified. Patients who were randomized but did not subsequently receive study treatment will be included in the analysis in the treatment group to which they were randomized.

3.3. Per Protocol Set (PPS)

The **PPS** is a subset of ITT and includes all patients who received at least one dose of study drug and had no major protocol deviations that may affect the efficacy assessment. During the blinded data review meeting, major protocol deviations will be identified and assessed as relevant for the primary efficacy endpoint before the database locks. Analyses of the PPS will be based on the randomly assigned treatment.

3.4. Safety Analysis Set (SST)

The **SST** includes all randomized patients who received at least 1 dose of study drug. Analyses using the SST will be conducted according to the actual treatment received. All safety analyses will be evaluated using the SST.

3.5. Pharmacokinetic Analysis Set (PKS)

The **PKS** includes all patients who received at least 1 dose of the study drug and had at least one quantifiable PK sample for PK Analysis.

4. GENERAL STATISTICAL CONSIDERATION

4.1. General Principles for Data Analysis

4.1.1. Multicenter Study

In this study the stratified block randomization is not done within centers. The analyses will be conducted by pooling data from all study centers and will not include study center as a covariate in the statistical modeling.

The study is randomized based on the stratification factor of region (China vs. ex-China). The region as randomized will be included in the statistical models as appropriate (see details in Section 6).

4.1.2. Examination of Subgroups

To determine whether the treatment effect is consistent across various subgroups, the selected key efficacy endpoints will be summarized descriptively for each of the following subgroups:

- Age Group (< 40 years, ≥ 40years, ≥ 40 - < 65 years, ≥ 65 years)
- Sex (Female vs. Male)
- Region Category 1 (China vs. ex-China)
- Region Category 2 (North American[a] vs. ex-North American)
- Region Category 3 (China vs. North American[a] vs. EU)
- Race Category 1 (Asian vs. non-Asian)
- Race Category 2 (Caucasian[b], non-Caucasian)
- Ethnicity (Hispanic or Latino vs. Not Hispanic or Latino)
- Lower extremity tumors vs. upper extremity tumors
- Knee vs Others
- TCGT type (localized-type vs. Diffuse-type vs. Unknown)
- ECOG PS (0 vs.1)
- Number of prior surgeries (0 vs. ⩾ 1)
- Prior systemic therapy (Yes vs. No)

a. North American includes United State and Canada
b. Caucasian includes White

The following demographic and baseline characteristic subgroups will be used for the analysis of select baseline and safety parameters within all the defined safety pools:

- Age Group (< 40 years, ≥ 40 - < 65 years, ≥ 65 years)
- Sex (Female vs. Male)
- BMI (<18.5kg/m^2, ≥ 18.5 - < 25 kg/m^2, ≥ 25 kg/m^2)

- Region Category 1 (China vs. ex-China)
- Region Category 2 (North American[a] vs. ex-North American)
- Region Category 3 (China vs. North American vs. EU)
- Race (White, Black or African American, Asian, Native Hawaiian or other Pacific Islander, American Indian or Alaska Native)
- Race Category 1 (Asian, non-Asian)
- ECOG PS (0 vs.≥1)
- Prior systemic therapy (Yes vs. No)
- Liver function[b] (Normal, Mild, Moderate, Severe)
- Renal function[c] (Normal, Mild, Moderate, Severe)

a. North American includes United State and Canada
b. The patients will be categorized based on National Cancer Institute Organ Dysfunction Working Group (NCI-ODWG) criteria
c. The patients will be categorized based on their creatinine clearance (CrCL) values (as estimated using the Cockcroft-Gault formula.

4.2. Data Handling Conventions and Definitions

4.2.1. Premature Withdrawal and Missing data

All available data from patients who were withdrawn from the study will be listed, and all available planned data will be included in summary tables and figures unless otherwise specified. Data from patients who withdraw after meeting the entry criteria but not treated will be only listed. Any data collected in the clinical database after a patient withdraws informed consent from all further participation in the trial should be removed from the database and not be included in the analysis data sets.

In general, missing data will not be imputed and only observed values will be analyzed unless otherwise specified. Any patient excluded from the summaries and/or statistical analyses will be documented along with the reason for exclusion in the clinical study report.

Partial missing dates for AE and concomitant medication will be imputed based on Appendix 1 of Section 11.1. Partial dates will be displayed as captured in patient listing displays.

4.2.2. Data Derivation

4.2.2.1. Age and BMI

Ages collected in the CRF page will be used in analyses. If only the year of birthday is collected, the birth date will be imputed as '30th June'. Age will be derived from the using the following formula date of informed consent-date of birth, rounded down to the nearest integer.

Age (year) = (date of informed consent — date of birth + 1)/ 365.25

BMI (kg/m^2) = ($weight/height^2$) where weight is in kg and height is in meter.

4.2.2.2. Category of Liver and Renal Function

Liver Function Test

Liver function test (LFT) abnormalities at baseline will be used in defining subgroups:

LFT Abnormalities[2] at Baseline:

- Normal: TBIL≤ULN, AST≤ULN
- Mild: TBIL ≤ULN, AST>ULN or TBIL>ULN, ≤1.5 ×ULN
- Moderate: TBIL>1.5 ×ULN, ≤3.0 ×ULN
- Severe: TBIL > 3.0×ULN

TBIL = Total bilirubin; AST = Aspartate aminotransferase; ULN = Upper limit of Normal

Renal Function Test

The estimated Glomerular Filtration Rate (eGFR) of creatinine clearance (CrCl) will be used to assess patients' renal function.

Baseline CrCl results will be categorized as follows used in defining subgroups:

- Normal: CrCl ≥ 90 mL/min
- Mild: 60 mL/min ≤ CrCl < 90 mL/min
- Moderate: 30 mL/min ≤ CrCl < 60 mL/min
- Severe: CrCl < 30 mL/min

CrCl will be estimated using the Cockcroft-Gault equation.

$$\text{CrCl (mL/min)} = \{([140\text{-age}] \times \text{weight}) / (\text{serum creatinine} \times 72)\} \times \text{sex}$$

where, age in years, weight in kg is the patient weight collected at the same visit or most recent prior to the collection of serum creatinine, serum creatinine in mg/dL, sex = 1 for male and 0.85 for female.

4.2.2.3. Baseline and Change from Baseline

For patients who were randomized to ABSK021 group at the beginning of trial, the last non-missing value or mean on or before the randomization will be used as a baseline for efficacy endpoints, while the last recorded value on or before the first Part 1 dose of study drug will serve as the baseline measurement for safety endpoints unless it is specified stated otherwise. If multiple planned assessments are collected on the same scheduled time, the average of these assessments will be used.

For patients who were given placebo treatment during Part 1 of the study and then switched to ABSK021 treatment in Part 2, their baseline values for both efficacy and safety parameters at Part 2 will be reassigned based on the non-missing values or means before they were given their first Part 2 dose. Baselines at Part 1 will follow the same rule described for patients who were randomized to ABSK021 group.

Absolute changes from baseline will be calculated as the post-baseline value minus the baseline value. Percentage changes from baseline will be calculated as the change from baseline divided by the baseline value, multiplied by 100. If either of values is missing, then change from baseline and percentage change from baseline will be set to missing.

4.2.3. Study Day and Duration

Where appropriate, the calculated study day or duration of each assessment or event will be presented with the assessment or event date on patient data listings, where study day will be determined as:

- If the assessment/event date < the reference start date, then
  Study Day = event start date (visit date, onset date, assessment date) - reference start date
- If the assessment/event date ⩾ the reference start date, then
  Study Day/Duration = event start date (visit date, onset date, assessment date) - reference start date + 1

Table 4.1 Reference start date for efficacy and safety data

	Study Part 1		**Study Part 2 and 3**	
	Efficacy	**Safety**	**Efficacy**	**Safety**
Patients randomized to ABSK021	Randomization date	First Part 1 dose	Randomization date	First Part 1 dose
Patient randomized to Placebo and switched to ABSK021	Randomization date	First Part 1 dose	First Part 2 dose	First Part 2 dose

The study day will be displayed in data listings along with the start dates of the event and reference.

The following unit conversion factor will be used to present duration in week, month or year.

- Weeks: A duration expressed in weeks will be calculated by dividing the duration in days by 7
- Months: A duration expressed in months will be calculated by dividing the duration in days by 30.4375
- Years: A duration expressed in years will be calculated by dividing the duration in days by 365.25

4.2.4. Definition of Analysis Windows

The window for the visits following baseline will be constructed in such a way that the upper limit of the interval falls halfway between the two visits (the lower limit of the first post-baseline

visit will be Day 2). If an even number of days exists between two consecutive visits then the upper limit will be taken as the midpoint value minus 1 day.

The following rules will be applied to assign actual patient visits to protocol-specified visits. For each protocol-specified study visit, a target study day will be identified to represent the corresponding visit along with a window around the target day. Windows will be selected in a non-overlapping fashion so that a collection date does not fall into multiple visit windows. If a patient has two or more actual visits in one visit window, the visit closest to the target day will be used for analysis. If two visits are equidistant from the target day, then the later visit will be used for analysis.

The visit window and the target study day from Part 1 for each protocol-specified visit are displayed in Table 4.2 and Table 4.3 (depending on the different visit schedules of different endpoints).

Table 4.2 Analysis Windows for PRO endpoints and Range of Motion

Study Period	Protocol Specified Visit	Target Day	Visit Window Range [min, max]		
			Worst Stiffness NRS, BPI Worst Pain NRS, PROMIS-PF and EQ-5D-5L	Range of Motion	NCI-PRO-CTCAE
Part 1	Baseline	1[a]	[-28, 1]	[-28, 1]	[-28, 1]
	Week 2 (C1D8)	8			[2,11]
	Week 3 (C1D15)	15			[12, 18]
	Week 4 (C1D22)	22			[19,25]
	Week 5 (C2D1)	29	[2, 43]		[26, 32]
	Week 6 (C2D8)	36			[33,39]
	Week 7 (C2D15)	43			[40, 46]
	Week 8 (C2D22)	50			[47, 53]

Study Period	Protocol Specified Visit	Target Day	Visit Window Range [min, max]		
			Worst Stiffness NRS, BPI Worst Pain NRS, PROMIS-PF and EQ-5D-5L	Range of Motion	NCI-PRO-CTCAE
	Week 9 (C3D1)	57	[44, 71]		[54, 71]
	Week 13 (C4D1)	85	[72, 99]	[2, 127]	[72, 99]
	Week 17 (C5D1)	113	[100, 127]		[100, 127]
	Week 21 (C6D1)	141	[128, 155]		[128, 155]
	Week 25 (C7D1)	169	[156, 1st dose date of Part 2 or 184]	[128, 1st dose date of Part 2 or 184]	[156, 1st dose date of Part 2 or 184]
Part 2	Week 25 (C1D1)[b]	169	[156, 1st dose date of Part2]		
	Week 29 (C2D1)	197	[1 day after 1st dose date of Part 2, 211]		
	Week 33 (C3D1)	225	[212, 239]		
	Week 37 (C4D1)	253	[240, 295]	[1 day after 1st dose date of Part 2, 295]	
	Week 49 (C7D1)	337	[296, 1st dose date of Part 3 or 352]	[296, 1st dose date of Part 3 or 352]	
Part 3	Week 61 (C4D1)	421	[1st dose date of Part 3 +1, 463]	[1st dose date of Part 3 +1, 463]	
	Week 73	505	[464, 548]	[464, 548]	

Study Period	Protocol Specified Visit	Target Day	Visit Window Range [min, max]		
			Worst Stiffness NRS, BPI Worst Pain NRS, PROMIS-PF and EQ-5D-5L	Range of Motion	NCI-PRO-CTCAE
	(C7D1)				
	Week 97 (C13D1)	673	[549, 757]	[549, 757]	
	Week 121	841	[758, 925]	[758, 925]	
	+ 24 weeks				

a. Day of first dose of study drug.
b. Part 2 baseline for placebo arm in Part 1.

Table 4.3 Analysis Windows for Safety Endpoints

Study Period	Protocol Specified Visit	Target Day	Visit Window Range [min, max]			
			Pregnancy test	Hematology / Urinalysis / Physical examination / ECOG	Blood chemistry / Myocardial Enzyme Test / Vital Signs / 12-lead ECG	Coagulation / Thyroid function tests / Echocardiography/MUGA
Part 1	Baseline	1a	[-28, 1]	[-28, 1]	[-28, 1]	[-28, 1]
	Week1 (C1D1)				1, 3h after 1st dose for ECG	
	Week 3 (C1D15)	15		[2, 22]	[2, 22]	
	Week 5 (C2D1)	29	[2, 43]	[23, 43]	[23, 36]	
	Week 7 (C2D15)	43			[37, 50]	
	Week 9 (C3D1)	57	[44, 71]	[44, 71]	[51, 71]	
	Week 13 (C4D1)	85	[72, 99]	[72, 99]	[72, 99]	[2, 127]
	Week 17 (C5D1)	113	[100, 127]	[100, 127]	[100, 127]	
	Week 21 (C6D1)	141	[128, 155]	[128, 155]	[128, 155]	
	Week 25 (C7D1)	169	[156, 1st dose date of Part 2 or 184]	[156, 1st dose date of Part 2 or 184]	[156, 1st dose date of Part 2 or 184]	[128, 1st dose date of Part 2 or 184]

Study Period	Protocol Specified Visit	Target Day	Visit Window Range [min, max]			
			Pregnancy test	Hematology / Urinalysis / Physical examination / ECOG	Blood chemistry / Myocardial Enzyme Test / Vital Signs / 12-lead ECG	Coagulation / Thyroid function tests / Echocardiography/MUGA
Part 2	Week 25 (C1D1)[b]	169	[156, 1st dose date of Part2]	[156, 1st dose date of Part2]	[156, 1st dose date of Part2]	
	Week 29 (C2D1)	197		[1 day after 1st dose date of Part 2, 211]	[1 day after 1st dose date of Part 2, 211]	
	Week 33 (C3D1)	225		[212, 239]	[212, 239]	
	Week 37 (C4D1)	253	[1 day after 1st dose date of Part 2, 295]	[240, 295]	[240, 295]	[1 day after 1st dose date of Part 2, 295]
	Week 49 (C7D1)	337	[296, 1st dose date of Part 3 or 352]	[296, 1st dose date of Part 3 or 352]	[296, 1st dose date of Part 3 or 352]	[296, 1st dose date of Part 3 or 352]
Part 3	Week 61 (C4D1)	421	[1 day after 1st dose date of Part 3, 463]	[1 day after 1st dose date of Part 3, 463]	[1 day after 1st dose date of Part 3, 463]	[1 day after 1st dose date of Part 3, 463]
	Week 73 (C7D1)	505	[464, 548]	[464, 548]	[464, 548]	[464, 548]
	Week 97 (C13D1)	673	[549, 757]	[549, 757]	[549, 757]	[549, 757]

Study Period	Protocol Specified Visit	Target Day	Visit Window Range [min, max]			
			Pregnancy test	Hematology / Urinalysis / Physical examination / ECOG	Blood chemistry / Myocardial Enzyme Test / Vital Signs / 12-lead ECG	Coagulation / Thyroid function tests / Echocardiography/MUGA
	Week 121	841	[758, 925]	[758, 925]	[758, 925]	[758, 925]
	+ 24 weeks					

a. Day of first dose of study drug.
b. Part 2 baseline for placebo arm in Part 1.

If a patient early discontinues from Part 1 and directly enters into Part 2, the following visit window in Part 2 and Part 3 will be recalculated based on the new target study day from Part 2 and displayed in Table 4.4 and Table 4.5 as below.

Table 4.4 Early entering Part 2 Analysis Windows for PRO endpoints and Range of Motion

Study Period	Protocol Specified Visit	Target Day	Visit Window Range [min, max]	
			Worst Stiffness NRS, BPI Worst Pain NRS, PROMIS-PF and EQ-5D-5L	Range of Motion
Part 2	Week 25 (C1D1)[a]	1	[-14, 1]	
	Week 29 (C2D1)	29	[2, 43]	
	Week 33 (C3D1)	57	[44, 71]	
	Week 37 (C4D1)	85	[72, 127]	[2, 127]
	Week 49 (C7D1)	169	[128, 1st dose date of Part 3 or 184]	[128, 1st dose date of Part 3 or 184]

Study Period	Protocol Specified Visit	Target Day	Visit Window Range [min, max]	
			Worst Stiffness NRS, BPI Worst Pain NRS, PROMIS-PF and EQ-5D-5L	Range of Motion
Part 3	Week 61 (C4D1)	253	[1st dose date of Part 3 +1, 295]	[1st dose date of Part 3 +1, 295]
	Week 73 (C7D1)	337	[296, 380]	[296, 380]
	Week 97 (C13D1)	505	[381, 589]	[381, 589]
	Week 121	673	[590, 757]	[590, 757]
	+ 24 weeks			

a. Day of first dose of Part 2

Table 4.5 Early entering Part 2 Analysis Windows for Safety Endpoints

Study Period	Protocol Specified Visit	Target Day	Visit Window Range [min, max]			
			Pregnancy test	Hematology / Urinalysis / Physical examination / ECOG	Blood chemistry / Myocardial Enzyme Test / Vital Signs / 12-lead ECG	Coagulation / Thyroid function tests / Echocardiography/MUGA
Part 2	Week 25 (C1D1)[a]	1	[-14,1]	[-14,1]	[-14, 1]	
	Week 29 (C2D1)	29		[2, 43]	[2,43]	
	Week 33 (C3D1)	57		[44, 71]	[44, 71]	

Study Period	Protocol Specified Visit	Target Day	Visit Window Range [min, max]			
			Pregnancy test	Hematology / Urinalysis / Physical examination / ECOG	Blood chemistry / Myocardial Enzyme Test / Vital Signs / 12-lead ECG	Coagulation / Thyroid function tests / Echocardiography/MUGA
	Week 37 (C4D1)	85	[2, 127]	[72, 127]	[72, 127]	[2, 127]
	Week 49 (C7D1)	169	[128, 1st dose date of Part 3 or 184]	[128, 1st dose date of Part 3 or 184]	[128, 1st dose date of Part 3 or 184]	[128, 1st dose date of Part 3 or 184]
Part 3	Week 61 (C4D1)	253	[1st dose date of Part 3 +1, 295]	[1st dose date of Part 3 +1, 295]	[1st dose date of Part 3 +1, 295]	[1st dose date of Part 3 +1, 295]
	Week 73 (C7D1)	337	[296, 380]	[296, 380]	[296, 380]	[296, 380]
	Week 97 (C13D1)	505	[381, 589]	[381, 589]	[381, 589]	[381, 589]
	Week 121	673	[590, 757]	[590, 757]	[590, 757]	[590, 757]
	+ 24 weeks					

a. Day of first dose of study drug.

4.2.5. PK Concentration Values below Quantification Limit (BQL)

The following rules will be applied to impute any BQL data prior to a descriptive summary and individual plots.

- All BQL values will be set to zero. These BQL will be listed as BQL and tabulated the number of BQL.

- When more than half (>50%) of the values at a single time point are BQL, mean and median values are reported as BQL. Standard deviation and %CV are not reported; maximum and minimum values are reported as observed (including BQL).

4.3. Data Reporting Convention

4.3.1. General Rules

Unless otherwise specified, tables summarizing disposition, demographics and baseline characteristics will also include an overall column for all patients combined.

In general, continuous variables will be summarized by number of patients with available data (n), mean, standard deviation (SD), median, minimum, and maximum values. Categorical variables will be summarized by the number (n) and percentage of patients in each category. For time-to-event variables, the 25th, median, 75th percentile of the time, and the corresponding 95% confidence intervals (CIs) will be estimated based on Kaplan-Meier method.

All collected data and any derived variables will be presented in patient data listings. A note will be added for any imputed data (e.g., imputed AE start date). Listings will be ordered by region, country, site, patient number, treatment group, planned visit, and assessment or event date, if applicable. The treatment group presented in listings will be based on randomized treatment for efficacy data and demographic and baseline data, and actual treatment received for safety and PK data unless otherwise noted.

The format of tables, figures and listings (TFLs) will be specified in the separate file of TFL mockup. Any rounding will be done after all calculations are made. Precision details refer to Appendix 2 of Section 11.2.

All data processing, summarization and analyses will be conducted using the SAS® system (SAS Institute Inc., Cary, NC) Version 9.4 or above.

4.3.2. Study Treatment Groups Presentation

The summary of study data will include all relevant data obtained throughout a specific study period for the following respective patient groups. The grouping of patients for efficacy and safety data will be primarily based on the randomized treatment and the actual treatment received in Part 1, respectively, and whether patients received at least one dose of open-label ABSK021 beyond Part 1. The groups included in a specific table, figure, and listing (TFL) depend on their primary purpose and analysis stages.

Table 4.6 Efficacy Data Display of Study Treatment Groups

Part	SAP		CSR Report	
	Description	**Data inclusion**	**TFL Display (Order)**	**Included in Analysis**
Part 1	Randomized Placebo in Part 1	Data from patients who were randomized to Placebo during Part 1 before their first dose in Part 2	Placebo P1 (1)	Primary analysis* Final analysis*
	Randomized ABSK021 in Part 1	Data from patients who were randomized to ABSK021 during Part 1 before their first dose in Part 2.	ABSK021 P1 (2)	Primary analysis Final analysis
Part 2 and Part 3	Placebo Switch to ABSK021	Part 2 and Part 3 data from patients who switched from blinded Placebo to open-label ABSK021 in Part 2	Switched to ABSK021 P2+P3 (3)	Primary analysis Final analysis
Part 1, 2 and Part 3	Randomized ABSK021 with longer follow up	Parts 1- 3 data up to the DCO from the patients who were randomized to ABSK021 in Part 1	ABSK021 P1-P3 (4)	Primary analysis Final analysis
Part 1, 2 and Part 3	All ABSK021 Treated	Parts 1-3 data up to the DCO from patients who received at least 1 dose of ABSK021	All ABSK021 Treated P1-P3 (5)	Primary analysis Final analysis End of study analysis

*Note: the primary and final analyses are defined in Section 2.

Table 4.7 Safety Data Display of Study Treatment Groups

Part	SAP		CSR Report	
	Description	**Data inclusion**	**TFL Display (Order)**	**Included in Analysis**
Part 1	Received Placebo in Part 1	Data from the patients who received blinded Placebo during Part 1 before their first dose in Part 2	Placebo P1 (1)	Primary analysis* Final analysis*
Part 1	Received ABSK021 in Part 1	Data from the patients who received blinded ABSK021 in Part 1 before their first dose in Part 2	ABSK021 P1 (2)	Primary analysis Final analysis
Part 2 and Part 3	Placebo Switch to ABSK021	Part 2 and Part 3 data from patients who switched from blinded Placebo to open-label ABSK021 in Part 2	Switched to ABSK021 P2+P3 (3)	Primary analysis Final analysis
Part 1, 2 and Part 3	Received ABSK021 with longer follow up	Parts 1- 3 data up to the DCO from the patients who received blinded ABSK021 in Part 1	ABSK021 P1-P3 (4)	Primary analysis Final analysis
Part 1, 2 and Part 3	All ABSK021 Treated	Parts 1-3 data up to the DCO from patients who received at least 1 dose of ABSK021	All ABSK021 Treated P1-P3 (5)	Primary analysis Final Analysis End of study analysis

*Note: the primary and final analyses are defined in Section 2.

5. BACKGROUND CHARACTERISTICS

5.1. Patient Disposition

The patient disposition summary will include the following groups:

- Screened patients: the total number of screened patients by overall will be presented for the following groups: screened, screening failure and primary reasons for screening failure.
- Enrolled patients: the total number of enrolled patients by overall will be presented for the following groups: enrolled patients (i.e., signed inform consent form and meet eligible criteria), enrolled but not randomized and primary reasons.
- Randomized patients: the total number of randomized patients by randomized treatment group and overall will be presented for the following groups: randomized patients, randomized but not received treatment and primary reasons.
- Patients entered Part 1: the total number and percentage of patients by randomized treatment group and overall will be presented for the following groups: received treatment, treatment completed, treatment discontinuation and the primary reasons, completed Part 1 study completed, early withdrawal from the study Part 1.
- Patients entered Part 2: the total number and percentage of patients by randomized treatment group and overall will be presented for the following groups: entered Part 2, not enter Part 2 and primary reasons, received treatment, treatment completed, treatment ongoing, treatment discontinuation and the primary reasons, Part 2 study completed, study ongoing, early withdrawal from the study Part 2.
- Patients entered Part 3: the total number and percentage of patients by randomized treatment group and overall will be presented for the following groups: entered Part 3, not enter Part 3 and primary reasons, received treatment, treatment completed, treatment ongoing, treatment discontinuation and the primary reasons, Part 3 study completed, early withdrawal from the study Part 3.

All these primary reasons for premature treatment and study discontinuations will be based on data collected from the end of treatment and the end of study CRF pages for each study part. The denominators for calculating all percentages will be the number of patients who received the study drug in each randomized treatment group within each study part.

The total number and percentage of patients by randomized treatment group and overall will be presented for analysis data sets (excluding screened set), including those patients who were excluded from any analysis set and primary reasons for exclusion, if applicable.

Patient eligibility information (including Initial Informed Consent Date, Protocol Version Number, Additional Optional Informed Consent Date, Eligibility criteria not met) will be listed for all screened patients.

Patient recruitment by region (China, US/North America, EU), country, and center will be summarized by randomized treatment based on ITT analysis set. Randomization details will be listed, including the data of randomization, randomization number and randomization strata, and any prematurely unblinding, date and reasons for early unblinding prior to completion of Part 1 treatment.

5.1.1. Protocol Deviations

Protocol deviations (PDs) will be tracked by the study team throughout the conduct of the study in accordance with the Protocol Deviation Management Plan. PDs will be categorized to Major PD and Minor PD.

PD data and their categories will be reviewed and agreed before freezing the database for each planned analysis.

All protocol deviations will be listed for the ITT. All major and minor protocol deviations will be summarized by treatment group based on PD Categories.

The number and percentage of patients in the ITT with any protocol deviation will be tabulated for the following categories by deviation category (as specified in the study deviation specification document) and by treatment group.

- Patients with at least one protocol deviation
- Patients with at least one major protocol deviation
- Patients excluded from per-protocol analysis set due to major PD

All protocol deviations will be listed with flags for severity (major, minor).

5.2. Demographics and Baseline Characteristics

All demographic data such as age, age groups, gender, race, and ethnicity will be summarized by treatment group for the ITT analysis set. Standard descriptive statistics will be presented for the continuous variables of:

- Age (years);
- Baseline weight (kg);
- Height (cm);
- BMI (kg/m^2).

The frequencies and percentages of patients will be presented for the categorical variables of subgroups as described in Section 4.1.2, plus the following group.

- Unresectability (Yes vs. No vs. Unable to confirm)

Other baseline measurements (e.g., PRO, lab, vital signs, ECG, etc.) will be summarized together with the post-baseline measurements.

5.3. Diagnosis of Tumor and Baseline Disease Characteristics

These data will be summarized by treatment group and overall for the ITT. A listing will be prepared as well.

Standard descriptive statistics will be presented for the continuous variables of:

- Time from initial diagnosis (years) [calculated as (date of informed consent – date of initial diagnosis + 1)/365.25, partial date of initial diagnosis will be imputed per rules in the Appendix 1 of Section 11.1 before calculation];
- Time from current diagnosis (days) [calculated as (date of informed consent – date of current diagnosis + 1), partial date of current diagnosis will be imputed per rules in the Appendix 1 of Section 11.1 before calculation];

The frequencies and percentages of patients will be presented for the categorical variables of:

- Initial Diagnosis

- TGCT Type
- Current Diagnosis
- Current Tumor Location
- Metastatic (Yes, No)

  - If Yes, Location of Metastasis

These categories in each item will be based on subcategories specified in the corresponding CRF page unless stated otherwise.

5.4. Medical History

Medical history will be coded using MedDRA Version 25.1 or higher. Medical history data will be summarized by randomized treatment group.

5.5. Prior, Concomitant and Subsequent Therapy

5.5.1. Definition and Classification

Prior and concomitant and subsequent therapy includes any medications for treating both cancer and non-cancer conditions, surgeries, and radiotherapies that patients received prior to the study start, during study treatment and throughout study follow-up period. Medications will be coded to Anatomical Therapeutic Classification (ATC) and preferred name (PN) using the latest version of World Health Organization Drug Dictionary (WHODD). Non-medication therapies will be coded using MedDRA.

Prior therapies are those medications, surgeries and radiotherapies received prior to the first dose of the study drug and ending before the first study drug dose of Part 1. Concomitant therapies are defined as of those were started on or after the first dose of the study drug but prior to the last dose of the study drug at the time of data cut-off + 30 days or started prior to the study drug and were ongoing during the treatment period. Subsequent anti-cancer therapies are those that ~~were~~ started after the last dose of the study drug. Medications, surgeries and radiotherapies with incomplete start and/or end dates will be imputed for classification of prior, concomitant and subsequent according to Appendix 1. Those with incomplete start and/or end dates will be assumed to be concomitant if it cannot be shown that the medication, surgery or radiation therapy was not taken during the treatment period. For any incomplete onset or end date, the imputed date and raw date will be listed.

5.5.2. Prior Anti-Cancer Therapy

The number and percentage of patients with any prior systemic anti-cancer therapy, intent, line of therapy, type of therapy, the best response, reason for discontinuation and disease progression or recurrence will be summarized by treatment group for the ITT analysis set. The duration of last prior systematic therapy and time from the date of last prior systemic therapy to the first dose of study drug may be descriptively summarized.

The number and percentage of patients under different number of prior surgery (i.e., none, ⩾1, ⩾2, ⩾3 etc.) and patients with or without radiation therapy will be summarized by treatment group and overall.

5.5.3. Prior and Concomitant Medication

The number and percentage of patients receiving any prior medication and concomitant medication will be summarized by treatment group and overall, by ATC (level 2) and PN for the ITT analysis set. If a patient took a medication multiple times, the patient will only be counted once within ATC and PN within ATC, respectively. The summary will be presented in a descending order of incidence of the ATC and PN within ATC across patients overall.

The concomitant medication used for adverse events of clinical interest (defined in Section 7.2.3) will be summarized by ATC (Level 2) and PN, separately.

Listings will be produced for prior and concomitant medication, prior and concomitant non-medication therapies with flags indicating prior or concomitant.

5.5.4. New Anti-Cancer Therapy

The number and percentage of patients receiving any new anti-cancer surgery, surgery location, surgery intent, pathological results after surgery and recurrence will be summarized by treatment group and overall.

Patients receiving any other new anti-cancer therapies other than surgery will be listed.

5.6. Study Drug Administration

This study drug administration summary will include all relevant safety data obtained throughout the on-treatment period. The summary will be prepared for each respective group within each planned analysis stage as outlined in Table 4.7 based on the Safety Analysis set.

Duration of drug exposure, cumulative actual dose, actual dose intensity (ADI), relative dose intensity (RDI), percentage of intended dose (PID), and number of patients with dose reductions, interruptions and their reasons will be descriptively summarized and listed for the Safety Analysis Set. Furthermore, the percentage of intended dose (PID) will be summarized by the planned visit.

Details of the derivations and summaries for the SST are provided in the following sections.

5.6.1. Duration of Study Drug Exposure

Exposure to the study drug will be defined as:

- Total duration of exposure (day) = min (last dose date, date of death, date of cut-off) – first dose date + 1
- Actual duration of exposure (day) = total planned exposure – total duration of dose interruptions

The total and actual duration of exposure for ABSK021 will be summarized descriptively The number and percentage of patients will be summarized by following categories of duration of drug exposure:

- < 1 month, >= 1 - < 3 months, >= 3 - < 6 months, >= 6 - < 9 months, >= 9 - < 12 months, >= 12 - 18 months, >= 18 - 24 months, >= 24 months
- >= 1 month, >= 3 months, >= 6 months, >= 9 months , >= 12 months, >= 18 months, >= 24 months

5.6.2. Study Drug Dose Intensity

Actual Cumulative Dose

Actual cumulative dose is defined as the total dose of the medication given during the study treatment exposure. All doses, including unscheduled doses, will be included in the calculation.

Actual cumulative dose (mg) = sum of (actual dose * actual frequency * duration per dose)

Dose Intensity and Relative Dose Intensity

Planned dose intensity (PDI) is defined as follows:

- PDI (mg/ day) = Assigned treatment dose arm (i.e., total dosage per day)

Actual dose intensity (ADI) is defined as follows:

- ADI (mg/ day) = Actual cumulative dose (mg) / Actual duration of exposure (day)

Relative dose intensity (RDI) is defined as follows:

- RDI (%) = (ADI / PDI) *100

The number and percentage of patients in the following categories for RDI will be presented for: > 0 - < 80%, 80% -100%, >100%- <=120%, > 120%.

Percentage of Intended Dose

Percentage Intended Dose (PID) will be calculated as:

- (Actual cumulative dose/ Intended cumulative dose) * 100

Intended cumulative dose is defined as planned dose intensity * total duration of exposure.

The number and percentage of patients in the following categories for PID will be presented for: > 0 - < 80%, 80% -100%, > 100%- <=120%, > 120%.

5.6.3. Study Drug Modification

Study drug modifications will be summarized by treatment group in the following categories: dose reduction, dose interruption, dose modification (dose reduction or dose interruption). The number of patients and percentage who had at least one of event in each category and the number of study drug modifications (0, 1, 2 etc.) will be prepared separately.

The reasons for study drug modifications as above are also summarized according to categories collected in CRF.

6. EFFICACY ANALYSIS

6.1. General Consideration of Efficacy Analysis

ITT will be used as the primary analysis set for the primary and secondary efficacy analyses. The PPS will be used for the supportive analysis of the primary efficacy endpoint. Patients will be assigned to the treatment group which they are randomized to in Study Part 1.

Patients will be randomized according to the stratification factor of China versus non-China sites. When appropriate, the stratification factor will be included in the statistical model for efficacy analyses with the strata information as entered in IWRS at the time of randomization. All statistical tests will be two-sided at the alpha of 0.05 significance level.

The study will include three sets of planned efficacy analyses as outlined below. Detailed definitions of treatment groups refer to Table 4.6.

- Primary Efficacy Analysis at the End of Study Part 1

  The primary efficacy analysis by the randomized treatment groups of ABSK021 and Placebo will be performed on efficacy data up to the end of Study Part 1 with the last assessment at Week 25 visit before the first dose of Part 2. Formal statistical inferences will be generated, and results from this set of analysis will be used as the key efficacy findings of this study. See Table 4.6 for details.

- Final Efficacy Analysis at the End of Study Part 2

  The final efficacy analysis for ABSK021 will be performed on all efficacy data collected in Part, Part 2 and Part 3 up to the DCO.

- End of Study Efficacy Analysis at the End of Study Part 3

  The long-term efficacy analysis for ABSK021 will be performed on all efficacy data collected in Parts 1 – 3.

For the primary analysis at the end of study Part 1, the baseline for an efficacy variable is the last non-missing value before randomization. In addition, baseline values for PRO variables are the mean of non-missing values within 14 days before randomization and non-missing values before the first dose of C1D1. The baseline for each PRO variable is defined in detail in their corresponding sections below. For the final analysis at the end of study part 2, for the randomized ABSK021 treatment group, the baseline will follow all baseline definitions in Part 1.

For Placebo Switched to the ABSK021 group, the baseline value will be derived from the measurement values recorded on or before the date of the first dose of ABSK021 in Part 2. Specifically, for both BIRC and Investigator's tumor assessments, a patient's last time point tumor assessment of Part 1 will be considered as the baseline for Part 2 (see details in Section 6.5.1). The last non-missing measurement recorded on or before the date of the first dose of ABSK021 in Part 2 will be used as the baseline for ROM, PROMIS-PF, and EQ-5D-5L efficacy endpoints. The average of at least 4 of 7 consecutive days before the last visit recorded on or before the date of the first dose of ABSK021 in Part 2 will be used as the baseline for BPI-Worst-Pain NRS and Worst-Stiffness-NRS. Post-baseline values will be derived as the data collected after the first dosing of ABSK021 in Part 2.

6.2. Testing Strategy and Multiplicity Adjustments

A hierarchical gatekeeping testing procedure will be applied to control family-wise Type-1 error rate at 2-sided alpha = 0.05 level over the primary testing treatment effect of primary and key secondary endpoints. These five key secondary efficacy endpoints will be tested in the prespecified order below:

1) 25-week ORR by BIRC based on TVS

2) Mean change from baseline in relative ROM of the affected joint at Week 25

3) Mean change from baseline in Worst-Stiffness-NRS score at Week 25

4) Mean change from baseline in BPI-Worst-Pain-NRS score at Week 25

5) Mean change from baseline in PROMIS PF scale score at Week 25

The treatment effect on the second endpoint #1 can only be evaluated if the statistically significant treatment effect on the 25-Week ORR by BIRC has been established at the 2-sided significance level of 0.05 ($P < 0.05$); and the treatment effect on the second endpoint #2 will be evaluated only if the statistically significant treatment effect on the second endpoint #1 has been established at the 2-sided significance level of 0.05. Similarly, all remaining secondary endpoints can be sequentially tested at a 2-sided significance level of 0.05 in the pre-specified order. A statistical test for any subsequent secondary endpoint (s) will not be performed if the treatment effect on any preceding secondary endpoint is not statistically significant.

Once an efficacy endpoint is found to be insignificant ($P \geq 0.05$), the formal testing procedure will stop. For all subsequent key efficacy endpoints, nominal p-values will be provided but will not be considered a formal test of hypotheses. Unless otherwise stated, all statistical tests will be conducted at the 2-sided significance level of 0.05.

No multiplicity adjustments will be applied to any other analyses of efficacy endpoints. Nominal p-values may be computed for these analyses as a measure of the strength of association between the endpoint and treatment effect rather than formal tests of hypotheses, if applicable.

6.3. Primary Efficacy Analysis at the End of Study Part 1

6.3.1. Definition of Primary Efficacy Endpoint

During Study Part 1, tumor scans are scheduled on Day 1/Week 13 and Day 1/Week 25. In this randomized study, there is no requirement for a confirmation scan during the subsequent MRI assessment to determine a Complete Response (CR) or Partial Response (PR) in accordance with RECIST v1.1. A blinded independent radiology committee (BIRC) at the CRO Clario will evaluate the radiographic images and provide Time Point Response (TPR) at Week 13 and Week 25, and cross-time point assessments including the Response Status, the Date of Progression (if applicable), and the Date of First Response (if applicable) according to RECIST v1.1. Further details can be found in Table 6.1 for the necessary clarifications to determine the Response Status. Details refer to the Blinded Independent Review Charter for this study.

The primary efficacy endpoint, BIRC-25-Week ORR, will be calculated as the proportion of patients classified as responders, i.e., those achieving either a Complete Response (CR) or Partial Response (PR) within 25 weeks, with all randomized patients as the denominator. The calculation will follow these rules in line with the data handling strategy for intercurrent events:

- If patients undergo dose modifications due to toxicity before the Week 25 visit, any response that occurs after the dose modification will be included in the calculation of ORR, aligning with the treatment policy strategy.
- If patients undergo treatment discontinuation without disease progression before the Week 25 visit, any response that occurs after treatment discontinuation will be included in the calculation of ORR, aligning with the treatment policy strategy.
- If patients receive an alternative anticancer therapy (including surgery) before the Week 25 visit, only responses that occurred before the event will be included in the calculation of ORR, aligning with the while-on-treatment policy strategy.

Table 6.1 Evaluation of Response Status – Part 1

TPR at Week 13	TPP at Week 25	Overall Response Status at Week 25 (i.e., BOR)
CR or PR	CR	Response (CR)
CR or PR	PD	Non-response (PD)
PR	non-CR/non-PD/non-NE	Response (PR)
SD	CR or PR	Response (CR or PR from Week 25)
SD	SD	Non-response (SD)
SD	PD	Non-response (PD)
CR, PR, SD, or NE	NE	Non-response (NE)
PD	Any	Non-response (PD)
NE	CR or PR	Response (CR or PR from Week 25)
NE	SD or PD	Non-response (SD or PD from Week 25)

TPR = time point response; CR = complete response; NE = Not evaluable (including MRI scan is not done); PD = progressive disease; PR = partial response; SD = stable disease

6.3.2. Handling of Dropouts and Missing Data

If patients withdraw consent from the study before the Week 25 visit, no further data will be collected, and only tumor scans up until the date of withdrawn consent will be included in the evaluation of tumor response. Missing MRI data will not be imputed. Patients with missing MRI data at the Week 25 visit will be considered as non-responders.

6.3.3. Primary Analysis of Primary Efficacy Endpoint

The primary BIRC-25-Week ORR based on RECIST v1.1 will be analyzed based on the ITT analysis set. The ORR will be compared at a two-sided significance level of 0.05 using Fisher's exact test. Fisher's exact test is considered the preferred method for this primary analysis, particularly due to the expectation that the number of anticipated response events might be less than 5 within stratified placebo groups based on the historical data.

The 95% confidence interval (CI) for the difference in ORR (ABSK021 - placebo) will be calculated using the Wilson method. The option RISKDIFF (CL = WILSON) will be used in SAS PROCFREQ procedure. The exemplary SAS code for the primary analysis is presented below.

```
PROC FREQ DATA= dataset;
CLASS treatment response;
TABLES treatment * response / ALPHA = 0.05 RISKDIFF (CL = WILSON);
EXACT FISHER;
RUN;
```

The number and percentage of patients in each category of Best Overall Response (BOR) by the BIRC (CR, PR, SD, non-CR/non-PD/non-NE, PD, or NE) will be presented by treatment group. A similar summary of the primary ORR for each treatment group along with their exact two-sided 95% CIs using the Clopper-Pearson method will be provided.

In addition, summaries will be performed on the BIRC-assessed tumor size of target lesions:

- The absolute values and percentage changes in target lesion tumor size from baseline by BIRC will be descriptively summarized and presented by visit and treatment group.
- The best percentage change from baseline (i.e., the maximum reduction from baseline or the minimum increase from baseline in the absence of a reduction) in the sum of diameters of target lesions will be descriptively summarized.
- The best percentage change from baseline in tumor size for each patient will also be presented graphically using waterfall plots. Reference lines at the +20% and -30% change in tumor size levels will be added to the plots, which correspond with the definitions of progression and 'partial' response respectively.

Detailed tumor assessment data from the BIRC will be listed for all randomized patients.

6.3.4. Sensitivity Analyses of the Primary Efficacy Endpoint

The following sensitivity analyses are planned to evaluate robustness of the primary analysis of the primary ORR based on the ITT analysis set.

- The treatment effect on the BIRC-25-Week ORR will be evaluated using Fisher's exact test with mid p-value correction (Lydersen et al[3,4], 2003, 2009). Fisher's exact test is considered the preferred method for this primary analysis, particularly due to the expectation that the number of anticipated response events might be less than 5 within stratified placebo groups based on the historical data. The corrected p-value will be calculated by subtracting half of the probability of the observed table from Fisher's p-value as shown below:
  Fisher's corrected p-value = 2- sided Fisher's p-value – (Table probability/2)
- The treatment effect on the BIRC-25-Week ORR will be evaluated using a stratified Cochran–Mantel–Haenszel (CMH) method to account for the stratified factor of region (China vs. non-China sites). The adjusted proportion difference (ABSK021- Placebo) and its 95% CI will be calculated using CMH-weighted method (Kim[5], 2013).

6.3.5. Supplemental Analyses of the Primary Endpoint

The ORR may be summarized by treatment group based on Per Protocol Set (PPS). This analysis will employ the same methodology used in the primary efficacy analysis, incorporating both Fisher's exact test with or without mid p-value correction and the stratified Cochran–Mantel–Haenszel (CMH) method. Furthermore, for the PPS, the number and percentage of patients in each category of Best Overall Response (BOR) by the BIRC (CR, PR, Non-CR/Non-PD/Non-NE, SD, PD, NE) will be presented. A similar summary of the primary ORR for each treatment group along with their exact two-sided 95% CIs using the Clopper-Pearson method will be provided.

6.3.6. Subgroup Analyses

Subgroup analyses for the primary BIRC-25-Week ORR will also be performed within the following subgroups to assess whether study drug exhibits a consistent treatment effect across each of these subgroups.

- Age Group (< 40years, ≥ 40years, ≥ 40 - < 65 years, ≥ 65 years)
- Sex (Female vs. Male)
- Region Category 1 (China vs. ex-China)
- Region Category 2 (China vs. North American vs. EU)
- Race Category 1 (Asian vs. non-Asian)
- Race Category 2 (Caucasian, non-Caucasian)
- Ethnicity (Hispanic or Latino vs. Not Hispanic or Latino)
- Lower extremity tumors vs. upper extremity tumors
- Knee vs. Others
- TCGT type (localized-type vs. Diffuse-type)
- ECOG PS (0 vs.1)

- Number of prior surgeries (0 vs. ≥ 1)
- Prior systemic therapy (Yes vs. No)

For each subgroup defined above, the same approach employed in the primary analysis of the primary ORR will be used. These subgroup analyses are considered exploratory in nature and will not involve adjustments for multiple comparisons. A forest plot illustrating the difference in ORR (ABSK021- Placebo) and its corresponding 95% CI (Wilson method) will be produced for each subgroup. If any subgroup contains an insufficient number (i.e. less than 5) of patients for meaningful analysis, the data for those subgroups will be presented descriptively rather than being subjected to statistical testing.

6.4. Key Secondary Efficacy Endpoints Analysis – Part 1

6.4.1. Definitions of Key Efficacy Endpoints

6.4.1.1. BIRC-25-Week ORR Based on Tumor Volume Score （TVS）

The BIRC-25-Week ORR based on TVS measures the proportion of patients who achieve a BOR of either CR or PR as assessed by BIRC within 25 weeks. This assessment is based on the tumor scans that are scheduled on Day 1/Week 13 and Day 1/Week 25 according to TVS criteria detailed in the protocol section 6.3.2. TVS-based tumor response assessments are similar to RECIST v1.1 with necessary clarifications described in Table 6.1 above. All MRI scans will be only read in a blind manner according to the procedures outlined in Blinded Independent Review Charter for this study.

Similarly, BIRC-25-Week ORR based on TVS will be calculated based on all randomized patients. The calculation of the ORR will follow the same rules outlined for the primary efficacy endpoint in terms of handling intercurrent events.

6.4.1.2. Relative Range of Motion (ROM)

For the assessment of the range-of-motion endpoint, raw measurements of the affected joint across different motion types will be performed using a goniometer and expressed in degrees (Protocol Section 6.3.3) at Screening, Week 13 and Week 25 visits in Study Part 1. To obtain a normalized assessment, measurements of a joint and motion type will be standardized to the full range of motion (ROM) for that joint, resulting in a relative value. As shown in Table 6.2, the standard full range of motion value for a given joint and motion type is calculated as the sum of standard reference values based on the American Medical Association disability criteria (Gerhardt JJ[6], 2002). Consequently, the relative ROM value for a given joint and motion type will be calculated as follows.

Relative ROM value (%) = 100 x (absolute ROM value)/ (ROM full range value)

For instance, if a patient's measurements for shoulder's extension and flexion are 40 degrees and 170 degrees respectively, and the full range value for that joint is 230 degrees, the relative ROM for shoulder Extension-0-Flexion type would be 91.3% (calculated as 100 * (40 + 170) / 230).

Table 6.2 Standard Reference Value of ROM Assessment

Location	Motion Type	Standard Reference Value (Degree)	Full Range of Motion (Degree)
Shoulder	Extension -0- Flexion	50 -0- 180	230
	Abduction -0- Adduction	160 -0- 30	190
	External Rotation -0- Internal Rotation	90 -0- 90	180
Elbow	Extension -0- Flexion	0-0-135	135
	(Hyper) Extension -0- Flexion	10-0-135	145
Wrist	Extension -0- Flexion	60 -0- 50	110
Thumb (MCP joint)	Extension -0- Flexion	0-0-60	60
Thumb (IPP joint)	Extension -0- Flexion	30-0-80	110
Fingers (MCP joint)	Extension -0- Flexion	20-0-90	110
Fingers (PIP joint)	Extension -0- Flexion	0-0-100	100
Fingers (DIP joint)	Extension -0- Flexion	0-0-100	100
Hip	Extension -0- Flexion	30 -0- 100	130
	Abduction -0- Adduction	45 -0- 30	75
	External Rotation -0- Internal Rotation	50 -0- 40	90
Knee	Extension -0- Flexion	0 -0-150	150
	(Hyper) Extension -0- Flexion	10 -0- 150	160
Ankle	Dorsiflexion -0- Plantar flexion	20 -0- 40	60

Location	Motion Type	Standard Reference Value (Degree)	Full Range of Motion (Degree)
Foot	Eversion -0- Inversion	20 -0- 30	50

MCP: Metacarpophalangeal joints. IPP: Interphalangeal joint. PIP: Proximal interphalangeal joints. DIP: Distal interphalangeal joints.

The last non-missing value of ROM on or before the randomization will be used as a baseline. If multiple records for the same joint location are measured by different motion types at the baseline, the motion type with the smallest (worst) relative ROM will be used as the baseline and this motion type will be used for evaluating the relative ROM subsequently. In the event of ties of more than one motion type at baseline, the average of relative ROM values for all such motion types will be calculated for each post-baseline ROM assessment. The post-baseline records will follow the same joint locations and motion types chosen at baseline.

The change from the baseline in ROM will be calculated for each scheduled visit by subtracting the baseline value from each post-baseline value. If either the baseline value or the post-baseline value is missing, the change will be set as missing.

6.4.1.3. Worst Stiffness Numeric Rating Scale (NRS)

Speck R. et al. (2019 [7]) described the content validity of the measures of these 2 PROs for patients with tenosynovial giant cell tumors (TGCT). Worst Stiffness NRS and PROMIS-PF have good reliability, validity, and responsiveness, and provide guidance for the interpretation of meaningful change in TGCT patients (Speck R. et al. 2020[8]). The Worst-Stiffness-NRS item is a one-item self-administered questionnaire assessing the "worst" stiffness in the last 24 hours. The NRS for this item ranges from 0 ("no stiffness") to 10 ("stiffness as bad as you can imagine"). The instrument will be measured at Screening, Week1, Week 5, Week 9, Week 13, Week 17, Week 21 and Week 25 visits during Study Part 1. Except for Week 1 (C1D1), patients are required to complete the Worst-Stiffness-NRS assessments for at least 4 of 7 consecutive days during each schedule visit. The assessment at Week 1 (C1D1) should be completed prior to the first dosing. Consequently, a minimum of 4 out of the consecutive 7 days will be required to compute the mean of NRS score for each post-dose visit; otherwise, it will be set to missing. The baseline will be derived based on the mean score of evaluations completed prior to randomization, specifically, at least 4 of 7 consecutive days during the 2 weeks prior to randomization, and on C1D1. If there are multiple consecutive 7-day assessments, the baseline will be calculated using the most recent 7-day assessments, in addition to C1D1.

The change from baseline will be calculated for each scheduled visit by subtracting the mean baseline value from each post-baseline mean value. If either the baseline value or the post-baseline value is missing, the change will be set as missing.

6.4.1.4. BPI Worst Pain Numeric Rating Scale (NRS)

The BPI-Worst-Pain-NRS item is a one-item self-administered questionnaire assessing the "worst" pain in the last 24 hours. The NRS for this item ranges from 0 ("no pain") to 10 ("pain as bad as you can imagine"). The instrument will be measured at Screening, Week1, Week 5, Week 9, Week 13, Week 17, Week 21 and Week 25 visit during Study Part 1. Except for Week 1

(C1D1), patients are required to complete the Worst-Pain-NRS assessments for at least 4 of 7 consecutive days during each scheduled visit. The assessment at Week 1 (C1D1) should be completed prior to the first dosing. Consequently, a minimum of 4 out of the consecutive 7 days will be required to compute the mean of NRS score for each post-dose visit; otherwise, it will be set to missing. The baseline will be derived based on the mean score of evaluations completed prior to randomization, specifically, at least 4 of 7 consecutive days during the 2 weeks prior to randomization, and on C1D1. If there are multiple consecutive 7-day assessments, the baseline will be calculated using the most recent 7-day assessments, in addition to C1D1.

The change from baseline will be calculated for each scheduled visit by subtracting the mean baseline value from each post-baseline mean value. If either the baseline value or the post-baseline value is missing, the change will be set as missing.

6.4.1.5. PROMIS Physical Function (PROMIS-PF) Scale

Physical function items relevant to the assessment of lower and upper limb function have been selected from the PROMIS-PF Bank v2.0. The customized scale includes two different sets of topics for the upper and lower extremities (protocol Appendix 12.6). A total of 13 items assessing lower extremity function will be used for patients with lower extremity tumors, and 11 items for assessing upper extremity function will be used for patients with upper extremity tumors. The PROMIS- PF will be measured at Screening, Week 1, Week 5, Week 9, Week 13, Week 17, Week 21 and Week 25 visits during Study Part 1.

The HealthMeasures Scoring Service Platform (see Appendix 3) will be used to score customized PROMIS items by utilizing the response pattern scoring approach, which is especially useful when there is missing data. The platform that uses response pattern scoring can produce a T-score for a respondent as long as 1 item has been answered. However, to produce a precise score using response pattern scoring, a patient must answer at least 4 items in a short form that includes 4 or more items.

The raw score (i.e., 1-5) for each item from each patient at each visit will be uploaded to the above-mentioned Service Platform for scoring. The steps to utilize the Service Platform are detailed in Appendix 3. The Platform will convert raw scores into a T-score for each patient at each visit. The T-score rescales the raw score into a standardized score with a mean of 50 and a standard deviation (SD) of 10. Therefore, a person with a T-score of 40 is one SD below the mean. The T-score ranges from 0 to 100, with a higher score indicating better physical function status.

The raw scores collected during the screening period (within 14 days before randomization) and on C1D1 will be converted to T-scores before calculating the baseline T-score average. The change from baseline will be calculated for each scheduled visit by subtracting the mean baseline value from each post-baseline mean value. The change will be set as missing if either the baseline or post-baseline values are missing.

6.4.2. Handling of Dropouts and Missing Data

To minimize missing data due to early treatment discontinuation and toxicity, patients are strongly encouraged to complete all visits until they withdraw from the study under the treatment

policy strategy. The following analyses will be conducted to assess the timings, patterns, and reasons for missing data from the study.

Tabular summaries will be presented to show the percentage of patients who discontinue their randomized treatment or withdraw from the study by treatment group. Kaplan-Meier plots will be used to present the time from the first dose to discontinuation of randomized treatment or withdrawal from the study, with further analysis split by treatment-related or non-treatment-related reasons for discontinuation. If necessary, additional exploratory analyses will be conducted to better understand the pattern of missing data based on these outputs.

For ORR-TVS, the same approach described for the primary efficacy ORR will be employed.

The primary analysis of the key secondary endpoints (excluding ORR-TVS) will use mixed models for repeated measurements (MMRM) that incorporate all available measurements from each patient under the missing completely at random (MCAR) and the missing at random (MAR) assumptions.

For sensitivity analysis of longitudinal continuous endpoints, including relative ROM, Worst-Stiffness-NRS, Worst-Pain-NRS, and PROMIS-PF T-scores, under missing not at random (MNAR), the control-based pattern mixture models (PMM), specifically, Jump to Reference (J2R), Copy Reference (CR) and the tipping point analysis (a type of PMM with delta adjustment) (Sande V. et. al. 2021[9]) will be performed. Details will be described in Section 6.4.4 below.

6.4.3. Primary Analysis of Key Secondary Endpoints

6.4.3.1. BIRC-25-Week ORR based on TVS

The BIRC-25-Week ORR based on TVS will be based on the ITT analysis set and primarily analyzed using the same method as the primary endpoint. The treatment effect on the ORR will be formally tested at the 2-sided significance level of 0.05 using Fisher's exact test only if the statistically significant treatment effect on the 25-Week ORR by BIRC has been established. Otherwise, p-value will be presented for the purpose of demonstrating trends.

The number and percentage of patients in each category of Best Overall Response (BOR) by the BIRC based on TVS (CR, PR, SD, PD, or NE) will be presented by treatment group. A similar summary of the ORR for each treatment group along with their exact two-sided 95% CIs using the Clopper-Pearson method will be provided.

6.4.3.2. Relative ROM at Week 25

The primary analysis of change from baseline in relative ROM at Week 25 will be conducted based on the ITT analysis set. A mixed model for repeated measures (MMRM) will be employed with use of all available data from baseline up to Week 25 visit. The MMRM model will include change from baseline as the dependent variable, and treatment, baseline, visit, stratification-factor of China vs. non-China sites, and treatment-by-visit interaction, baseline-by-visit interaction, joint-type-category (knee, ankle and others) as fixed effects. For the model, the restricted maximum likelihood (REML) estimation will be used along with Kenward-Roger

approximation for estimation of the degrees of freedom. To account for the variability in measurements within patients, an unstructured variance-covariance matrix will be used. If the fit of the unstructured covariance structure fails to converge, compound symmetry will be employed. The least squares mean (LSMEAN) for each treatment, and the treatment difference (ABSK021 – Placebo) in LSMEAN along with its' 95% CI and p-value will be provided for scheduled visits at Week 13 and Week 25. A formal statistical comparison between treatment groups will only be made at Week 25 following the prespecified order in the hierarchical gatekeeping testing procedure. In addition, a line plot showing LSMEAM of change from baseline along with standard error (SE) bars over time in Part 1 will be produced by treatment group.

The derived relative ROM and the change from baselines at each scheduled visit will be summarized descriptively by the treatment group. A line plot showing observed mean along with standard error bars over time will be produced by treatment group.

In addition, a table will be prepared to summarize the distribution of the mean scores used for analysis.

6.4.3.3. Worst-Stiffness-NRS at Week 25

The change from baseline in Worst-Stiffness-NRS at Week 25 will be analyzed based on the ITT analysis set using a mixed model for repeated measures (MMRM) with use of all data from baseline up to Week 25 visit. The model will include change from baseline as the dependent variable, and treatment, baseline, visit, stratification-factor of China sites vs. non-China sites, and treatment-by-visit interaction and baseline-by-visit interaction as fixed effects. For the model, the restricted maximum likelihood (REML) estimation will be used; an unstructured variance-covariance matrix will be used to model the within-patient error and the Kenward-Roger approximation will be used to estimate the degrees of freedom. If the fit of the unstructured covariance structure fails to converge, the following covariance structures will be tried in order until convergence is reached: toeplitz with heterogeneity, autoregressive with heterogeneity, toeplitz, autoregressive and compound symmetry. The least squares mean (LSMEAN) for each treatment, and treatment difference (ABSK021 – Placebo) in LSMEAN along with its' 95% CI and p-value will be provided for each scheduled visit. A formal statistical comparison between treatment groups will only be made at Week 25 following the prespecified order in the hierarchical gatekeeping testing procedure. In addition, a line plot showing LSMEAM of change from baseline along with standard error (SE) bars over time in Part 1will be produced by treatment group.

Descriptive statistics will be used to summarize the Worst-Stiffness-NRS and the change from baselines at each scheduled visit by treatment group. A line plot showing observed mean along with SE bars over time will be produced by treatment group.

In addition, a table will be prepared to summarize the distribution of the mean scores used for analysis.

6.4.3.4. BPI-Worst-Pain-NRS at Week 25

The change from baseline in BPI-Worst-Pain-NRS will be primarily analyzed by employing the same method described for the Worst-Stiffness-NRS. A formal statistical comparison between treatment groups will only be made at Week 25 following the prespecified order in the hierarchical gatekeeping testing procedure. In addition, a line plot showing LSMEAN of change from baseline along with standard error (SE) bars over time in Part 1will be produced by treatment group.

Descriptive statistics will be used to summarize the BPI-Worst-Pain-NRS and the change from baselines by treatment group at each scheduled visit. A line plot showing observed mean along with SE bars over time will be produced by treatment group.

In addition, a table will be prepared to summarize the distribution of the mean scores used for analysis.

6.4.3.5. PROMIS PF T-score at Week 25

The change from baseline in PROMIS PF T-score will be primarily analyzed by employing the same method described for the Worst-Stiffness-NRS. All valid T-scores from the Service platform will be included in the model. Any non-valid score will be set to missing. For a T-score to be considered valid, a minimum of 4 non-missing items are utilized for deriving the score. A formal statistical comparison between treatment groups will only be made at Week 25 following the prespecified order in the hierarchical gatekeeping testing procedure. In addition, a line plot showing LSMEAM of change from baseline along with standard error (SE) bars over time in Part 1 will be produced by treatment group.

Descriptive statistics will be used to summarize the PROMIS PF T-score and the change from baselines by treatment group at each scheduled visit. A line plot showing observed mean along with SE bars over time will be produced by treatment group.

In addition, a table will be prepared to summarize the distribution of the T-score.

6.4.4. Sensitivity Analysis of Key Secondary Endpoints

6.4.4.1. Sensitivity Analyses of BIRC-25-Week ORR based on TVS

Sensitivity Analyses of BIRC-25-Week ORR based on TVS will follow the same approach described for the primary efficacy ORR in Section 6.3.4.

6.4.4.2. Sensitivity Analyses of Continuous Key Secondary Endpoints

Sensitivity analyses will be performed to examine the robustness of the results obtained from the primary analysis of continuous key primary endpoints under the treatment policy estimand using missing not at random (MNAR) assumption. These sensitivity analyses will consider possible departures from the underlying assumptions, specifically under MNAR scenarios. The following control-based pattern mixture models (PMM) will be used.

Jump to Reference (J2R):

In this sensitivity analysis, intermittent missing values (i.e., Week 13 value is missing while Week 25 value is available) will be imputed using the Markov Chain Monte Carlo (MCMC) methodology, which assumes a multivariate normal distribution over all variables included in the imputation model. This imputation will be done by treatment group, using the MI procedure in SAS 9.4. The variables to be used in the imputation are treatment, corresponding baseline values, and values observed at all double-blind visits (Week 13 and Week 25). Then all the monotone missing values will be multiply-imputed using the imputation model built from the Placebo group, but only from the point of missing time point (i.e., jump to reference). Once the completed data sets are formed, the same MMRM analysis model for the primary analysis as specified in sections 6.4.3.2, 6.4.3.3, 6.4.3.4, 6.4.3.5 will be applied to each completed set, and the inference will be drawn using Rubin's combination rules (SAS proc MIANALYZE).

Copy Reference:

In this sensitivity analysis, like the J2R imputation method, intermittent missing values will be imputed using the Markov Chain Monte Carlo (MCMC) methodology. Then all the monotone missing values will be multiply-imputed using the imputation model built from the Placebo group, i.e., assuming the missing data in the ABSK021 treatment group will have a profile that equals the profile of the Placebo group for all time points (i.e., a copy-reference imputation). Once the completed data sets are formed, the same MMRM analysis model for the primary analysis as specified in sections 6.4.3.2, 6.4.3.3, 6.4.3.4, 6.4.3.5 will be applied to each completed set, and the inference will be drawn using Rubin's combination rules (SAS proc MIANALYZE).

Tipping point:

In this sensitivity analysis, like the above imputation methods, intermittent missing values will be imputed using the Markov Chain Monte Carlo (MCMC) methodology. Then for the ABSK021 treatment group, add a range of constant delta (penalties) directly onto the imputed values, specifically reductions for ROM and PROMIS-PF T-score and increases for stiffness and pain scores, to worsen the treatment effect. It is desired to impose a fixed and definite set of quantities to encapsulate the change in efficacy associated with missing for the ABSK021 treatment group while missing for the Placebo treatment group is following the MAR assumption.

The analysis finds a tipping point in this spectrum of assumptions of delta, at which statistically significant conclusions change from being favorable ($p<0.05$) to the ABSK021 treatment to being unfavorable ($p>0.05$).

Patient Disposition and Summary of Missing Data for PROs

Table summarizing PRO data completeness will be provided. The denominator used to evaluate data completeness differs based on the PRO objective. When the PRO objective is clinical benefit (i.e., ROM, Worst-Stiffness, BPI-Worst-Pain, and PROMIS-PF), the percentage for available data rate and each missing category will be calculated using the randomized population as the fixed denominator.

Within the tables, counts and percentages of patients who completed and did not complete the PRO measure will be calculated with the accompanying reasons for missing observations (including intercurrent event categories) based on collected data.

6.4.5. Supportive Analyses of Key Secondary Endpoints

ROM and PROs under Different Strategy to the Intercurrent Event

For intercurrent event of receiving an alternative anti-tumor therapy (including surgery), while on treatment policy will be used for ROM and PROs as the sensitivity analyses. The similar MMRM model as the primary analysis of ROM and PROs will be used. The data after receiving an alternative anti-tumor therapy (including surgery) will not be included in the analyses.

BPI-30

A BPI-30 responder will be defined as a patient who (i) experienced a decrease of at least 30% in the mean BPI Worst Pain NRS item AND (ii) did not experience a 30% or greater increase in narcotic analgesic use, comparing data collected in concomitant medication during a 7-day period at the end of Part 1 compared with baseline values collected prior to the first dose of study treatment. Patients who do not provide data for the endpoint will be considered to be non-responders. Analgesic use will also be recorded in BPI visit and linked to concomitant medication page. If a patient didn't use analgesic medication at baseline but used it at post-baseline, it's regarded as a 30% or greater increase in narcotic analgesic use. If a patient used the same analgesic medication at baseline and post-baseline, the change in analgesic use should be calculated by the change in patients' mean analgesic use that calculated by multiplying the daily dose unit by the number of units taken, averaged by the number of days with available data in concomitant medication. If a patient used different analgesic medications at baseline and post-baseline, whether a 30% or greater increase in analgesic use will be judged by medical review case by case.

The BPI-30 responders will be summarized by treatment group based on ITT. The treatment comparison of BPI-30 will be based on Fisher' exact test.

Responder Summary of COAs (i.e.,ROM and PRO) by Clinical Thresholds

Descriptive by-treatment responder summary of the COAs change from baseline at Week 25 by thresholds of clinically meaningful within-patient change will be performed, such as 5%, 6.7%, 10%, 15%, 20%, and 25% for ROM; 1 point improvement for Worst Stiffness and 3 points improvement for PROMIS-PF T-scores (Speck et al., 2020[8]), and 1 point improvement for BPI (Dworkin et al., 2008[10]).

eCDF and ePDF Plots

Empirical cumulative distribution function (eCDF) and empirical probability density function (ePDF) of COAs (ROM, Worst Stiffness, BPI and PROMIS) changes from baseline at Week 25 for each treatment group will be prepared.

Correlation Analyses for Key Secondary Endpoints

To investigate the correlation between the treatment effect on tumor responses and improvements in key secondary clinical outcome assessments (COA), including change from baseline to Week 25 in relative ROM, Worst-Stiffness-NRS, Worst-Pain-NRS, and PROMIS-PF T-scores (named change below), the following descriptive summaries and subgroup analyses will be carried out at Week 25. For the tumor response assessment, BIRC evaluations based on the RECIST v1.1 criteria will be used, unless specified otherwise.

- A scatter plot of percent change in tumor size at X-axis and change in each COA at Y-axis with fitting in a nonparametric local regression (Loess) line and 95% CI will be provided, respectively.
- Spearman's rank-order correlation coefficients will be provided between the percentage change in tumor size and change for each COA, respectively.
- A waterfall plot of the change by tumor responder status by treatment group for each COA, respectively.
- Responder subgroup analysis of the change by tumor responder status will be performed. The evaluation will use a similar approach for the primary analysis of each COA endpoint, with the addition of three fixed factors (subgroup, subgroup-treatment interaction, subgroup-treatment-visit interaction) into the corresponding MMRM model.

6.4.6. Subgroup Analysis of Key Secondary Endpoints

Subgroup analysis of key secondary endpoints will follow the subgroups classification for primary efficacy endpoint in Section 6.3.6.

The evaluation for binary endpoint will use the same approach as the primary efficacy endpoint ORR.

The evaluation for continuous endpoints will use a similar approach for the primary analysis of each key secondary endpoint, with the addition of three fixed factors (subgroup, subgroup-by-treatment and subgroup-by-treatment-by-time) into the corresponding MMRM model.

6.5. Other Secondary Efficacy Analyses at the End of Study Part 1

6.5.1. Tumor Responses by Investigators

In addition to tumor responses by BIRC based on both RECIST v1.1 and TVS, the best overall response (BOR) based on RECIST v 1.1 will be determined programmatically using overall time point responses by investigators up until the last evaluable visit response at Day 1 of Week 25 visit prior to or on the date of (i) radiographic progression as defined by RECIST v1.1; or (ii) withdrawal of consent to obtain scans; or (iii) receiving subsequent anti-cancer therapy or surgery, whichever is earlier.

Investigator-25-Week ORR will be calculated as the proportion of patients with either CR or PR (i.e., responders) as the response by the Investigator within 25 Weeks based on RECIST v1.1 criteria. The calculation will follow the same rule as the primary endpoint of BIRC-25-Week ORR. The response status within 25 weeks (i.e., BOR) will follow Table 6.1 based on the tumor response at Week 13 and Week 25.

The number and percentage of patients in each category of BOR based on investigators' tumor assessments (CR, PR, SD, non-CR/non-PD, PD, or NE) will be presented for the ITT analysis set. The associated exact 95% two-sided CIs using Clopper-Pearson method will be computed for ORR.

Detailed tumor assessment data from investigators will be listed along with derived response for all randomized patients.

The best percent change from baseline in the sum of diameters of target lesions will be summarized using descriptive statistics (n, mean, standard deviation, median, minimum, and maximum). Additionally, the best percent change from baseline in the sum of diameters of target lesions of individual patients and BOR will be plotted by treatment groups using a waterfall plot. A swimmer plot of the overall response and treatment duration will be provided.

The Investigator-25-Week ORR based on RECIST v1.1 will be based on the ITT analysis set and analyzed using the same method as the primary endpoint.

To compare discrepancy of response assessments, a summary of concordance between the BICR and Investigator review of radiographic images will be provided.

6.5.2. Duration of Response (DOR)

In general, the DOR for responders is defined as the time interval between the date of the first documentation of objective response (CR/PR) and the date of first documentation of radiographic disease progression, or death, whichever occurs earlier.

DOR (months) = [date of event/censoring – date of the first qualifying response + 1]/30.4375

Patients who have no record of documented disease progression or death at the time of data cutoff date for a planned analysis will be censored according to the rules detailed in Table 6.3.

The following DORs will be derived for all patients with an objective response in Part 1.

- DOR by BIRC based on RECIST v1.1
- DOR by BIRC based on TVS
- DOR by Investigator based on RECIST v1.1

Analysis of DOR will be conducted for patients with an objective response in the ITT analysis set. The 25th, median, 75th percentile of the time, and their associated 95% CIs (Brookmeyer and Crowley[11], 1982) will be estimated using Kaplan-Meier method adjusting for censoring. In addition, the probability of DOR at landmarks of month 3, 6, 9 and 12 and their corresponding two-sided 95% CIs (Kalbfleisch[12], 1980) will be provided.

Furthermore, the number of patients who have an ongoing response with DOR ≥ 6 months and ≥ 12 months and the number of patients on follow-up will be summarized by treatment group.

Table 6.3 Censoring Rules for DOR

Situation	Date of Progression or Censoring	Outcome
Documented progression or death	Date of progression (or death)	Progressed
Death without documented progression between scheduled tumor assessments	Date of death	Progressed
No documented progression and no death	Date of last adequate tumor assessment	Censored
Documented progression or death after exactly one missing assessment	Date of progression (or death)	Progressed
Documented progression or death after more than one consecutive missed assessment visits	Date of last adequate tumor assessment prior to the first missed visit	Censored
Treatment discontinuation for toxicity or other reason (i.e., not due to documented progression)	Ignore the treatment discontinuation and follow situations above (ITT approach)	As per above situation
New anticancer treatment started prior to documented progression	Date of last adequate tumor assessment prior to the start of new anticancer treatment	Censored

6.5.3. European Quality of Life 5 Dimensions – 5 Levels (EQ-5D-5L)

The EQ-5D-5L are administered at Screening, Week 1, Week 5, Week 9, Week 13, Week 17, Week 21, and Week 25 during Part 1. The instrument consists of 2 pages: the EQ-5D descriptive system and the EQ visual analogue scale (VAS). A copy of the questionnaire is provided in protocol Appendix Section 12.7.

The questionnaire will be analyzed according to the recommendations of the authors (www.euroqol.org). The descriptive system contains 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), each with 5 response categories (no problems, slight problems, moderate problems, severe problems, and extreme problems). The number and percentage of patients in each response category will be tabulated by visit by treatment group. Percentages are based on the number of patients for whom an assessment is provided at the respective visit.

The EQ visual analogue scale is a visual scale from 0-100 to record a respondent's overall self-rated health state. The respondent is asked to mark an 'X' on the scale then record the corresponding number; 0 refers to the worst possible health state, 100 refers to the best possible

health state. The EQ VAS will be summarized for the observed response and change from baseline by visit by treatment group.

The change from baseline in EQ-VAS score at Week 25 will be analyzed using similar MMRM models as for BPI. Missing data will not be imputed. The least squares mean (LSMEAN) for each treatment, and treatment difference (ABSK021 – Placebo) in LSMEAN along with its' 95% CI and p-value will be provided for each scheduled visit.

6.6. Final Efficacy Analyses at the End of Study Part 2

6.6.1. Long-Term ORRs and DORs

To evaluate the treatment effect beyond 24 weeks, all three ORRs will be thoroughly analyzed based on the ITT analysis set.

- BIRC-ORR based on RECIST v1.1
- BIRC-ORR based on TVS
- Investigator-ORR based on RECIST v1.1

To minimize biased evaluations beyond Part 1, the BIRC and Investigator will remain blinded to the treatment assignment after unblinding study for the Part 1 primary analysis.

The BIRC will provide two sets of tumor response assessments based on the Part 1 and Part 2 baselines for all patients irrespective of treatment assignment in Part 1, respectively. In the case of patients who have been switched to the ABSK021 treatment after receiving a placebo, their Part 2 BIRC-ORR evaluations based on their Part 2 baseline tumor assessments will be included in the analysis of ORR. On the other hand, the patients who were randomized to ABSK021 in Part 1 will remain unchanged.

However, investigators will evaluate the patient's tumor responses based on the initial assessment done in Part 1, regardless of their treatment assignment. For this reason, for those patients who have been switched to the ABSK021 treatment after receiving a placebo, their overall visit response by RECIST v1.1 during subsequent tumor assessment visits after Part 1 will be determined programmatically based on the Part 2 baseline measurements of tumor size and subsequent information collected on the electronic case report form (eCRF) regarding the investigator's measurement of target lesions (TL), non-target lesions (NTLs) and new lesions. On the other hand, for patients who were randomized to ABSK021 in Part 1, investigators' evaluations of tumor responses for each visit will be used in the analysis.

These ORRs and associated BORs by BIRC and Investigator will be descriptively summarized by randomized treatment group sequence as defined in Section 4.3.2 along with their exact two-sided 95% CIs using the Clopper-Pearson method.

Corresponding DORs for above-mentioned ORRs will be derived. These DORs will be summarized in the same approach as described above.

6.6.2. Other Final Efficacy Analyses

In addition to the long-term ORR and DOR mentioned above, other assessments beyond Part 1 will be summarized for long-term efficacy at each visit until completion of the study unless otherwise specified.

- Change from baseline in relative ROM;
- Change from baseline in Worst-Stiffness-NRS scores;
- Change from baseline in BPI Worst Pain NRS scores;
- Change from baseline in PROMIS-PF scales;
- EQ-5D-5L health profile in 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression)
- Change from baseline in EQ-5D-5L VAS

If patients switched to the ABSK021 treatment group after receiving a placebo in Part 1, their baselines will be derived from the latest non-missing assessments prior to or on the date of the first dose in Part 2. Descriptive statistics will be provided for each randomized treatment group sequence as defined in Section 4.3.2. These include the number of observations, mean, standard deviation, 95%CI, median, minimum, and maximum for continuous endpoints; and frequencies and percentages with 95% CI using the Clopper-Pearson method for binary endpoints.

7. SAFETY ANALYSIS

7.1. General Consideration of Safety Analysis

Safety analyses will be conducted for each planned analysis based on the data group specified in Table 4.7.

Safety evaluations will be based on the incidence, severity, and type of AEs, and changes in the patient's ECOG, vital signs, ECG findings, clinical laboratory results (hematology, blood chemistry, urinalysis, myocardial enzyme test, coagulation, thyroid function tests), Echocardiography/MUGA, and NCI-PRO-CTCAE (Part 1 only). Missing safety data will generally not be imputed, unless otherwise specified.

Safety summaries will include only on-treatment assessments/events. An on-treatment assessment is defined as any assessment that is evaluated between the date of first dose of study drug and the date of last dose of study drug + 30 days. If the last dose date of study drug is missing, any assessment occurring after the start of study drug will be considered as on-treatment.

Safety analyses will be based on the Safety Analysis Set. All available safety data up to each planned data-cutoff date will be summarized by treatment group. This summary will include all relevant safety data obtained throughout the on-treatment period for each respective group as outlined in Table 4.7.

In the evaluation of safety data for those patients switching from the blinded Placebo group to the open-label ABSK021 group in Part 2, the baseline value will be determined based on the last non-missing measurement recorded before the date of the first ABSK021dose in Part 2.

Generally, no inferential statistical testing is planned on the safety data, unless otherwise specified. Visit-based safety summaries will present data at each scheduled analysis visit that defined in Section 4.2.4, maximum/minimum post-baseline. Any worst-case (including maximum and minimum) summaries will be based on extreme values falling in the category from both scheduled and unscheduled visits during the on-treatment period. Patients with High and Low values for Normal Ranges at any post-baseline visits for safety parameters will be counted in both the High and Low categories of "Any visit post-baseline" row of related summary tables. This will also apply to relevant Potential Clinical Importance summary tables.

Listings will display all values collected at each time point along with change from baseline for continuous variables. If the average of replicates is derived for any analysis (i.e., ECG), the corresponding listings will include both replicates and average. Listings will include both planned visit and derived analysis visit with a flag for the record included in summaries. For any incomplete AE onset or end date, the imputed date and raw date will be listed.

7.2. Adverse Events

7.2.1. Coding and Grading of AEs

AEs will be collected from the time of signed informed consent throughout the treatment period and up to 30 days after the last dose of study treatment. All AEs will be coded to primary System Organ Class (SOC) and Preferred Term (PT) according to the latest version of Medical Dictionary for Regulatory Activities (MedDRA). The severity (toxicity grades 1-5) of AEs will be graded according to the National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) version 5.0 by investigators.

7.2.2. Classifications of AEs

Treatment-emergent adverse event (TEAE) is defined as any AE that occurs after administration of the first dose of study drug and through 30 days (inclusive) after the last dose of study drug, any event that is present at baseline but worsens in severity or is considered treatment-related by the Investigator after administration of the first dose of study drug. AEs with start/end dates that are partially/completely missing will be imputed according to the specifications in Appendix 1 to classify TEAEs.

Any TEAE will be defined as drug-related for related relationship to study drug judged by the investigator. Every attempt will be made to obtain complete information for AEs regarding severity (i.e., CTCAE Grade) and relationship to drug; however, in the rare case of missing data, missing AE relationship will be classified as "Related" in summary tables. The non-imputed raw data will be presented in AE listings.

7.2.3. Adverse Events of Clinical Interest (AECI)

The AECI categories will be identified by the following search criteria per Standard MedDRA Queries (SMQs)/Company MedDRA Queries (CMQs) in Appendix 4. The study medical team will provide the preferred terms for identifying Adverse Events of Clinical Interest (AECIs)

before the primary database lock (DBL) for the end of Part 1 analysis. Any additional preferred terms identified after the primary DBL will be added as an addendum to the SAP.

7.2.4. AE Summary

Only TEAEs will be summarized by treatment group and pre-defined group (see Section 7.1 for details). All AEs will be listed with a flag of prior-treatment, TEAE, post-treatment (if any).

First, a high-level overall summary table for overall TEAEs will be produced by presenting the number and percentage of patients with at least one specific TEAE in any of the following categories:

- Any TEAEs
- Any drug-related TEAEs
- Any TEAEs with CTCAE grade >=3
- Any drug-related TEAEs with CTCAE grade >=3
- Any TEAEs resulting in death
- Any drug-related TEAEs resulting in death
- Any serious TEAEs
- Any drug-related serious TEAEs
- Any TEAEs leading to dose reduction
- Any TEAEs leading to dose interruption
- Any TEAEs leading to dose reduction/interruption
- Any TEAEs leading to discontinuation of treatment
- Any drug-related TEAEs leading to discontinuation of treatment

For overview summary of TEAE, the Exposure-Adjusted-Incidence-Rate (EAIR) will also be provided for each item. The EAIR per 100 patient year is defined as follows:

- $EAIR = 100 * \frac{n_{TEAE}}{(\sum_{i=1}^{n_{TEAE}} t_i + \sum_{j=1}^{n_{noTEAE}} t_j)/365.25}$
- $t_i = (date\ of\ first\ occurrence - date\ of\ first\ exposure) + 1$

If cutoff date is earlier than end of treatment:

- $t_j = (earliest\ of\ (date\ of\ last\ exposure/cutoff + 30, date\ of\ death) - date\ of\ first\ exposure) + 1$

If cutoff date is later than end of treatment:

- $t_j = (earliest\ of\ (date\ of\ last\ exposure + 30, date\ of\ death, date\ of\ cutoff) - date\ of\ first\ exposure) + 1$

Where n_{TEAE} is the number of patients with specific AE. t_i is the exposure time up to the first occurrence of event for these patients with specific AE. n_{noTEAE} is the number of patients who did not experience specific AE and t_j is the time at risk for them.

The EAIR and its 95% Confidence Intervals (CIs) [if 4 or more patients with specific TEAE in any of the treatment groups] will be presented (Xin He, et. al., 2015)[13].

In addition, the number of events will be summarized for the following AE categories:

- Serious TEAEs including incidence of SAEs based on individual components of the SAE criteria
- TEAE leading to discontinuation of study drug
- TEAE leading to dose modification, including subcategories of dose interruption, and dose reduction

TEAEs in selected categories above will be summarized by presenting the number and percentage of patients having at least one TEAE in each primary SOC or each PT as outlined under the table below. A patient with multiple occurrences of an TEAE associated with a specific SOC or PT within a SOC will be counted only once within the SOC or PT within a SOC, respectively.

For TEAE summaries presented by CTCAE grade, a patient with multiple grades for a TEAE will be summarized under the maximum grade recorded for the event. For drug-related TEAE summaries, a patient with multiple causalities for a TEAE will be considered related as long as assessed as related at least on one occasion by investigators.

In TEAE summaries the primary SOC and PT within a SOC will be presented in descending order based on the incidence and PT alphabetical order across all patients. The incidence rate of AE will be calculated using the number of patients within each specified treatment group as the denominator.

The following table summarizes the planned analyses, including selective subgroup analyses.

TEAEs Tabulated by:	**Any TEAE**	**TR-TEAE**	**TESAE**	**TR-TESAE**	**AECI**
System Organ Class and Preferred Term	▲[a, b, d]	▲[a, b, d]	▲	▲	
Preferred Term	▲[a,c]	▲[a,c]			▲[a]
PT and Maximum CTCAE Grade	▲	▲			▲

TR-TEAE: treatment-related TEAE; TESAE: Treatment emergent SAE; TR-TESAE: treatment-related TESAE.
[a] Repeat for: of grade 3 or higher
[b] Repeat for: leading to discontinuation of treatment, dose reduction/dose interruption/modification.
[c] Repeat for: Resulting in death.
[d] Selective Subgroups

In addition, summaries of most common TEAEs and most common TEAEs with any CTCAE grade 3 or 4, showing all events that occur in at least 5% of patients in any treatment group will be summarized by preferred term. This 5% cut-off may be modified after review of the data.

For AECI, an overview summary table will be provided including number and percentage of patients having at least one AECI event, at least one AECI with maximum CTCAE grade, at least one ongoing AECI event, at least one resolved AECI event, at least one serious/non-serious AECI event and at least one action taken (dose reduction/interruption, discontinuation of treatment) to AECI events. Time to first onset any grade / grade >=3 and resolution of AECI events will be summarized descriptively.

The following AE listings will be flagged with TEAE, prior-treatment or post-treatment.

- All serious AEs

- AEs leading to discontinuation of any study drug
- AEs leading to deaths
- All AECIs

Additionally, overdose (MedDRA HLT "Overdoses NEC") and medication errors (MedDRA SMQ (narrow) "Medication errors") will be summarized by PT. Drug interactions (MedDRA HLT "Interactions" will be summarized by PT. Drug abuse (MedDRA SMQ (narrow) "Drug abuse and dependence") will be summarized by PT. Withdrawal and rebound includes MedDRA HLT "Withdrawal and rebound effects".

7.3. Clinical Laboratory Evaluations

7.3.1. Laboratory Parameters

Clinical laboratory tests include hematology, blood chemistry, urinalysis, myocardial enzymes, coagulation, thyroid function test for Parts 1 and 2 and virology (screening only). Clinical laboratory tests will be measured according to Schedule of Study Assessments in protocol.

Laboratory data from all local laboratories will be included in the analyses. The default convention for reporting of laboratory units will be standard international (SI) units. If a lab value is reported using an inequality symbol e.g., less than (<) a certain value, or greater than (>) a certain value, the given numeric value will be used in the summary. Data will be presented in listings with their inequality symbol.

All lab results will be graded by the low/normal/high classifications based on each specific laboratory normal reference ranges. All applicable lab results will be graded to toxicity grades according to NCI CTCAE v5.0 when possible. Both scheduled and unscheduled during on-treatment period assessments will be used to identify the maximum post-baseline toxicity grade. Any laboratory parameter for which CTCAE toxicity grades' criteria include clinical inputs will be graded based on numerical criteria.

For continuous parameters, changes from baselines in hematology, blood chemistry, myocardial enzymes, coagulation, and thyroid function tests will be calculated. The maximum and minimum post-baseline will be derived over the entire on-treatment post-baseline period. Urinalysis data will be categorized as negative (0), positive (+), or strongly positive (++, +++, or >+++) at each timepoint.

For the liver function tests AST, ALT, ALP, total bilirubin (TBL), the fold of the local lab upper limit of the normal (ULN) range will be calculated for each data point as Fold = Value / ULN.

7.3.2. Laboratory Data Summary

All laboratory data will be included in the analyses. The default convention for reporting of laboratory units will be standard international (SI) units. If a lab value is reported using an inequality symbol e.g., less than (<) a certain value, or greater than (>) a certain value, the given numeric value will be used in the summary.

In general, summaries will include all laboratory assessments during the on-treatment period. The minimum and maximum values and their corresponding change from baseline for each

patient over the entire treatment period for each hematology and chemistry laboratory parameter will also be derived.

Observed values and changes from baseline from continuous laboratory parameters will be summarized by treatment groups for each scheduled analysis visit, overall maximum and minimum post-baseline during the on-treatment period. If a patient has multiple laboratory assessments in an interval of interest, the maximum CTCAE grade will be reported.

Shift summaries in CTCAE grade from baseline to the worst post-baseline value by treatment groups will be produced for selected laboratory parameters, including at least hemoglobin, WBC count, ANC, absolute lymphocyte count, platelet count, AST, ALT, bilirubin, creatinine, alkaline phosphatase, INR, PTT, amylase, CK, lipase, and electrolytes (bi-directional) for overall on-treatment period. The shift tables will present baseline and worst post-baseline value (for bi-directional laboratory tests, summarization will be according to specified direction), as applicable for each parameter and will include patients with both baseline and at least one post-baseline data. Any assessment for which CTCAE toxicity grades are not available, will not be included in any analyses for which toxicity grades are required. Instead, these parameter results will be graded by clinically significant results based on the investigators judgements.

Laboratory parameters graded by NCI CTCAE will be summarized using shift tables. The number and percentage of patients graded according to CTCAE v5.0 will be presented as shift table of:

- Patients with Baseline versus their worst post-Baseline grade (for bi-directional laboratory tests, summarization will be according to specified direction)

Any graded abnormality that occurs following the initiation of the study drug and represents at least a 1-grade increase from the baseline assessment is defined as a treatment-emergent abnormality. Patients with laboratory treatment-emergent abnormality in CTCAE grade will be summarized for all grades and grade 3 or 4. Subgroup analyses will also be performed.

Time to the first onset of laboratory treatment-emergent abnormality in CTCAE will be summarized descriptively.

To evaluate the status of missing data, for AST, ALT, alkaline phosphatase, total bilirubin, gamma-glutamyl transferase, international normalized ratio and creatinine, the figures of proportion of patients remaining with missing and existing laboratory data records will be provided. The figures evaluate the actual data obtained during the trial rather than the planned study procedures as stated in the protocol.

For hepatotoxicity screening, the following categories of abnormal hepatic laboratory values will be summarized by treatment group for any occurrence among all on-treatment period assessments, including scheduled and unscheduled values.

- Criterion 1: ALT and/or AST $\geq$ 3xULN, ALP $<$ 2xULN or missing, and Total Bilirubin $\geq$ 2xULN
- Criterion 2: ALT and/or AST $\geq$ 3xULN, and total bilirubin $\geq$ 2xULN
- Criterion 3: ALT and/or AST $\geq$ 3xULN, ALP $<$ 2xULN, and total bilirubin $\geq$ 1.5xULN
- Criterion 4: ALT $\geq$ 3xULN and ALP or Gamma-glutamyl Transferase (GGT) $\geq$2xULN and TBL $\geq$ 2xULN
- Criterion 5: ALP or GGT $\geq$ 2xULN and TBL $\geq$2xULN

For these categories, patients are counted as having any post-baseline total bilirubin equal to or exceeding 2xULN or 1.5xULN within 15 days after a post-baseline ALT or AST equal to or exceeding 3xULN, and ALP <2xULN or not. Or within 15 days after a post-baseline ALP or GGT equal to or exceeding 2xULN, and ALT ≥ 3xULN or not.

The eDish plot is intended to quickly identify cases of possible serious hepatocellular DILI, each patient is plotted based on their maximum post-baseline total bilirubin and transaminase (ALT or AST, whichever is higher). Each value is expressed as multiples of ULN on logarithmic scales. Dashed lines in this plot represent total bilirubin and transaminase cutoffs of 2xULN and 3xULN, respectively and are based on the criterion 1 above.

A line plot of liver function tests, including ALT, AST, ALP and total bilirubin, over time will also be presented for each individual patient who meet the criterion 1 above.

In addition, the following categories of liver function abnormality levels will be derived and summarized by treatment groups. No subgroup analysis will be conducted.

- For ALT, AST, ALT and/or AST: ≥ 3 and < 5×ULN, ≥ 5 and < 10×ULN, ≥ 10 ×ULN, ≥ 20 ×ULN
- Total bilirubin: ≥ 1.5 and <2×ULN, ≥ 2 and < 3×ULN, ≥ 3 ×ULN
- ALP: ≥ 1.5 and < 2×ULN, ≥ 2 and < 3×ULN, ≥ 3×ULN

And renal function abnormality levels will be summarized following the categories:

- Creatinine: ≥ 1.5 and < 2 × baseline, ≥ 2 and < 3 × baseline, ≥ 3 × baseline
- eGFR change from baseline: ≥ 25% and < 50% decrease, ≥ 50% and < 75% decrease, ≥ 75% decrease

All laboratory results will be listed with values flagged to show the corresponding CTCAE grades and the classifications relative to the laboratory reference ranges. In addition, a separate listing will be prepared for patients who met the criterion 1.

7.4. Vital Signs

7.4.1. Vital Sign Parameters

Vital signs including pulse rate, respiration rate, body temperature, systolic blood pressure (SBP), diastolic blood pressure (DBP); weight and height will be measured according to Schedule of Study Assessments in the protocol.

Change from baseline for each parameter will be derived. The potentially clinically significant findings of vital signs and weight will be defined based on criteria in Table 7.1. Both scheduled and unscheduled assessments will be used to identify the worst post-baseline values.

Table 7.1 Potentially Clinically Significant Criteria for Vital Signs and Weight

Vital Sign Parameter	Potentially Clinically Significant Criteria	
	Value Categories	Change from Baseline Categories

SBP	≥120 mmHg	Increase ≥40 mmHg
	≥ 140 mmHg	Decrease ≥40 mmHg
	≥ 160 mmHg	
	≥ 180 mmHg	
	< 90 mmHg	
	≥ 120 mmHg and increase from baseline ≥ 20 mmHg	
DBP	> 90 mmHg	Increase ≥ 20 mmHg
	> 110 mmHg	Decrease ≥40 mmHg
	≥ 120 mmHg	
	≤ 60 mmHg	
	≥ 80 mmHg and increase from baseline ≥ 20 mmHg	
Pulse Rate	≥ 110 bpm	Increase ≥40 bpm
	≤ 50 bpm	Decrease ≥40 bpm
	≥ 120 bpm and increase from baseline ≥ 15 bpm	
	≤ 50 bpm and increase from baseline ≥ 15 bpm	
Weight	n/a	Increase ≥10%
		Decrease ≥10%
Respiratory Rate	> 20 breaths/min	
	< 12 breaths/min	

Note: bpm = beats per minute

7.4.2. Vital Sign Summary

The summaries will include all vital sign data collected during the on-treatment period. Observed values and changes from baseline for each vital sign parameter will be summarized by treatment groups (Section 7.1) for each scheduled analysis visit, overall maximum and minimum post-baseline during on-treatment period.

The number and percentage of patients who met any potentially clinically significant criteria for each vital sign parameter as described in Table 7.1 will be summarized by treatment groups for the overall on-treatment period.

All observed vital sign values and change from baseline with the abnormality flag will be listed.

7.5. ECG Data

7.5.1. ECG Parameters

The standard 12-lead ECG value for heart rate, PR interval, RR interval, QRS interval, QT, QTcF, and investigator's clinical interpretation of 12-lead ECG (normal or abnormal) will be collected according to Schedule and Assessment Table in the protocol.

QTcF values collected on CRF page will not be used for analysis, instead, derived QTc values will be the basis for reporting. Corrected QTcF and QTcB will be calculated as follows from uncorrected QT and RR (converted from collected msec to sec) as follows:

Bazett's method QTcB = Uncorrected QT/ $(\text{RR Interval})^{1/2}$

Fridericia's method QTcF = Uncorrected QT/ $(\text{RR Interval})^{1/3}$

Change from baseline in each ECG parameter will be derived. ECG potentially clinically significant criteria are listed in Table 7.2 below. Both scheduled and unscheduled assessments will be used to identify the worst post-baseline values.

Table 7.2 Potentially Clinically Significant Criteria for ECG Parameters

ECG Parameter	Potentially Clinically Significant Criteria	
	Value Categories	**Change from Baseline Categories**
Heart Rate	>120 beats/min	n/a
	>120 beats/min and change from baseline > 25% increase	
	<50 beats/min	n/a
	<50 beats/min and change from baseline > 25% decrease	
RR Interval	>1200 msec	n/a
	<500 msec	n/a
PR Interval	≥ 210 msec	n/a
	⩾ 210 msec and change from baseline > 25% increase	
QRS	≥ 120 msec	n/a

ECG Parameter	Potentially Clinically Significant Criteria	
	Value Categories	**Change from Baseline Categories**
	≥ 120 msec and change from baseline > 25% increase	
	≤ 50 msec	n/a
QT Interval	≥ 500 msec	n/a
	≤ 300 msec	n/a
Calculated QTcB & QTcF Intervals	≥ 500 msec	Increase ≥ 30 msec
	≥ 480 and < 500 msec	Increase ≥ 30 and < 60 msec
	≥ 450 and < 480 msec	Increase ≥ 60 msec
	≤ 300 msec	
	≥ 500 msec and increase from baseline ≥ 60 msec	

7.5.2. ECG Summary

The summaries will include ECG heart rate, PR interval, RR interval and QT collected from CRF and derived QTcB and QTcF during the on-treatment period.

Observed values and changes from baseline for each ECG parameter will be summarized by treatment groups (Section 7.1) for each scheduled analysis visit, overall maximum and minimum post-baseline during on-treatment period.

The number and percentage of patients who met any potentially clinically significant criteria for each ECG parameter as described in section above will be summarized by treatment groups for the overall treatment period.

A shift table will be produced to display investigator's clinical interpretation of the 12-lead ECG normal, abnormal – not clinically significant, abnormal – clinically significant and not done. The shift tables will present baseline and worst post-baseline observation including unscheduled values during the overall treatment period.

All ECG parameters ECG heart rate, PR interval, RR interval and QT collected from CRF and derived QTcB and QTcF and change from baseline with the abnormality flags along with investigator's clinical interpretation will be listed.

7.6. ECOG Data

The Eastern Cooperative Oncology Group (ECOG) performance score (0-5) will be collected according to Schedule and Assessment Table in the protocol.

The number and percentage of patients will be summarized for each score by schedule analysis visit. A shift table presenting baseline to worst post-baseline will also be prepared.

A listing of ECOG scores for all patients including change from baseline at each planned timepoint will be provided.

7.7. NCI-PRO-CTCAE Items (Part 1 only)

NCI-PRO-CTCAE is a patient-reported outcome (PRO) measurement system developed to evaluate symptomatic toxicity in patients. Ten relevant NCI-PRO-CTCAE items which represent 6 symptomatic toxicities drawn from the CTCAE were included in this study based on the frequency of symptomatic AEs observed in early clinical data. Each AE includes up to 3 discrete questions, separately representing the frequency, severity, interference, amount and/or presence with daily activities of the event (Table 7.3). In addition, a free-text question is added as the last question to mitigate concerns for missing important symptom items.

Table 7.3 PRO-CTCAE Attributes, Item Structures, and Item Scores

Frequency	Severity	Interference	Amount	Presence / Absence
In the last 7 days, how often did you have ____?	In the last 7 days, what was the severity of your ____ at its worst?	In the last 7 days, how much did ____ interfere with your usual or daily activities?	In the last 7 days, did you have any ____?	In the last 7 days, did you have any ____?
• Never (0) • Rarely (1) • Occasionally (2) • Frequently (3) • Almost constantly (4)	• None (0) • Mild (1) • Moderate (2) • Severe (3) • Very severe (4)	• Not at all (0) • A little bit (1) • Somewhat (2) • Quite a bit (3) • Very much (4)	• Not at all (0) • A little bit (1) • Somewhat (2) • Quite a bit (3) • Very much (4)	• No (0) • Yes (1)

PRO-CTCAE responses are scored from 0 to 4 (or 0/1 for Presence/Absence). Specifically, if the response provided to the first question in a PRO-CTCAE item set is the lowest response on the scale (e.g., 'Never' for frequency, or 'None' for severity), scores for the conditionally branched PRO-CTCAE items should be scored/coded as 0 and are not to be coded as missing values.

The number and percentage of patients reporting each response category (i.e., None, Mild, Moderate, Severe, and Very severe for Severity) ratings within each PRO-CTCAE item will be

tabulated by treatment group. To graphically visualize the PRO-CTCAE item, butterfly charts (as included in FDA's Project Patient Voice[14] will be produced at each scheduled visit.

For items that are rated on a 5-point Likert scale (i.e., excluding Presence/Absence), the following displays will be prepared by treatment group:

- The raw rating and change from baseline by category will be summarized at each schedule analysis visit.
- The maximum baseline-adjusted response ratings with each PRO-CTCAE item will be summarized by treatment groups. Baseline-adjusted PRO-CTCAE values are the observed value at the visit if higher (worsen) than the baseline value, and 0 otherwise.

Patient disposition for each PRO-CTCAE item will be tabulated and plotted based on completion rate which will be calculated where the denominator is the number of patients expected to complete the PRO measure at the designated PRO assessment timepoint. Summaries for completion rate and missing data rate will include reasons for missing data.

In addition, summary tables and corresponding plots will provide for individual item and/or summary scores, as appropriate.

7.8. Other Safety Assessments

Echocardiography/ MUGA will be summarized by treatment groups. Observed values and changes from baseline for continuous variables will be summarized by treatment groups (Section 7.1) for each scheduled analysis visit, overall maximum and minimum post-baseline during on-treatment period, and last-value-on-treatment.

Number and percentage of patients will be summarized for echocardiography/MUGA overall interpretation by schedule analysis visit.

The pregnancy test will be listed.

8. PHARMACOKINETIC DATA ANALYSES

Pharmacokinetics will be evaluated for the pharmacokinetic analysis set and conducted in randomized ABSK021 treatment group and placebo switched to ABSK021 treatment group under the guidance of Abbisko Clinical Pharmacology Group. Below analysis will be performed for ABSK021 and its metabolite.

Concentration-time data will be listed to include both actual and planned time points, deviation time between actual and planned time, and actual blood sampling time relative to dose.

Descriptive summary of concentrations-time data will be provided for each nominal time including number of patients with available data (n), number of subjects with imputed value (BQL assigned to zero); arithmetic mean (Mean), standard deviation (SD); coefficient of variation (%CV); median (Median), maximum (Max), minimum (Min) and geometric mean (Geo Mean), geometric %CV.

PK concentrations will also be presented in the by-patient data listings.

9. CHANGE FROM ANALYSIS PLANNED IN PROTOCOL

Plan specified in protocol	Change in SAP and rationales
NA	NA

10. REFERENCE

1. Tap WD, Gelderblom H, Palmerini E, et al. Pexidartinib versus placebo for advanced tenosynovial giant cell tumour (ENLIVEN): a randomised phase 3 trial. Lancet. 2019;394(10197):478–487.
2. Ramalingam, S. S., Kummar, S., Sarantopoulos, J., Shibata, S., LoRusso, P., Yerk, M., ... & Egorin, M. J. (2010). Phase I study of vorinostat in patients with advanced solid tumors and hepatic dysfunction: a National Cancer Institute Organ Dysfunction Working Group study. Journal of clinical oncology, 28(29), 4507.
3. Lydersen S, Laake P. Power comparison of two-sided exact tests for association in 2×2 contingency tables using standard, mid p, and randomized test versions. Statist. Med. 2003; 22:3859–3871
4. Lydersen S, Fagerland MW, Laake P. Recommended tests for association in 2 x 2 tables. Statist Med. 2009;28:1159–75.
5. Kim Y, Won S. (2013) Adjusted proportion difference and confidence interval in stratified randomized trials. PharmaSUG; Paper SP-04.
6. Gerhardt JJ, Cocchiarella L, Lea RD. The Practical Guide to Range of Motion Assessment. Chicago, IL. Amer Med Assoc Press; 2002.
7. Gelhorn HL, Ye X, Speck RM, et al. The measurement of physical functioning among patients with Tenosynovial Giant Cell Tumor (TGCT) using the Patient-Reported Outcomes Measurement Information System (PROMIS). Journal of Patient-reported Outcomes. 2019 Feb;3(1):6. DOI: 10.1186/s41687-019-0099-0. PMID: 30714073; PMCID: PMC6360193.
8. Speck RM, Ye X, Bernthal NM, Gelhorn HL. Psychometric properties of a custom Patient-Reported Outcomes Measurement Information System (PROMIS) physical function short form and worst stiffness numeric rating scale in tenosynovial giant cell tumors. J Patient Rep Outcomes. 2020 Jul 16;4(1):61. doi: 10.1186/s41687-020-00217-6. PMID: 32676941; PMCID: PMC7366525.
9. Van De Sande M, Tap WD, Gelhorn HL, Ye X, Speck RM, Palmerini E, Stacchiotti S, Desai J, Wagner AJ, Alcindor T, Ganjoo K, Martín-Broto J, Wang Q, Shuster D, Gelderblom H, Healey JH. Pexidartinib improves physical functioning and stiffness in patients with tenosynovial giant cell tumor: results from the ENLIVEN randomized clinical trial. Acta Orthop. 2021 Aug;92(4):493-499. doi: 10.1080/17453674.2021.1922161. Epub 2021 May 12. PMID: 33977825; PMCID: PMC8382018.
10. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, Haythornthwaite JA, Jensen MP, Kerns RD, Ader DN, Brandenburg N, Burke LB, Cella D, Chandler J, Cowan P, Dimitrova R, Dionne R, Hertz S, Jadad AR, Katz NP, Kehlet H, Kramer LD, Manning DC, McCormick C, McDermott MP, McQuay HJ, Patel S, Porter L, Quessy S, Rappaport BA, Rauschkolb C, Revicki DA, Rothman M, Schmader KE, Stacey BR, Stauffer JW, von Stein T, White RE, Witter J, Zavisic S. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain. 2008 Feb;9(2):105-21. doi: 10.1016/j.jpain.2007.09.005. Epub 2007 Dec 11. PMID: 18055266.
11. Brookmeyer R, Crowley J. (1982). A confidence interval for the median survival time. Biometrics. 29-41.
12. Kalbfleisch, J.D. and Prentice, R.L. (1980) The statistical analysis of failure time data. John Wiley and Sons, New York.

13. He X, Chen L, Lei L, Xia HA, Lee MLT (2015) A Simple Method for Estimating Confidence Intervals for Exposure Adjusted Incidence Rate and Its Applications to Clinical Trials. J Biom Biostat 6: 238.
14. U.S. Food and Drug Administration. FDA perspective: patient self-reporting in the evaluation of cancer drug tolerability. https://www.fda.gov/media/148468/download. Accessed January 18, 2024.

11. APPENDICES

11.1. Appendix 1: Partial Date Imputation Rule

Completely Missing or Partially Missing AE Date

AE start date

- If the start date has month and year but day is missing, the first dose date will be used if the month and year is the same as the first dose date and the end date is greater than the first dose or completely missing, otherwise, the first day of the month will be used.
- If the start date has year, but day and month are missing, then the first dose date will be used if the year is the same as the first dose date and the end date is greater than the first dose or completely missing, otherwise 1st January will be used.
- If the start date is completely missing, then it will be imputed as the first dose date of study drug if the end date is equal or greater than the first dose date or completely missing, otherwise ICF signed date will be used.

After imputation, the imputed start date will be compared with ICF signed date. If the imputed start date is earlier than ICF signed date, the start date will be imputed with ICF signed date instead.

AE end date

- If the end date has month and year but day is missing, the last day of the month will be used.
- If the end date has year, but day and month are missing, the 31st December will be used.
- If the end date is completely missing, the last dose date + 30 days will be used if the AE occurs or worsens after the first dose date of study drug; If the AE occurs prior to the first dose date of study drug, the end date will be imputed with the first dose date.

After the imputation, the imputed AE end date will be compared against the AE start date and the death date for patients who died. If the imputed AE end date is earlier than the start date, then the AE start date will be used to impute the end date instead. If the date is later than the death date, the date of death will be used to impute the end date instead.

Completely Missing or Partially Missing Medication/Non-Medication Therapy Date

Start date

- If the start date has month and year but day is missing, the first dose date will be used if the month and year is the same as the first dose date and the end date is greater than the first dose or completely missing, otherwise, the first day of the month will be used.
- If the start date has year, but day and month are missing, then the first dose date will be used if the year is the same as the first dose date and the end date is greater than the first dose or completely missing, otherwise 1st January will be used.
- If the start date is completely missing, then it will be imputed as the first dose date of study drug if the end date is equal or greater than the first dose date or completely missing, otherwise ICF signed date will be used.

After the imputation, the imputed start date will be compared with ICF signed date, if available. If the imputed start date is earlier than ICF signed date and the collected partial part is the same as ICF signed date, the start date will be imputed with ICF signed date instead.

End date

- If the end date has month and year but day is missing, the last day of the month will be used.
- If the end date has year, but day and month are missing, the 31st December will be used.
- If the end date is completely missing, the last dose date + 30 days will be used if the medication/non-medication therapy is taken after the first dose date of study drug; If it is taken prior to the first dose date of study drug, the end date will be imputed with the first dose date.

After the imputation, the imputed end date will be compared against the start date and the death date for patients who died. If the imputed end date is earlier than the start date, then the start date will be used to impute the end date instead. If the date is later than the death date, the date of death will be used to impute the end date instead.

Partially Missing Initial Diagnosis Date

- If the date has month and year but day is missing, the first day of the month will be used.
- If the date has year, but day and month are missing, 1st January will be used.

Completely Missing or Partially Missing Subsequent Anti-Cancer Therapy Date

- If the start date has month and year but day is missing, the first day of the month will be used if the partial part is later than the earliest of the end date and the later one among the first PD date + 1 and the last dose date +1, or the last day of the month will be used if the partial part is earlier than the earliest of the end date and the later one among the first PD date + 1 and the last dose date +1.
- If the start date has year, but day and month are missing, 1st January will be used if the partial part is later than the earliest of the end date and the later one among the first PD date + 1 and the last dose date +1, or 31st December will be used if the partial part is earlier than the earliest of the end date and the later one among the first PD date + 1 and the last dose date +1.

- If the start date is completely missing or the partial part is as same as the earliest of the end date and the later one among the first PD date + 1 and the last dose date +1, then it will be imputed as the earliest of the end date and the later one among the first PD date + 1 and the last dose date +1.

Completely Missing or Partially Missing Death Date

- If the date has month and year but day is missing, the first day of the month will be used if the partial part is greater than last survival date.
- If the date has year, but day and month are missing, 1st January will be used if the partial part is greater than last survival date.
- If the date is completely missing or the partial part is as same as the last survival date, then use the last survival date + 1 to impute.

11.2. Appendix 2: Precision Format

Rounding conventions for presentation of summary statistics will be based on the precision of the variable (n) of summarization, as it is collected in its rawest form (i.e. on the electronic case report form [eCRF] or as provided within an external file) and are outlined in the table below with the additional rule of no more than 4 decimal places, and will not happen before reporting. Derived time related variable will keep two decimal places. Other derived data has one more decimal place than the original data.

Descriptive statistics for derived data follow the principles below.

Table 11.1 Summary Table of Precision Format

Statistics	Decimal Digits
Mean / Median	n +1
SD / SE	n +2
Minimum / Maximum	n
%-value	1; If the percentage is 0, then no percentage is given; If the percentage is 100, the percentage remains an integer
Geometric Mean	n + 1
Coefficient of variation (%CV)	1
Ratio of Geometric mean (GMR)	2
Confidence interval for GMR	2
P-value	4; if p-value < 0.0001 then display it as '<0.0001'; p-value > 0.9999 then display it as '> 0.9999'

For pharmacokinetic parameters descriptive statistics values up to, but not including 1000, display to 3 significant figures and when values >= 1000, the whole integer is reported.

t_{max} is to be reported to 2 decimal places, half-life is to be reported to 1 decimal place.

11.3. Appendix 3: PROMIS- PF Scoring

The Health Measures Scoring Service (HM-SS) will be used to calculate PROMIS-PF T-score based on individual item raw scores. The HM-SS is a software application which automates and facilitates IRT calculated scoring of short form data from the four HealthMeasures measurement systems including PROMIS. The HM-SS accepts a user's data file and produces an output file containing IRT scores, i.e. Theta, T-score and standard error (SE). It is located at: https://www.assessmentcenter.net/ac_scoringservice .

A simple registration is required prior to the first use of the HM-SS. In the upper right-hand corner of the HM-SS homepage there are two hyperlinks which allow users to access the HM-SS user manual and input file template. The Input file template should be used as guide of how users should structure their dataset. Please follow the several steps and selections in Table 11.2 prior to upload of a data input file.

Table 11.2 Upload Procedure in HealthMeasures Scoring Service

Step 1: Type of Short Form Selection	***Selection***
What type of short form instrument will you be scoring during this upload?	Custom Short Form
Step 2: Instrument Search	
Select Measurement System	PROMIS
Select Respondents	Adult
Select Domain	Physical Function
Step 3: Original Instrument/ Calibration Sample Selection	
Select your item bank from which items originated	PROMIS Bank v2.0 - Physical Function
Select calibration sample to be used during scoring:	PROMIS 1 Wave 1 with Extension (default)
Step 4: Enter File Name and Verify Email Address	
Step 5: File selection & Upload	

The final step is to upload the input file. The file type must be a .csv. CSV stands for comma separated values. It is important to follow the Input File Template (Table 11.3) from the HM-SS homepage.

Table 11.3 Input File Template

PIN	Assmnt	<Insert Item ID>	<Insert Item ID>	<Insert Item ID>
<Insert PIN>	<insert Assmnt value>	<insert response score>	<insert response score>	<insert response score>
<Insert PIN>	<insert Assmnt value>	<insert response score>	<insert response score>	<insert response score>
Example (delete these rows before uploading input file):				
PIN	Assmnt	EDANX01	EDANX02	EDANX05
1400018	1	2	2	3
1400019	1	2	1	2

11.4. Appendix 4: Adverse Events of Clinical Interest (AECI)

AECI Categories	Criteria
Elevation of aminotransferase	MedDRA PTs: Alanine aminotransferase abnormal, Alanine aminotransferase increased, Aspartate aminotransferase abnormal, Aspartate aminotransferase increased, AST/ALT ratio abnormal, Hepatic enzyme abnormal, Hepatic enzyme increased, Hepatic function abnormal, Hypertransaminasaemia, Liver function test abnormal, Liver function test increased, Transaminases abnormal, Transaminases increased, AST to platelet ratio index increased, Mitochondrial aspartate aminotransferase increased.
Hepatotoxicity	MedDRA SMQ (narrow): Drug related hepatic disorders – comprehensive search, excluding PTs in AECI of Elevation of aminotransferase. MedDRA SMQ (narrow): Biliary tract disorders
Elevation of blood creatine phosphokinase	MedDRA PTs: Blood creatine phosphokinase increased Blood creatine phosphokinase abnormal Blood creatine phosphokinase BB increased

	Blood creatine phosphokinase MB abnormal Blood creatine phosphokinase MB increased Blood creatine phosphokinase MM increased
Myopathy and Myocardial injury	MedDRA SMQ (narrow): Rhabdomyolysis/myopathy MedDRA SMQ (narrow): Noninfectious myocarditis/ pericarditis MedDRA SMQ (narrow): Myocardial infarction
Elevation of pancreatic enzymes	MedDRA PTs: Amylase abnormal, Amylase increased, Lipase abnormal, Lipase increased, pancreatic enzymes abnormal, pancreatic enzymes increased, Lipase urine increased, Hyperamylasaemia, Hyperlipasaemia, Pancreatic enzyme abnormality
Renal impairment	MedDRA SMQ: Acute renal failure (narrow) MedDRA SMQ: Chronic kidney disease (narrow)
PERIORBITAL OEDEMA*	MedDRA PTs: Eye oedema, Eye swelling, Eyelid oedema, Orbital oedema, Orbital swelling, Periorbital oedema, Periorbital swelling, Swelling of eyelid
Oedema (excluding periorbital oedema)	MedDRA PTs: PERIPHERAL OEDEMA*: Oedema Swelling Peripheral swelling Oedema peripheral Generalised oedema Gravitational oedema Localised oedema Skin oedema FACE OEDEMA*: Face oedema Swelling face
RASH*	MedDRA HLT: Rashes, eruptions and exanthems NEC MedDRA HLT: Dermatitis and eczema
Dyslipidaemia	MedDRA SMQ (narrow): Dyslipidaemia
Torsade de pointes/QT prolongation	MedDRA SMQ (narrow and broad): Torsade de pointes/QT prolongation

- The MedDRA terms were selected based on MedDRA 27.0.

- AECI categories and search strategies may be updated based on emerging safety data from ABSK021 clinical development program or outcomes from consultations with health authority.