Æterna Zentaris

Doc. Ref. No.: AEZS-108-050_SAP_4 / FINAL

March 10, 2017

Statistical Analysis Plan

Study AEZS-108-050 / Phase III

Randomized controlled study comparing AEZS 108 with doxorubicin as second line therapy for locally advanced, recurrent or metastatic endometrial cancer

SAP Approval

The undersigned have reviewed the format and content of this prospective statistical analysis plan (SAP) and have approved it for use to analyze the AEZS-108-050 study data.

Coordinating investigator

David Scott Miller, M.D.,
F.A.C.O.G., F.A.C.S.

Date, Signature

MAR 16 2017

Sponsor, Aeterna Zentaris

- Study Manager / Sr. Medical Advisor

Herbert Sindermann, PhD

16. Mar 2017

Date, Signature

- Medical Monitor

Nicola Ammer, M.D.

16. März 2017

Date, Signature

Biostatistician, IDDI

Katrien Verschueren, Ir. M.Sc.

17MAR 2017

Date, Signature

Document History

Version	Date	Changes made since previous version
0.01	November 15, 2012	First version
1.0	March 13, 2013	Specified two key secondary efficacy variables (PFS and ORR) that will be tested using hierarchical procedures to protect overall Type I Error and deleted time to progression as a secondary efficacy endpoint.
1.01	March 13, 2014	Section 8 added to describe the analysis of data coming from the PK sub-study. Update of Flow chart from protocol version 3.1.
1.02	December 19, 2014	Reference to latest version of protocol Study objectives adapted to reflect the latest version of protocol's objectives Details on PK analyses removed because this will be taken up in separate supplements. In SAP only summary is presented
1.03	January 23, 2015	Details on analysis of QoL included
2	March 23, 2015	Inclusion of definition of modified ITT population (used for sensitivity analysis on OS primary endpoint) Second final version
3	September 26, 2016	Removal of LHRH receptor related analysis due to non-availability of corresponding receptor assay and data Inclusion of sub-analyses for setting of prior chemotherapy (adjuvant vs advanced). Populations (safety and ITT) are also specified for summary table Inclusion of sub-analyses for key secondary efficacy endpoints Inclusion of analysis of prophylactic/symptomatic antiemetic/antiallergic concomitant treatment Patient disposition summaries will also include number of subjects in modified ITT and PP (next to ITT and safety populations) More detailed specification of how Study Treatment Exposure summary will be made Supportive and sensitivity analyses on primary efficacy endpoint OS is redefined: -different censoring; - definition of different covariates used in exploratory analyses of the effect of baseline characteristics and other covariates (see section 7.1.3 for details) Supportive and sensitivity analyses on secondary efficacy endpoints ORR and PFS is redefined (identical to analyses on OS) In section 9.1. Adverse events, certain details on content of AE listings is given 9.2. Laboratory assessments: for WBC, neutrophils (absolute value) and thrombocytes extra shift table only considering cycle 1 lab results is foresee 9.5. Cardiac assessments is extended and more detailed (both for ECG and LVEF)
4	March 10, 2017	In 4. General conventions: -add AEZS-108 and Doxorubicin will be used; -elaborate on definition of baseline

In 5.1. Disposition: correct typo in country abbreviations per territory

In 6.1. Prior In 6.1. Treatment for Endometrial Cancer: remove ITT and explain how to analyze

In 6.2. Prior and Concomitant Medications: explain how to analyze

In 6.3. Study treatment exposure: addition of summaries will be made on patients receiving prophylactic antiemetic treatment per treatment cycle, and patients receiving anti-allergic treatment per treatment cycle

Efficacy analysis:

For backward selection in multivariate Cox model and logistic regression model: the significance level to retain a variable in the model will be set at 10%

Sensitivity analysis for ORR: logistic regression models will be used. For ORR, in addition: duration of overall response, duration of stable disease and duration of follow-up will be analyzed

PFS will be analyzed in the subset of patients from MITT that have CR, PR or SD as best overall response

Slight adaptation in description for mixed model approach for analysis of questionnaires

Adverse event analysis:

Clarification of time on treatment (date of last dose before AE onset – C1D1) and time since last dose (start date of AE – date of last dose before AE onset).

Addition: If treatment-relatedness is missing, AE will be considered as RELATED.

Extra AE tables by PT only (>= certain cut-off)

Addition of tables for serious TEAEs grade 3/4 and 5

ECG: update of formulas for calculation of QTcB and QTcF

Addition of Appendix E: list of TLFs

List of Abbreviations

Abbreviation	Definition
AE	adverse event
ANC	absolute neutrophil count
ALT	alanine aminotransferase
AST	aspartate aminotransferase
BSA	body surface area
BUN	blood urea nitrogen
CR	complete response
CT	Computerized Axial Tomography
CTCAE	Common Terminology Criteria for Adverse Events
°C	degrees Celsius
DSMB	Data and Safety Monitoring Board
ECOG	Eastern Cooperative Oncology Group
FDA	Food and Drug Administration
ICH	International Conference on Harmonization
ITT	intent-to-treat
MITT	modified intent-to-treat
LLOQ	lower limit of quantification
MedDRA	Medical Dictionary for Regulatory Activities
Mg	Milligram
MRI	magnetic resonance imaging
NC	no change
NCI	National Cancer Institute
ORR	overall response rate
OS	overall survival
PD	progressive disease
PFS	progression-free survival
PK	Pharmacokinetics
PR	partial response
PT	MedDRA Preferred Term
SAE	serious adverse event
SOC	MedDRA System Organ Class
TEAE	treatment-emergent adverse event

Table of Contents

1 INTRODUCTION

This statistical analysis plan (SAP) is based on:

- Protocol AEZS-108-050, protocol version 3.2, dated September 12, 2014
- FDA guidance on cancer trial endpoints (Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics)
- EMA Guideline on the evaluation of anticancer medicinal products in man (EMA/CHMP/205/95)
- ICH guidelines E9 (Statistical Principles for Clinical Trials).

The purpose of this document is to provide details on study populations and on how the variables will be derived, how missing data will be handled, and how censoring procedures will be applied to time to event related variables as well as details on statistical methods to be used to analyze the safety and efficacy data from the study.

The document may evolve over time, for example, to reflect the requirements of protocol amendments or regulatory requests. However, the final SAP must be finalized, approved by the Sponsor, and placed on file before the database is locked prior to the final analysis. Deviations from the approved plan will be noted in the clinical study report.

2 STUDY DESCRIPTION

2.1 Study Objectives

The primary objective of this study is:

- Compare the overall survival (OS) of patients treated with AEZS-108 to the OS of patients treated with doxorubicin.

The secondary objectives include:

- Compare efficacy based on progression-free survival (PFS), objective response rate (ORR) and clinical benefit rate (CBR).
- Compare safety.
- Determine the impact of these regimens on patient-reported quality of life during and for up to 1 year after completion of study treatment.
- Assess the pharmacokinetics and exposure-response relationships of AEZS-108, doxorubicin and doxorubicinol in a PK sub-study at selected sites and by sparse PK sampling, respectively.
- Assess the acute effects on electrocardiographic parameters (e.g., QT/QTc interval) within PK sub-study.

2.2 Study Design

This is an open-label, randomized, active-controlled, two-arm Phase III study to compare the efficacy and safety of AEZS-108 and doxorubicin. The study will include about 500 patients with advanced, recurrent or metastatic endometrial cancer who have

progressed and who have received one chemotherapeutic (i.e., cytotoxic) regimen with platinum and taxane (either as adjuvant or first line treatment).

Patients will be centrally randomized in a 1:1 ratio to receive treatment with either AEZS-108 (Arm A) or doxorubicin (Arm B).

During ongoing treatment, response will be evaluated every 3 cycles; earlier reassessments should be scheduled to verify a response (at least 4 weeks after first observation of the response) or in case of suspected progression. Patients, who have gone off-treatment for reasons other than progression, will be reassessed every 12 weeks until progression. All patients will be followed-up for survival.

On a regular basis, at intervals no longer than 6 months, results from safety analyses will be submitted to an independent Data and Safety Monitoring Board (DSMB) that will advise the Sponsor of potentially critical findings.

The final analysis will be performed after about 384 deaths have been observed. There will be two planned interim analyses; the first will be for futility only, the second will be for safety and efficacy.

Based on the availability of the assay for LHRH receptor expression in tumor specimens, subgroup analyses stratified for extent of LHRH receptor expression was planned to assess the predictive value of the LHRH receptor assay. Due to non-availability of a validated assay, no such stratified analyses can be performed.

2.3 Randomization and Blinding

This is an open-label study.

3 ANALYSIS POPULATIONS

3.1 Intent-to-Treat Population

The Intent-to-Treat (ITT) population will consist of all randomized patients. Analyses of this population will assign patients the treatment they were scheduled to receive, regardless of any errors of dosing or dose modifications. Patients who received treatment in the group that is different from the one to which they have been randomized to, will be noted in the CSR.

3.2 Modified Intent-to-Treat Population

A modified Intent-to-Treat (MITT) population will consist of all randomized patients, who received at least one dose of study treatment, i.e. the MITT population will exclude patients who never received study treatment.

3.3 Safety Population

The safety population will include all randomized patients who received at least one dose of study treatment. In the safety analyses, patients will be included in the treatment arm that they have actually received. Patients who are included in a treatment group that is different from the one to which they have been randomized will be noted.

3.4 Per Protocol Population

The Per-Protocol (PP) population will include all randomized patients without major protocol deviations. What constitutes a major protocol deviation will be determined and documented prior to database lock. Protocol deviations will also be presented in the clinical study report.

4 GENERAL CONVENTIONS

Unless otherwise stated, all analyses will be performed using SAS Version 9.2 or higher and all hypothesis tests will be conducted at a two-sided significance level of 0.05. P-values will be presented with 3 decimals and p-values that are less than 0.001 will be presented as < 0.001.

Continuous (non-survival related) data will be summarized using descriptive statistics: number of observations (n), mean, standard deviation (SD), median, minimum, and maximum. Frequencies and percentages will be used to summarize categorical (discrete) data. Presentations of categorical data will generally suppress percentages for items where the count is zero in order to draw attention to the nonzero counts. In general, mean, standard deviation, median, 25th and 75th percentiles, minimum, maximum, and percentages will be presented with one decimal.

Unless otherwise stated, confidence intervals, when presented, will be constructed at the two-sided 95 % level. For binomial variables, the 95 % confidence intervals will be constructed using the normal approximation without continuity correction.

In the tables the following indication for treatment will be used: AEZS-108 and Doxorubicin. Data listings will show data presented in the summary tables by study drug, country, center and patient number.

4.1 Definition of Baseline

In general, the last observed measurement prior to first administration of study treatment will be considered the baseline measurement. In practice, for questionnaires, LVEF and ECG baseline is defined as screening visit. For laboratory parameters, vital signs and ECOG cycle 1 day 1 (C1D1) is considered as baseline. If C1D1 assessment is not available or if assessment on C1D1 is not before first treatment, screening value will be taken as baseline value. If date of first treatment is missing, screening value will be taken as baseline value.

4.2 Definition of Time

For the purpose of summarizing efficacy data, time will be defined relative to the date of randomization. Unless otherwise stated, for visits (or events) that occur on or after randomization, time is calculated as:

time (days) = visit date (event date) – date of randomization + 1.

For visits (or events) that occur prior to randomization, time is calculated as:

time (days) = visit date (event date) – date of randomization to study treatment.

For listings (such as for adverse events) of the quantity 'days since first/last dose' is defined as:

days since first/last dose = event date – date of first/last dose.

Events that occur on the same day as the first/last dose will therefore be described as occurring zero days from the first/last dose. In most cases, listings will include the number of days since first/last dose of AEZS-108/comparator (doxorubicin).

4.3 Tumor Assessment Schedule

Tumor assessments will occur every 3 cycles (± 7 days) during ongoing treatment, then if required beyond the end of treatment every 3 months (± 7 days) while the patient is on study. The last assessment will occur when approximately 384 randomized patients have died.

4.4 Missing and Partial Data

In general, other than for partial dates, missing data will not be imputed and will be treated as missing. The algorithms for imputation of partial dates vary depending upon the parameter and are presented in Section 12.4 (Appendix D - Imputation Rules for Missing Dates.)

4.5 Software

Unless otherwise stated, all analyses will be conducted using SAS Version 9.2 or higher.

4.6 Changes to Planned Analyses

Draft versions of the SAP will be numbered sequentially as Version x-1.0i. The final approved version will be numbered as Version x.0. The Clinical Study Report will document any change made after database lock.

5 DESCRIPTION OF THE STUDY POPULATIONS

The paragraphs below specify presentations of data for both the ITT and safety populations. If, however, the two populations are almost the same (i.e., safety population is 95 % or more than the ITT population), then only the ITT tables will be presented. A summary of differences between the two populations will be noted in the CSR.

5.1 Disposition

Patient disposition summaries will be presented by treatment arm and will include the number of patients enrolled, randomized, the number and percentage of randomized patients in the ITT, the modified ITT, the PP, and the safety populations. The summaries will also include the reasons for permanent discontinuation of study treatment and study. Disposition by investigational sites and by territory will also be presented. Trial territories are defined as follows:

- North America: US, CA;
- W/N-Europe/Israel: AT, BE, DE, DK, ES, FI, GB, IE, IL, IT, NL, NO, PL;
- Eastern Europe: BG, BA, BY, CZ, RU, UA, RO.

5.2 Demographic and Baseline Characteristics

A summary of demographics and baseline characteristics will be presented by treatment arm for the ITT and safety populations.

The demographic characteristics consist of age, age category (defined below), sex, ethnicity, race, and Eastern Cooperative Oncology Group (ECOG) performance status. Baseline height and weight will also be presented using standard descriptive statistics.

Age (years) will be calculated as (date of informed consent – date of birth) / 365.25.

Age categories: < 65 years and 65 years or older, will be presented using frequencies and percentages.

The number and percentage of patients' ethnicity as well as reported race category will also be reported.

5.3 Medical History

A data listing will present medical history. If useful for interpretation of the results of the study, a table will summarize the data.

5.4 History of Endometrial Cancer

Tables will present information regarding the patients' history of endometrial cancer, including time since initial diagnosis to randomization, primary diagnosis. Summaries will be prepared for the ITT and safety populations and will display the information by treatment arm and overall. All details of endometrial cancer history (including origin of tumor specimen) will also be listed.

5.5 Baseline Comparisons

The two treatment arms will be assessed descriptively for comparability of demographic and baseline characteristics. Data to be evaluated will include at least age, sex, race, and components of disease severity assessment and disease status (e.g., locally advanced, recurrent or metastatic endometrial cancer).

6 TREATMENTS AND MEDICATIONS

6.1 Prior Treatment for Endometrial Cancer

Use of prior treatment for endometrial cancer will be presented as subject frequency counts and percentages for the safety population. Descriptive statistics for type of drugs per patient will also be shown. Participants are only counted once for each type of therapy.

Prior chemotherapy will be coded using the World Health Organization (WHO) dictionary and summarized by treatment arms for the safety population;

- By WHO Drug base substance preferred name using frequency counts and percentages
- By regimen number using frequency counts and percentages on unique combinations of WHO Drug base substance preferred names.

For platinum and taxane based treatment the use as adjuvant or advanced treatment will be specified.

6.2 Prior and Concomitant Medications

All medications recorded on the CRFs will be coded using the WHO dictionary. Prior and concomitant medications will be summarized by treatment arm in the safety population by anatomical therapeutic chemical (ATC) Class Level 4 and WHO Drug base substance preferred name. Participants are only counted once for each preferred term.

Prior medications are defined as medications with stop dates occurring before the date of first administration of any study treatment component. Concomitant medications are defined as medications with start dates occurring on or after the date of first administration of any study treatment component and no more than 30 days after the last administration of any study treatment. Medications with start and stop dates that bracket the date of first administration of any study treatment will be summarized as both prior and concomitant medications.

For the purpose of summarizing prior and concomitant medications, incomplete medication start and stop dates will be imputed as detailed in Section 12.4. Based on imputed start and stop dates, medications that clearly stopped prior to date of first administration of any study treatment component will be included in the prior medications table, and medications that clearly started on or after date of first administration of any study treatment component will be included in the concomitant medications table. All other medications will be included in both the prior and concomitant medications tables.

6.3 Study Treatment Exposure

Study treatment exposure will be summarized by treatment arm in the ITT and safety populations.

The following will be summarized by treatment arm using descriptive statistics:

- number and percentage of patients who received at least one dose (‘cycle’)
- duration of exposure per patient, calculated as (date of last dose [start date of last cycle] – date of first dose [start date of first cycle] + 1),
- number of doses (cycles) received per patient,
- cumulative dose (in mg/m^2) received per patient,
- patients receiving prophylactic antiemetic treatment per treatment cycle, and
- patients receiving anti-allergic treatment per treatment cycle.

In addition, a table will be made summarizing duration of cycle 1 administration and duration of administration – all cycles combined.

6.4 Study Treatment Modifications

Treatment modifications will be summarized by treatment arm in the ITT and safety populations for each randomly assigned study treatment.

The number and percentage of patients with doses delayed (withheld) and doses reduced, and reasons for these dose changes will be summarized by treatment cycle. The total number of dose delays and reductions will be shown.

6.5 Subsequent Anticancer Therapy

A summary of the number and percentage of patients receiving subsequent anticancer therapy after discontinuation of study treatment will be presented by treatment arm in the ITT population.

7 EFFICACY ANALYSES

7.1 Primary Efficacy Outcome – Overall Survival

The primary efficacy outcome is overall survival (OS).

7.1.1 OS: Definition

OS is defined as the time from randomization to death for any cause during the study. Patients without confirmed vital status are censored on the day they are last known alive. Every reasonable effort will be made to establish vital status of each patient on the last day of the study, that is, the clinical cutoff date established after approximately 384 patients have died. OS will be calculated as follows:

For those who die during the study,

overall survival (days) = date of death – randomization date + 1;

For those who are alive or vital status cannot be determined at the end of the study,

overall survival (days) = date of censoring – randomization date + 1, with censor indicator indicating a censor.

7.1.2 OS: Primary Analysis

The primary analysis of OS will be performed using the ITT population.

The primary efficacy analysis will be performed using the ITT population. The final analysis, which is event-based, will be conducted after approximately 384 randomized patients have died. In the primary analysis, a log-rank test with an overall two-sided Type I error rate of 0.05 after taking the interim analyses into account will be used to compare OS between the two treatment arms via a SAS lifetest procedure. Kaplan-Meier estimates will be used to calculate median OS and the 95 % confidence interval of the median OS. The proportion of patients alive at six and 12 months (from randomization date) and the 95 % confidence intervals for these estimated proportions, if appropriate, will be presented.

A Cox model with treatment effects will be used to estimate the hazard ratio and perform hypothesis testing. The estimated hazard ratio and the 95 % confidence interval of the hazard ratio will be presented.

7.1.3 Supportive and Sensitivity Analyses

To assess the robustness of the study results, supportive analyses of OS will be performed using modified ITT and per protocol population. In addition, the following sensitivity OS analyses will be performed, if appropriate:

Patients will be censored at the time when they received other non-protocol required anticancer treatment, if applicable.

Patients will be censored at the time when they drop out of the study due to AEs.

Exploratory analyses of the effect of baseline characteristics and other covariates will be conducted using Cox regression models. Cox models with treatment (forced into the model) and different (see listed underneath) covariates each at the time, and a global Cox model with treatment (forced into the model) and different covariates using backward selection will be made. The significance level to retain a variable in the model will be set at 10%. The following covariates will be considered:

- Age (entered as continuous covariate)
- Endometrial cancer type: Type I vs Type II
  Type I: endometrioid, squamous, adenocarcinoma (recorded under 'other')
  Type II: serous, clear cell, mixed, carcinosarcoma (recorded under 'other')
  (additional histology variants captured under 'other', will be assigned to Type I or II as medically appropriate)
- Histology: endometrioid vs serous vs clear cell vs squamous cell carcinoma vs other
- Tumor grade: Gx vs G1 vs G2 vs G3 vs G4 vs other
- Prior radiotherapy (yes vs no)
- Time since diagnosis (< 12 months vs $\geq$ 12 months)
- Stage of disease at study entry (advanced or recurrent disease vs metastatic disease)
- Potential center-effects
  Territory (North America vs W/N-Europe vs Eastern Europe)
  - North America: US, CA
  - W/N-Europe/Israel: AT, BE, DE, DK, ES, FI, GB, IE, IL, IT, NL, NO, PL
  - Eastern Europe: BG, BA, BY, CZ, RU, UA, RO
  Recruitment volume ('low recruiters' vs 'high recruiters')
  - low recruiter: < 10 patients/site
  - high recruiter: $\geq$10 patients/site

7.2 Secondary Efficacy Outcomes

The study has two key secondary efficacy outcomes: progression free survival (PFS) and overall response rate (ORR).

7.2.1 ORR and Clinical Benefit: Definitions

The ORR for each treatment arm will be estimated as the proportion of responders, defined as a patient whose best overall response is PR or better during the treatment period. All responses must be confirmed at least 4 weeks after the initial response is seen. Tumor assessments will occur every 3 cycles (± 7 days) during ongoing treatment then, if required beyond the end of treatment, every 3 months (± 7 days) while the patient is on study. The last assessment will occur either when progression is confirmed or when approximately 384 randomized patients have died.

Clinical benefit is defined as having stable disease or better lasting for at least 9 weeks.

7.2.2 ORR and Clinical Benefit: Primary Analysis

The ITT population will be used to test the effects of AEZS-108 and comparator on ORR.

Hypothesis testing between the two treatment arms will be performed using a Mantel-Haenszel test. The corresponding odds ratio will be also estimated. The 95 % confidence interval of the ORR will be reported via normal approximation without continuity correction.

Clinical benefit rate will be analyzed using the same methods for the ORR analyses.

7.2.3 ORR: Supportive and Sensitivity Analysis

ORR will also be analyzed for the modified ITT and PP population, and logistic regression models analogous as defined for OS (see Section 7.1.3) will be made.

In addition, duration of overall response and duration of stable disease will be analyzed by means of time-to-event analysis (see details in section 7.1.2). Duration of overall response is defined for subjects with CR or PR as best overall response as duration from the first time when criteria for a response (PR or CR) were met until first observation of PD. If PD is not reached, the subject is censored at date of last visit with adequate assessment.

Duration of stable disease is defined as duration from start of treatment until first observation of PD for patients with SD as best overall response. If PD is not reached, the subject is censored at date of last visit with adequate assessment. If subject did not receive treatment, date of randomization will be taken as start date.

Moreover, duration of follow-up will be summarized and median duration will also be estimated by means of K-M analysis (i.e. reversing OS censoring).

7.2.4 PFS: Definition

PFS is defined as the number of days between randomization and the date of progression or death that occurred up to the end of study. PFS for patients who progress and then subsequently die will be the number of days between randomization and the date of progression is first observed.

Progression should be confirmed as specified. PFS data will be censored for subjects whose PFS status cannot be determined. The table below summarized rules of determining the dates of progression and censoring:

Situation	Date of Progression or Censoring	Outcome
No baseline tumor assessments	Randomization	Censored
Progression documented between scheduled visits	Date of the documented progression	Progressed
No progression	Date of last visit with adequate assessment	Censored
Treatment discontinuation for undocumented progression	Date of last visit with adequate assessment	Censored
Treatment discontinuation for toxicity or other reason	Date of last visit with adequate assessment	Censored
New anticancer treatment started	Date of last visit with adequate assessment	Censored
Death between adequate assessment visits	Date of death	Progressed
Death or progression after more than one missed visit	Date of last visit with adequate assessment	Censored

Once the dates of progression, survival, and censoring are derived, PFS will be calculated as follows:

For those who died or progressed up to the end of the study,

duration of PFS (days) = date of progression or death due to any case – randomization date + 1;

for those who are censored,

duration of PFS (days) = date of censoring – randomization date + 1, with censor indicator indicating a censor.

7.2.5 PFS: Primary Analysis

The primary analysis of PFS will be performed using the ITT population and based on the data from the study.

Hypothesis testing between the two treatment arms will be performed using a log-rank test with an overall two-sided 0.05 level of significance via a model including treatment effects.

For each treatment arm the median duration of PFS and the proportion of patients alive and progression-free at six and 12 months will be estimated using the Kaplan-Meier method. For each estimate, a 95 % confidence interval will be reported. The hazard ratio and its 95 % confidence interval will also be reported.

7.2.6 PFS: Supportive and Sensitivity Analysis

As a supportive analysis, PFS will be analyzed based on the modified ITT and PP populations. The same Cox models as defined for OS (see Section 7.1.3) will be made. In

addition, PFS will be analyzed in the subset of patients from MITT that have CR, PR or SD as best overall response.

7.2.7 Quality of Life

Change from baseline in Quality of life will be analyzed via mixed model analysis with treatment effect, visit, its interaction, baseline quality of life score and its interaction with visit as covariates. The treatment effect and treatment difference will be estimated and the 95 % confidence interval of the estimates will be presented.

7.2.7.1 QoL Scoring

The following sections describe the scoring algorithms used for both the QLQ-C30 and QLQ-EN24 questionnaires. Scoring procedures are similar for both questionnaires and can be found in the EORTC QLQ-C30 Scoring Manual, version 3.0 (Fayers et al. 2001) and the EN24 summary document. All scales and the single-item measures range from 0 to 100.

7.2.7.1.1 QLQ-C30

For all scales, calculate the raw score (RS) of a scale using the mean of the item scores in the scale as follows: $RS = (S_1 + S_2 + \ldots + S_n) / n$

where S_i: i=1, ..., n are the item scores and n is the number of items with valid scores, assuming the number of items with valid scores meets the minimum requirement as specified in Table 1 or this score will be assumed missing.

Use a linear transformation to standardize the raw score in order that scores will range from 0-100: Global Health Status/QOL = {(RS-1)/range} * 100

Functional Scales = {1- ((RS-1)/range)} * 100

Symptom Scales = {(RS-1)/range} * 100

where range for each scale is defined in Table 1.

Table 1: QLQ-C30 scales and scoring details

	Number of Items	Item Range [a]	Item Numbers	Minimum Not Missing
Global Health Status/QOL	2	6	29,30	1
Functional Scales				
Physical Functioning	5	3	1-5	3
Role Functioning	2	3	6,7	1
Emotional Functioning	4	3	21-24	2
Cognitive Functioning	2	3	20,25	1
Social Functioning	2	3	26,27	1
Symptom Scales / Items				
Fatigue	3	3	10,12,18	2
Nausea / Vomiting	2	3	14,15	1

	Number of Items	Item Range [a]	Item Numbers	Minimum Not Missing
Pain	2	3	9,19	1
Dyspnoea	1	3	8	N/A
Insomnia	1	3	11	N/A
Appetite Loss	1	3	13	N/A
Constipation	1	3	16	N/A
Diarrhoea	1	3	17	N/A
Financial Difficulties	1	3	28	N/A

[a] Range is the difference between the maximum possible value of the raw score and the minimum possible value.

QLQ-C30 scale scores range from 0 to 100 with higher scores representing a better health state for the functional scales and lower scores representing a better health state for symptom scores.

7.2.7.1.2 QLQ-EN24

For all scales, calculate the raw score (RS) of a scale using the mean of the item scores in the scale as follows: $RS = (S_1 + S_2 + \ldots + S_n) / n$

where Si: i=1, …, n are the item scores and n is the number of items with valid scores, assuming the number of items with valid scores meets the minimum requirement as specified in Table 2 or this score will be assumed missing.

Use a linear transformation to standardize the raw score in order that scores will range from 0-100: Functional Scales = {1- (RS-1)/range)} * 100

Symptom Scales = {(RS-1)/range)} * 100

where range for each scale is defined in Table 2.

Table 2: QLQ-EN24 scales and scoring details

	Number of Items	Item Range [a]	Item Numbers	Minimum Not Missing
Functional Scales				
Sexual Interest	1	3	49	N/A
Sexual Activity	1	3	50	N/A
Sexual Enjoyment	1	3	54	N/A
Symptom Scales / Items				
Lymphoedema	2	3	31-32	1
Urological symptoms	4	3	34-37	2

	Number of Items	Item Range [a]	Item Numbers	Minimum Not Missing
Gastrointestinal symptoms	5	3	38-42	3
Poor body image	2	3	47–48	1
Sexual/vaginal problems	3	3	51–53	2
Pain in back and pelvis	1	3	33	N/A
Tingling/numbness	1	3	43	N/A
Muscular pain	1	3	44	N/A
Hair loss	1	3	45	N/A
Taste change	1	3	46	N/A

[a] Range is the difference between the maximum possible value of the raw score and the minimum possible value.

A high score for the functional scales represents a high level of functioning, while a high score for the symptom scales represents a high level of symptoms or problems. Symptoms related to sexual/vaginal problems (including items 51-53) are optional.

7.2.7.1.3 Missing items

The method used for both health-related quality of life (HRQL) instruments requires that for multi-item scales, at least half of the items from the scale have been answered. The missing items are imputed based on the average of those items that are answered. None of the single item measures can be imputed using this method of imputation. In the event of missing full assessments, none of the subscales will be scored or imputed. Tables 1 and 2 detail the number of required items for the QLQ-C30 subscales and QLQ-EN24 subscales respectively.

7.2.7.2 Mixed Model Analysis

7.2.7.2.1 Hypothesis

The hypothesis for the HRQL endpoint – change over time in the QLQ-C30 GHS/QOL is as follows:

H0: There is no difference in health-related quality of life based on the EORTC QLQ C-30 GHS/QOL scale between AEZS-108 and doxorubicin treated subjects.

Ha: Treatment with AEZS-108 is associated with better health-related quality of life based on the EORTC QLQ C-30 GHS/QOL scale as compared to treatment with doxorubicin.

The same set of hypotheses holds for QLQ-EN24.

HRQL assessments are performed at the following visits:

Screening, end of cycles 3 and 6, end of therapy (3-5 weeks after last dose) and at post therapy follow-up (every 3 months).

7.2.7.2.2 Missingness

Missingness can be classified in terms of missing items (i.e. specific item(s) not completed on an otherwise completed form) or missing forms (i.e. entire form not completed). Furthermore, missing forms can be classified in terms of intermittent missing (i.e. missing assessment(s) between two non-missing assessments) or monotone missing (i.e. missing all assessments after a certain time point possibly due to subject dropout or treatment discontinuation).

In this study, as the HRQL data capture is scheduled up to the end of therapy and post therapy visits and subjects are expected to experience certain disease- or treatment-related morbidity that will lead to discontinuation of study treatment, missing assessments are inevitable. Also missing full assessments even for the subjects still on treatment can occur despite the best effort to minimize those. The underlying reasons for missing data are unknown but, in general, 3 different missing data mechanisms are considered (Rubin 1976; Little and Rubin, 1987; Fairclough, 2002).

- **Missing Completely at Random (MCAR)**
  Missingness is assumed to be completely unrelated to the HRQL scores. In general, this assumption is not plausible in oncology studies and will therefore not be assumed for the analyses.
- **Missing at Random (MAR)**
  Missingness depends only on observed data (e.g. prior observed HRQL scores or observed covariates). This is the assumption underlying the mixed modelling approach described here.
- **Missing Not at Random (MNAR)**
  Missingness depends on the HRQL scores that are missing. Since we have not observed the missing HRQL scores, it is not possible to test if the missingness of HRQL data depends on the value they would have had if not missing. However, the possibility of missing not at random may not be completely ruled out.

7.2.7.2.3 Description of mixed model with assumptions

Changes over time in QLQ-C30 Global Health Status/QOL scores and in QLQ-EN24 scores will be compared between treatment groups using a restricted maximum likelihood-based mixed model for repeated measures (MMRM) data under the missing at random (MAR) assumption (Mallinckrodt et al, 2008).

The dependent variable of this model will be the change of QLQ-C30 Global Health Status/QOL score or the QLQ-EN24 score measured at the end of cycles 3 and 6, at the end of therapy and at post therapy follow-up as compared to baseline (denoted by HRQL). The model will include the fixed, categorical effects of treatment (trt), visit (visit), and treatment-by-visit interaction (trt*visit). As there is only a limited amount of time points per subject and given the possibility that not all time points are equally spaced in time, time is entered into the model as the categorical variable visit. The random subject effects

are not explicitly formulated in the model form, but will be modelled as part of the within-subject error correlation structure. For this purpose, an unstructured covariance matrix will be used, allowing for different variances / covariances across subjects. If the analysis using the unstructured covariance matrix fails to converge, the following covariance structures will be tested: heterogeneous Toeplitz, heterogeneous compound symmetry, Toeplitz and compound symmetry. The covariance structure converging to the best fit, as determined by Akaike's information criterion, will be used as the primary analysis. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom. The least square means for each treatment group and the treatment difference will be presented together with the 95% confidence interval and p-value.

The pseudo SAS code of the primary model is as follows:

```
proc mixed data = HRQLdataset method = reml;
  class PatID trt visit;
  model HRQL = trt visit trt*visit baseline baseline*visit / noint ddfm = kr;
  repeated r / subject = PatID type=un;
run;
```

where PatID represents the individual subject id.

The mixed model will provide estimates of the treatment effect and of the treatment difference at each visit until end of treatment, together with their 95% confidence intervals.

7.3 Interim Analyses

The Data and Safety Monitoring Board (DSMB) will review data periodically throughout the trial. The DSMB may recommend stopping the trial for safety at any time or for futility at the interim looks. Two planned interim analyses will be conducted, one for futility only and one for both safety and efficacy. The first for a futility analysis only will be conducted at 1/3 of information time (i.e., approximately at 128 death event time). The interim efficacy/safety analysis is planned at 50 % of information time (when approximately 192 death events take place).

In the interim efficacy/safety analysis, OS, the primary efficacy outcome of the study, will be analyzed using a two-sided log-rank test. The Lan-DeMets implementation of the O'Brien Fleming boundary will be employed to control the overall Type I error rate at 0.05. The nominal p-values for the planned second interim and the final analyses will be 0.003051 and 0.049002, respectively.

The interim analysis may not be binding. However, if the interim analysis results demonstrate strong evidence of efficacy, the DSMB may request a meeting with the FDA to discuss potential early termination of the study.

7.4 Adjustments for Multiplicity

One non-binding efficacy interim analysis is planned for this study. Lan-DeMets implementation of the O'Brien-Fleming alpha-spending function will be used to control

the overall Type I error rate at 0.05. The study has one primary efficacy endpoint; no p-value adjustment for multiple endpoints is required.

Two key secondary efficacy variables, PFS and ORR will also be analyzed for potential labeling claims. Hierarchical testing procedures will be used to control the Type I Error at the 0.05 level. Specifically, the superiority of the active treatment will be tested at a two-sided significance level of 0.05 for OS first, then for PFS, and then for ORR. The superiority claim for an efficacy outcome can be made only if all previous claims are successfully made.

7.5 Power and Sample Size Justification

Approximately 384 events of deaths will be required to achieve 80 % power to detect a treatment difference at the overall two-sided 0.05 significance level after taking the planned efficacy/safety interim look into account. It is expected that approximately 500 patients will be enrolled during an estimated 24-month recruitment period and will then be followed for 12 months to observe a total of approximately 384 death events. In the sample size calculation, it is assumed that the median OS is 12 months for the group with investigational treatment (AEZS-108) and 9 months for the control group (doxorubicin). The sample size calculation has taken the interim analyses into consideration.

Based on published data on the pharmacokinetics of doxorubicin, a sample size of 20 patients for each arm is proposed, to provide a more detailed description than is currently available for the key pharmacokinetic parameters of AEZS-108, doxorubicin, and doxorubicinol. Based on preliminary pharmacokinetic data available from an ongoing study in bladder cancer patients, a variability of AEZS-108 similar to that for doxorubicin may be expected. Since collection of only a limited number of samples over a limited period (about 2-3 terminal elimination half-lives) is feasible within this sub-study, the comparison of the metabolite profiles will be orienting / exploratory / descriptive in nature.

8 PHARMACOKINETICS

A PK sub-study at selected sites and sparse PK sampling at all sites is performed to assess the pharmacokinetics and exposure-response relationships of AEZS-108, doxorubicin and doxorubicinol. A SAP supplement contains all details of non-compartmental analysis that will be performed on the PK sub-study samples and is not part of this main SAP. All details of population PK analysis (using all available PK samples) will be described in a second SAP supplement that is also no part of this main SAP.

9 SUMMARIES OF MEASURES OF SAFETY

Safety analyses will be performed for the safety population. Safety evaluations will be based on the incidence, intensity, and type of adverse events, as well as on clinically significant changes in the patient's physical examination, vital signs, and clinical laboratory results. Safety variables will be tabulated and presented by study drug actually received.

Because there is no pre-specified safety outcome defined in terms of AEs, clinically relevant laboratory parameters, or vital signs, any formal comparisons between the treatment arms with respect to specific safety parameters will be post-hoc.

9.1 Adverse Events

Each AE and SAE term recorded on the case report forms (CRFs) will be mapped to a preferred term (PT) using the MedDRA dictionary. The investigator will classify the severity of AEs and SAEs using the NCI CTCAE v4.0 and will assess the relationship of each event to the study treatment.

All AEs and SAEs occurring on study will be listed by treatment drug, center, and patient. Four additional listings will be made: listing of AEs associated with dropout (i.e. Drug withdrawn on CRF page), listing of AEs leading to dose reduction (i.e. Dose reduced on CRF page), listing of AEs with fatal outcome and listing of deaths. AE listings will contain (not limited to): last dose before event (mg/m2), time on treatment (date of last dose before AE onset – C1D1) and time since last dose (start date of AE – date of last dose before AE onset). The listing of deaths will include for each death case: overall survival (duration), reason for end of treatment, time on treatment (C1D1 – date of last dose), time since last dose (start date – date of last dose) and best response.

The frequency and percentages of patients with treatment-emergent adverse events (TEAEs) will be tabulated by system organ class (SOC) and PT, where treatment-emergent is defined as any AE that;

- occurs after randomization and through the end of the study or up through 30 days after the last dose of study treatment,
- is considered treatment-related (i.e., considered at least possibly related to the study treatment) regardless of the start date of the event, or
- is present before randomization but worsens in intensity or the investigator subsequently considers treatment-related (i.e., considered at least possibly related to any component of the study treatment).

For the purpose of calculating treatment emergence and inclusion in summary tables, incomplete onset dates will be imputed as detailed in Appendix D. If treatment-relatedness is missing, AE will be considered as RELATED. Summaries will display incidence by study drug (AEZS-108 or doxorubicin) received and total incidence, and PTs within each SOC will appear in decreasing order of total incidence. At each level of summarization, a patient will be counted only once for each AE, SOC, or PT experienced within that level.

Related TEAEs, serious TEAEs, Grade 3 or higher TEAEs, Grade 3/4 serious TEAEs, Grade 5 serious TEAEs, related serious TEAEs, related Grade 3 or higher TEAEs, and TEAEs resulting in discontinuation or dose modification of study treatment will be similarly summarized. Summaries of TEAEs by relationship to each of the study treatments will also be prepared.

At each level of summarization, a patient will be counted only once for each AE, SOC, or PT experienced within that level. In the summation for AE severity, within each level of AE, SOC, or PT experienced, the one with the highest severity will be included. In the

summation for AE's relationship to the study drug, within each level of AE, SOC, or PT experienced, the one with the closest relationship to the study drug will be included.

Several TEAE tables with AEs sorted by PT (>= certain cut-off) by decreasing frequency (in the AEZS-108 arm), irrespective of SOC, will be made in addition (see Appendix E – List of Tables/Graphs and Listings).

9.2 Laboratory Assessments

Laboratory data for hematology and serum chemistry tests will be reported in International Units. Individual values outside the central laboratory reference ranges will be identified (by "H" for high and "L" for low) in the data listings displaying the absolute values for each patient.

9.2.1 Variables

The following Hematology, Chemistries, Urinalysis laboratory evaluations will be performed:

Panel name	**Laboratory parameters included**
Hematology	leukocytes, erythrocytes, hemoglobin, hematocrit, thrombocytes (platelets); differential white blood cell count (neutrophils, lymphocytes, monocytes, basophils, eosinophils)
Chemistries - Enzymes - Substrates - Electrolytes	 AST/SGOT, ALT/SGPT, gamma-GT, alkaline phosphatase, LDH total protein, total bilirubin, creatinine, urea, uric acid sodium, potassium, calcium
Urinalysis	glucose*, protein*, microalbumin**
Hormones	luteinizing hormone (LH), follicle stimulating hormone (FSH)
Pregnancy test	ß-HCG (in serum; for women of childbearing potential only)

* In case there is no possibility for evaluation of the parameters provided above semiquantitative dipstick analysis will be accepted.

** In cases where 24-hour urine is not available to determine microalbumin, the microalbumin/creatinine ratio may be used instead.

9.2.2 Statistical Analysis

Continuous laboratory test results will be summarized descriptively by study drug received for actual values and for changes from baseline. Visits to be summarized include all scheduled post-Cycle 1 Day 1 visits. Unscheduled visit data is not captured in summary statistics.

Shift from baseline tables for laboratory parameters will be presented. Categories will be based on CTCAE grade (where applicable) or by high/low flags (where CTCAE grades are not defined). Data will be analyzed at all post-Cycle 1 Day 1 visits, by worst grade post-Cycle 1 Day 1. Laboratory tests that have high and low abnormalities will be summarized separately for each direction (e.g., hypocalcemia and hypercalcemia). Worst post-baseline

values also include unscheduled visits. For WBC, neutrophils (absolute value) and thrombocytes extra shift table only considering cycle 1 lab results will be made in addition.

Abnormalities in laboratory tests that the investigator considered clinically significant (relevant) were to be recorded and summarized as AEs.

9.3 Vital Signs

9.3.1 Variables

The following vital signs will be summarized:

- Blood pressure (systolic and diastolic, mmHg)
- Oral temperature (Celsius)
- Pulse rate (beats/min)
- Respiration rate (breaths/min)
- Body weight (kg)

9.3.2 Statistical Analysis

Summary tables of vital signs and change from baseline will be presented for all scheduled visits where vital signs were assessed. All recorded vital sign data will be listed.

9.4 Physical Examination

Physical examination data will be presented in a data listing.

9.5 Cardiac Assessments

9.5.1 Left Ventricular Ejection Fraction (LVEF)

LVEF (ECHO or MUGA) data (raw values and change from baseline) will be summarized descriptively (arithmetic mean, standard deviation, median, range interquartile values) by treatment at all available timepoints. In case only a range is specified for an LVEF measurement: use the mid-point (arithmetic mean) of the specified range as the substitute for the not-available point-value when calculating “Raw value result” and “Change”.

Moreover, a shift table will be made analyzing the data at all scheduled post-Cycle 1 Day 1 visits, by worst value post Cycle 1 Day 1 until end of study treatment (EOT) visit, and until end of post therapy follow-up. The following categories will be considered: FROM (%) 50-55, 55-60, 60-65, 65-70, 70-75 and >75 to <35, 40-35, 45-40, 50-55, 55-60, 60-65, 65-70, 70-75, >75. For LVEF-values specified by range the mid-point of the range should be used for the assignment. In addition, at the end of C3, C6, C7, C8, end of treatment and at post-therapy follow-up, number of subjects (and percentage) with absolute decline from baseline of 10% or more or absolute value of LVEF <50%, will be presented. Moreover, cumulatively at any time after C1D1, the following will be assessed: (number of subjects and percentage of):

- Absolute decline by 10% or more,
- Absolute decline by 15% or more,

- Absolute value of LVEF <50,
- Absolute value of LVEF <45%,
- Absolute decline from baseline of 10% or more or absolute value of LVEF <50%,
- Absolute decline from baseline of 15% or more or absolute value of LVEF <45%

A plot with mean LVEF (mean of raw LVEFE values) evolution over time –by treatment arm (including indication of number of patients present at each time point) will also be made.

9.5.2 Electrocardiogram (ECG)

For all 12-ECGs recorded as part of the core study, the global classification of clinically significant (relevant) abnormalities will be summarized. For all continuous ECG variables, descriptive statistics (arithmetic mean, standard deviation, median, range interquartile values) will be used for summaries of all continuous ECG variables and their changes from baseline.

Categorical analysis will be performed (by means of frequency tables, worst post-cycle 1 day 1 until post-therapy FU result) with respect to QTcF and other cardiac intervals for the following categories:
Absolute values:

- QTc ≤ 450 ms
- QTc > 450 ms and ≤ 480 ms
- QTc > 480 ms and ≤ 500 ms
- QTc > 500 ms
- PR > 220 ms
- QRS > 110 ms

Change from baseline in cardiac intervals:

- ΔQTc increase ≤ 30 ms
- ΔQTc increase > 30 ms and ≤ 60 ms
- ΔQTc increase > 60 ms
- ΔPR > 25%
- ΔQRS > 25%

The heart rate corrected QTc interval will be calculated according to

Bazett's formula ($QTcB = \frac{QT}{\sqrt{RR}}$),

or Fridericia's formula ($QTcF = \frac{QT}{\sqrt[3]{RR}}$),

where QTc is the QT interval corrected for rate, and RR is the interval from the onset of one QRS complex to the onset of the next QRS complex, measured in miliseconds.

RR interval (sec) = 60 / HR; with Heart Rate (HR) in beats per minute.

For patients participating in the PK sub-study, descriptive analyses of results and changes from baseline will be performed. More details can be found in the SAP supplement for the PK/QTc sub-study.

9.6 ECOG Performance Status

ECOG Performance Status will be summarized as shift from baseline tables by visit using frequencies and percentages at all scheduled visits where performance status was assessed.

10 IDENTIFICATION AND SUMMARY OF PROTOCOL DEVIATIONS

All protocol deviations will be classified as major or minor. All protocol deviations that can be extracted from numeric or coded study data as well as treatment compliance will be summarized. Protocol deviation classification will be determined and documented prior to data base lock.

10.1 Data Quality Assurance

Accurate and reliable data collection will be ensured by verification and cross check of the CRFs against the investigator's records by the study monitor (source document verification) and by the maintenance of a drug-dispensing log by the investigator. Collected data will be entered into a computer database and subject to electronic and manual quality assurance procedures.

11 REFERENCES

Fairclough, DL. Design and analysis of quality of life studies in clinical trials. Chapman & Hall/CRC. 2002.

Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A, on behalf of the EORTC Quality of Life Group. *The EORTC QLQ-C30 Scoring Manual (3rd Edition).* Published by: European Organisation for Research and Treatment of Cancer, Brussels 2001.

Lan K.K.G., D.D., *Discrete sequential boundaries for clinical trials*. Biometrika, 1883. **70**: p.659-63.

Little, RJA and Rubin, DB. Statistical Analysis with Missing Data. New York: Wiley. 1987.

Mallinckrodt, C.H., Lane, P.W., Schnell, D., Peng Y, and Mancuso, J.P. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. Drug Information Journal, Vol. 42, pp.303-319, 2008.

Rubin, DB. Inference and missing data. Biometrika 63:581-592, 1976.

12 APPENDICES

12.1 Appendix A - Time and Events Schedule

Procedures	**Screen**	**Cycle 1**		**Cycle 2 to 9**	**End of therapy**	**Post Therapy Follow up**
Visit / Cycle Day	**Within 4 wks before C1 D1**	**D1**	**D8 + D15 (+D22 + D29** if retreatment delayed**)**		**3-5 weeks after last dose**	**Every 3 months**
1)	**2)**		**3)**	**4)**	**5)**	**6)**
Informed consent	**X**					
Demographics	**X**					
Medical history	**X**					
Physical examination	**X**					
Interval history and symptom-directed exams		**X**		**D1**	**X**	
Clinical assessments: Height (screen only), Weight	**X**	**X**		**D1**	**X**	
ECOG performance status	**X**	**X**		**D1**	**X**	**X** *(until PD)*
Blood pressure, pulse	**X**	**X**	**X**	**D1**	**X**	**X** *(until PD)*
Adverse events		**Cumulative recording** [7] *(whenever observed/applicable)*				**X**[9]
Concomitant treatments	**X**	**X**		**D1**	**X**	**X**[10]
Clinical laboratory tests						
Hematology, Chemistries, and Urinalysis	**X**	**X**	**X**[11]	**D1** [4]	**X**	
LH, FSH	**X**			**D1**	**X**	
Pregnancy test (ß-HCG)	**X** [8]					
Tumor tissue specimens collection	**X** [14]					
Pharmacokinetics (sub-study)		**X**[12]	**D2+D3**[12]			
Cardiac function test LVEF (ECHO or MUGA)	**X** [2A]			**End of C3 , C6, C7, C8**[16]	**X**	**X** [9]
ECG (12-lead)	**X**			**D1**	**X**	**X** [9]

Procedures	**Screen**	**Cycle 1**		**Cycle 2 to 9**	**End of therapy**	**Post Therapy Follow up**
Tumor imaging/evaluation	**X** [2B)]			**End of C3 + C6**	**X**	**X**
Quality of Life questionnaire	**X**			**End of C3 + C6**	**X**	**X**
Survival follow-up						**X**
Patient registration	**X**					
Eligibility criteria review	**X**	**X**[15)]				
Randomization[13)]	**D-1 to D-4**[13)]					
Study treatment		**X**		**D1**		

1) A window of +/- 3 days will be allowed for scheduled visits/actions, +/- 7 days for 3 monthly follow-ups. If a procedure or visit falls on a public holiday, the patient should be scheduled for the next working day.
2) Pre-study evaluations dating back more than 2 weeks prior to treatment start are acceptable.

   2A) LVEF: In absence of intercurrent chemotherapy or cardiac medical history, results dating back up to 8 weeks will be acceptable.

   2B) Tumor imaging: Assessments of indicator lesions up to 4 weeks prior to treatment start (C1D1) will be considered valid for baseline evaluation.
3) Patients who have not recovered sufficiently from treatment-related adverse events within 3 weeks may have their retreatment delayed by up to 2 weeks; patients who cannot be retreated 5 weeks after previous dose, will not receive further study treatment but will still be followed for outcome.
4) Start of a new cycle is planned every 3 weeks unless treatment delay is required due to residual toxicity. Per footnote 1, a window of up to 3 days is allowed to schedule tests prior to the anticipated day of retreatment, to assure sufficient recovery from treatment-related adverse events ('toxicity') prior to dosing.
5) A full End-of-Therapy assessment will be also required whenever a patient is being withdrawn from the study. The reason(s) for discontinuation will be documented.
6) Patients discontinuing study treatment for reasons other than disease progression will be reassessed at 3-month intervals to assess the time to treatment failure (progression / death).

   Follow-up until progression will be suspended in patients in whom a new cancer treatment has been started without prior documentation of progressive disease.

   Information on survival status may be captured by telephone if no follow-up visit is scheduled for other reasons.
7) If treatment was discontinued because of a drug-related adverse event (i.e., causality at least possibly related), the patient will be followed up as clinically indicated to assess and document the outcome.
8) Only required in women of child-bearing potential
9) Only a patient developing signs of cardiac failure during post-treatment follow-up will have LVEF and ECG evaluated; a left ventricular dysfunction of CTCAE grade 4 will be reported as an SAE.
10) Only anticancer treatments that a patient received during post treatment follow-up for survival will be recorded.

11) For tests performed by a laboratory not affiliated to study site (e.g. mid-cycle controls by the patient's home doctor) copies of the external laboratory reports will be provided to the sponsor and results recorded in the CRF.

12) The pharmacokinetic sub-study will be conducted at selected investigational sites. Procedures are described in Appendix 7 of the protocol.

13) Every effort should be made to start treatment within 4 days of randomization.

14) Specimens from FFPE fresh biopsies if FFPE archival tumor specimens are not available.

15) Except for Day 1 clinical laboratory tests

16) Recommended to be conducted also at end of Cycle 7 and 8 (i.e., before start of Cycle 8 and Cycle 9, respectively) for both arms. Results are to be captured on the eCRF if there is an abnormal finding.

12.2 Appendix B - Progressive Disease and Response Algorithm

12.2.1 Introduction

This algorithm is designed to determine the following:

- Whether a patient's disease has progressed and, if so, the date of progression
- The censoring date for patients whose disease never progressed

12.2.2 Response Evaluation Criteria in Solid Tumors (RECIST 1.1)

Details can be found in Protocol Section 7.1.

12.2.3 Date of Progression or Censoring

The date of death is the documented date of death. The date of progression (non death) will be determined as follows:

Progression is first observed based on	Date of progression
Unscheduled test prescribed by the investigator based on patients clinical presentation	The date the test is performed
• Scheduled assessment per protocol	
Assessment done on scheduled visit date + 7 days (Visit Window)	The date of the assessment
• Assessment done > 7 days after scheduled visit date (Note: this case is not relevant for this trial.)	The date of the next scheduled assessment date after the date of last test indicating no progression. For example, if a patient has an assessment done at Jth scheduled assessment visit with a result no progression and has the next assessment done at (J+4)th scheduled assessment visit (i.e., skipping 4 scheduled assessment visit) with a result showing progression, the date of progression will be the date of (J+1)st scheduled assessment visit.

12.3 Appendix C - ECOG Performance Status Criteria

ECOG Performance Status Scale		Karnofsky Performance Scale	
Grade	**Descriptions**	**Percent**	**Description**
0	Normal activity. Fully active, able to carry on all pre-disease performance without restriction.	100	Normal, no complaints, no evidence of disease.
		90	Able to carry on normal activity; minor signs or symptoms of disease.
1	Symptoms, but ambulatory. Restricted in physically strenuous activity, but ambulatory and able to carry out work of a light or sedentary nature (e.g., light housework, office work).	80	Normal activity with effort; some signs or symptoms of disease.
		70	Cares for self, unable to carry on normal activity or to do active work.
2	In bed <50 % of the time. Ambulatory and capable of all self-care, but unable to carry out any work activities. Up and about more than 50 % of waking hours.	60	Requires occasional assistance, but is able to care for most of his/her needs.
		50	Requires considerable assistance and frequent medical care.
3	In bed >50 % of the time. Capable of only limited self-care, confined to bed or chair more than 50 % of waking hours.	40	Disabled, requires special care and assistance.
		30	Severely disabled, hospitalization indicated. Death not imminent.
4	100 % bedridden. Completely disabled. Cannot carry on any self-care. Totally confined to bed or chair.	20	Very sick, hospitalization indicated. Death not imminent.
		10	Moribund, fatal processes progressing rapidly.

5	Dead.	0	Dead.

12.4 Appendix D - Imputation Rules for Missing Dates

Terminology in this appendix: first dose = first administration of either AEZS-108 (Arm A) or doxorubicin (Arm B).

The date of event (death/progression) is the documented date of event. The missing date will be handled as follows:

What is missing in event date	Imputed value is
Year	
• If year is missing or cannot be confirmed to be earlier than the month of last assessment	the year of last assessment
• If, without considering year, the date of the event is confirmed to be earlier than the date of the last assessment (e.g., January vs. November or May 15 vs. May 1)	the year of last assessment+1
Month	
• If month of the event is missing or is confirmed to be the same as the year of the last assessment	Month of the last assessment unless the day of death is earlier than the day of the last assessment (5th vs. 10th). In this case the month will be imputed as the month after the month of the last assessment
• If month of the event is confirmed to be after the year of the last assessment	January
Day	1, unless the resulting imputed date is earlier than the last assessment date. In this case the imputed day is the day of the last assessment

Adverse Event

- If onset date is completely missing, onset date is set to date of first dose.
- If (year is present and month and day are missing) or (year and day are present and month is missing):
  - If year = year of first dose, then set month and day to month and day of first dose.
  - If year < year of first dose, then set month and day to December 31st.
  - If year > year of first dose, then set month and day to January 1st.

- If month and year are present and day is missing:
    - If year = year of first dose and
        - if month = month of first dose then set day to day of first dose date.
        - if month < month of first dose then set day to last day of month.
        - if month > month of first dose then set day to 1st day of month.
    - If year < year of first dose then set day to last day of month.
    - If year > year of first dose then set day to 1st day of month.
- For all other cases, set onset date to date of first dose.

Concomitant Medications/Medical History

- If start date is completely missing, start date will not be imputed.
- If (year is present and month and day are missing) or (year and day are present and month is missing), set month and day to January 1.
- If year and month are present and day is missing, set day to 1st day of month.
- If end date is completely missing, end date will not be imputed.
- If (year is present and month and day are missing) or (year and day are present and month is missing), set month and day to December 31.
- If year and month are present and day is missing, set day to last day of the month.

12.5 Appendix E – List of Tables/Graphs and Listings

List of Statistical Tables

Number	**Description** **CODE: GREEN = topline; needed first**
Table 14.1.1.1	Patient disposition
Table 14.1.1.2	Patient disposition by investigational site
Table 14.1.1.3	Patient disposition by territory
Table 14.1.2	Major protocol deviations
Table 14.1.3.1	Patients demographics and baseline characteristics (ITT)
Table 14.1.3.2	Patients demographics and baseline characteristics (safety)
Table 14.1.4.1	History of endometrial cancer (ITT)
Table 14.1.4.2	History of endometrial cancer (safety)
Table 14.1.5	Prior treatment for endometrial cancer
Table 14.1.6.1	Prior medications
Table 14.1.6.2	Concomitant medications
Table 14.1.7.1	Treatment exposure (ITT)
Table 14.1.7.2	Treatment exposure (safety)
Table 14.1.7.3	Treatment – mode of administration (ITT)
Table 14.1.8.1	Treatment modifications (ITT)
Table 14.1.8.2	Treatment modifications (safety)
Table 14.1.9	Post-study treatment anticancer therapy
Table 14.2.1	Overall survival: primary analysis (ITT)
Table 14.2.2	Overall survival: sensitivity analysis 1 (PP)
Table 14.2.3	Overall survival: sensitivity analysis 2 (MITT)

Table 14.2.4	Overall survival: sensitivity analysis 3 (ITT, censoring at time subject drops out due to AE)
Table 14.2.5.1	Overall survival: sensitivity analysis 4: part 1 (Cox models for treatment and each covarariate separately)
Table 14.2.5.1	Overall survival: sensitivity analysis 4: part 2 (multivariate Cox model)
Table 14.2.6	Overall response rate: primary analysis
Table 14.2.7	Overall response rate: sensitivity analysis 1 (PP)
Table 14.2.8	Overall response rate: sensitivity analysis 2 (MITT)
Table 14.2.9.1	Overall response rate: sensitivity analysis 3: part 1 (logistic regression models for treatment and each covarariate separately)
Table 14.2.9.2	Overall response rate: sensitivity analysis 3: part 2 (multivariate logistic regression model)
Table 14.2.10	Duration of overall response
Table 14.2.11	Duration of stable disease
Table 14.2.12	Duration of follow-up
Table 14.2.13	Progression-free survival: primary analysis
Table 14.2.14	Progression-free survival: sensitivity analysis 1 (PP)
Table 14.2.15	Progression-free survival: sensitivity analysis 2 (MITT)
Table 14.2.16	Progression-free survival: sensitivity analysis 3 (MITT – selection of patients with CR, PR or SD as best overall response)
Table 14.2.17.1	Progression-free survival: sensitivity analysis 4: part 1 (Cox models for treatment and each covarariate separately)
Table 14.2.17.2	Progression-free survival: sensitivity analysis 4: part 2 (multivariate Cox model)
Table 14.2.18.1	Quality of Life: C30 scores raw data summaries
Table 14.2.18.2	Quality of Life: EN24 scores raw data summaries
Table 14.2.18.3	Quality of Life: individual questions C30 raw data summaries

Table 14.2.18.4	Quality of Life: individual questions EN 24 raw data summaries
Table 14.2.19.1	Quality of Life: C30 scores change from baseline summaries
Table 14.2.19.2	Quality of Life: EN24 scores change from baseline summaries
Table 14.2.19.3	Quality of Life: individual questions C30 change from baseline summaries
Table 14.2.19.4	Quality of Life: individual questions EN 24 change from baseline summaries
Table 14.2.20.1	Quality of Life: mixed model analysis on C30 scores change from baseline
Table 14.2.20.2	Quality of Life: mixed model analysis on EN24 scores change from baseline
Table 14.3.1.1	Summary of occurrence of treatment emergent adverse events (TEAE)
Table 14.3.1.2	Treatment emergent adverse events (TEAE) – by SOC and PT
Table 14.3.1.3	Treatment emergent adverse events (TEAE) (showing PT >= 5%) – by PT
Table 14.3.1.4	Treatment emergent adverse events (TEAE) – by SOC, PT and severity
Table 14.3.1.5	Related treatment emergent adverse events (TEAE) – by SOC and PT
Table 14.3.1.6	Related treatment emergent adverse events (TEAE) (showing PT >= 5%) – by PT
Table 14.3.1.7	Summary of TEAEs by relationship to each study treatment –by SOC and PT
Table 14.3.1.8	Summary of TEAEs by combined relationship to each study treatment – by SOC and PT
Table 14.3.1.9	Serious treatment emergent adverse events (TEAE) – by SOC and PT
Table 14.3.1.10	Serious treatment emergent adverse events (TEAE) (showing PT >= 2%) – by PT
Table 14.3.1.11	Serious treatment emergent adverse events (TEAE) – by SOC, PT and severity
Table 14.3.1.12	Related serious treatment emergent adverse events (TEAE) – by SOC

	and PT
Table 14.3.1.13	Related serious treatment emergent adverse events (TEAE) (showing PT >= 1.8%) – by PT
Table 14.3.1.14	Grade 3 or 4 serious treatment emergent adverse events (TEAE) – by SOC and PT
Table 14.3.1.15	Grade 5 serious treatment emergent adverse events (TEAE) – by SOC and PT
Table 14.3.1.16	Grade 3 or higher treatment emergent adverse events (TEAE) – by SOC and PT
Table 14.3.1.17	Grade 3 or higher treatment emergent adverse events (TEAE) (showing PT >= 2%) – by PT
Table 14.3.1.18	Related grade 3 or higher treatment emergent adverse events (TEAE) – by SOC and PT
Table 14.3.1.19	Related grade 3 or higher treatment emergent adverse events (TEAE) (showing PT >= 1.8%) – by PT
Table 14.3.1.20	Treatment emergent adverse events (TEAE) resulting in discontinuation of study treatment – by SOC and PT
Table 14.3.1.21	Treatment emergent adverse events (TEAE) resulting in dose reduction of study treatment – by SOC and PT
Table 14.3.1.22	Treatment emergent adverse events (TEAE) resulting in interruption of study treatment administration – by SOC and PT
Table 14.3.2.5	Deaths
Table 14.3.2.6	Deaths on study treatment (within 30 days after last dose of study treatment)
Table 14.3.4.1	Quantitative laboratory test results in SI units, hematology
Table 14.3.4.2	Shift from baseline, hematology
Table 14.3.4.3	Cycle 1 WBC, neutrophiles, and thrombocytes, shift from baseline, CTCAE grades
Table 14.3.4.4	Quantitative laboratory test results in SI units, serum chemistry
Table 14.3.4.5	Shift from baseline, serum chemistry

Table 14.3.4.6	Quantitative laboratory test results in SI units, urinalysis
Table 14.3.4.7	Shift from baseline (quantitative), urinalysis
Table 14.3.4.8	Dipstick test results, urinalysis
Table 14.3.4.9	Quantitative laboratory test results in SI units, hormones
Table 14.3.4.10	Shift from baseline, hormones
Table 14.3.4.11	Pregnancy test results
Table 14.3.4.12	Vital signs
Table 14.3.4.13	LVEF summaries
Table 14.3.4.14	LVEF shift from baseline – part 1
Table 14.3.4.15	LVEF shift from baseline – part 2
Table 14.3.4.16	Interpretation of 12 lead ECG results
Table 14.3.4.17	ECG summaries
Table 14.3.4.18	ECG categorical analysis
Table 14.3.4.19	ECOG performance status shift from baseline

List of Graphs

Number	Description
Figure 14.2.1	Kaplan Meier curve for OS: primary analysis
Figure 14.2.2	Kaplan Meier curve for OS: sensitivity analysis 1 (PP population)
Figure 14.2.3	Kaplan Meier curve for OS: sensitivity analysis 2 (MITT population)
Figure 14.2.4	Kaplan Meier curve for OS: sensitivity analysis 3 (ITT population – censoring at time subject drops out due to AE)
Figure 14.2.5	Kaplan Meier curve for PFS: primary analysis
Figure 14.2.6	Kaplan Meier curve for PFS: sensitivity analysis 1 (PP population)
Figure 14.2.7	Kaplan Meier curve for PFS: sensitivity analysis 2 (MITT population)
Figure 14.2.8	Kaplan Meier curve for PFS: sensitivity analysis 3 (MITT – selection of patients with CR, PR or SD as best overall response)
Figure 14.3.1	LVEF evolution over time

List of Listings

Number	Description
Listing 16.2.1.1	Patient disposition
Listing 16.2.2	Major protocol deviations
Listing 16.2.4.1	Patients demographics
Listing 16.2.4.2	Patients baseline characteristics
Listing 16.2.4.3	History of endometrial cancer
Listing 16.2.4.4	Prior treatment for endometrial cancer
Listing 16.2.4.5	Prior medications
Listing 16.2.4.6	Concomitant medications
Listing 16.2.5	Treatment exposure (including modifications)
Listing 16.2.6.1	OS: time to event
Listing 16.2.6.2	Overall response rate
Listing 16.2.6.3	PFS: time to event
Listing 16.2.6.4.1	Quality of Life: C30 individual questions
Listing 16.2.6.4.2	Quality of Life: C30 scores
Listing 16.2.6.4.3	Quality of Life: EN24 individual questions
Listing 16.2.6.4.4	Quality of Life: EN24 scores
Listing 16.2.7	Adverse events
Listing 16.2.7.3	AEs associated with dropout
Listing 16.2.7.4	AEs leading to dose reduction
Listing 16.2.7.7	AEs with fatal outcome
Listing 16.2.7.8	Deaths

Listing 16.2.8.1	Laboratory test results: hematology
Listing 16.2.8.2	Laboratory test results: serum chemistry
Listing 16.2.8.3	Laboratory test results: urinalysis
Listing 16.2.8.4	Vital signs
Listing 16.2.8.5	LVEF
Listing 16.2.8.6	ECG
Listing 16.2.8.7	ECOG performance status