STATISTICAL ANALYSIS PLAN	
Study Code	D9571C00001
Edition Number	3.0
Date	25-Sep-2024

A Phase I/II Open-label, Multi-center Study to Assess Safety, Tolerability, Pharmacokinetics and Preliminary Efficacy of AZD7789, an anti-PD-1 and anti-TIM 3 Bispecific Antibody, in Patients with Relapsed or Refractory Classical Hodgkin Lymphoma

TABLE OF CONTENTS

TITLE PAGE 1
TABLE OF CONTENTS 2
LIST OF ABBREVIATIONS 5
AMENDMENT HISTORY 7
1 INTRODUCTION 8
2 CHANGES TO PROTOCOL PLANNED ANALYSES 8
3 DATA ANALYSIS CONSIDERATIONS 8
3.1 Timing of Analyses 8
3.2 Analysis Populations 9
3.3 General Considerations 11
3.3.1 General Study Level Definitions 11
3.3.2 Visit Window 13
3.3.3 Handling of Unscheduled Visits 14
3.3.4 Multiplicity/Multiple Comparisons 14
3.3.5 Handling of Protocol Deviations in Study Analysis 14
3.3.6 Missing Dates 15
3.3.7 Sample Size 16
3.3.7.1 Dose escalation: mTPI-2 (Part A) 17
3.3.7.2 Accelerated Titration Design (ATD) 17
3.3.7.3 Part B Dose Expansion 18
4 STATISTICAL ANALYSIS 18
4.1 Study Population 18
4.1.1 Participant Disposition and Completion Status 18
4.1.1.1 Definitions and Derivations 18
4.1.1.2 Presentation 19
4.1.2 Analysis Sets 19
4.1.2.1 Definitions and Derivations 19
4.1.2.2 Presentation 19
4.1.3 Protocol Deviations 20
4.1.3.1 Definitions and Derivations 20
4.1.3.2 Presentation 20
4.1.4 Demographics 20
4.1.4.1 Definitions and Derivations 20
4.1.4.2 Presentation 20
4.1.5 Baseline Characteristics 20
4.1.5.1 Definitions and Derivations 20
4.1.5.2 Presentation 20
4.1.6 Disease Characteristics 21
4.1.6.1 Definitions and Derivations 21
4.1.6.2 Presentation 21
4.1.7 Medical History 21

4.1.7.1 Definitions and Derivations 21
4.1.7.2 Presentation 21
4.1.8 Prior and Concomitant Medications/Procedures 22
4.1.8.1 Definitions and Derivations 22
4.1.8.2 Presentation 22
4.2 Endpoint Analyses 23
4.2.1 Primary Safety Endpoints: AEs, immune mediated AEs and serious AEs, laboratory evaluations, vital signs, ECG results Definition and DLTs 26
4.2.2 Primary Endpoint: ORR 26
4.2.2.1 Definition and Derivations 26
4.2.2.2 Primary Analysis of Primary Endpoint: ORR 27
4.2.2.3 Sensitivity Analysis of ORR Endpoint 27
4.2.2.4 Supportive Secondary Analysis of ORR Endpoint 27
4.2.2.5 Supportive Secondary Analysis: Change in Target Lesion Tumor Size 28
4.2.2.6 Subgroup Analyses 28
4.2.3 Primary Endpoint: Complete Response Rate (CRR) 29
4.2.3.1 Definition and Derivations 29
4.2.3.2 Primary Analysis of Primary Endpoint: CRR 29
4.2.3.3 Sensitivity Analysis of Primary Endpoint: CRR 29
4.2.3.4 Supportive Analysis of CRR Endpoint 29
4.2.3.5 Subgroup Analyses 29
4.2.4 Secondary Endpoint: Duration of Response (DoR) 30
4.2.4.1 Definition 30
4.2.4.2 Derivations 30
4.2.4.3 Analysis of Secondary Endpoint: DoR 30
4.2.4.4 Supportive Analysis of DoR Endpoint 30
4.2.5 Secondary Endpoint: Duration of Complete Response (DoCR) 30
4.2.5.1 Definition 31
4.2.5.2 Derivations 31
4.2.5.3 Analysis of Secondary Endpoint: DoCR 31
4.2.5.4 Supportive Analysis of DoCR Endpoint 31
4.2.6 Secondary Endpoint: Progression-free Survival (PFS) 31
4.2.6.1 Definition 32
4.2.6.2 Derivations 32
4.2.6.3 Analysis of Secondary Endpoint: PFS 34
4.2.6.4 Supportive Analysis of PFS Endpoint 34
4.2.7 Secondary Endpoint: Overall Survival (OS) 34
4.2.7.1 Definition 34
4.2.7.2 Derivations 34
4.2.7.3 Handling of Dropouts and Missing Data 36
4.2.7.4 Primary Analysis of Secondary Endpoint (Overall Survival) 36
4.2.7.5 Sensitivity Analysis of OS Endpoint 36
4.2.8 Other Endpoints 37
4.2.9 Pharmacodynamic Endpoint 37
4.3 Pharmacokinetics 37
4.3.1 Serum Pharmacokinetic Concentration Data 37

4.3.2 Serum Pharmacokinetic Parameter Data 38
4.3.3 Graphical presentation of Pharmacokinetic Concentration and Parameter Data 40
4.3.4 Precision and Rounding Rules for Pharmacokinetic Concentration and Parameter Data 41
4.4 Immunogenicity 41
4.5 Safety Endpoints Analyses 42
4.5.1 Exposure 43
4.5.1.1 Definitions and Derivations 43
4.5.1.2 Presentation 44
4.5.2 Adverse Events 44
4.5.2.1 Definitions and Derivations 44
4.5.2.2 Presentation 47
4.5.3 Deaths 48
4.5.3.1 Presentation 48
4.5.4 Clinical Laboratory, Blood Sample 48
4.5.4.1 Definitions and Derivations 48
4.5.4.2 Presentations 49
4.5.5 Clinical Laboratory, Urinalysis 49
4.5.5.1 Definitions and Derivations 49
4.5.5.2 Presentations 49
4.5.6 Vital Signs 49
4.5.6.1 Definitions and Derivations 49
4.5.6.2 Presentations 50
4.5.7 Electrocardiogram 50
4.5.7.1 Definitions and Derivations 50
4.5.7.2 Presentations 51
4.5.8 Echocardiogram/Multiple-gated Acquisition Scan (ECHO/MUGA) 52
4.5.8.1 Definitions and Derivations 52
4.5.8.2 Presentations 52
4.5.9 B-symptoms 52
4.5.9.1 Definitions and Derivations 52
4.5.9.2 Presentations 52
4.5.10 Other Safety Assessments 52
4.5.10.1 Definitions and Derivations 52
4.5.10.2 Presentations 53
5 INTERIM ANALYSES 53
6 REFERENCES 55
7 APPENDICES 56

LIST OF ABBREVIATIONS

Abbreviation or Specialized Term	Definition
AE	Adverse event
AESIs	Adverse event of special interest
ATC	Anatomical Therapeutic Chemical
BICR	Blinded independent central review
BMI	Body mass index
BOR	Best overall response
BSR	Baseline scaled ratio
CI	Confidence interval
COVID-19	Coronavirus disease 2019
CPS	Clinical Pharmacology Scientist
CRF	Case Report Form
CRR	Complete Response Rate
CSP	Clinical Study Protocol
CSR	Clinical Study Report
CTCAE	Common Terminology Criteria for AEs
CV	Coefficient of variation
DCO	Data cut-off
DLT	Dose-limiting toxicity
DoCR	Duration of Complete Response
DoR	Duration of Response
ECG	Electrocardiogram
ECHO	Echocardiogram
ECOG	Eastern Cooperative Oncology Group
FAS	Full analysis set
ICF	Informed consent form
imAE	Immune-mediated adverse event
IP	Investigational Product
IPD	Important Protocol Deviation
ITT	Intention-To-Treat
LLOQ	Lower limit of quantification
LRV	Lower reference value
LVEF	Left ventricular ejection fraction

MedDRA	Medical Dictionary for Regulatory Activities
MR	Minor response
MRI	Magnetic resonance imaging
MTD	Maximum Tolerated Dose
mTPI	Modified toxicity probability interval
MUGA	Multiple-gated acquisition scan
OAE	Other significant adverse events
ORR	Objective Response Rate
OS	Overall Survival
PAVA	Pool adjacent violators algorithm
PD	Progressive Disease
PFS	Progression-free Survival
PK	Pharmacokinetics
PT	Preferred Term
QTcF	Corrected QT interval, using Fridericia's formula
RP2D	Recommended phase 2 dose
RECIL	Response-evaluation Criteria in Lymphoma
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SD	Stable disease
SOC	System Organ Class
Std Dev	Standard deviation
TEAEs	Treatment emergent adverse events
TV	Target value
WHO	World Health Organization

AMENDMENT HISTORY

CATEGORY Change refers to:	Date	Description of change	In line with CSP?	Rationale
N/A	Click or tap to enter a date.	Initial approved SAP	N/A	N/A
Presentation	21 Jun 2024	Updated analysis sets in section 4.1.2	Yes	CSP updated to version 4.0
Presentation	21 Jun 2024	Timing of analysis changed in section 1	Yes	CSP updated to version 5.0
N/A	02 Jul 2024	New approved version 2.0	Yes	N/A
Presentation	25 Sep 2024	Updated tumor assessment visit window	Yes	Correction
N/A	25 Sep 2024	New approved version 3.0	Yes	N/A

1 INTRODUCTION

The purpose of this document is to give details for the statistical analysis of study D9571C00001. The Statistical Analysis Plan (SAP) should not be read in isolation but in conjunction with the Clinical Study Protocol (CSP) version 5 (dated 07 May 2024) for details of study conduct.

If there is a discontinuation of the study, with no subjects starting a cohort or a part of the study, then the summary or analysis specified in this SAP for that cohort or part of the study will not be presented. The final data cutoff for the end of study and final analysis will take place when the last subject completes the last study visit. Following the end of study, subjects may transition to the post-trial access program to order to provide continued access to sabestomig (AZD7789).

2 CHANGES TO PROTOCOL PLANNED ANALYSES

Not applicable.

3 DATA ANALYSIS CONSIDERATIONS

3.1 Timing of Analyses

There are multiple planned data cut-offs (DCOs) for this study, consisting of interim analyses, a primary analysis, and a final analysis. The timing of the interim analyses may be combined where appropriate.

Refer to Section 5 for further details of the planned interim analyses.

The data cut-off for the final analysis will be taken once the last participant has transitioned to the post-trial access program.

Analysis milestone	Timing
Interim Safety Analysis (Cohort B1 & Cohort B2)	After approximately CCI participants have received the first dose at CCI prior to data-cut off. For Cohort B1, participants treated at the RP2D in Part A will be included as part of the CCI participants.
Efficacy Interim Analysis (Cohort B1)	After approximately CCI participants in each respective cohort are included in the interim response-evaluable population.

	Subsequent interim analyses may be performed after every CCI additional participants for Cohort B1 get included in the response-evaluable population.
Efficacy Interim Analysis (Cohort B2)	After approximately CCI participants are included in the interim response-evaluable population. Subsequent interim analyses may be performed after every CCI additional participants for Cohort B2 get included in the response-evaluable population.
Primary Analysis	At CCI after the final enrolled participant received first dose of sabestomig.
Final Analysis	May be performed CCI after the last participant begins treatment or when Astrazeneca stops the study, whichever occurs first.

3.2 Analysis Populations

All participants who receive any amount of study intervention will be included in the Safety analysis set. For the safety and PK analyses, participants will be classified according to dose level they actually received. For all efficacy analyses, and for baseline and demography, participants will be classified according to the dose they were assigned to (ie, the planned dose level).

The analysis sets will be summarized and any exclusions from analysis sets will be presented in data listings that will include:

- participants excluded from the analysis set and
- data excluded from any analysis (e.g.: data censored at time of intra-participant dose escalation).

The following populations are defined:

Table 1 Populations for Analyses

Population/Analysis set	Description	Endpoint/Output
Enrolled	All participants who sign the Informed Consent Form (ICF), or whose legally authorized representative[a] sign the Informed Consent Form.	Disposition
Full/Safety	All participants who receive any amount of study intervention.	Exposure Safety Baseline and demography PK concentrations and parameters Listings PFS OS
ITT	Participants assigned to study intervention.	Sensitivity analyses
Interim safety	All participants who received the first dose at least CCI to data cut-off.	Safety at interim
DLT-evaluable	Participants enrolled in the dose escalation phase who have received sabestomig and completed the DLT evaluation period (defined as 28 days after the start of study intervention) or who experienced any DLT.	DLT
Response-evaluable	All dosed participants who had measurable disease at baseline.	ORR CRR BOR Duration of response Duration of complete response
Interim response-evaluable	All dosed participants who had measurable disease at baseline and received first dose at CCI prior to data cut-off.	Efficacy at interim
PK	All participants who received at least 1 dose of study intervention with at least 1 reportable concentration.	PK concentrations PK parameters
Pharmacodynamics	All participants who received at least 1 dose of study intervention with at least 1 reportable pharmacodynamic measurement.	Pharmacodynamic endpoints
Immunogenicity	All participants who received at least 1 dose of study intervention with at least 1 reportable immunogenicity measurement.	Immunogenicity endpoints

Population/Analysis set	Description	Endpoint/Output
CCI [redacted]	CCI [redacted]	CCI [redacted]

BOR: best overall response; CRR: Complete Response Rate; DLT: dose-limiting toxicity; DoR: Duration of Response; DoCR: Duration of Complete Response; ITT: Intention to Treat; ORR: Objective Response Rate; OS: Overall Survival; PFS: Progression-free Survival; PK: pharmacokinetic(s).

[a] Legally authorized representative will be required to sign a statement of informed consent and informed assent from the participant (as appropriate) that meets the requirements of 21 CFR 50, local regulations, ICH guidelines Health Insurance Portability and Accountability Act requirements, where applicable, and the IRB/IEC or study center.

Assigned to study intervention will not include participants determined to be a screen failure.

Individual PK concentration and parameter data for any participants not included in the PK analysis set or excluded from the descriptive summary tables, figures, and/or inferential statistical analyses – eg, due to important protocol deviations that might affect PK – will be included in the listings and flagged with an appropriate footnote.

DoR is reported for the subset of participants with objective response.

DoCR is reported for the subset of participants with complete response.

Data summaries may be presented on more than one analysis population, as appropriate.

3.3 General Considerations

Unless stated otherwise, data will be presented by dose level in Part A and by cohort in Part B. For Cohort B1, participants treated at the recommended phase 2 dose (RP2D) in Part A will be included.

3.3.1 General Study Level Definitions

The general principles described below are followed throughout the study:

- Continuous endpoints will be summarized by the number of observations, mean, standard deviation (Std Dev), median, upper and lower quartiles (as applicable), minimum, and maximum. For data that requires log-transformation, it is more appropriate to present geometric mean, coefficient of variation (CV), median, minimum and maximum. Categorical endpoints are summarized by frequency counts and percentages for each category.
- If data are available for less than 3 participants, where presented, only minimum, maximum and number of observations will be presented.
- In general, summaries will be presented by study part and by dose level or cohort. Dose level groups with only single participant will only be listed.
- Unless otherwise stated, percentages will be calculated out of the analysis set total (excluding efficacy and exposure) and by dose level or cohort as appropriate. Percentages will not be presented for zero counts.

- For continuous data, descriptive summary statistics (mean, median, standard deviation, standard error, confidence intervals) will be rounded to 1 additional decimal place compared to the reported data. Minimum and maximum are displayed with the same accuracy as the reported data.
- For categorical data, percentages are rounded to 1 decimal place.
- Time to event variables will be presented using the Kaplan-Meier (KM) methodology where appropriate, including median time calculated from the KM curves.
- For summaries at the participant level, all values will be included, regardless of whether they appear in a corresponding visit-based summary.
- SAS® version 9.4 (as a minimum) will be used for all analyses. The pharmacokinetic (PK) parameters of the serum concentration data for sabestomig will be derived using non-compartmental methods in Phoenix® WinNonlin® Version 8.3 or higher (Certara).
- Baseline is the last non-missing value obtained prior to the first dose/administration of any study treatment and any information taken after first dose/administration of study treatment is regarded as post-baseline information. If two visits are equally eligible to assess participant status at baseline (eg, screening and baseline assessments both on the same date prior to first dose/administration with no washout or other intervention in the screening period), the average is taken as the baseline value. For non-numeric laboratory tests (ie, some of the urinalysis parameters) where taking an average is not possible then the best value is taken as baseline as this is the most conservative. In the scenario where there are two assessments on Day 1 prior to first dose, one with time recorded and the other without time recorded, the one with time recorded is selected as baseline. Where safety data are summarized over time, study day is calculated in relation to date of first treatment. For assessments on the day of first dose where time is not captured, a nominal pre-dose indicator, if available, serves as sufficient evidence that the assessment occurred prior to first dose. Assessments on the day of the first dose where neither time nor a nominal pre-dose indicator are captured is considered prior to the first dose if such procedures are required by the protocol to be conducted before the first dose. If no value exists before the first dose/administration, then the baseline value is treated as missing.
- In all summaries, change from baseline endpoints will be calculated as the post-treatment value minus the value at baseline. The percentage change from baseline will be calculated as (post-baseline value - baseline value) / baseline value × 100. For any endpoint subjected to log transformation, the change from baseline calculated and summarized on the log scale will be back-transformed and presented as a 'baseline scaled ratio' (BSR). Percentage change will be then calculated as (BSR - 1) × 100.
- Unless stated otherwise, two-sided confidence intervals are produced at 95%.

- For the purposes of summarizing safety data assessed at visits, in addition to baseline data, only on treatment data will be included in the summary tables. On treatment data will be defined as data after the first dose of IP and with assessment date up to and including the date of last IP + 90 days, and prior to the start of any subsequent cancer therapy.
- For efficacy assessments, figures may only be produced for dose levels or cohorts with more than 10 participants.

3.3.2 Visit Window

For safety and tumor assessments, visit windows are defined for any presentations that summarise values by visit. The following conventions apply:

For tumor assessments:

- The protocol assigned windows for tumor assessments will be used to assign the result to a particular visit.

For safety assessments:

- Visit windows will be exhaustive so that data recorded at any timepoint has the potential to be summarized. Inclusion within the visit window will be based on the actual date and not the intended date of the visit.

- All unscheduled visit data have the potential to be included in the summaries.

- The window for the visits following baseline will be constructed in such a way that the upper limit of the interval falls half way between the two visits (the lower limit of the first post-baseline visit is Day 2). If an even number of days exists between two consecutive visits then the upper limit is taken as the midpoint value minus 1 day.

- If treatment cycle delays are experienced, the planned dosing visit windows will be modified during the programming to anchor to the actual visit date where intervention was administered on to adjust for these treatment cycle delays. The same rules apply to safety visits of other types (non-dosing visits) where the safety visit is moved due to a treatment cycle delay.

- Visit windowing will be done separately for each assessment based on the schedule of events specific to that assessment. See Appendix B for details.

- For summaries showing the maximum or minimum values, the maximum/minimum value recorded on treatment will be used (regardless of where it falls in an interval).

- Listings should display all values contributing to a time point for a participant.
- For visit based summaries, if there are more than one value per participant within a visit window then the closest value to the scheduled visit date will be summarized, or the earlier, in the event the values are equidistant from the nominal visit date. If there is still a tie, preference to the nominal visit at which the results were reported at will be given. The listings will highlight the value for the participant that contributed to the summary table, wherever feasible. Note: in summaries of extreme values, all post baseline values collected will be used including those collected at unscheduled visits regardless of whether or not the value is closest to the scheduled visit date.
- For summaries at a participant level, all values will be included, regardless of whether they appear in a corresponding visit based summary, when deriving a participant level statistic such as a maximum.

3.3.3 Handling of Unscheduled Visits

Refer to Section 3.3.2.

3.3.4 Multiplicity/Multiple Comparisons

Not applicable.

3.3.5 Handling of Protocol Deviations in Study Analysis

Important protocol deviations (IPDs) are those deviations from the protocol likely to have an impact on the perceived efficacy and/or safety of study intervention.

IPDs may include, but are not limited to the following:

- Written informed consent not obtained prior to mandatory study specific procedures, sampling and analyses
- Participants who did not meet inclusion criteria or met exclusion criteria, and received study intervention
- No baseline Lugano assessments on or before the date of first dose
- Participants who received prohibited concomitant medications during study period
- Participants who met study intervention discontinuation criteria but continued study intervention, and potentially had major impact to safety of participants according to clinical judgement

Protocol deviations will be reviewed on a case-by-case basis by AstraZeneca to determine their level of importance. Deviations considered to be important will be listed and discussed

in the CSR as appropriate. All decisions on importance will be made ahead of database lock and will be documented as part of the Data Review Meeting minutes prior to the primary analysis being performed.

A detailed list of protocol deviations to be considered are defined in the study specific Protocol Deviation Specifications.

None of the deviations will lead to participants being excluded from any analysis populations described in the SAP, unless otherwise specified. If a deviation is serious enough to have a potential impact on the primary analysis, sensitivity analyses may be performed. The need for such a sensitivity analysis will be determined following review of the protocol deviations ahead of database lock and will be documented prior to the analysis being conducted.

A list of all protocol deviations will be reviewed and decisions regarding how to handle these deviations in the analyses will be documented by the study team physician, clinical pharmacology scientist and statistician prior to database lock.

3.3.6 Missing Dates

When partial dates exist in the data there are some general conventions to be applied when the month or day are missing. These are described in the PHUSE guidance on partial dates, which are outlined below:

If the whole date is missing, it is more difficult to follow a general principle and these are reviewed within the study and decided on how to be handled. General guidance for completely missing dates are provided below, but the guidance is assessed as necessary within the study.

Generally, the imputation of dates is used to decide if an observation is treatment emergent for adverse events (AEs) or concomitant medications. The imputed dates should not be used to calculate durations, where the results would be less accurate.

The following are the guidelines used when partial dates are detected in the study:

- For missing diagnostic dates (eg,: disease diagnosis), if day and/or month are missing use 01 and/or Jan. If year is missing, put the complete date to missing.
- For missing AE and concomitant medication start dates, the following is applied:
    a. Missing day - impute the 1st of the month unless month is the same as month of the first dose of study drug then impute first dose date.
    b. Missing day and month - impute 1st January unless year is the same as first dose date then impute first dose date.

c. Completely missing - impute first dose date unless the end date suggests it could have started prior to this in which case impute the 1st January of the same year as the end date.

d. Imputed start date should be no later than the end date.

- For missing AE and concomitant medication end dates, the following is applied:

a. Missing day - impute the last day of the month unless both the month and the year are the same as the last dose date or the analysis data cut-off date then impute the last dose date or the analysis data cut-off date.

b. Missing day and month - impute 31st December unless the year is the same as the last dose date or the analysis data cut-off date then impute the last dose date or the analysis data cut-off date.

c. Completely Missing – need to look at whether the AE/medication is still ongoing before imputing a date and also when it started in relation to study drug. If the ongoing flag is missing then assume that AE is still present / medication is still being taken (ie, do not impute a date). If the AE/medication has stopped and start date is prior to first dose date then impute first dose date. Or if it started on or after first dose date then impute to the last dose of study drug date + 1.

- Flags are retained in the database indicating where any programmatic imputation has been applied, and in such cases for AEs and concomitant medications, any durations would not be calculated.
- If a participant is known to have died where only a partial death date is available then the date of death is imputed as the latest of the last date known to be alive +1 from the database and the death date, using the available information provided and applying the following:
  - For Missing day only – using the 1st of the month.
  - For Missing day and Month – using the 1st of January.

3.3.7 Sample Size

Approximately 180 participants will be treated with sabestomig, with up to 52 participants in Part A (Dose Escalation) and CCI participants in Part B (Dose Expansion).

Additional participants may be enrolled if additional dose levels, expansion cohorts, treatment schedules are explored, or participants require replacement for any reason.

Further details are provided in section 9.2 of the protocol.

3.3.7.1 Dose escalation: mTPI-2 (Part A)

The cohort size (3-12) in Part A is based on mTPI-2 dose escalation method and clinical consideration. The modified toxicity probability interval-2 (mTPI-2) employs a simple beta-binomial Bayesian model (Guo et al, 2017

). The prior distribution for all dose levels is Beta(1,1). The posterior distribution for all dose levels is Beta(1+ *a*, , 1+ *b*), where *a*, and *b* are the number of participants with and without a dose limiting toxicity (DLT) at the current dose level, respectively. The posterior density of the toxicity probability is divided into multiple intervals with equal length. These intervals are categorized as underdosing (below), proper dosing (equivalent), and overdosing (above) in terms of toxicity. The underdosing interval corresponds to a dose escalation, overdosing corresponds to a dose de-escalation, and proper dosing corresponds to staying at the current dose. Given an interval and a probability distribution, the unit probability mass of that interval is defined as the probability of the interval divided by the length of the interval. The design for the dose-escalation phase of the study uses a target (DLT) rate of [redacted]% and an equivalence interval [redacted] for dose-escalation/de-escalation decisions as well as maximum tolerated dose (MTD) determination. A dose level is considered unsafe, with no additional participants enrolled at that dose level, if it has an estimated 80% or more probability of exceeding the target DLT rate of [redacted] with at least [redacted] participants treated and evaluated at that dose level.

After the escalation phase is completed, DLT rates at each dose level are estimated by isotonic regression (Ji et al, 2010). The weighted least squares regression model conditional on monotonic non-decreasing DLT rates with increasing dose and use the empirical (observed) DLT rates at each dose level as responses and sample sizes at each dose level as weights, along with the pool adjacent violators algorithm (PAVA) to estimate the DLT rate at each dose level using available software (eg, Cytel EAST or the function pava() from the R package 'ISO'). Given the DLT estimates for each dose level, the MTD is selected from all tried dose levels that have not been previously declared to be unsafe with a "de-escalate to the previous lower dose and the current dose is never used again due to unacceptable toxicity" (DU) decision according to the mTPI-2 decision table. With this constraint, the MTD is determined as the dose level with the DLT estimate closest to the target toxicity level of [redacted]%.

For specific dose escalation rules, refer to protocol Section 6.6.1.2.

3.3.7.2 Accelerated Titration Design (ATD)

For dose levels < 225 mg (Cohorts 1 to 4), dose escalation will follow an ATD and will be single participant cohorts. Cohorts 5, 6, 7, and the optional last cohort, Cohort 8, will follow an mTPI-2 and will consist of 3 to 12 participants. The enrolled participants must complete the DLT evaluation period before a dose escalation decision is made by the SRC. At any

ATD dose level, dose escalation will switch from ATD dose escalation to the mTPI-2 dose escalation algorithm if the initially enrolled participant has any of the following during the DLT evaluation period:

- Any DLT.
- Any ≥ Grade 2 TEAE.

Once the mTPI-2 algorithm is triggered, all further dose escalation will follow mTPI-2 algorithm.

3.3.7.3 Part B Dose Expansion

In Part B approximately CCI participants (Cohort B1) and CCI participants (Cohort B2) with measurable disease at baseline will be assessed for ORR and CRR respectively. Based on a CCI the sample sizes for each cohort have been primarily chosen to CCI A lower reference value (LRV) of CCI% and target value (TV) of CCI% was used for the ORR in Cohort B1. In Cohort B2, an LRV of CCI% and a TV of CCI% was used for the CRR.

Assuming a true ORR of CCI%, for cohort B1, the sample size of CCI gives a CCI of CCI%. Assuming a true CRR of CCI%, for cohort B2, the sample size of CCI gives a CCI CCI rate of CCI%.

For details, please refer to Sections 9.2 and 9.5 of the protocol.

4 STATISTICAL ANALYSIS

This section provides information on definitions, derivation and analysis/data presentation per domain.

4.1 Study Population

The domain study population covers participant disposition, analysis sets, protocol deviations, demographics, baseline characteristics medical history, prior and concomitant medication and study drug compliance.

4.1.1 Participant Disposition and Completion Status

4.1.1.1 Definitions and Derivations

Participants enrolled/screened is defined as agreement to participate in the clinical study following completion of the informed consent process by the participant or their legally authorized representative. Completion of study is defined in protocol Section 4.4.

4.1.1.2 Presentation

Participant disposition will be summarized and listed for all participants enrolled by dose level/cohort as defined by the current relevant tables, figures, listings (TFL) standards. Screen failures and reason for screen failures will be reported for the overall category only. The number and percentage of participants for the following categories will be summarized if applicable:

- Participants screened;
- Screen failures;
- Participants assigned to treatment;
- Participants assigned to treatment, but were not treated;
- Participants who started treatment;
- Participants ongoing treatment at data cut-off (DCO);
- Participants who withdrew;
- Participants who discontinued treatment;
- Participants ongoing study at DCO.

Disposition due to global/country situation may be summarized The number and percentage of participants for the following summaries will be presented:

- Participants who discontinued treatment due to global/country situation;
- Participants who withdrew from study due to global/country situation.

The study disruptions due to the global/country situation are also summarized as a separate table.

4.1.2 Analysis Sets

4.1.2.1 Definitions and Derivations

For the definitions of each analysis set, refer to Section 3.2.

4.1.2.2 Presentation

The analysis sets will be summarized by dose level /cohort. Any exclusions from analysis sets will be listed.

4.1.3 Protocol Deviations

4.1.3.1 Definitions and Derivations

Protocol deviations are defined in Section 3.3.5.

4.1.3.2 Presentation

The incidence of IPDs will be summarized by deviation categories for the full analysis set (FAS). The number and percentage of participants in the following categories will be summarized:

- Number of participants with at least 1 IPD
- If applicable, number of participants with at least 1 global/country related IPD
- If applicable, number of participants with at least 1 IPD, excluding global/country related IPDs.

A listing will be provided with IPD details.

4.1.4 Demographics

4.1.4.1 Definitions and Derivations

Age will be grouped into the following categories:

- Age group (standard): < 18, ≥ 18 - < 65, and ≥ 65 years.
- Age group (disease setting): ≥ 16 - < 25, ≥ 25 - < 50, ≥ 50 - < 65, ≥ 65 - < 75, and ≥ 75 years.

Each race category counts participants who selected only that category.

4.1.4.2 Presentation

Demographics will be summarized and listed based on the FAS by dose level/ cohort. The following will be summarized: age, age group, sex, race, ethnicity, and country.

4.1.5 Baseline Characteristics

4.1.5.1 Definitions and Derivations

Body mass index (BMI) will be derived as:

$$\text{BMI}\left(\frac{kg}{m^2}\right) = \frac{weight(kg)}{height(m)^2}$$

4.1.5.2 Presentation

Baseline characteristics will be listed and summarized for the FAS by dose level/cohort. The following will be summarized: height (cm), weight (kg) and BMI (kg/m^2).

Participant characteristics may further be summarized separately for all participants in the FAS who had confirmed or suspected COVID-19 infection. If less than 5 participants had confirmed or suspected COVID-19 infection, this may be presented in a listing rather than being summarized.

4.1.6 Disease Characteristics

4.1.6.1 Definitions and Derivations

Time from original diagnosis to first dose will be calculated as: (Date of first dose – date of original diagnosis) +1 / (365.25/12).

4.1.6.2 Presentation

Disease characteristics at baseline will be summarized and listed for all participants in the FAS by dose level/cohort as defined by the current relevant TFL standards.

Summaries will be produced that present:

- Eastern Cooperative Oncology Group (ECOG) performance status
- Hodgkins Lymphoma Subtype (histology)
- Hodgkins Lymphoma Subtype
- Hodgkins Lymphoma status after the last line of therapy
- Disease stage (revised Ann Arbor classification)
- Ann Arbor classification subdivision
- Presence of bulky disease
- Presence of extra nodal disease

4.1.7 Medical History

4.1.7.1 Definitions and Derivations

Medical history will be coded using the latest version of the Medical Dictionary for Regulatory Activities (MedDRA).

4.1.7.2 Presentation

Medical history and relevant surgery will be listed and summarized for the FAS set by dose level/cohort as defined by the current TFL standards.

Summaries of the number and percentage of participants who have had prior disease-related treatments, including autologous HSCT will be presented.

The number and percentage of participants who have had prior anti-cancer therapy for cHL will be summarized by Anatomical Therapeutic Chemical (ATC) classification and generic drug name, coded by World Health Organization (WHO) – Drug dictionary (WHODrug Global B3 Sep 2021 or later).

Summaries for the number and percentage of participants who had a certain number of lines of therapies and the best response on most recent line of therapy will be produced.

Summaries on participants' medical history (prior and ongoing) by System Organ Class (SOC) and Preferred Term (PT) will be produced.

4.1.8 Prior and Concomitant Medications/Procedures

4.1.8.1 Definitions and Derivations

For the purpose of inclusion in prior and/or concomitant medication, procedures or therapy summaries, incomplete medication or radiotherapy start and stop dates are imputed as detailed in Section 3.3.6.

Prior medications, concomitant and post-treatment medications or procedures are defined based on imputed start and stop dates as follows:

- Prior medications/procedures are those taken prior to study treatment with a stop date prior to the first dose of study treatment.
- Concomitant medications/procedures are those with a stop date or ongoing on or after the first dose date of study treatment, and must have started prior to or during treatment so there is at least one day in common with the study treatment.
- Post-treatment medications/procedures are those with a start date after the last dose date of study treatment plus 90 days.

4.1.8.2 Presentation

The number and percentage of participants who took prior and concomitant medications or had concomitant procedures will be summarized by ATC classification codes and the generic term coded by WHO Drug Global B3 Sep 2021 or later for the FAS by dose level/cohort.

All prior, concomitant and post study intervention medication data will be listed.

Missing coding terms will be listed and summarized as "Not coded".

In addition, post study drug - anti-cancer therapies for cHL post study drugs – HSCT and post study drugs – radiotherapies will be listed as well as summarized (number and percentage of subjects).

4.2 Endpoint Analyses

This section covers details related to the endpoint analyses such as primary, secondary, other endpoints including sensitivity and supportive analyses.

Table 2 Endpoint Analyses

Statistical category	Endpoint	Population	Population level summary (analysis)	Details in section
Part A Dose Escalation **Objective 1: To assess the safety and tolerability of sabestomig in participants with r/r cHL**				
Primary	AEs, imAEs, SAEs, AEs leading to discontinuation of sabestomig, Laboratory data changes, vital signs changes and ECG changes	Safety analysis set	Descriptive statistics on safety endpoints	4.5.2 (AEs) 4.5.4 and 4.5.5 (laboratory data) 4.5.6 (vital signs) 4.5.7 (ECG)
Primary	Incidence of dose-limiting toxicities	Safety analysis set	Dose-limiting toxicities	4.5.10.1 DLT
Part B Dose Expansion (all) **Objective 2: To assess the safety and tolerability of sabestomig in participants with r/r cHL**				
Primary	AEs, imAEs, SAEs, AEs leading to discontinuation of sabestomig Laboratory data changes, vital signs changes and ECG changes	Safety analysis set	Descriptive statistics on safety endpoints	4.5.2 (AEs) 4.5.4 and 4.5.5 (laboratory data) 4.5.6 (vital signs) 4.5.7 (ECG)
Part B Dose Expansion (B1) **Objective 3: To assess the preliminary antitumor activity of sabestomig in participants with r/r cHL (anti-PD-1/PD-L1 exposed)**				
Primary	Objective Response Rate [a]	Response evaluable set	ORR and 95% CI	4.2.2
Part B Dose Expansion (B2) **Objective 4: To assess the preliminary antitumor activity of sabestomig in participants with r/r cHL (anti-PD-1/PD-L1 naïve)**				
Primary	Complete Response Rate [a]	Response evaluable set	CR rate and 95% CI	4.2.3
Part A Dose Escalation **Objective 5: To assess the preliminary antitumor activity of sabestomig in participants with r/r cHL**				
Secondary	Complete Response Rate [b]	Response evaluable set	CR rate and 95% CI	4.2.3

Table 2 Endpoint Analyses				
Statistical category	**Endpoint**	**Population**	**Population level summary (analysis)**	**Details in section**
Secondary	Objective Response Rate [b]	Response evaluable set	OR rate and 80% CI	4.2.2
Secondary	Duration of Response [b]	Response evaluable set	Median DoR and its two-sided 95% CI estimated using Kaplan-Meier method	4.2.4
Secondary	Duration of Complete Response [b]	Response evaluable set	Median DoCR and its two-sided 95% CI estimated using Kaplan-Meier method	4.2.5
Secondary	Progression Free Survival at 12 and 24 months [b]	Full analysis set	Median PFS and its two-sided 95% CI estimated using Kaplan-Meier method	4.2.6
Secondary	Overall Survival at 12 and 24 months	Full analysis set	Median OS and its two-sided 95% CI estimated using Kaplan-Meier method	4.2.7
Part B Dose Expansion **Objective 6: To further assess the preliminary antitumor activity of sabestomig in participants with r/r cHL**				
Secondary	Duration of Response [a]	Response evaluable set	Median DoR and its two-sided 95% CI estimated using Kaplan-Meier method	4.2.4
Secondary	Duration of Complete Response [a]	Response evaluable set	Median DoCR and its two-sided 95% CI estimated using Kaplan-	4.2.5

Table 2 Endpoint Analyses

Statistical category	Endpoint	Population	Population level summary (analysis)	Details in section	
				Meier method	
Secondary	Progression Free Survival at 12 and 24 months [a]		Full analysis set	Median PFS and its two-sided 95% CI estimated using Kaplan-Meier method	4.2.6
Secondary	Overall Survival at 12 and 24 months		Full analysis set	Median OS and its two-sided 95% CI estimated using Kaplan-Meier method	4.2.7
Part A Dose Escalation and Part B Dose Expansion **Objective 7:** • **To assess the PK of sabestomig in participants with r/r cHL** • **To assess the immunogenicity of sabestomig in participants with r/r cHL**					
Secondary	PK of sabestomig		PK analysis set	Descriptive statistics on PK endpoints	4.3
Secondary	Immunogenicity of sabestomig		Immunogenicity analysis set	Descriptive statistics of anti-drug antibodies	4.4

[a] Based on Blinded Independent Central Review (BICR) assessments based on Modified Lugano criteria 2014 (Chelson et al 2014).

[b] Based on investigator assessments (as entered in the database) based on Modified Lugano criteria 2014 (Chelson et al 2014) and RECIL criteria (Younes et al, 2017).

The images of CCI participants treated with RP2D in part A will undergo a retrospective central review to be combined with Part B BICR assessments for efficacy analyses.

cHL: classical Hodgkin Lymphoma; DoR: Duration of Response; DoCR: Duration of Complete Response; OS: Overall Survival; PFS: Progression-free Survival; PK: pharmacokinetic(s); RECIL: Response-evaluation Criteria in Lymphoma; r/r cHL: relapsed/refractory cHL.

Additionally, supportive secondary efficacy analyses for part B , as done for part A, including objective and complete response rate, duration of response, and PFS at 12 and 24 months will be based on Investigator assessments based on Lugano criteria for malignant lymphoma (Chelson et al 2014) and RECIL criteria (Younes et al, 2017). All efficacy results will be listed.

All efficacy analyses will be presented by dose level/cohort.

4.2.1 Primary Safety Endpoints: AEs, immune mediated AEs and serious AEs, laboratory evaluations, vital signs, ECG results Definition and DLTs

Analysis methods for the primary safety endpoints are described in Section 4.5.

4.2.2 Primary Endpoint: ORR

The primary efficacy endpoint for Part B Dose Expansion (B1) is ORR assessed by BICR based on modified Lugano criteria for lymphoma. ORR is a secondary endpoint for Part A Dose Escalation assessed by Investigator based on modified Lugano criteria and RECIL criteria.

4.2.2.1 Definition and Derivations

Objective Response (OR) is defined as a Best Overall Response (BOR) of CR or PR that occurs prior to the initiation of subsequent anticancer treatment and prior to progression. Objective Response Rate (ORR) is defined as the percentage of participants with objective response.

Data obtained from first dose until progression, or last evaluable assessment in the absence of progression, will be included in the assessment of ORR, regardless of whether the participant withdraws from therapy. Participants who discontinue treatment without a response, receive subsequent anti-lymphoma therapy and then respond will not be included as responders in the ORR (ie, the visit contributing to a response must be prior to subsequent therapy for the participant to be considered as a responder).

The ORR will be based on the BICR assessment, using modified Lugano criteria for lymphoma including data of all scans for Part B, regardless of whether it was scheduled or not.

ORR for Part A will be defined similarly but taking Investigator responses and based on modified Lugano criteria and RECIL criteria. Data will be recorded in the eCRF.

Best overall response

Best Overall Response (BoR) will be determined using disease assessments at different evaluation time points from the date of the first dose of study treatment until documented disease progression or start of any further anti-cancer therapy, whichever is earlier. Clinical deterioration will not be considered as documented disease progression. Categories of BOR will be based on modified Lugano criteria using the following response criteria: CR, PR, stable disease [SD], progressive disease [PD], and not evaluable [NE], and RECIL criteria using the following criteria: CR, PR, minor response [MR], SD, PD, and NE.

BoR will be calculated based on Investigator assessed data for part A efficacy analysis based on RECIL criteria, and based on BICR assessed data for part B analysis. The images of the CCI participants dosed at RP2D in part A will undergo a retrospective assessment by BICR whose result will be combined with other BICR-based part B data for Part B analysis.

For participants without documented progression or subsequent therapy, all available response designations will contribute to the BOR determination. For purpose of analysis, if a participant receives one dose and discontinues the study without assessment or receives subsequent therapy prior to assessment, this participant will be counted in the denominator (as non-responder).

4.2.2.2 Primary Analysis of Primary Endpoint: ORR

Summaries will be produced that present the number and percentage of participants with a response (CR/PR) based upon the number of participants in the response-evaluable analysis set. ORR will be presented with 95% CI based on exact binomial proportions.

BoR will be summarized by number of participants and percentage for each category (CR, PR, MR [for RECIL criteria], SD, PD, and NE). No formal statistical analyses are planned for BoR.

4.2.2.3 Sensitivity Analysis of ORR Endpoint

As sensitivity analysis, if the number of participants in FAS is different from that in Response Evaluable Set, a summary of CCI [redacted] may be presented. CCI [redacted] CCI [redacted] will be used in this analysis. For the computation of ORR, participants with CCI [redacted] will be included in the FAS and will be considered non-responders.

4.2.2.4 Supportive Secondary Analysis of ORR Endpoint

A supportive secondary analysis of ORR for Part B will be presented using Investigator assessments and based on modified Lugano criteria and RECIL. Data recorded in eCRF will be used for this analysis.

4.2.2.5 Supportive Secondary Analysis: Change in Target Lesion Tumor Size

Change in target lesion tumor size will be presented as supportive secondary analysis for Part A and Part B.

4.2.2.5.1 Definition and Derivations: Change in Target Lesion Tumor Size

Target lesions are measurable tumor lesions. Baseline is defined to be the last evaluable assessment prior to starting treatment.

Target lesion tumor size for lymph nodes and extra-lymphatic sites at any timepoint is defined as the sum of the product of the perpendicular diameters for multiple lesions. When a lesion is too small to measure, 5 mm x 5 mm will be assigned as the product of the perpendicular diameters.

The percentage change in target lesion size at each timepoint for which data are available will be obtained for each participant, taking the difference between the sum of the perpendicular diameters (SPD) of target lesions at each timepoint and the sum of the product of the perpendicular diameters of target lesions at baseline multiplied by 100 (ie, [timepoint – baseline]/baseline x 100).

4.2.2.5.2 Analysis of Supportive Data: Change in Target Lesion Tumor Size

Only participants included in the response-evaluable analysis set will be included in summaries of change in tumor size.

The target lesion tumor size (SPD of lymph node) and percentage change from baseline in target lesion tumor size may be summarized using descriptive statistics and presented at each timepoint.

Additionally, 'spider' plots of percentage change from baseline in target lesion size by participants will be presented. A graphical summary of the best percentage change in target lesion tumor size will be presented in a vertical bar chart with each participant's percentage change from baseline to nadir displayed as a vertical bar, with colour coding that indicates best response obtained ("waterfall" plot).

4.2.2.6 Subgroup Analyses

The following subgroup analysis may be performed for ORR (primary efficacy variable):

- Participants who are brentuximab naïve versus participants who are brentuximab exposed. Exposure to brentuximab will be identified from the participant's concomitant medications record.

4.2.3 Primary Endpoint: Complete Response Rate (CRR)

The primary efficacy endpoint for Part B Dose Expansion (B2) is CRR assessed by BICR based on modified Lugano criteria for lymphoma.

CRR is a secondary endpoint for Part A Dose Escalation assessed by Investigator based on modified Lugano criteria and RECIL criteria.

4.2.3.1 Definition and Derivations

CRR is defined as the percentage of participants with CR, with the denominator defined as the number of participants in the response-evaluable analysis set.

Data obtained until progression, or last evaluable assessment in the absence of progression, will be included in the assessment of CRR, regardless of whether the participant withdraws from therapy. Participants who discontinue treatment without a response, receive subsequent anti-lymphoma therapy and then respond will not be included as responders in the CRR.

The CRR will include data of all scans, regardless of whether it was scheduled or not.

CRR as a secondary endpoint for Part A will be defined similarly but taking Investigator responses and based on modified Lugano criteria and RECIL criteria. Data will be recorded in the eCRF.

4.2.3.2 Primary Analysis of Primary Endpoint: CRR

CRR will be calculated and binomial exact two-sided CIs at 95% will be presented for participants in the response-evaluable analysis set.

4.2.3.3 Sensitivity Analysis of Primary Endpoint: CRR

As sensitivity analysis, if the number of participants in FAS is different from that in Response Evaluable Set, a summary of CCI [redacted] may be presented. CCI [redacted] CCI [redacted] will be used in this analysis. For the computation of CRR, participants with CCI [redacted] will be included in the FAS and will be considered non-responders.

4.2.3.4 Supportive Analysis of CRR Endpoint

A supportive analysis of CRR for Part B will be performed using Investigator responses and based on modified Lugano criteria and RECIL. Data will be recorded in eCRF.

4.2.3.5 Subgroup Analyses

The following subgroup analysis may be performed for CRR (primary efficacy variable):

- Participants who are brentuximab naïve versus participants who are brentuximab exposed.

4.2.4 Secondary Endpoint: Duration of Response (DoR)

DoR is a secondary endpoint for Part A Dose Escalation and Part B Dose Expansion. Responses assessed by Investigator for Part A using modified Lugano and RECIL criteria and by BICR for Part B using modified Lugano.

4.2.4.1 Definition

DoR is defined as the time from the date of first documented objective response (CR or PR) until date of first documented disease progression or death (by any cause in the absence of disease progression).

4.2.4.2 Derivations

DoR (months) = (date of progression free survival [PFS] event [progression/death] or censoring – date of first objective response + 1) / (365.25/12).

The end of response should coincide with the date of progression or death from any cause used for the PFS endpoint. The time of the initial response is defined as the latest of the dates contributing towards the first visit response of CR or PR. If a participant does not progress following a response, then their DoR is censored on the PFS censoring date. Only participants who have achieved objective response are evaluated for DoR.

4.2.4.3 Analysis of Secondary Endpoint: DoR

The analysis of DoR will include all participants in the response-evaluable analysis set. Only participants who achieved objective response (CR or PR) will be included in the summaries of DoR. Kaplan-Meier plots of DoR with at least CCI responders per dose level may be presented. The median DoR and 2 sided 95% CI will be estimated using the Kaplan-Meier method. In addition, percentage of participants remaining in response at 3, 6, 12, 18 and 24 months after initial response may also be presented. These time points may be adjusted according to actual data observed in the study without amendment to this SAP. Reasons for censoring for DoR will also be summarized.

Swimmer plots by participant (including non-responders) with symbols to display each participant's study status will be presented.

4.2.4.4 Supportive Analysis of DoR Endpoint

A supportive analyses of DoR for Part B will be presented using Investigator responses and based on modified Lugano criteria and RECIL. Data will be recorded in eCRF.

4.2.5 Secondary Endpoint: Duration of Complete Response (DoCR)

DoCR is a secondary endpoint for Part A Dose Escalation and Part B Dose Expansion. Responses assessed by Investigator for Part A using modified Lugano and RECIL criteria; and by BICR for part B using modified Lugano.

4.2.5.1 Definition

DoCR is defined as the time from the date of first documented complete response (CR) until date of first documented relapse/disease progression per Lugano and RECIL criteria as assessed by Investigator (or by BICR for Part B) at local site or death (by any cause in the absence of disease progression).

4.2.5.2 Derivations

DoCR (months) = (date of progression free survival [PFS] event [progression/death] or censoring – date of first complete response + 1) / (365.25/12).

The end of response should coincide with the date of progression or death from any cause used for the PFS endpoint. The time of the initial response is defined as the latest of the dates contributing towards the first visit response of CR. If a participant does not progress following a response, then their DoCR is censored on the PFS censoring date. Only participants who have achieved objective response are evaluated for DoCR.

4.2.5.3 Analysis of Secondary Endpoint: DoCR

The analysis of DoCR will include all participants in the response-evaluable analysis set. Only participants who achieved complete response will be included in the summaries of DoCR.

Kaplan-Meier plots of DoCR with at least [redacted] responders per dose level may be presented if data permitted. The median DoCR and 2 sided 95% CI will be estimated and graphically presented using the Kaplan-Meier method. In addition, percentage of participants remaining in response at 3, 6, 12, 18 and 24 months after initial response may also be presented. These time points may be adjusted according to actual data observed in the study without amendment to this SAP. Reasons for censoring for DoCR will also be summarized.

Swimmer plots by participant (including non-complete responders) may be created to visualise when complete response begins and ends and when study intervention is withdrawn.

4.2.5.4 Supportive Analysis of DoCR Endpoint

A supportive analyses of DoCR for Part B will be presented using Investigator responses and based on modified Lugano criteria and RECIL. Data will be recorded in eCRF.

4.2.6 Secondary Endpoint: Progression-free Survival (PFS)

PFS is a secondary efficacy endpoint for Part A Dose Escalation and Part B Dose Expansion. Responses assessed by Investigator for Part A using modified Lugano and RECIL criteria; and assessed by BICR for part B, using modified Lugano.

4.2.6.1 Definition

PFS is defined as the time from first dose until the earlier of the date of first documented disease progression, per Lugano classification (and per RECIL) as assessed by the investigator for Part A (by BICR per Lugano classification for Part B), or death (by any cause in the absence of disease progression or subsequent anticancer treatment or bone marrow transplantation in response).

4.2.6.2 Derivations

PFS (months) = (date of PFS event [progression/death] or censoring – date of first dose + 1) / (365.25/12).

Participants who have not progressed or died at the time of analysis, or have unknown status, are censored at the time of the latest date of assessment from their last evaluable disease assessment. However, if the participant progresses or dies immediately after two or more consecutive missed visits, the participant is censored at the time of the latest evaluable disease assessment prior to the two missed visits. Note: a NE visit is not considered as a missed visit.

If the participant has no evaluable disease assessments post-baseline or does not have baseline tumor assessment data they are censored at Day 1, unless they die within 57 weeks (i.e, 2 missed visit after Year 1 whereby disease assessments are every 6 months [2 x 26 weeks + 5 week for late visit = 57 weeks]) then the death qualifies as a PFS event. Participants who are censored at Day 1 are included in the FAS and the analysis but they do not contribute to any risk set. For participants who met one or more censoring conditions, PFS will be censored according to the earliest censoring condition; for participants who met none of the censoring conditions, PFS will be calculated by the earliest PFS event date.

A summary of censoring rules and the date of PD/death or censoring are given in Table 4 . Note that censoring overrides event in certain specified cases.

Table 4 Summary of Censoring Rules for PFS

Situation	Date of PD/Death or Censoring	PFS Outcome
Documented Progressive Disease (PD) or death in the absence of progression	Date of earliest documentation of PD or date of death	Event
Either no assessment at baseline or no evaluable assessments post-baseline AND death prior to Year 1 scheduled post-baseline disease assessment	Date of death	Event
Either no assessment at baseline or no evaluable assessments post-baseline AND no death prior to Year 1 scheduled post-baseline disease assessment	Date of first dose (Day 1)	Censored
PD or death (in the absence of progression) 57 weeks since previous disease assessments	Last evaluable progression-free disease assessment prior to missed assessments	Censored
At least one post-baseline tumor assessment, On-going with neither PD nor death at the time of analysis or lost to follow-up or withdrawn consent	Date of last evaluable disease assessment	Censored
Initiation of subsequent anticancer, excluding BM transplantation treatment prior to PD or death	Date of last evaluable disease assessment prior to initiation of subsequent anticancer treatment	Censored

Abbreviations: BM bone marrow, PD, progressive disease; PFS, progression-free survival

If response is defined as CR or PR, provided the participant did not receive any new anti-cancer treatment for Hodgkin lymphoma from the last dose of sabestomig and start of the preconditioning for the BM transplantation. If the participant received any new treatment (before the BM transplantation), he would follow the censoring rules for new treatment for Hodgkin lymphoma. The information will be collected in the EDC.

Radiotherapy consolidation for a participant in complete response is not considered a subsequent anticancer treatment.

The PFS time is always derived based on scan/assessment dates, not visit dates. Disease assessments/scans contributing towards a particular visit may be performed on different dates. The following rules are applied:

- The date of progression is determined based on the earliest of the dates of the component that triggered the progression.

Note: for target lesions only the latest scan date is recorded out of all scans performed at that assessment for the target lesions and similarly for non-target lesions only the latest scan date is recorded out of all scans performed at that assessment for the non-target lesions.

Duration of follow-up for PFS is applicable only for PFS censored participants and is defined as follows:

Duration of follow-up for PFS in censored participants (months) = (date of PFS event [progression/death] or censoring – date of first dose + 1) / (365.25/12).

4.2.6.3 Analysis of Secondary Endpoint: PFS

The main analysis of PFS is based on the FAS. The number and percentage of participants experiencing a PFS event (broken down by type of event/censoring) and Kaplan-Meier plots of PFS, for dose levels with at least [redacted] participants may be presented. The median PFS and its two-sided 95% CI will be estimated using the Kaplan-Meier method (if participant numbers allow).

The intervention status at progression of participants at the time of analysis will be summarized. This includes the number (%) of participants who were on intervention at the time of progression, the number (%) of participants who discontinued study intervention prior to progression, the number (%) of participants who have not progressed and were on intervention or discontinued intervention.

A summary of the duration of follow-up for PFS is included using median (range). This will be presented for censored participants (including all types of PFS censoring).

The proportion of participants alive and progression free at 3, 6, 12, 18 and 24 months and associated two-sided 95% CI will be estimated using the Kaplan-Meier method. Other landmark rates may be considered per sponsor discretion.

4.2.6.4 Supportive Analysis of PFS Endpoint

A supportive analyses of PFS for Part B will be presented using Investigator responses and based on modified Lugano criteria and RECIL. Data will be recorded in eCRF.

4.2.7 Secondary Endpoint: Overall Survival (OS)

OS is a secondary efficacy endpoint.

4.2.7.1 Definition

OS is defined as the time from the date of first dose until death due to any cause regardless of whether the participant withdraws from study therapy or receives another anti-cancer therapy.

4.2.7.2 Derivations

OS (months) = (date of death or censoring – date of first dose + 1) / (365.25/12).

Any participant not known to have died at the time of analysis will be censored based on the last recorded date on which the participant was known to be alive. Participants known to be alive or dead after the data cut-off for analysis will be censored at the DCO. Participants lost

to follow-up will be censored at the date the participant is last know known to have been alive.

Note: Survival follow-up calls will be made in the week following the date of DCO for the analysis, and if participants are confirmed to be alive or if the death date is post the DCO date these participants will be censored at the date of DCO. The status of ongoing, withdrawn (from the study) and "lost to follow-up" participants at the time of the final OS analysis should be obtained by the site personnel by checking the participant's notes, hospital records, contacting the participant's general practitioner and checking publicly-available death registries. In the event that the participant has actively withdrawn consent to the processing of their personal data, the vital status of the participant can be obtained by site personnel from publicly available resources where it is possible to do so under applicable local laws.

Note: For any OS analysis performed prior to the final OS analysis, in the absence of survival calls being made, it may be necessary to use all relevant eCRF fields to determine the last recorded date on which the participant was known to be alive for those participants still on intervention (since the SURVIVE module is only completed for participants off intervention if a survival sweep is not performed). The last date for each individual participant is defined as the latest among the following dates recorded on the eCRFs:

- AE start, stop and change in severity dates
- Admission and discharge dates of hospitalization
- Study intervention date
- End of intervention date
- Concomitant medication start and stop dates
- Laboratory test dates
- Date of vital signs
- Disease assessment dates on eCRF
- Start and stop dates of alternative anticancer intervention
- Date last known alive on survival status eCRF
- End of study date

Duration of follow-up for OS is reported separately for censored participants and non-censored participants is defined as follows:

Duration of follow-up for OS (months) = (date of death or censoring (date last known to be alive) – date of first dose + 1) / (365.25/12).

4.2.7.3 Handling of Dropouts and Missing Data

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

- For Missing day only – using the 1st of the month of death.
- For Missing day and Month – using the 1st of January of the year of death.

If there is evidence of death but the date is entirely missing, it is treated as missing, ie, censored at the last known alive date.

4.2.7.4 Primary Analysis of Secondary Endpoint (Overall Survival)

The analysis of OS is based on the FAS. The number and percentage of participants experiencing an OS event and Kaplan-Meier plots, for dose levels with at least CCI participants of OS may be presented. The median OS and two-sided 95% CI will be estimated using the Kaplan-Meier method (if participant numbers allow).

Summaries of the number and percentage of participants who have died, those still in survival follow-up, those lost to follow-up and those who have withdrawn consent will be provided.

A summary of the duration of follow-up for OS will be included using median (range). This will be presented for censored participants.

The proportion of participants alive at 6, 12, 18 and 24 months and associated two-sided 95% CI will be estimated using the Kaplan-Meier method.

4.2.7.5 Sensitivity Analysis of OS Endpoint

A sensitivity analysis may be performed using Response Evaluable Set. A sensitivity analysis may also be conducted to assess the impact of CCI [redacted]. That is, participants who had an CCI [redacted] This sensitivity analysis may only be performed CCI [redacted] A sensitivity analysis may also be conducted to assess the impact of CCI [redacted] CCI That is, participants who had a CCI [redacted]

4.2.8 Other Endpoints

The analysis of the exploratory research endpoints will be handled outside this SAP and reported separately from the Clinical Study Report (CSR).

4.2.9 Pharmacodynamic Endpoint

Analysis of pharmacodynamic data will be described in a separate analysis plan.

4.3 Pharmacokinetics

This section covers details related to pharmacokinetics (PK) endpoints and analyses.

4.3.1 Serum Pharmacokinetic Concentration Data

Serum concentrations for each scheduled time-point are summarized by dose level in Part A and by cohort in Part B using appropriate descriptive statistics. Any issues considered to impact the PK data may result in the exclusion of concentration data from the PK analysis and/or exclusion of parameters and concentrations from the PK descriptive statistics and/or inferential statistical analyses, the reason(s) for exclusion(s) will be documented by the PK Scientist. These will be agreed between the PK Scientist and the AstraZeneca Clinical Pharmacology Scientist prior to the final PK analysis and handover of parameters to programming and identified in the PK flagging file and/or PK handover document. However, any excluded data will be presented in the listings.

For Part B where intensive PK sample are collected for the China population, a full set of PK NCA parameters will be provided separately from the Global and Japanese population for the final PK analysis.

The following descriptive statistics are presented for serum concentrations:

- n
- n below LLOQ
- geometric mean (gmean)
- geometric coefficient of variance (%) (gCV)
- arithmetic mean (mean)
- arithmetic standard deviation (Std Dev)
- median
- minimum (min)
- maximum (max)

The gmean is calculated as $\exp(\mu)$, where μ is the mean of the data on the natural log scale.

The gCV is calculated as $100 \times \text{sqrt}[\exp(s^2)-1]$, where s is the Std Dev of the data on the natural log scale.

Where required for plots: The gSD is calculated as exp(σ), where σ is the standard deviation of the data on the natural log scale. The gmean ± gSD (gmean-gSD and gmean+gSD) are calculated as exp[μ ± s].

Protocol scheduled times are used to present the PK concentration summary tables and corresponding gmean concentration-time figures.

Handling of Non-Quantifiable Concentrations

Individual concentrations below the LLOQ of the bioanalytical assay are reported as NQ in the listings with the LLOQ defined in the footnotes of the relevant TFLs. Individual serum concentrations that are Not Reportable are reported as NR and those that are missing are reported as NS (No Sample) in the listings. Serum concentrations that are NQ, NR or NS are handled as follows for the provision of descriptive statistics:

- Any values reported as NR or NS are excluded from the summary tables and corresponding figures.
- At a time point where less than or equal to 50% of the concentration values are NQ, all NQ values are set to the LLOQ, and all descriptive statistics are calculated accordingly.
- At a time point where more than 50% (but not all) of the values are NQ, the gmean and gCV% are set to Not calculable (NC). The maximum value is reported from the individual data, and the minimum and median are set to NQ.
- If all concentrations are NQ at a time point, no descriptive statistics are calculated for that time point. The gmean, minimum, median and maximum are reported as NQ and the gCV% as NC.
- The number of values below LLOQ (n < LLOQ) are reported for each time point together with the total number of collected values (n).

Three observations > LLOQ are required as a minimum for a serum concentration or PK parameter (eg, Cmax) to be summarized. Two observations > LLOQ are presented as minimum and maximum with the other summary statistics as NC.

4.3.2 Serum Pharmacokinetic Parameter Data

The pharmacokinetic (PK) parameters of the serum concentration data for sabestomig will be derived using non-compartmental methods in Phoenix® WinNonlin® Version 8.3 or higher (Certara).

PK analysis is, where data allow, carried out using actual elapsed times determined from the PK sampling and dosing times recorded in the database. If actual elapsed times are missing, nominal times may be used at the discretion of the AZ PK Scientist with approval from the

AZ Clinical Pharmacology Scientist (CPS). Nominal sampling times may be used for any agreed interim PK parameter calculations.

For each PK sampling period, serum concentrations that are non-quantifiable (NQ) from the time of pre-dose sampling (t=0) up to the time of the first quantifiable concentration is set to a value of zero. After this time point, NQ serum concentrations are set to missing for all concentration profiles. Where 2 or more consecutive concentrations are NQ at the end of a profile, the profile is deemed to have terminated and therefore any further quantifiable concentrations are set to missing for the calculation of the PK parameters unless it is considered to be a true characteristic of the profile of the drug.

If an entire concentration-time profile is NQ, the profile is excluded from the PK analysis.

Cmax, Crough, tmax, and tlast are taken directly from the concentration-time profiles.

Refer to Appendix C for the calculation of PK parameters to be estimated, where possible and appropriate, from the PK serum concentration.

PK parameter Listings

All reportable PK parameters, including individual diagnostic and lambda z related parameters, are listed for each participant, for each analyte separately.

PK parameter descriptive statistics

All primary and secondary PK parameters are summarized for each analyte by dose level in Part A and by cohort in Part B using appropriate descriptive statistics.

The descriptive statistics for PK parameters are presented as follows:

- Cmax, Ctrough, AUCinf, AUClast: present n, gmean, arithmetic mean of non log-transformed data (mean), arithmetic standard deviation (Std Dev), gCV(%), median, min and max.
- t½λz, CL, Vss, Vz: present n, arithmetic mean of non log-transformed data (mean), arithmetic standard deviation (Std Dev), median, min and max.
- Cmax/D, Ctrough/D, AUCinf/D, AUClast/D, λz: present n, gmean, gCV(%), mean, Std Dev, median, min and max.
- tmax and tlast: present n, median, min and max.
- Diagnostic parameters (eg, λz lower, λz upper, λzN, Rsq, Rsq adj, λz span ratio, and AUCextr): present n, arithmetic mean, Std Dev, gmean, gCV%, median, min and max.

Three values are required as a minimum for PK parameters to be summarized. Two values are presented as a min and max with the other summary statistics as NC.

If one or more values for a given parameter is zero, then no geometric statistics are calculated for that parameter and the results for geometric statistics are set to NA (Not Applicable).

4.3.3 Graphical presentation of Pharmacokinetic Concentration and Parameter Data

All mean (arithmetic mean and/or gmean) plots or combined plots showing all participants by treatment are based on the PK analysis set. Individual plots by participant are based on the safety analysis set.

For consistency, the serum concentration values used in the mean (arithmetic mean and/or gmean) data graphs are those given in the descriptive statistics summary table for each time point.

For gmean concentration-time plots, NQ values are handled as described for the descriptive statistics; if the geometric mean is NQ, the value plotted is zero for linear plots and missing for semi-logarithmic plots. Any gmean±gSD error bar values that are negative are truncated at zero on linear concentration-time plots and omitted from semi-logarithmic plots.

For individual plots, serum concentrations which are NQ prior to the first quantifiable concentration are set to a value of zero (linear plots only). After the first quantifiable concentration, any NQ serum concentrations are regarded as missing.

Data permitting, the following figures may be presented as appropriate:

- Figures for the gmean serum concentration-time data (with ±gSD error bars) presented on both linear and semi-logarithmic scales using scheduled post-dose time as follows:
  - By PK Day with all dose level/cohort overlaid on the same plot
- Individual participant serum concentration-time data graphically presented on both linear and semi-logarithmic scales using actual time post-dose as:
  - Combined individual plots by dose level/cohort with all participants overlaid on the same plot for each dose level/cohort
- Box-plot of pharmacokinetic parameters versus dose level/cohort:
  - Cmax/D, AUCinf/D and AUClast/D at C1D1
  - Cmax/D and Ctrough/D at C3D1 (dose escalation) or C4D1 (dose expansion)
- Individual and geometric mean of dose normalized pharmacokinetic parameters versus dose level/cohort:
  - Cmax/D, AUCinf/D and AUClast/D at C1D1
  - Cmax/D and Ctrough/D at C3D1 (dose escalation) or C4D1 (dose expansion)

4.3.4 Precision and Rounding Rules for Pharmacokinetic Concentration and Parameter Data

PK concentration data

PK concentration data listings present to the same number of significant figures as the data received from the bioanalytical laboratory (usually but not always to 3 significant figures) and against the same units as received.

PK concentration descriptive statistics present 4 significant figures with the exception of the min and max which present 3 significant figures and n and n<LLOQ which present as integers.

PK parameter data

PK parameter listings are presented according to the following rules:

- Cmax, Ctrough, Cmax/D and Ctrough/D: present to the same number of significant figures as received from the bioanalytical laboratory
- tmax, tlast, λz lower and λz upper: present as received in the data, usually to 2 decimal places
- AUCinf, AUClast, AUCinf/D, AUClast/D, AUCextr, λz, t½λz, CL, Vss, Vz, Rsq, Rsq adj, λz span ratio: present to 3 significant figures
- λzN: present as an integer (no decimals)

The descriptive statistics for PK parameter data are presented to 4 significant figures with the exception of the min and max which are presented to 3 significant figures apart from the following:

- tmax, and tlast: present as received in the data, usually to 2 decimal places
- number of values (n): present as an integer

4.4 Immunogenicity

Serum samples for ADA assessments will be conducted utilizing a tiered approach (screen, confirm, titer), and ADA data will be collected at scheduled visits shown in the CSP. Blood samples collected outside of the protocolled window are summarized at the closest nominal time point that does not already have a value. ADA result from each sample will be reported as either positive or negative. If the sample is positive, the ADA titer will be reported as well. A participant is defined as being ADA-positive if a positive ADA result is available at any time, including baseline and all post-baseline measurements; otherwise ADA negative.

All participants who received at least 1 dose of study intervention with at least 1 reportable immunogenicity measurement will be included in the immunogenicity analysis set. The

number and percentage of ADA-evaluable participants that fall into the immunogenicity analysis set in the following categories will be provided by levels/cohorts:

- Anti-drug antibody (ADA) positive at baseline and/or post-baseline visits. The percentage of these participants in a population is known as ADA prevalence.
- Treatment-induced ADA positive (positive post-baseline and not detected at baseline).
- Treatment-boosted ADA positive (baseline ADA titer that was boosted greater than the variability of the assay (ie, $\geq$ 4-fold increase).
- Treatment-emergent ADA positive (either treatment-induced ADA positive or treatment-boosted ADA positive). The percentage of these participants in a population is known as ADA incidence.
- Participants who are treatment-emergent ADA negative (Participants who are ADA positive but not fulfilling the conditions for treatment-emergent ADA positive).
- ADA positive at baseline and at least one post-baseline assessment.
- ADA positive at baseline and not detected post-baseline.
- Persistent positive, defined as ADA negative at baseline and having at least 2 post-baseline ADA positive measurements with $\geq$ 16 weeks between first and last positive, or an ADA positive result at the last available post baseline assessment. For studies less than 16 weeks in duration, only the last post-baseline requirement applies.
- Transient positive, defined as ADA negative at baseline and at least one post- baseline ADA positive measurement and not fulfilling the conditions for persistently positive.

The potential impact of ADAs on PK, pharmacodynamics, and safety will be assessed if data allow. These analyses will be described separately from the SAP and presented separately from the main CSR.

4.5 Safety Endpoints Analyses

The domain safety covers exposure, AEs, clinical laboratory, vital signs, electrocardiograms (ECGs), echocardiograms, ECOG performance status and B-symptoms.

Tables will be provided for the safety analysis set, listings are provided for all participants or the safety analysis set depending on the availability of data.

All safety analysis will be presented by dose level/cohort.

4.5.1 Exposure

4.5.1.1 Definitions and Derivations

- Duration of exposure is defined by the last date of actual dosing (ie, a dose > 0 mg was given) in the last cycle plus 21 days minus the date of first treatment study intervention plus 1.
    - For participants who die whilst on study treatment or if a DCO occurs, duration of exposure is defined as date of death/DCO (whichever occurs first) minus the date of first treatment plus 1 day.
    - Therefore: Duration of exposure (weeks) = [min(last dose date where dose > 0 mg + 21 – 1, date of death, date of DCO) – first dose date + 1]/7 .
- The duration of exposure in cycles is calculated as:

Duration of exposure (cycles) = the number of cycles in which at least one portion of intervention was administered (ie, dose > 0 mg). If a cycle is prolonged due to toxicity, this is still counted as one cycle.

- Actual duration of exposure is the number of sabestomig doses administered (ie, a dose >0mg was given) multiplied by 21 days (due to the fact that sabestomig as ordinarily dosed on every 3 weeks schedule).
    - Therefore: Actual duration of exposure in weeks (weeks) = the number of sabestomig doses administered multiplied by 3.
- Dose intensity of study intervention is addressed by considering relative dose intensity (RDI), where RDI is the percentage of the actual dose delivered relative to the intended dose through to treatment discontinuation. More specifically, RDI is defined as follows:
    - RDI = 100% × d/D, where d is the actual cumulative dose delivered up to the actual last day of dosing and D is the intended cumulative dose up to the actual last day of dosing. D is the total dose that would be delivered, if there were no modification to dose or schedule.
- Duration of treatment cycle delays will be derived for treatment cycles indicated as being delayed on the eCRF. For each individual dose, the duration of treatment cycle delay is the number of days the dose was received outside of the original planned dosing schedule. Overall duration of treatment cycle delays will be calculated as the sum of all individual treatment cycle delays during the study. For example, assume the eCRF indicate that there was a treatment cycle delay (question "Treatment cycle delayed" indicated as "Yes" by investigator) then the duration of the individual delay will be [date first dose received after

delay – date last dose received before delay – 21]. For example, if dose is given on days 1 and 24 with the eCRF indicating a delay, the duration of the delay = 24 – 1 – 21 = 2 days.

4.5.1.2 Presentation

Duration of exposure to study intervention in weeks and cycles will be summarized by descriptive statistics. Dose intensity and duration of treatment cycle delays will be summarized by descriptive statistics. Exposure to IP, ie, total amount of study drug received will be listed for all participants. Exposure swimmer plot(s) may be produced, with a line presented for each participant to display relevant exposure and disposition details.

Dosing deviations for study intervention will be summarized with reasons for deviations for the following category: treatment cycle delays. Dosing delays will be derived based on the scheduled dosing dates using the previous dose given as reference. Treatment cycle delays will be summarized.

Analysis of treatment cycle delays may be performed if greater than 20% of participants experience a treatment cycle delay.

4.5.2 Adverse Events

4.5.2.1 Definitions and Derivations

The MedDRA (using the latest or current MedDRA version) will be used to code AEs. AEs will be graded according to the National Cancer Institute Common Terminology Criteria for AEs (CTCAE) (using the CTCAE version 5.0).

Treatment emergent adverse events (TEAEs) are all AEs which onset or worsen in severity following the first administration of IP within the duration of the treatment period, up to and including 90 days after the last dose of study treatment as per the study safety follow-up period but prior to subsequent cancer therapy. Worsening in severity (intensity) is determined by comparison with the pre-treatment CTCAE grade of the AE recorded closest to the start of dosing.

When assigning AEs to the relevant phase of the study, the following rules apply and any deviations must be agreed by the study team:

- Pre-treatment phase: All AEs with a start date after signing the informed consent form, but before the first dose of study treatment that do not subsequently go on to worsen during the treatment emergent phase.

- Treatment emergent phase: All AEs (starting or worsening) on or after the first dose of study intervention and within 90 days post-treatment safety follow-up after last dose of study treatment or but prior to subsequent cancer therapy, whichever occurs first.

- Post-treatment phase: All AEs starting more than 90 days after discontinuation of study treatment or on or after the initiation of subsequent anti-cancer therapy, whichever occurs first.

For rules on missing or partial dates, see Section 3.3.6.

Serious Adverse Events (SAEs)

A serious adverse event (SAE) is any AE that:

- Results in death
- Is immediately life-threatening
- Requires inpatient hospitalization or prolongation of existing hospitalization
- Results in persistent or significant disability/incapacity
- Is a congenital anomaly/birth defect in offspring of the participant
- Is an important medical event that may jeopardize the participant or may require medical treatment to prevent one of the outcomes listed above

AEs of special interest

An adverse event of special interest (AESI) is one of scientific and medical interest specific to understanding of the IP and may require close monitoring and rapid communication by the investigator to the sponsor. The AESIs for sabestomig include events with a potential inflammatory or immune-mediated mechanism and which may require more frequent monitoring and/or interventions such as steroids, immunosuppressants and/or hormone replacement therapy. Additionally, IRRs and events of similar presentation are considered AESIs for sabestomig. These AESIs are being closely monitored in clinical studies with sabestomig.

An immune-mediated adverse event (imAE) is defined as an AESI that is associated with drug exposure, is consistent with an immune-mediated mechanism of action and has no clear alternate etiology. Serologic, immunologic, and histologic (biopsy) data, as appropriate, should be used to support an imAE diagnosis. Appropriate efforts should be made to rule out neoplastic, infectious, metabolic, toxin, or other etiologic causes of the imAE.

The imAEs that may be observed with sabestomig (based on the available nonclinical and clinical data, review of the cumulative literature, reported toxicities for the same class of

agents, and biological plausibility) and are considered AESIs for this study include, but are not limited to:

- Respiratory disorders: Pneumonitis, ILD
- Gastrointestinal disorders: Non-infectious diarrhea, colitis, including bowel perforation, hematochezia, and ileus
- Endocrinopathies: Hyperthyroidism and hypothyroidism (including thyrotoxicosis), adrenal insufficiency, diabetes mellitus, pituitary inflammation (hypophysitis), and hyperpituitarism and hypopituitarism
- Hepatic disorders: Elevated transaminases and hepatitis
- Renal disorders: Nephritis, nephropathy, elevated serum creatinine, acute renal/kidney injury, and glomerulonephritis/sclerosis
- Cardiac disorders: Myocarditis, pericarditis, bradycardia, tachycardia, arrhythmia, ischemia, heart failure, and acute coronary syndrome
- Skin disorders: Rash, pruritus, dermatitis, vitiligo, psoriasis, SJS/TEN, bullous pemphigoid, and lichenoid dermatitis
- Other rare imAEs including encephalitis, hemophagocytic lymphohistiocytosis, encephalitis, lymphocytosis, myositis/polymyositis, uveitis, iridiocyclitis, uritis, arthritis, Guillain-Barre syndrome, myasthenia gravis, vasculitis, hemolytic anemia, aplastic anemia, anemia, and sarcoidosis

Other AEs which are considered to be AESIs with sabestomig include, but are not limited to:

- Infusion-related reactions; hypersensitivity including anaphylactic/anaphylactoid and other allergic reactions
- Immune complex disease and vasculitis
- Cytokine release syndrome; tumor lysis syndrome

Other categories may be added, or existing terms may be modified as necessary. An AstraZeneca medically qualified expert after consultation with the Global Patient Safety Physician has reviewed the AEs of interest and identified which higher-level terms and which preferred terms contribute to each AESI. Further reviews may take place prior to database lock to ensure any further terms not already included are captured within the categories. Preferred terms used to identify AESIs will be listed prior to database lock.

Other significant adverse events (OAE)

No OAEs are planned to be defined for this study.

4.5.2.2 Presentation

All TEAEs will be summarized and listed. AEs which are not treatment emergent will only be listed based on safety analysis set.

TEAEs will be counted once for each participant for calculating percentages of participants experiencing TEAEs. In addition, if the same TEAE occurs multiple times within a particular participant, the highest severity and level of relationship observed will be reported. For tables by MedDRA system organ class (SOC) and MedDRA preferred term (PT), participants with multiple TEAEs will be counted once for each SOC/PT.

An overall summary table of the number of participants experiencing each category of AEs will be produced. The number of participants experiencing TEAEs by MedDRA SOC and PT will be summarized and the type of incidence, severity, and relationship to IP will be summarized. Further splits by CTCAE grade, causal relationship to IP and AEs with Grade 3 or higher will also be also summarized.

Separate tables will present dose limiting toxicities, AEs leading to discontinuation, SAEs, IP-related adverse events.

Details of any deaths will be summarized and listed. AEs leading to death are also summarized.

SAEs

SAEs will be summarized as described above for the TEAEs.

AEs of special interest

Grouped summary tables of certain MedDRA PTs will be produced and may also show the individual PTs which constitute each AESI grouping. Groupings will be based on PTs provided by the medical team prior to database lock, and a listing of the PTs in each grouping will be provided.

Summaries of the above-mentioned grouped AE categories include number and percentage (%) of participants who have:

- At least one AESI presented by event outcome
- At least one AESI causally related to IP

- At least one AESI leading to discontinuation of study intervention

A summary of total duration (days) of AESI will be provided for events which have an end date, and this may be supported by summaries of ongoing AESIs at death and, separately, at data cut-off.

ImAEs

ImAEs will be summarized as described above for the TEAEs.

4.5.3 Deaths

4.5.3.1 Presentation

A summary of deaths will be provided with the number and percentage of participants categorized as:

- Related to disease under investigation only
- AE outcome = death only
- Both related to disease under investigation and with AE outcome = death
- AE with outcome = death > 30 days after last study intervention
- Other deaths

A corresponding listing will also be produced.

4.5.4 Clinical Laboratory, Blood Sample

4.5.4.1 Definitions and Derivations

Laboratory tests are grouped according to chemistry, hematology and coagulation. Laboratory parameters will be assessed at baseline as well as throughout the study.

For chemistry and hematology parameters, laboratory abnormalities with toxicity grades according to the NCI CTCAE version 5.0 will be derived.

Laboratory variables that will be measured are detailed in Table 17 of the CSP.

Change from baseline in hematology, clinical chemistry and coagulation endpoints will be calculated for each post-dose visit. CTC grade is calculated at each visit for tests with CTC grading defined. Maximum on-intervention CTC will be also calculated. Absolute values will be compared to the local laboratory reference range and classified as low (below range),

normal (within range or on limits of range) and high (above range). All values classified as high or low will be flagged on the listings.

Liver Function Parameters

Number and percentage of participants with elevated post-baseline ALT, AST, or Total Bilirubin will be tabulated. Individual participant data where elevated ALT or AST plus total bilirubin fall into the "Potential Hy's law" will be summarized and listed.

4.5.4.2 Presentations

The change in each laboratory parameter from baseline to each post-baseline visit may be summarized graphically.

Laboratory abnormalities occurring from the start of IP administration to the last assessment on treatment will be presented. Worst toxicity grade, rates of grade 3 or higher toxicity, and grade shifts of 2 or more from baseline to the maximum grade may be presented. Summaries indicating hyper- and hypo- directionality of change are produced, where appropriate. Laboratory parameters that cannot be graded via NCI CTCAE are summarized with frequencies of post-baseline laboratory values categorized as low (L), normal (N), or high (H) using laboratory normal ranges compared to baseline.

Listings are provided for all laboratory results.

4.5.5 Clinical Laboratory, Urinalysis

4.5.5.1 Definitions and Derivations

Urinalysis variables that will be measured are detailed in Table 17 of CSP.

CTC grade will be calculated for urinalysis at each visit. Maximum on treatment CTC will be also calculated. Absolute values will be compared to the local laboratory reference range and classified as low (below range), normal (within range or on limits of range) and high (above range). All values classified as high or low are flagged in the listings.

4.5.5.2 Presentations

Shift tables ("Negative", "Trace", "Positive", "0", "+", "++", "+++") from baseline to worst on-intervention results may be produced for urinalysis.

Listings will be provided for all laboratory results including urinalysis.

4.5.6 Vital Signs

4.5.6.1 Definitions and Derivations

Body temperature, blood pressure, pulse rate, and respiratory rate and pulse oximetry will be assessed at baseline and throughout the study.

Normal ranges for vital signs are presented in Table 5.

Table 5 Vital Sign Normal Ranges

VS Test	Outside AZ defined reference range lower limit if	Outside AZ defined reference range upper limit if
BMI	< 18	>25
Temperature (C)	< 36	> 37.5
Systolic Blood Pressure (mmHG)	< 90	> 140
Diastolic Blood Pressure(mmHG)	< 60	> 90
Pulse rate	< 60	> 100
Respiratory rate (min)	< 12	> 20
Pulse rate	< 60	> 100
Oxygen saturation (%)	< 94	NA

4.5.6.2 Presentations

Vital signs will be summarized by study visit which may include descriptive statistics for the value of the parameters and the changes from baseline.

Absolute values and change from baseline for vital signs data at each timepoint may be presented using box plots.

4.5.7 Electrocardiogram

4.5.7.1 Definitions and Derivations

Electrocardiogram (ECG) parameters will be assessed at baseline as well as throughout the study. ECG parameters include: PR, RR, QRS, QT and QTcF. The QTcF is considered as the primary correction method to assess participant cardiac safety.

The ECGs will be performed in triplicate (all 3 ECGs to be performed within a 5-minute time period, at least 1 minute apart). For triplicate ECGs, the mean of the three ECG assessments will be used to determine the value at that time point. Baseline of ECG is the mean of last three ECG assessments prior to first dose. If only two ECG assessments prior to first dose are available, then take the average of these last two ECG assessment prior to first dose as baseline of ECG; if only one ECG assessments prior to first the dose is available, then take the last one ECG assessment prior to the first dose as baseline of ECG.

From these resting 12-lead ECGs values of the QT and RR intervals and the QT interval corrected for heart rate using Fridericia's correction (QTcF) is derived using the following formula:

- QTcF = QT/RR^ (1/3) where RR is in seconds

The values of QTcF (milliseconds) are re-derived from the values of RR and QT during the creation of the reporting database.]

The notable ECG interval values while on treatment are:

- Maximum QTcF intervals > 450 milliseconds, > 480 milliseconds, and > 500 milliseconds.
- Maximum changes from baseline in QTcF > 30, >60, and > 90 milliseconds.

Normal ranges for ECG are presented in Table 6.

Table 6 ECG Normal Ranges

ECG Parameter	Outside AZ defined reference range lower limit if	Outside AZ defined reference range upper limit if
Heart rate (bpm)	< 40	> 100
RR (msec)	< 600	> 1200
PR (msec)	< 120	> 200
QRS (msec)	< 60	> 109
QT (msec) and QTcF (msec)	< 320	> 450

4.5.7.2 Presentations

ECG parameters will be summarized using descriptive statistics by visit and change from baseline in ECG endpoints are calculated for each post-dose visit. Absolute values and change from baseline for ECG data at each timepoint may be presented using box plots.

The number and percentage of participants having notable ECG interval values while on treatment are summarized.

4.5.8 Echocardiogram/Multiple-gated Acquisition Scan (ECHO/MUGA)

4.5.8.1 Definitions and Derivations

Left ventricular ejection fraction (LVEF) will be measured by ECHO or MUGA scan at screening, end of treatment (EOT), and as clinically indicated during the study.

4.5.8.2 Presentations

Absolute value, change from baseline and percentage change from baseline will be presented for LVEF.

Box plots of absolute value and change from baseline over time may also be presented

The overall evaluation (normal/abnormal) will be summarized over time.

4.5.9 B-symptoms

4.5.9.1 Definitions and Derivations

B-symptoms are constitutional symptoms defined as any one or more of the following disease-related symptoms or signs:

- Unintentional weight loss of 10% or more within the previous 6 months
- Fevers > 100.5°F or 38.0°C for ≥ 2 weeks without other evidence of infection, or
- Night sweats for > 1 month without evidence of infection.

4.5.9.2 Presentations

B-symptoms results will be listed for all participants and visits.

4.5.10 Other Safety Assessments

4.5.10.1 Definitions and Derivations

Maximum Tolerated Dose (MTD) Evaluation

See Section 9.4.3.1 of the CSP for the definition of a DLT.

After the escalation phase is completed, DLT rates at each dose level will be estimated by isotonic regression (Ji et al, 2010). The weighted least squares regression model conditional on monotonic non-decreasing DLT rates with increasing dose and use the empirical (observed) DLT rates at each dose level as responses and sample sizes at each dose level as weights, along with the pool adjacent violators algorithm (PAVA) to estimate the DLT rate at each dose level using available software (eg, Cytel EAST or the function pava() from the

R package 'ISO'). Given the DLT estimates for each dose level, the MTD will be selected from all tried dose levels that have not been previously declared to be unsafe with a "de-escalate to the previous lower dose and the current dose will be never used again due to unacceptable toxicity" (DU) decision according to the mTPI-2 decision table. With this constraint, the MTD is determined as the dose level with the DLT estimate closest to the target toxicity level of [redacted]%.

In the case of dose levels with estimated toxicity of equal distance (tied dose levels) from the target toxicity of [redacted]%, the following approach is used (Ji et al, 2010) among all tied dose levels the highest dose level with target toxicity [redacted]% will be selected, unless all tied dose levels have estimated toxicity [redacted]%, in which case the lowest dose level will be selected.

4.5.10.2 Presentations

The MTD evaluation will be based on the DLT evaluable population. The number and percentage of participants with DLT during the dose escalation phase will be presented by treatment groups.

5 INTERIM ANALYSES

Efficacy interim analyses

Interim analyses will be performed in a continuous manner for Part B [redacted]

For each cohort, interim analyses will begin after approximately [redacted] participants in each respective cohort are included in the interim response-evaluable population. For Cohort B1, participants treated at the RP2D in Part A will be included.

Following this initial interim analysis, subsequent interim analyses may be performed after every [redacted] additional participants in Cohort B1 and every [redacted] additional participants in Cohort B2 get included in the interim response-evaluable population, until all participants are enrolled and evaluated. The interim analyses will be conducted to determine futility, where the stopping criteria for futility will be based on the interim response-evaluable set. For any interim futility analyses, an objective response is defined as a CR or PR as per the Modified Lugano criteria and a complete response is defined as a CR as per the Modified Lugano criteria.

Recruitment will not be paused while the participants required for the interim analysis are evaluated. Any decision to stop recruitment in a specific cohort will be at the discretion of AstraZeneca and will be based on emerging efficacy, safety, and tolerability.

Analysis of PFS/OS may be performed at interim analyses using the Interim Response Evaluable population.

Analyses will follow the methodology outlined in this SAP. Outputs presented for the interim analyses will concentrate on ORR, CRR and BoR but may include other outputs, such as PFS/OS. The list of tables, listings and figures required for interim analyses will be flagged in the list of TLFs.

Interim Safety analyses

Interim safety analysis will be conducted in Part B after approximately CCI participants have received the first dose for at least CCI to data cut-off, using the interim safety population. Data for both B1 and B2 will be combined. For cohort B1, participants in Part A who are treated at the RP2D will also be included. The study recruitment may be paused, pending investigation by the sponsor in discussion with SRC, if at least one of the following events occur:

- CCI
  - CCI
  - CCI
  - CCI
  - CCI
  - CCI
  - CCI
- CCI

6 REFERENCES

Cheson et al 2014
Cheson BD, Fisher RI, Barrington SF, Cavalli F, Schwartz LH, Zucca E, et al. Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. J Clin Oncol. 2014;32(27):3059-3068. doi: 10.1200/JCO.2013.54.8800.

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

Guo et al, 2017
Guo W, Wang SJ, Yang S, Lynn H, Ji Y. A Bayesian interval dose-finding design addressing Ockham's razor: mTPI-2. Contemp Clin Trials. 2017;58:23-33.

Ji et al, 2010
Ji Y, Liu P, Li Y, Bekele BN. A modified toxicity probability interval method for dose-finding trials. Clin Trials. 2010;7(6):653-63.AddTextHere

CCI [redacted]

Ribas et al, 2005
Ribas A, Camacho LH, Lopez-Berestein G, Pavlov D, Bulanhagui CA, Millham R, et al. Antitumor activity in melanoma and anti-self responses in a phase I trial with the anti-cytotoxic T lymphocyte-associated antigen 4 monoclonal antibody CP-675,206. J Clin Oncol. 2005;23(35):8968-77.

Younes et al, 2017
Younes A, Hilden P, Coiffier B, Hagenbeek A, Salles G, Wilson W, et al. International Working Group consensus response-evaluation criteria in lymphoma (RECIL 2017). Ann Oncol 2017;28:1436-47.

7 APPENDICES

Appendix A Response Evaluation Using the Modified Lugano (Cheson et al, JCO 2014) and RECIL (Younes et al, Ann Oncol 2017) Criteria

Measurability of Tumor Lesions

Tumor lesions will be categorized as follows:

Baseline: Baseline assessments are those assessments performed as close as possible to the start of study treatment. Baseline scans will be used to confirm eligibility based on the presence of measurable disease and as reference to assess disease response to study treatment. No anti-lymphoma treatment (including palliative RT) should be implemented between the baseline scans and the first planned dose.

Nodal vs Extranodal: a lesion can be categorized as:

- Nodal (a lymph node or a nodal mass), or
- Extranodal (a lesion located in other organs, including liver and bone marrow; Tonsils, Waldever's ring, and spleen are considered nodal tissue).

Measurable Lesions – a measurable lesion is a lesion that can be clearly measured in at least two perpendicular dimensions by CT/MRI (longest diameter [Ldi] and shortest diameter [Sdi]). The Ldi and Sdi will be measured in the transverse plane. A lesion is considered measurable if :

- For nodal lesions: Ldi > 1.5 cm,
- For extranodal lesion (eg hepatic, lung nodules): Ldi > 1 cm,
- The Ldi and the Sdi should be measured on the same slice.

Bulky disease – the presence of a bulky disease is captured by the Ldi on CT scan. For HL, bulky is defined as a single nodal mass, of 10cm or greater than a third of the transthoracic diameter at any level of thoracic vertebrae.

Target Lesions – At baseline, all lesions up to a maximum of 6 measurable nodal or extranodal lesions should be identified as target lesions at baseline and followed throughout the study. It may be the case that, on occasion, the largest lesion does not lend itself to reproducible measurement in which circumstance the next largest lesion which can be measured reproducibly should be selected. Target lesions should be identified from different body regions representative of the participant's overall disease burden and include mediastinal and retroperitoneal disease, if involved. A bulky lesion can also be selected as target lesion. Target lesions must meet the measurability criteria and show FDG-avid uptake. Lesions visible on PET but not measurable on CT/MRI should be assigned as non-target lesions. When selecting the target lesions, consider the following;

– Target lesions should be identified from different body regions representative of the participant's overall disease burden and include mediastinal and retroperitoneal disease, if involved.

– A bulky lesion can also be selected as target lesion.

– Target lesions should be selected from those of the largest size that can be measured reproducibly.

– In certain anatomical sites (inguinal, axillary, and portocaval), normal lymph nodes may exist in a very narrow, elongated form. Clinical judgment should be used when selecting these lesions are target lesions.

– Because of the variable anatomical coverage of scans at different timepoints, it is preferrable not to select as target lesions nodal lesions located in the higher cervical and/or lower inguinal districts.

For the selected target lesions, the sum of the product of the perpendicular diameters (SPD) will be calculated with the percentage change from baseline for assessment of response and nadir for assessment of progression.

Non-target Lesions – All other lesions (including nodal, extranodal, and assessable disease) not selected as target lesions, as well as truly non-measurable sites of disease should be followed as non-measurable disease (eg. Cutaneous, gastrointestinal, bone, spleen, liver, kidney, pleural or pericardial effusions, ascites) and should be factored into the overall response assessment. Non-target lesions will be documented at baseline and throughout the study qualitatively (for example: present, absent/normalized /clear progression). Measurement of these lesions is not required to be documented in the eCRF.

Organ Involvement

Spleen involvement: Spleen will be considered to be normal if size of its vertical length (cranial-caudal measurement) is $\leq$ 13 cm. Spleen vertical length will be assessed at screening and all subsequent response-evaluations. Splenic nodules should be considered as extra-nodal lesions and treated as target, non-target or new lesions are appropriate.

Liver involvement: Given variability in body habitus and the impact of numerous medical conditions, liver size by physical examination or CT scan is not a reliable measure of hepatic involvement by lymphoma. For these reasons, liver size will not be collected in eCRFs. The presence of a diffuse or focal uptake with or without focal or disseminated nodules support liver involvement. Intrahepatic lesions should be considered as target, non-target, or new lesion as applicable.

Bone Marrow involvement: Bone marrow involvement by lymphoma documented by biopsy and/or aspirate or by PET will be reported on the eCRF as positive, negative or unknown.

Other, skin and soft tissue lesions: Skin lesions histologically proven for lymphoma should be preferably selected as Non-target lesions due to variability in measurement on skin photography. If skin is the only site of measurable disease, colour photographs including a ruler should be submitted to BICR (as applicable) and stored in medical records. Measurements should be reported in eCRFs at each disease response assessment if these lesions are selected as target lesions

New Lesions – Lesions which were not present at the baseline, but are visible at a subsequent timepoint:

Nodal lesion of > 1.5 cm in any axis,

Extranodal lesion of > 1 cm in any axis.

In case of appearance of a new extranodal lesion ≤ 1 cm, a biopsy confirmation of lymphoma is always preferable to rule out a benign origin, unless not feasible (lesions too small) or if the procedure is medically contraindicated.The appearance of a new lesion, even if all other lesions are decreasing should be considered progression.

Staging and Classification

Stage: Extent of disease should be described at baseline using the Revised Ann Arbor classification (Table 2).

Symptoms: according to the absence (A) or presence (B) of disease related symptoms. Only participants with HL need to be assigned the designations A or B because symptoms only direct treatment in that disease.

Table 2. Revised Staging System for Primary Nodal Lymphomas

Stage	Involvement	Extranodal (E) Status
Limited		
I	One node or a group of adjacent nodes	Single extranodal lesions without nodal involvement
II	Two or more nodal groups on the same side of the diaphragm	Stage I or II by nodal extent with limited contiguous extranodal involvement
II bulky*	II as above with "bulky" disease	Not applicable
Advanced		
III	Nodes on both sides of the diaphragm; nodes above the diaphragm with spleen involvement	Not applicable
IV	Additional noncontiguous extralymphatic involvement	Not applicable

NOTE. Extent of disease is determined by positron emission tomography–computed tomography for avid lymphomas and computed tomography for nonavid histologies. Tonsils, Waldeyer's ring, and spleen are considered nodal tissue.

*Whether stage II bulky disease is treated as limited or advanced disease may be determined by histology and a number of prognostic factors.

Modified Lugano Response-evaluation Criteria (Cheson et al, 2014)

Table 3. Revised Criteria for Response Assessment

Response and Site	PET-CT–Based Response	CT-Based Response
Complete	Complete metabolic response	Complete radiologic response (all of the following)
Lymph nodes and extralymphatic sites	Score 1, 2, or 3* with or without a residual mass on 5PS†	Target nodes/nodal masses must regress to ≤ 1.5 cm in LDi
	It is recognized that in Waldeyer's ring or extranodal sites with high physiologic uptake or with activation within spleen or marrow (eg, with chemotherapy or myeloid colony-stimulating factors), uptake may be greater than normal mediastinum and/or liver. In this circumstance, complete metabolic response may be inferred if uptake at sites of initial involvement is no greater than surrounding normal tissue even if the tissue has high physiologic uptake	No extralymphatic sites of disease
Nonmeasured lesion	Not applicable	Absent
Organ enlargement	Not applicable	Regress to normal
New lesions	None	None
Bone marrow	No evidence of FDG-avid disease in marrow	Normal by morphology; if indeterminate, IHC negative
Partial	Partial metabolic response	Partial remission (all of the following)
Lymph nodes and extralymphatic sites	Score 4 or 5† with reduced uptake compared with baseline and residual mass(es) of any size	≥ 50% decrease in SPD of up to 6 target measurable nodes and extranodal sites
	At interim, these findings suggest responding disease	When a lesion is too small to measure on CT, assign 5 mm × 5 mm as the default value
	At end of treatment, these findings indicate residual disease	When no longer visible, 0 × 0 mm
		For a node > 5 mm × 5 mm, but smaller than normal, use actual measurement for calculation
Nonmeasured lesions	Not applicable	Absent/normal, regressed, but no increase
Organ enlargement	Not applicable	Spleen must have regressed by > 50% in length beyond normal
New lesions	None	None
Bone marrow	Residual uptake higher than uptake in normal marrow but reduced compared with baseline (diffuse uptake compatible with reactive changes from chemotherapy allowed). If there are persistent focal changes in the marrow in the context of a nodal response, consideration should be given to further evaluation with MRI or biopsy or an interval scan	Not applicable
No response or stable disease	No metabolic response	Stable disease
Target nodes/nodal masses, extranodal lesions	Score 4 or 5 with no significant change in FDG uptake from baseline at interim or end of treatment	< 50% decrease from baseline in SPD of up to 6 dominant, measurable nodes and extranodal sites; no criteria for progressive disease are met
Nonmeasured lesions	Not applicable	No increase consistent with progression
Organ enlargement	Not applicable	No increase consistent with progression
New lesions	None	None
Bone marrow	No change from baseline	Not applicable
Progressive disease	Progressive metabolic disease	Progressive disease requires at least 1 of the following
Individual target nodes/nodal masses	Score 4 or 5 with an increase in intensity of uptake from baseline and/or	PPD progression:
Extranodal lesions	New FDG-avid foci consistent with lymphoma at interim or end-of-treatment assessment	An individual node/lesion must be abnormal with: LDi > 1.5 cm and Increase by ≥ 50% from PPD nadir and An increase in LDi or SDi from nadir 0.5 cm for lesions ≤ 2 cm 1.0 cm for lesions > 2 cm In the setting of splenomegaly, the splenic length must increase by > 50% of the extent of its prior increase beyond baseline (eg, a 15-cm spleen must increase to > 16 cm). If no prior splenomegaly, must increase by at least 2 cm from baseline New or recurrent splenomegaly
Nonmeasured lesions	None	New or clear progression of preexisting nonmeasured lesions

(continued on following page)

Table 3. Revised Criteria for Response Assessment (continued)

Response and Site	PET-CT–Based Response	CT-Based Response
New lesions	New FDG-avid foci consistent with lymphoma rather than another etiology (eg, infection, inflammation). If uncertain regarding etiology of new lesions, biopsy or interval scan may be considered	Regrowth of previously resolved lesions A new node > 1.5 cm in any axis A new extranodal site > 1.0 cm in any axis; if < 1.0 cm in any axis, its presence must be unequivocal and must be attributable to lymphoma Assessable disease of any size unequivocally attributable to lymphoma
Bone marrow	New or recurrent FDG-avid foci	New or recurrent involvement

Abbreviations: 5PS, 5-point scale; CT, computed tomography; FDG, fluorodeoxyglucose; IHC, immunohistochemistry; LDi, longest transverse diameter of a lesion; MRI, magnetic resonance imaging; PET, positron emission tomography; PPD, cross product of the LDi and perpendicular diameter; SDi, shortest axis perpendicular to the LDi; SPD, sum of the product of the perpendicular diameters for multiple lesions.

"A score of 3 in many patients indicates a good prognosis with standard treatment, especially if at the time of an interim scan. However, in trials involving PET where de-escalation is investigated, it may be preferable to consider a score of 3 as inadequate response (to avoid undertreatment). Measured dominant lesions: Up to six of the largest dominant nodes, nodal masses, and extranodal lesions selected to be clearly measurable in two diameters. Nodes should preferably be from disparate regions of the body and should include, where applicable, mediastinal and retroperitoneal areas. Non-nodal lesions include those in solid organs (eg, liver, spleen, kidneys, lungs), GI involvement, cutaneous lesions, or those noted on palpation. Nonmeasured lesions: Any disease not selected as measured, dominant disease and truly assessable disease should be considered not measured. These sites include any nodes, nodal masses, and extranodal sites not selected as dominant or measurable or that do not meet the requirements for measurability but are still considered abnormal, as well as truly assessable disease, which is any site of suspected disease that would be difficult to follow quantitatively with measurement, including pleural effusions, ascites, bone lesions, leptomeningeal disease, abdominal masses, and other lesions that cannot be confirmed and followed by imaging. In Waldeyer's ring or in extranodal sites (eg, GI tract, liver, bone marrow), FDG uptake may be greater than in the mediastinum with complete metabolic response, but should be no higher than surrounding normal physiologic uptake (eg, with marrow activation as a result of chemotherapy or myeloid growth factors).

†PET 5PS: 1, no uptake above background; 2, uptake ≤ mediastinum; 3, uptake > mediastinum but ≤ liver; 4, uptake moderately > liver; 5, uptake markedly higher than liver and/or new lesions; X, new areas of uptake unlikely to be related to lymphoma.

Deauville 5-point scale: The use of PET/CT is standard for FDG-avid lymphomas and whenever aggressive transformation is suspected. Variation in FDG uptake in a nodal or extranodal sites indicative for lymphoma will be visually assessed using the Deauville 5-point scale. The scale ranges from 1 to 5, where 1 is best and 5 is the worst. Each FDG-avid (or previously FDG-avid) lesion is rated independently.

Deauville 5-point scale

Score	Description
Score 1	No uptake above background
Score 2	Uptake ≤ mediastinum
Score 3	Uptake > mediastinum, but ≤ liver
Score 4	Uptake moderately > liver [a]
Score 5	Uptake markedly higher than liver and/or new lesion [a]
X (New)	Areas of uptake unlikely to be related to lymphoma [b]

[a] Barrington et al (2014) suggest the following: "The terms moderately and markedly were not defined initially, because there were insufficient data to define scores quantitatively. Meanwhile, it is suggested according to published data that score 4 be applied to uptake greater than the maximum SUV in a large region of normal liver and score 5 to uptake 2 × to 3 × greater than the maximum SUV in the liver."

[b] Barrington SF, Mikhaeel NG, Kostakoglu L, et al: Role of imaging in the staging and response assessment of lymphoma: Consensus of the International Conference on Malignant Lymphomas Imaging Working Group. J Clin Oncol 32:3048-3058, 2014
SUV: standardized uptake value.

It is recognized that in Waldeyer's ring or extranodal sites with high physiologic uptake or with activation within spleen or bone marrow (eg, following chemotherapy or G-CSF treatment), the uptake may be greater than normal mediastinum and/or liver. In this circumstance, complete CR may be inferred if uptake at sites of initial involvement is no greater than surrounding normal tissue even if the tissue has high physiologic uptake. In presence of new areas of FDG uptake that are unlikely to be related to lymphoma but are thought to be of inflammatory or infectious origin by the Investigator, the reporter should assign 'X' in addition to the Deauville score. For example, if there is a complete resolution of all uptake with no abnormal nodes but new lesions likely to be related to another etiology occur, the total score should be of DS 1X rather than DS 5. Additionally, these lesions should be tracked and should not be used to determine the PET response at subsequent assessments.

Evaluation of Overall Response

The table below provides overall responses for all possible combinations of tumor responses in target and non-target lesions with or without the appearance of new lesions.

Evaluation of Overall Response Using the Modified Lugano Criteria

Target Lesions	Non-target Lesions	New Lesions	Overall Response
CR	CR	No	CR
No target lesion [a]	CR	No	CR
CR	Not evaluable [b]	No	PR
CR	Non-CR/ non-PD	No	PR
PR	Non-PD and not evaluable [b]	No	PR
Stable disease	Non-PD and not evaluable [b]	No	Stable disease
Not all evaluated	Non-PD	No	Not evaluable
No target lesion [a]	Not all evaluated	No	Not evaluable
No target lesion [a]	Non-CR/ non-PD	No	Non-CR/ non-PD
PD	Any	Yes/No	PD
Any	Unequivocal PD	Yes/No	PD
Any	Any	Yes	PD
No target lesion [a]	Unequivocal PD	Yes/No	PD
No target lesion [a]	Any	Yes	PD

a Defined as no target lesion at baseline.

b Not evaluable is defined as either when no or only a subset of lesion measurements are made at an assessment.

Abbreviation: CR: complete response; PD: progressive disease; PR: partial response.
Modified Lugano: Response-evaluated Criteria

Integrated Radiological Overall Response

PET-based Overall Response	CT/MRI-based Overall Response	Radiological Overall Response
CMR	Any [b]	CR
PMR	Any [b]	PR
NMR	Any [b]	SD
UNK [a]	CR/PR/SD + prior PET response of CMR (as long as the CT response remains stable/improves as compared to the previous timepoint when PET was available)	CR
UNK [a]	CR/PR + prior PET PMR or no PET (as long as the CT response remains stable/improves as compared to the previous timepoint when PET was available)	PR
UNK [a]	CR/PR + prior PET NMR	SD
UNK [a]	Any response except for PD (but has worsened from the previous timepoint when PET was available)	UNK
UNK [a]	UNK	UNK
UNK [a]	PD	PD
PMD	Any	PD

a PET not done/not available

b Depending on the histology of the lymphoma. For lymphomas with variable avidity, CT PD could lead to an overall assessment of PD

CMR: complete metabolic response; CR: complete response; NMR: no metabolic response; PD: progressive disease; PMD: progressive metabolic disease; PR: partial response; SD: stable disease; UNK: unknown.

Overall Response Assessment

Radiological Overall Response	Bone Marrow Aspirate/Biopsy	Clinical findings (Physical Examination/ Lesion Biopsy)	Overall Response Assessment
CR	Negative (+ 28 days from radiological assessment) or negative at baseline	No evidence of PD	CR
CR	Positive at baseline and positive (no new involvement) + 28 days from radiological assessment; or positive at the previous timepoint and not repeated	No evidence of PD	PR
PR/SD	Any but new or recurrent	No evidence of PD	PR/SD

	involvement		
PD	Any	Any	PD
Any	New or recurrent BM involvement	Any	PD
Any	Any	New or recurrent lymphoma manifestation	PD

BM: bone marrow; CR: complete response; PD: progressive disease; PR: partial response; SD: stable disease; UNK: unknown.

Additional response assessment guidelines:

- Nodes or Extranodal lesions that split when disease is responding: If a confluent nodal mass splits into several discrete nodes, the individual product of the perpendicular diameters (PPDs) of the nodes should be summed together to represent the PPD of the split lesion; this PPD is added to the sum of the PPDs of the remaining lesions to measure response. If subsequent growth of any or all of these discrete nodes occurs, the nadir of each individual node is used to determine progression (as if each individual node was selected as a target lesion at baseline).

  Nodes or extranodal lesions that become confluent: If a group of target lymph nodes becomes confluent, the PPD of the current confluent mass should be compared with the sum of the PPDs of the individual nodes, with more than 50% increase in the PPD of the confluent mass compared with the sum of individual nodes necessary to indicate progressive disease. The LDi and shortest diameter are no longer needed to determine progression.

- The presence of residual symptoms in the absence of detectable disease by imaging does not preclude the designation CR.

International WG Consensus Response evaluation Criteria in Lymphoma (Younes et al 2017 [RECIL 2017])

Table 1: RECIL 2017 Response Categories based on assessment of target lesions

Table 1. RECIL 2017: Response categories based on assessment of target lesions

	% Change in sum of diameters of target lesions from nadir				
	CR	**PR**	**MR[a]**	**SD**	**PD**
% change from baseline	• Complete disappearance of all target lesions and all nodes with long axis <10mm. • ≥30% decrease in the sum of longest diameters of target lesions (PR) with normalization of FDG-PET	≥30% decrease in the sum of longest diameters of target lesions but not a CR	≥10% decrease in the sum of longest diameters of target lesions but not a PR (<30%)	<10% decrease or ≤ 20% increase in the sum of longest diameters of target lesions	• >20% increase in the sum of longest diameters of target lesions • For small lymph nodes measuring <15 mm post therapy, a minimum absolute increase of 5 mm and the long diameter should exceed 15 mm • Appearance of a new lesion
FDG-PET	Normalization of FDG-PET (Deauville score 1-3)	Positive (Deauville score 4-5)	Any	Any	Any
Bone marrow involvement	Not involved	Any	Any	Any	Any
New lesions	No	No	No	No	Yes or No

CR, complete response; CT, computerized tomography; FDG-PET, [^{18}F]2-fluoro-2-deoxy-D-glucose; MR, minor response; PD, progression of disease; PR, partial response; SD, stable disease.
[a]A provisional category.

Table 2. Calculating sum of diameters to include small responsive lymph nodes

Target lesions	Baseline measurement (long axis; cm)	Nadir actual measurement (cm) method 1	Nadir normalized measurement (cm) method 2
Lesion 1	1.6	0.9	0 (resolved)
Lesion 2	1.7	1.4	1.4
Lesion 3	2	1.8	1.8
Sum of diameters	5.3	4.1	3.2
% change from baseline	N/a	23	40
Response designation	N/A	Minor response	Partial response (or CR if PET is negative)

CR, complete response; PET, positron emission tomography.

Appendix B Time windows

B 1 Tumor assessments

Timepoint	Naming convention in outputs	Window
Screening	Screening	-28 – 1
At least 10 days after sabestomig administration in Cycle CCI	CCI	CCI
At least 10 days after sabestomig administration in Cycle CCI	CCI	CCI
At least 10 days after sabestomig administration in Cycle CCI	CCI	CCI
1 year after date of first infusion	After 1 year	314 – 455 [a]
1 year and 6 months after date of first infusion	After 1.5 years	456 – 638 [a]
2 years after date of first infusion	After 2 years	639 – 821 [a]
2.5 years after date of first infusion Q6M until EoS	After 2.5 years ...	822 – 913 [a] ...

Any scan not within the above windows will be assigned to an unscheduled timepoint.

[a] If there are multiple scans in this period, the earliest scan will be assigned to the timepoint.

B 3 Time windows for Part A

Visit	Window					
	Hematology | Chemistry	**Physical examination | Vital Signs**	**Coagulation**	**ECG**	**B-symptoms**	**Thyroid function**
Cycle 1 Day 1	1	1	1	1	1	1
Cycle 1 Day 3	2 - 5					
Cycle 1 Day 8	6 - 11	2 - 11	2 - 11	2 - 15		
Cycle 1 Day 15	12 - 18	12 - 18	12 - 18			2 - 18
Cycle 2 Day 1	19 - 32	19 - 25	19 - 32	16 - 25	2 - 32	19 - 32
Cycle 2 Day 8		26 - 36		26 - 36		
Cycle 3 Day 1	33 - 53	37 - 53	33 - 53	37 - 53	33 - 53	33 - 53
Cycle 4 Day 1	54 - 74	54 - 74	54 - 74	54 - 74	54 - 74	54 - 74
Cycle 5 Day 1	75 - 95	75 - 95	75 - 95	75 - 95	75 - 95	75 - 95

B 3 Time windows for Part B

Visit	Window					
	Hematology | Chemistry	**Physical examination | Vital Signs**	**Coagulation**	**ECG**	**B-symptoms**	**Thyroid function**
Cycle 1 Day 1	1	1	1	1	1	1
Cycle 1 Day 8	6 - 11	2 - 11	2 - 11	2 - 15		
Cycle 1 Day 15	12 - 18	12 - 18	12 - 18			2 - 18
Cycle 2 Day 1	19 - 32	19 - 25	19 - 32	16 - 25	2 - 32	19 - 32
Cycle 2 Day 8		26 - 36		26 - 36		
Cycle 3 Day 1	33 - 53	37 - 53	33 - 53	37 - 53	33 - 53	33 - 53
Cycle 4 Day 1	54 - 74	54 - 74	54 - 74	54 - 74	54 - 74	54 - 74
Cycle 5 Day 1	75 - 95	75 - 95	75 - 95	75 - 95	75 - 95	75 - 95

Appendix C PK Parameter Derivation

Pharmacokinetic parameter definitions and derivations are described in the table below. The following PK Parameters are calculated and summarized for Cycle 1 Day1 and for China population with intensive PK sample collected:

Parameter Symbol (used in CSP/SAP/ CSR)	Definition	Derivation
AUClast	Area under serum concentration-time curve from time 0 to the last quantifiable concentration	Using the linear up/log down trapezoidal rule
AUCinf	Area under serum concentration-time curve from time 0 to infinity	Calculated by AUC and then extrapolated by Clast/λz to infinity
Cmax	Maximum observed concentration	
Ctrough	Observed lowest concentration before the next dose is administered	
tmax	Time to reach maximum observed concentration	
t½λz	Terminal elimination half-lide	ln 2/λz
CL	Total body clearance	Dose/AUCinf
Vss	Volume of distribution at steady state	MRT x CL
Vz	Volume of distribution based on the terminal phase	CL/λz
AUClast/D	Dose normalized AUClast	AUClast/Dose
AUCinf/D	Dose normalized AUCinf	AUCinf/Dose
Cmax/D	Dose normalized Cmax	Cmax/Dose
Ctrough/D	Dose normalized Ctrough	Ctrough/Dose

Where appropriate, additional parameters may be calculated.

For Cycle 3, Day 1 (dose escalation) and Cycle 4, Day 1 (dose expansion), both Cmax and Ctrough will be determined as described in the table above.

In addition, the following diagnostic PK Parameters are calculated and summarized for Cycle 1 Day1 and for China population with intensive PK sample collected:

Parameter Symbol (used in CSP/SAP/ CSR)	**Definition**	**Derivation**
AUCextr	Extrapolated area under the curve from tlast to infinity, expressed as percentage of AUCinf	
λz	Terminal elimination rate constant	Estimated from linear regression of the terminal part of the log concentration versus time curve
λz lower	Lower (earlier) time used for λz determination	
λz upper	Upper (later) time used for λz determination	
λzN	Number of data points used for λz determination	
λz span ratio	Time period over which λz was determined as ratio of t½λz	λz period (ie, λz upper- λz lower) /t½λz
Rsq	Statistical measure of fit for the regression used for λz determination	
Rsq adj	Statistical measure of fit for the regression used for λz determination adjusted for the number of used data points	
tlast	Time of last quantifiable concentration	

Signature Page for VV-RIM-02733273 System v8.0
D9571C00001 Statistical Analysis Plan Edition 3

Approve: Document Level Task Verdict: Approved	PPD proval 30-Sep-2024 11:51:46 GMT+0000

Approve: Document Level Task Verdict: Approved	PPD 30-Sep-2024 13:19:53 GMT+0000

Signature Page for VV-RIM-02733273 System v8.0
D9571C00001 Statistical Analysis Plan Edition 3