STATISTICAL ANALYSIS PLAN

Study Title:	A Phase 3 Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Efficacy and Safety of GS-5745 Combined with mFOLFOX6 as First Line Treatment in Patients with Advanced Gastric or Gastroesophageal Junction Adenocarcinoma
Name of Test Drug:	Andecaliximab (GS-5745)
Study Number:	GS-US-296-1080
Protocol Version:	Amendment 6
Protocol Date:	06 March 2017
Analysis Type:	Interim and Final Analysis
Analysis Plan Version:	2.0
Analysis Plan Date:	03 October 2018
Analysis Plan Author	PPD

CONFIDENTIAL AND PROPRIETARY INFORMATION

TABLE OF CONTENTS

LIST OF IN-TEXT TABLES

LIST OF IN-TEXT FIGURES

LIST OF ABBREVIATIONS

ADA	Anti-Drug Antibody
AE	Adverse Event
ALK	Anaplastic Lymphoma Kinase
ALT	Alanine Aminotransferase
ANC	Absolute Neutrophil Count
AST	Aspartate Aminotransferase
AUC	Area Under the Drug Concentration Over Time Curve
BLQ	Below the Limit of Quantitation
CI	Confidence interval
CNS	Central Nervous System
CSR	Clinical Study Report
CR	Complete Response
CRC	Colorectal Cancer
CTCAE	Common Terminology Criteria for Adverse Events
CV	Coefficient of Variation
DCR	Disease Control Rate
DLT	Dose Limiting Toxicity
DOR	Duration of Response
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic Case Report Form(s)
EOI	End of Infusion
GLSM	Geometric Least Squares Mean
HLT	High-level Term
IHC	Immunohistochemistry
INR	International Normalized Ratio
LLOQ	Lower Limit Of Quantitation
MedDRA	Medical Dictionary for Regulatory Activities
MTD	Maximum Tolerated Dose
NN	non-CR/non-PD
NSCLC	Non-Small Cell Lung Cancer
ORR	Objective Response Rate
OS	Overall Survival
PD-L1	Programmed death ligand 1
PFS	Progression-Free Survival
PK	Pharmacokinetics
PR	Partial Response
PT	Preferred Term

Q1	First Quartile
Q2W	Once Every Two Weeks
Q3	Third Quartile
Q3W	Once Every Three Weeks
QRS	Electrocardiographic deflection between the beginning of the Q wave and termination of the S wave representing time for ventricular depolarization
QT	Electrocardiographic interval between the beginning of the Q wave and termination of the T wave representing the time for both ventricular depolarization and repolarization to occur
QTc	Corrected QT interval
QTcB	Corrected QT interval based on Bazett's formula
QTcF	Corrected QT interval based on Fridericia's formula
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	Serious Adverse Event
SD	Stable Disease
SOC	System Organ Class
STD	Standard Deviation
TTR	Time To Response
ULN	Upper Limit of the Normal Range
WHO	World Health Organization

PHARMACOKINETIC ABBREVIATIONS

AUC_{last}	area under the concentration versus time curve from time zero to the last quantifiable concentration
AUC_{tau}	area under the concentration versus time curve over the dosing interval
AUC_{inf}	area under the concentration versus time curve from time zero extrapolated to infinity
$\%AUC_{exp}$	percentage of the area under the curve extrapolated to infinity observed from last quantifiable concentration to infinity
C_{last}	last observed quantifiable concentration of the drug
C_{max}	maximum observed concentration of drug
C_{tau}	observed drug concentration at the end of the dosing interval
C_0	initial concentration, given only for bolus IV models.
CL	total body clearance calculated as Dose divided by AUC_{inf}
V_{SS}	estimate of the volume of distribution at steady state
V_z	volume of distribution based on the terminal phase
$t_{1/2}$	estimate of the terminal elimination half-life of the drug, calculated by dividing the natural log of 2 by the terminal elimination rate constant (λ_z)
T_{last}	time (observed time point) of C_{last}
T_{max}	time (observed time point) of C_{max}
λz	terminal elimination rate constant, estimated by linear regression of the terminal elimination phase of the concentration of drug versus time curve

1. INTRODUCTION

This document describes the statistical analysis methods and data presentations to be used in the summary and analysis of data from Study GS-US-296-1080. Related documents are the study protocol and case report forms.

1.1. Study Objectives

The primary objective of this study is as follows:

- To compare the efficacy of andecaliximab (GS-5745) versus placebo in combination with mFOLFOX6 as measured by overall survival (OS)

The secondary objectives of this study are as follows:

- To compare the efficacy of andecaliximab versus placebo in combination with mFOLFOX6 as measured by progression-free survival (PFS)
- To compare the efficacy of andecaliximab versus placebo in combination with mFOLFOX6 as measured by objective response rate (ORR) per Response Evaluation Criteria In Solid Tumors version 1.1 (RECIST v1.1)
- To compare the safety of andecaliximab versus placebo in combination with mFOLFOX6

The exploratory objectives of this study are as follows:

1.2. Study Design

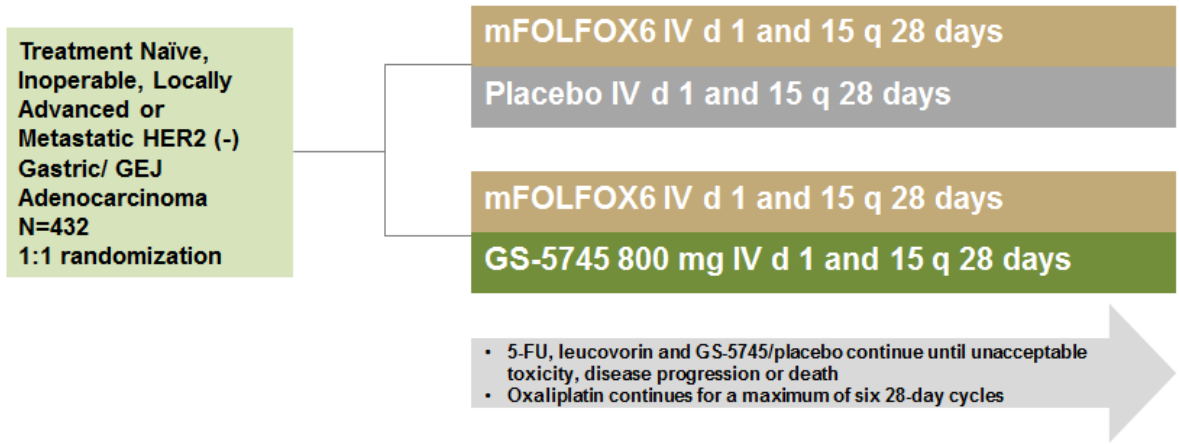
This is a Phase 3, randomized, double-blind, multicenter study of andecaliximab combined with mFOLFOX6 in subjects with untreated gastric and GEJ adenocarcinoma. A total of 430 eligible subjects with advanced gastric and GEJ cancer will be randomized in a 1:1 manner to receive mFOLFOX6 plus andecaliximab or mFOLFOX6 plus placebo. Treatment assignment will be stratified by ECOG status (0 or 1), geographic region (Latin America or other participating countries), and primary tumor site (gastric or GEJ).

Computed tomography (CT) or magnetic resonance imaging (MRI) scans will be performed every 8 weeks to evaluate response to treatment by RECIST v1.1.

Dosage and frequency will be as follows:

- mFOLFOX6 on Days 1 and 15 of each 28-day treatment cycle for a total of 6 cycles followed thereafter by leucovorin (LV) and 5-fluorouracil (5-FU) dosing on Days 1 and 15 of each 28-day treatment cycle until disease progression. The mFOLFOX6 dosing regimen will consist of *l*-LV 200 mg/m^2 or *dl*-LV 400 mg/m^2 and oxaliplatin 85 mg/m^2 followed by bolus 5-FU 400 mg/m^2 and a 46-hour infusion of 5-FU 2400 mg/m^2.
- Andecaliximab /Placebo 800 mg every 2 weeks until disease progression

Figure 1-1. Study Schema



An independent data monitoring committee (IDMC) will review the progress of the study and perform interim reviews of safety data. Safety review by the IDMC will be performed when the first 60 subjects have completed 4 treatment cycles. Thereafter, review of safety data will be performed at regular intervals as described in the DMC charter. In addition, the IDMC will meet after approximately 33.3% and approximately 66.7% of the expected number of OS events has occurred to review the results from the futility and efficacy interim analysis, respectively.

1.2.1. Inclusion Criteria

Subjects must meet all of the following inclusion criteria to be eligible for participation in this study.

1) Male or female ≥18 years of age

2) Histologically confirmed adenocarcinoma of the stomach or GEJ with inoperable, locally advanced or metastatic disease, not amenable to curative therapy

   Adenocarcinoma of the GEJ is defined as tumors that have their center within 5 cm proximal and distal of the anatomical esophagogastric junction as described in Siewert's classification system

3) Eastern Cooperative Oncology Group (ECOG) ≤ 1

4) Measurable disease or non-measurable but evaluable disease, according to RECIST v1.1. Subjects with peritoneal disease would generally be regarded as having evaluable disease and allowed to enter the trial

5) Subjects not receiving anticoagulant medication must have an international normalized ratio (INR) ≤ 1.5 and activated partial thromboplastin time (aPTT) ≤ 1.5 X upper limit of normal (ULN)

   The use of full-dose oral or parenteral anticoagulants is permitted as long as the INR or aPTT is within therapeutic limits (according to the medical standard in the institution) and the subject has been on stable dose of anticoagulants for at least 1 week at the time of randomization.

6) Adequate hematologic function:

   a) neutrophils ≥ 2.0 x 10^9/L

   b) platelets ≥ 100 x 10^9/L

   c) hemoglobin ≥ 9 g/dL

7) Adequate hepatic function:

   a) Direct or total bilirubin ≤ 1.5 x ULN.

   b) ALT and AST ≤ 2.5 x ULN, in case of liver metastases ≤ 5 x ULN

8) Creatinine clearance (CL_{cr}) should be ≥ 30 mL/min based on the Cockroft -Gault formula. Subjects with a CL_{cr} just below 30 mL/min may be eligible if a measured CL_{cr} (based on 24 hour urine collection or other reliable method) is ≥ 30 mL/min

9) For female subjects of childbearing potential, willingness to use a protocol-recommended method of contraception from the screening visit throughout the study treatment period, for 90 days following the last dose of study drug (GS-5745/placebo), and for 4 months after the last dose of oxaliplatin or 6 months after the last dose of 5-FU whichever occurs later unless the subject chooses continuous heterosexual abstinence as a lifestyle-choice

10) For male subjects of reproductive potential, willingness to use a protocol-recommended method of contraception and to refrain from sperm donation from the start of study drug, throughout the study treatment period, for 90 days after administration of the last dose of any study drug, and for 6 months after the last dose of oxaliplatin or 6 months following the last dose of 5-FU whichever occurs later

11) Breastfeeding females must agree to discontinue nursing before study drug administration

12) In the judgment of the investigator, participation in the protocol offers an acceptable benefit-to-risk ratio when considering current disease status, medical condition, and the potential benefits and risks of alternative treatments for the subject's cancer

13) Willingness to comply with scheduled visits, drug administration plan, imaging studies, laboratory tests, other study procedures, and study restrictions

14) Evidence of a signed informed consent prior to implementation of any protocol specific procedure

1.2.2. Exclusion Criteria

Subjects who meet *any* of the following exclusion criteria are not to be randomized in this study.

1) Previous chemotherapy for locally advanced or metastatic gastric or GEJ cancer. Subjects may have received prior neoadjuvant or adjuvant chemotherapy as long as it was completed at least 6 months prior to randomization

2) Human Epidermal Growth Factor Receptor 2 (HER2)-positive gastric cancer (primary tumor or metastatic lesion). HER2-positivity is defined as either IHC3+ or IHC2+/ISH+ (ISH positivity is defined as a HER2:CEP17 ratio of ≥2.0.)

3) Patients who have received palliative radiation and have not recovered from all acute, reversible effects.

4) Uncontrolled intercurrent illness including, but not limited to, active uncontrolled infection, active gastrointestinal bleeding, uncontrolled cardiac arrhythmia, or psychiatric illness/social situation that would limit compliance with study requirements as judged by treating physician

5) History of a concurrent or second malignancy except for adequately treated local basal cell or squamous cell carcinoma of the skin, cervical carcinoma in situ, superficial bladder cancer, asymptomatic prostate cancer without known metastatic disease and with no requirement for therapy or requiring only hormonal therapy and with normal prostate-specific antigen for ≥ 1 year prior to randomization, adequately treated Stage 1 or 2 cancer currently in complete remission, or any other cancer that has been in complete remission for ≥ 5 years

6) Major surgery, defined as any surgical procedure that involves general anesthesia and a significant incision (ie, larger than what is required for placement of central venous access, percutaneous feeding tube, or biopsy), within 28 days of first dose of study drug

7) Known positive status for human immunodeficiency virus (HIV)

8) Known acute or chronic-active infection with hepatitis B virus (HBV) or hepatitis C virus (HCV)

9) Peripheral neuropathy ≥ Grade 2 according to National Cancer Institute Common Terminology Criteria for Adverse Events version 4.03 (NCI CTCAE v.4.03)

10) Chronic daily treatment with oral corticosteroids (dose of > 10 mg/day methylprednisolone equivalent). Inhaled steroids and short courses of oral steroids for anti-emesis or as an appetite stimulant are allowed

11) Pregnant or breastfeeding women (pregnancy needs to be excluded by testing of beta-human chorionic gonadotropin [β-hCG])

12) Known or suspected central nervous system metastases

13) Known dihydropyrimidine dehydrogenase-deficiency (special screening not required)

14) Known alcohol or drug abuse or any other medical or psychiatric condition which contraindicates participation in the study

15) Documented myocardial infarction or unstable/uncontrolled cardiac disease (ie, unstable angina, congestive heart failure [New York Heart Association > Class II]) within 6 months of randomization

16) Active tuberculosis or history of latent tuberculosis that has not been treated

17) Any chronic medical condition that, in the opinion of the Investigator, would make the subject unsuitable for the study or would prevent compliance with the study protocol.

18) Serious systemic fungal, bacterial, viral, or other infection that is not controlled or requires intravenous antibiotics

19) Experimental medical treatment within 28 days prior to randomization

20) Known hypersensitivity to any of the study drugs or excipients or to Chinese hamster ovary cell products or to recombinant human or humanized antibodies**,** or known allergic reactions to products that contain platinum

21) History of long QT syndrome or whose corrected QT interval (QTc) measured using Fridericia's formula ($QTcF = QT/RR^{0.333}$) at screening is prolonged (> 450 ms for males and > 470 ms for females)

22) Subjects with potassium, magnesium or calcium less than the lower limit of normal (LLN); electrolyte replacement is permitted during screening

1.3. Sample Size and Power

Assuming a median OS time for the mFOLFOX6 + placebo group of 11.5 months, 286 OS events are needed to detect a hazard ratio (HR) of 0.70 with 85% power at a 2-sided significance level of 0.05, given 1 efficacy interim at 66.7% information. With an accrual period of 18 months, minimum follow-up of 18 months, and a 10% annual dropout rate, a total sample size of 430 subjects (215 subjects per treatment group) is needed to observe the required 286 OS events within the 36-month time frame.

As the targeted number of OS events is large (~286), if the null hypothesis of the primary end point of OS is rejected, it will convincingly demonstrate clinical treatment effect and would provide a narrow confidence interval. Based on the number of deaths and the assumed treatment effect on OS in the protocol, the expected 95% confidence interval on the HR of OS between the 2 treatment groups is ~ (0.627, 0.996). At the time of final analysis and assuming OS is significant, PFS will be tested at one-sided type I error of 0.016. Assuming that the hazard ratio in PFS is 0.7, which is expected to be on par or better than the treatment effect in OS, and that median PFS in control is 9 months, a sample size of 430 subjects (322 PFS events) will provide 85% power under the planned study enrolment (18 months), follow-up duration (18 month), and drop-out rate (annually 10%).

2. TYPE OF PLANNED ANALYSES

2.1. Interim Analyses

An independent data monitoring committee (IDMC) will review the progress of the study and perform reviews of safety data. The IDMC safety review will follow the schedule defined in the study protocol and the DMC charter. The IDMC will provide a recommendation to Gilead regarding whether the nature, frequency, and severity of adverse effects associated with study treatment indicate that the early termination of the study is in the best interests of the participants, the study should continue as planned, or the study should continue with modifications.

The IDMC will also review the unblinded efficacy data at the 2 planned interim analyses and make recommendations based on the pre-specified efficacy boundaries.

The first formal interim analysis is for futility and will be conducted when ~33% of expected 286 OS events have occurred. Based on the analysis results, the IDMC may recommend terminating the study due to futility if the predictive power on the OS endpoint in the intent to treat (ITT) analysis set is below 14%; otherwise, the study will be continued. Predicative power is defined as a weighted average of the conditional power; the weighting function is determined by the sampling distribution of the observed hazard ratio based on the data at the interim analysis. The futility decision rule is considered to be non-binding. There is no intent to claim efficacy therefore no alpha will be spent at the futility interim.

The second formal interim analysis is for efficacy and will be conducted when ~67% of expected 286 OS events have occurred. The type I error rate for testing OS will be controlled by the alpha spending function with O'Brien-Fleming type boundaries. Based on the assumed information time (67%) for the interim analysis, the OS endpoint will be tested at a one-sided alpha level of 0.006 at the second interim analysis. The actual significance levels used will be adjusted based on the actual number of OS events that occur. If the null hypothesis of OS endpoint is not rejected, the study will continue to final analysis. If the null hypothesis of OS endpoint is rejected, the IDMC could recommend early stop due to efficacy, and the secondary endpoints will be tested in sequence. The PFS endpoint will be tested at a one-sided alpha level of 0.016. Only if the PFS endpoint is rejected will the ORR endpoint be tested at a one-sided alpha level of 0.016.The actual significance levels used will be adjusted based on the actual number of OS events that occur. The stopping boundaries at each efficacy analysis time are provided in Table 2-1. Additional details are provided in Section 3.5 and Appendix 3.

Table 2-1. Stopping Boundaries for Efficacy Analyses

Efficacy Analysis	Event (%)	One-sided Decision Boundary		
		OS	PFS (After OS boundary is crossed)	ORR (After PFS boundary is crossed)
Interim	191 (66.7%)	0.006	0.016	0.016
Final	286 (100%)	0.023	0.016	0.016

2.2. Final Analysis

The final analysis for the primary endpoint will be conducted after approximately 286 OS events have occurred. Once all outstanding data queries have been resolved, the database will be locked, the blind will be broken, and the efficacy analysis of the study will be performed.

At the final analysis, the type I error rate for testing OS will be controlled by the alpha spending function with O'Brien-Fleming type boundaries. The OS event point will be tested at a one-sided alpha level of 0.023 at 286 OS events. The actual significance levels used will be adjusted based on the actual number of overall events that occur. If the OS endpoint is rejected, the PFS endpoint will be tested at a one-sided alpha level of 0.016. If the PFS endpoint is also rejected, the ORR endpoint will be tested at a one-sided alpha level of 0.016. The actual significance levels used will be adjusted based on the actual number of OS events that occur.

2.3. Follow-up Analysis

After the final analysis, additional supplemental analyses of efficacy and safety may be performed to satisfy regulatory requirements or to perform long-term efficacy (eg, OS) and follow-up safety assessments.

2.4. Changes from Protocol-Specified Analysis

Primary (OS) and secondary (PFS and ORR) efficacy endpoints may be examined in subgroups defined by baseline tumor programmed death ligand 1 (PD-L1) status (positive vs. negative) (Section 3.4). This analysis was not specified in the protocol. It is added prior to the study unblinding for the primary analysis. Coded archival samples will be provided to the CLIA/CAP-accredited testing laboratory. Both the laboratory and pathologist at the testing laboratory will be blinded to treatment and clinical outcome.

3. GENERAL CONSIDERATIONS FOR DATA ANALYSIS

Data will be summarized using descriptive statistics. For categorical variables, the number and percentage of subjects in each category will be presented; for continuous variables, the number of subjects (n), mean, standard deviation (StD) or standard error (SE), median, first quartile (Q1), third quartile (Q3), minimum, and maximum will be presented.

3.1. Analysis Sets

Analysis sets define the subjects to be included in an analysis. Analysis sets and their definitions are provided in this section. The analysis set will be identified and included as a subtitle of each table, figure, and listing.

For each analysis set, the number and percentage of subjects eligible for inclusion, as well as the number and percentage of subjects who were excluded and the reasons for their exclusion, will be summarized by treatment group.

A listing of reasons for exclusion from analysis sets will be provided by subject.

3.1.1. Intent-to-Treat Analysis Set

The ITT Analysis Set includes all subjects who were randomized in the study. This is the primary analysis set for efficacy analyses.

3.1.2. Per-protocol Analysis Set

The per-protocol (PP) analysis set includes subjects in the ITT and excludes those who meet any one of the following criteria:

- No documented diagnosis of histologically confirmed adenocarcinoma of the stomach or GEJ with inoperable, locally advanced or metastatic disease, not amenable to curative therapy
- Received previous chemotherapy for locally advanced or metastatic gastric or GEJ cancer
- ECOG > 1 at screening
- HER2 positive
- Did not receive study treatment
- Compliance with study drug < 75%

The PP analysis set is the secondary analysis for efficacy analysis.

3.1.3. Safety Analysis Set

The Safety Analysis Set includes all subjects who received at least 1 dose of andecaliximab/placebo. This is the primary analysis set for safety analyses.

3.1.4. Pharmacokinetic Analysis Set

The Pharmacokinetic (PK) Analysis Set will include all randomized subjects who received at least 1 dose of andecaliximab throughout the study or 1 dose of 5-FU or oxaliplatin on Cycle 2 Day 1, and have at least 1 nonmissing postdose concentration value reported by the PK laboratory. This is the primary analysis set for the summary of plasma concentrations of andecaliximab, 5-FU and oxaliplatin.

3.1.5. Pharmacokinetic Substudy Analysis Set

CCI [redacted]

3.1.6. Immunogenicity Analysis Set

The Immunogenicity Analysis Set will include all randomized subjects who received at least 1 dose of andecaliximab and have at least 1 nonmissing postdose antidrug antibody (ADA) status reported. This is the primary analysis set for all immunogenicity analyses.

3.2. Subject Grouping

For analyses based on the ITT Analysis Set, subjects will be grouped according to the treatment to which they were randomized.

For analyses based on the PP Analysis Set and Safety Analysis Set, subjects will be grouped according to the randomized treatment except when their actual treatment differs from randomized treatment for the entire treatment duration. In this case, subjects will be grouped based on actual treatment received.

For the PK Analysis Set, subjects will be grouped according to the actual treatment they received.

3.3. Strata and Covariates

Subjects will be randomly assigned to treatment groups via the interactive web response system (IWRS) in a 1:1 ratio using a stratified randomization schedule. Stratification will be based on the following variables:

- Geographic regions: Latin America vs All Other Countries
- ECOG status: 0 vs 1
- Tumor site: Gastric vs GEJ

If there are discrepancies in stratification factor values between the IWRS and the clinical database, the baseline values recorded in the clinical database will be used for analyses.

Efficacy endpoints will be evaluated using stratification factors as covariates or stratification variables for analyses, as specified in Section 6.

In the situation where there is insufficient information in a stratum (ie, if there are < 6 subjects or there is no informative event in a stratum by combined treatment arms), pooling of the stratum may be considered with the smallest adjacent stratum for stratified analyses; the smallest stratum is defined as that stratum having the fewest number of subjects or the fewest number of events in case the former is a tie and the adjacent stratum is defined as a stratum having 2 factors of the 3 at the same level and the other factor at an adjacent level.

3.4. Examination of Subject Subsets

Primary (OS) and secondary (PFS and ORR) efficacy endpoints will be examined in the following subgroups:

- Stratification factors:
  - Geographic regions: Latin America vs All Other Countries
  - ECOG status: 0 vs 1
  - Tumor site: Gastric vs GEJ
- Extent of Disease
  - Locally Advanced
  - Metastatic

    Note that CRF did not collect extent of disease directly and this will be derived as follows. Subjects with disease stage III at screening will be considered “locally advanced” and those with disease stage IV will be considered as “metastatic”.
- Measurability at Baseline
  - Measurable
  - Non-measurable
- Baseline tumor PD-L1 status
  - Positive: PD-L1 expression if Combined Positive Score (CPS) ≥ 1, using PD-L1 immunohistochemistry (IHC) 22C3 pharmDx assay per its label.
  - Negative: PD-L1 expression if CPS < 1, using PD-L1 IHC 22C3 pharmDx assay per its label.

    Note that inevaluable samples will be considered as missing and excluded from the subgroup analysis.

- Age group
  - < 65
  - >= 65
- Gender
  - Male
  - Female
- Race
  - White
  - Non-White

Adverse events (AEs) and lab abnormalities will be examined in the following subgroups:

- Age group
  - < 65
  - >= 65
- Gender
  - Male
  - Female
- Race
  - White
  - Non-White

3.5. Multiple Comparisons

A gate keeping approach will be employed to protect the family wise type I error for the study when testing the primary and secondary endpoints. Assuming the efficacy interim and final analyses are conducted when 191 and 286 OS events have occurred, the testing of the efficacy endpoints will be carried out as follows:

- At the efficacy interim analysis, the OS event point will be tested at a one-sided alpha level of 0.006. If the OS endpoint is not rejected, the study will continue to final analysis. If the OS endpoint is rejected, the IDMC could recommend early stop for efficacy. Furthermore, the PFS endpoint will be tested at a one-sided alpha level of 0.016. Only if the PFS endpoint is rejected will the ORR endpoint be tested at a one-sided alpha level of 0.016.

- At the final analysis, the OS event point will be tested at a one-sided alpha level of 0.023. If the OS endpoint is rejected, the PFS endpoint will be tested at a one-sided alpha level of 0.016. If the PFS endpoint is also rejected, the ORR endpoint will be tested at a one-sided alpha level of 0.016.

The actual significance levels used will be adjusted based on the actual number of OS events that occur.

Overall, this testing strategy employs O'Brian-Fleming type boundary for the primary OS endpoint and the Pocock type boundary for the secondary PFS/ORR endpoints. The testing strategy controls the overall one-sided family-wise type I error to be at 0.025, equivalent to two-sided 0.05 by appropriately adjusting for multiplicity in the efficacy interim and final analyses. Detailed statistical justification of the testing strategy can be found in the Appendix 3.

3.6. Handling of Dropouts, Missing Data and Outliers

3.6.1. Missing Data

A missing data point for a given study visit may be due to any of the following reasons:

- A visit occurred in the window but data were not collected or were unusable
- A visit did not occur in the window
- A subject permanently discontinued from the study before reaching the window

In general, values for missing data will not be imputed unless methods for handling missing data are specified.

For missing last dosing date of study drug, imputation rules are described in Section 4.2.1. The handling of missing or incomplete dates for disease diagnosis is described in Section 5.3.1, for death in Section 6.1, for new anticancer therapy in Section 6.2.1.1, for AE onset in Section 7.1.5.2, and for prior and concomitant medications in Section 7.4 . For subjects without any pre-dose laboratory values, the baseline lab toxicity grade will be treated as Grade 0 for the summary of graded laboratory abnormalities.

3.6.2. Outliers

Outliers will be identified during the data management and data analysis process, but no sensitivity analyses will be conducted. All data will be included in the data analysis.

Addition details are provided in Section 3.7.

3.7. Data Handling Conventions and Transformations

3.7.1. Data Handling for Laboratory Data

Laboratory data that are continuous in nature but are less than the lower limit of quantitation (LLOQ) or above the upper limit of quantitation will be imputed as follows:

- A value that is 1 unit less than the limit of quantitation will be used for calculation of descriptive statistics if the data is reported in the form of "< x" (x is considered the limit of quantitation). For example, if the values are reported as < 50 and < 5.0, then values of 49 and 4.9 will be used for calculation of summary statistics, respectively. However, for direct bilirubin, a value of "< 0.1" will be treated as 0.05 for calculation of summary statistics.

- A value that is 1 unit above the limit of quantitation will be used for calculation of descriptive statistics if the data is reported in the form of "> x" (x is considered the limit of quantitation). For example, if the values are reported as > 50 and > 5.0, then values of 51 and 5.1 will be used for calculation of summary statistics, respectively.

- The limit of quantitation will be used for calculation of descriptive statistics if the data is reported in the form of "≤ x" or "≥ x" (x is considered as the limit of quantitation).

3.7.2. Data Handling for PK Data

For PK plasma /blood concentrations and analysis of PK parameters natural logarithmic transformation will be used. Plasma concentration values that are below the limit of quantitation (BLQ) will be presented as "BLQ" in the concentration data listing. Values that are BLQ will be treated as 0 at predose time points, and one-half the value of LLOQ at postdose time points for determination of summary and order statistics.

The following conventions will be used for the presentation of summary and order statistics:

- If at least 1 subject has a concentration value of BLQ for the time point, the minimum value will be displayed as "BLQ."

- If more than 25% of the subjects have a concentration data value of BLQ for a given time point, the minimum and Q1 values will be displayed as "BLQ."

- If more than 50% of the subjects have a concentration data value of BLQ for a given time point, the minimum, Q1, and median values will be displayed as "BLQ."

- If more than 75% of the subjects have a concentration data value of BLQ for a given time point, the minimum, Q1, median, and Q3 values will be displayed as "BLQ."

- If all subjects have concentration data values of BLQ for a given time point, all order statistics (minimum, Q1, median, Q3, and maximum) and summary statistics will be displayed as "BLQ."

PK parameters that are BLQ will be imputed as one-half LLOQ before log transformation or statistical model fitting.

3.8. Analysis Visit Windows

3.8.1. Definition of Baseline and Study Day

Baseline for efficacy is defined as the last observation within 42 days on or prior to randomization, unless otherwise specified. In the rare cases where the screening assessments are after the randomization date but prior to the first dose of andecaliximab/placebo, the assessments will be considered as baseline.

Baseline for safety is defined as the last observation within 42 days prior to first dose of andecaliximab/placebo, unless otherwise specified.

Study Day 1 for the efficacy analysis will be defined as the day of randomization.

Study Day 1 for the safety analysis will be defined as the day for the first dose of andecaliximab/placebo.

Other study days will be derived as follows:

- Assessment Date – Study Day 1 + 1 if Assessment Date >= Study Day 1
- Assessment Date – Study Day 1 if Assessment Date < Study Day 1

3.8.2. Baseline and Analysis Windows

Subject visits might not occur on protocol-specified days. Therefore, for the purpose of analysis, observations will be assigned to analysis windows.

For post baseline parameters that will be summarized by visit, an evaluation that occurs within the visit window will be considered valid for that evaluation. The visit windows are specified in Table 3-1, Table 3-2, and Table 3-3. The lowest and highest study days are inclusive.

Table 3-1 Post Baseline Visit Windows for Chemistry, Hematology and Vital Signs

Analysis Visit	Relative Day	Lowest Day (inclusive)	Highest Day (inclusive)
Week 2	15	1 (after first dose)	21
Week 4	29	22	35
Week 6	43	36	49
Week X=8, 10, …	7*x+1	7x-6	7*x + 7

Table 3-2 Post Baseline Visit Windows for ECG, QoL and Urinalysis

Analysis Visit	Relative Day	Lowest Day (inclusive)	Highest Day(inclusive)
Week 4	29	1 (after first dose)	43
Week 8	57	44	71
Week 12	85	72	99
Week X=16, 20, …	7*x+1	7x-12	7*x + 15

Table 3-3 Post Baseline Visit Windows for Tumor Assessment

Analysis Visit	Relative Day	Lowest Day (inclusive)	Highest Day (inclusive)
Week 8	57	2	85
Week 16	113	86	141
Week 24	169	142	197
Week X=32, 40, …	7*x+1	7x-26	7*x + 29

3.8.3. Selection of Data in the Event of Multiple Records in an Analysis Window

Depending on the statistical analysis method, single values may be required for each analysis window. For example, change from baseline by visit usually requires a single value per visit.

If multiple, valid, non-missing values exist in an analysis window, records will be chosen based on the following rules if a single value is needed:

For baseline:

- If there are multiple records with the same time or no time recorded on the same day, the average (arithmetic or geometric mean, as appropriate) will be used for continuous variables; the value with the lowest severity will be used for categorical variables.

For post-baseline:

- If more than one assessment occurs during the same nominal visit, select the record closest to the scheduled day for that visit.
- If there are 2 assessments that are equidistant from the scheduled day, the later record will be selected.
- The last measurement will be used if multiple measurements are taken on the same day; if the measurements occur at the same time or time cannot be determined, the average will be taken for continuous variables and the most conservative value will be taken for categorical variables.

4. SUBJECT DISPOSITION

4.1. Subject Enrollment and Disposition

A summary of subject enrollment will be provided by treatment arm and overall. The summary will present the number and percentage of subjects enrolled. For each column, the denominator for the percentage calculation will be the total number of subjects enrolled for that column.

A similar enrollment table will be provided by each stratum in the stratified randomization. The denominator for the percentage of subjects in the stratum will be the total number of enrolled subjects within that stratum. If there are discrepancies in the value used for stratification assignment between the IWRS/IVRS and the clinical database, the value collected in the clinical database will be used for the summary. A listing of subjects with discrepancies in the value used for stratification assignment between the IWRS and the clinical database at the time of data finalization will be provided.

The randomization schedule used for the study will be provided as an appendix to the clinical study report (CSR)

A summary of subject disposition will be provided by treatment group in the ITT analysis set. The following categories will be included:

- Screened
- Randomized
- Treated with andecaliximab/placebo
- Discontinued all treatment drugs
- Discontinued all chemotherapy components
- Andecaliximab/placebo + mFOLFOX6 (all components) ongoing
- Andecaliximab/placebo + any chemo ongoing
- Andecaliximab/placebo ongoing
- Andecaliximab/placebo discontinued
  - Reasons for discontinuation
- Treated with mFOLFOX6
  - 5-FU bolus ongoing
  - 5-FU bolus discontinued
    - Reasons for discontinuation of 5-FU bolus

  - 5-FU infusion ongoing
  - 5-FU infusion discontinued
    - Reasons for discontinuation of 5-FU infusion
  - Oxaliplatin ongoing
  - Oxaliplatin discontinued
    - Reasons for discontinuation of oxaliplatin
  - Leucovorin ongoing
  - Leucovorin discontinued
    - Reasons for discontinuation of leucovorin
- Ongoing in lesion assessment
- Discontinued lesion assessment
  - Reasons for lesion assessment discontinuation
- Ongoing in study (including long term OS follow up)
- Completed study (including long term OS follow up)
- Discontinued study (including long term OS follow up)
  - Did not enter long term OS follow up
    - Reasons
  - Reasons for discontinuing long term OS follow up

The denominator for the percentage calculation will be the total number of subjects in the ITT Analysis Set corresponding to that column. In addition, the total number of subjects who were randomized and the number of subjects in each of the disposition categories listed above will be depicted using a flowchart.

The following by-subject listings will be provided by subject identification (ID) number in ascending order to support the above summary tables:

- Reasons for discontinuing study treatment, lesion assessment, or study.
- Reasons for screen failure (will be provided by screening ID number in ascending order)

4.2. Extent of Exposure

Extent of exposure to study drug will be examined by assessing the total duration of exposure to study drug, total number of doses, and the level of adherence to the study drug specified in the protocol.

4.2.1. Exposure to Andecaliximab/Placebo

A listing of andecaliximab/placebo administration will be provided by subject ID number (in ascending order) and visit (in chronological order), including dosing date/time, planned dosage, actual dosage administered, infusion outcome and reason for dose reduction (if applicable). Exposure to andecaliximab/placebo will be summarized by treatment arm for the safety analysis set.

4.2.1.1. Duration of Exposure

Duration of exposure to andecaliximab/placebo will be summarized using descriptive statistics (N, mean, standard deviation [StD], median, minimum and maximum). Duration of exposure to study drug will be defined as (last dose end date – first dose start date + 14), regardless of temporary interruptions in study drug administration, and will be expressed in weeks using up to 1 decimal place (eg, 4.5 weeks).

If the last study drug dosing end date is missing, the last study drug start date will be used.

4.2.1.2. Number of Doses

The number of doses subjects were exposed to study drug will be summarized using descriptive statistics, as well as the number and percentage of subjects exposed to a given cycle category. A subject is said to have received a dose of andecaliximab/placebo if he/she received any of the planned dose of andecaliximab/placebo.

4.2.1.3. Average Dose

The average dose in mg, which is defined as (total actual dose amount received in mg) / (number of doses given), will be summarized using descriptive statistics.

4.2.1.4. On-Treatment Adherence

The level of on-treatment adherence to the study drug regimen will be determined by the total amount of study drug administered relative to the total amount of study drug expected to be administered during a subject's actual on-treatment period based on the study drug regimen. Investigator-prescribed interruption and dose reductions as specified in the protocol will be taken into account.

The level of on-treatment adherence will be expressed as a percentage using the following formula:

$$\text{On-Treatment Adherence (\%)} = \left(\frac{\text{Total Amount of Study Drug Administered}}{\text{Study Drug Expected to be Administered on Treatment}} \right) \text{x } 100$$

Descriptive statistics for the level of on-treatment adherence with the number and percentage of subjects belonging to adherence categories (eg, < 75%, ≥ 75 to < 90%, ≥ 90%) will be provided by treatment group for the Safety Analysis Set. No formal statistical testing is planned.

4.2.2. Exposure to mFOLFOX6

Exposure to individual components of mFOLFOX6 will be listed and summarized in a similar manner as for andecaliximab/placebo. mFOLFOX6 received prior to study enrollment and after the start of other anti-cancer therapy at study completion/discontinuation will not be included.

4.3. Protocol Deviations

Subjects who did not meet the eligibility criteria for study entry but enrolled in the study will be summarized regardless of whether they were exempted by the sponsor or not. The summary will present the number and percentage of subjects who did not meet at least 1 eligibility criterion and the number of subjects who did not meet specific criteria by treatment group based on the ITT Analysis Set. A by-subject listing will be provided for those subjects who did not meet at least 1 eligibility (inclusion or exclusion) criterion. The listing will present the eligibility criterion (or criteria if more than 1 deviation) that subjects did not meet and related comments, if collected.

Protocol deviations occurring after subjects entered the study are documented during routine monitoring. The number and percentage of subjects with important protocol deviations by deviation reason (eg, non-adherence to andecaliximab/placebo, violation of select inclusion/exclusion criteria) will be summarized by treatment group for the ITT Analysis Set. A by-subject listing will be provided for those subjects with any protocol deviation.

5. BASELINE DATA

5.1. Demographics

Subject demographic variables (ie, age, sex, race, and ethnicity) will be summarized by treatment group and overall using descriptive statistics for age, and using number and percentage of subjects for sex, race, and ethnicity. Age at baseline is calculated in years at the date of randomization. The summary of demographic data will be provided for the ITT analysis set.

In addition, similar summaries by stratum of the stratified randomization will be provided.

A by-subject demographic listing, including the informed consent date, will be provided by subject ID number in ascending order.

5.2. Other Baseline Characteristics

Other baseline characteristics include body weight (in kg), height (in cm), and body mass index (BMI; in kg/m^2). These baseline characteristics will be summarized by treatment group and overall using descriptive statistics for continuous variables and using number and percentage of subjects for categorical variables. For baseline body weight, height, and BMI, descriptive statistics will also be presented by sex in the same table. The summary of baseline characteristics will be provided for the ITT analysis set.

A by-subject listing of other baseline characteristics will be provided by subject ID number in ascending order.

5.3. Medical History

General medical history data will be collected at screening and listed only. General medical history data will not be coded.

A by-subject listing of general medical history will be provided by subject ID number in ascending order.

5.3.1. Disease History

A summary of disease history will be provided by treatment arms for the ITT analysis set. Variables to be summarized include:

- Primary tumor site
- Histology of gastric cancer
- Disease stage at screening
- Differentiation at screening
- Time since diagnosis (weeks)

These variables will be summarized using number and percentages. The denominator for the percentage of histology of gastric cancer is the number of subjects with gastric cancer. The denominators for the percentage of other variables are the numbers of subjects in the ITT analysis set corresponding to that column.

A by-subject listing of disease history will be provided by subject ID number in ascending order.

In deriving the time since diagnosis, all partial dates of diagnosis will be identified, and the partial dates will be imputed as follows:

- If day is missing but the month and year are available, then the imputed day will be the first day of the month;
- If day and month are missing but year is available, then the imputed day and month will be 01Jan;
- Partial date will not be imputed if the year is missing.

5.4. Prior Chemotherapy and Radiotherapy

Prior (neoadjuvant or adjuvant) chemotherapy and radiotherapy will be summarized using descriptive statistics by treatment arm using ITT analysis set.

A by-subject listing of prior chemotherapy and radiotherapy will be provided by subject ID number in ascending order.

6. EFFICACY ANALYSES

Efficacy analysis will be performed based on the ITT Analysis Set. The investigator assessments will be considered as primary for analyses of efficacy endpoints.

6.1. Primary Efficacy Endpoint

6.1.1. Definition of the Primary Efficacy Endpoint

The primary efficacy endpoint of this study is OS, defined as the time from the date of randomization to death from any cause. Subjects who do not have an event will be censored at the last time that subject was known to be alive.

Every attempt will be made to ensure that complete death dates are recorded. In those rare instances where complete death dates are not recorded, the following algorithm will be used:

- If day is missing but the month and year are available, then the imputed day will be the midpoint of the month or the last assessment date + 1, whichever is later.
- If day and month are missing but year is available, then the imputed day and month will be 01Jan or the last day of the latest month that the subject was known to be alive if they have the same year, whichever is later.

6.1.2. Statistical Hypothesis for the Primary Efficacy Endpoint

The primary efficacy hypothesis to be tested is that there is no difference between Treatment Group (Group T) and Control Group (Group C) in OS. Using $S_T(t)$ and $S_C(t)$ to denote the OS survival distribution functions of Treatment Group (Group T) and Control Group (Group C), respectively, the statistical hypotheses to be tested in this study will be:

H_0: $S_T(t) = S_C(t)$

H_1: $S_T(t) > S_C(t)$ (Group T is superior to Group C in terms of OS)

6.1.3. Analysis of the Primary Efficacy Endpoint

OS in months will be calculated by (date of event/censoring – date of randomization + 1) / 30.4375.

The analysis of OS will be performed using the Kaplan-Meier method for the ITT analysis set. The OS distribution of the 2 treatment arm will be compared using the log-rank test, stratified by the stratification factors at randomization. Medians, Q1, Q3, the proportion of subjects who are alive at 6, 12, 18, and 24 months from randomization will be provided along with the corresponding 95% CI. Kaplan-Meier curves will be provided by treatment arm.

In addition, the hazard ratio (HR) between the2 treatment arms and its 95% CI will be estimated using the Cox proportional hazards regression model, adjusted for the stratification factors at randomization. The analysis strategy for the situation where there is insufficient information in a stratum is detailed in Section 3.3.

Follow-up time for OS will be summarized as a continuous variable with descriptive statistics by treatment arm.

- For subjects who discontinued study due to lost to follow-up or withdraw of consent, duration of OS follow-up = Date of last follow up – randomization date +1
- For the remaining subjects (including those who die), duration of OS follow up = Date of analysis cut-off date – randomization date +1

6.1.4. Sensitivity Analysis of the Primary Efficacy Endpoint

To assess the robustness of the primary OS results, the following exploratory sensitivity analyses will be performed:

6.2. Secondary Efficacy Endpoint

6.2.1. Definition of the Secondary Efficacy Endpoints

6.2.1.1. Progression-Free Survival

Progression-free survival is defined as the time interval from the date of randomization to the earlier of the first documentation of definitive disease progression or death from any cause. All scans, whether scheduled or unscheduled, will be considered.

The date of definitive progression will be the time point at which progression is first identified by relevant radiographic imaging data. Data will be censored on the date of last adequate tumor assessment for subjects (including assessments with a not evaluable [NE] outcome):

- who do not have disease progression or die before study discontinuation, or
- who start new anticancer therapy prior to documented disease progression or death, or
- who have ≥ 2 consecutive missing tumor assessments before disease progression or death

If a subject does not have a baseline tumor assessment, then PFS will be censored at Study Day 1, regardless of whether or not definitive progression or death has been observed.

PFS in months will be calculated by (date of event/censoring – date of randomization + 1) / 30.4375.

When imaging examinations for one visit are conducted on various dates, the following rules apply for the calculation of the assessment date:

- The response date will be the last date associated with that particular imaging time point.
- The progression date will be the first date associated with that particular imaging time point.

When the date of initiation of anticancer therapy other than the study treatment is incomplete or missing, the following algorithm will be followed:

- If the day is missing but the month and year are available, then the imputed day will be the last day of the month.
- If day and month are missing but year is available, then the imputed day and month will be 01Jan or the last day of the month for the last adequate disease assessment if they have the same year, whichever is later.

6.2.1.2. Objective Response Rate

Objective response rate is defined as the proportion of subjects with best overall response during andecaliximab therapy of complete response (CR) or partial response (PR) based on RECIST 1.1. Subjects, who do not have sufficient baseline or on-study tumor status information to be adequately assessed for response status (ie, those with best overall response of NE or no disease [ND]) or received anticancer therapy other than the study treatment prior to achieving CR or PR, will be considered as non-responders and will be included in the denominators in calculations of response rates.

6.2.2. Statistical Hypothesis for the Secondary Efficacy Endpoints

6.2.2.1. Progression-free survival

The secondary efficacy hypothesis to be tested is that there is no difference between Treatment Group (Group T) and Control Group (Group C) in PFS. Using $F_T(t)$ and $F_C(t)$ to denote the progression-free survival distribution functions of Treatment Group (Group T) and Control Group (Group C), respectively, the statistical hypotheses to be tested in this study will be:

H_0: $F_T(t) = F_C(t)$

H_1: $F_T(t) > F_C(t)$ (Group T is superior to Group C in terms of PFS)

6.2.2.2. Objective Response Rate

Another secondary efficacy hypothesis to be tested is that there is no difference between Treatment Group (Group T) and Control Group (Group C) in ORR. Using R_T and R_C to denote the objective response rate of Treatment Group (Group T) and Control Group (Group C), respectively, the statistical hypotheses to be tested in this study will be:

H_0: $R_T = R_C$

H_1: $R_T > R_C$ (Group T is superior to Group C in terms of ORR)

6.2.3. Analysis for the Secondary Efficacy Endpoints

6.2.3.1. Progression-free survival

The analysis of PFS will be performed using the Kaplan-Meier method for the ITT analysis set. The PFS distribution of the 2 treatment arms will be compared using the log-rank test, stratified by the stratification factors at randomization. Medians, Q1, Q3 and the proportion of subjects who are free from disease progression and alive at 6, 12, and 18 months from randomization will be provided along with the corresponding 95% CI. Kaplan-Meier curves will be provided by treatment arm.

In addition, the HR between the 2 treatment arms and its 95% CI will be estimated using the Cox proportional hazards regression model, adjusted for the stratification factors at randomization.

Follow-up time for PFS will be summarized as a continuous variable with descriptive statistics by treatment arm.

- For subjects who are censored from PFS and have discontinued from tumor assessment, duration of PFS follow up = Date of Censoring – randomization date +1
- For the remaining subjects (including subjects with progressive disease [PD] based on RECIST 1.1), duration of PFS follow up = Date of analysis cut-off date – randomization date +1

6.2.3.2. Objective Response Rate

The ORR will be compared between the 2 treatment arms using the Cochran-Mantel-Haenszel test adjusted for the stratification factors based on the ITT analysis set. ORR and the corresponding 95% CI based on Clopper-Pearson method of each treatment arm will be presented. The odds ratio comparing the 2 treatment arms adjusted for the stratification factors will be presented along with 95% CI. The 2-sided 95% CI of difference for ORR between the treatment and placebo will be calculated based on stratum-adjusted CMH proportion {Koch 1989}.

6.2.4. Sensitivity Analysis of the Secondary Efficacy Endpoints

To assess the robustness of the secondary endpoint results, the following exploratory sensitivity analyses will be performed:

CCI

6.3. Exploratory Efficacy Endpoints

6.3.1. Definition of the Exploratory Efficacy Endpoints

CCI

CCI [redacted]

[redacted]

6.3.2. Analysis of the Exploratory Endpoints

CCI [redacted]

[redacted]

CCI [redacted]

CCI [redacted]

7. SAFETY ANALYSES

Safety will be evaluated for all subjects in the Safety Analysis Set through assessment of clinical laboratory tests, physical examination findings, 12-lead ECG abnormalities, vital signs measurements, and by the documentation of AEs. Concomitant medication intake will also be recorded and coded to the corresponding WHO drug term and displayed in data listings.

All safety data collected at baseline and post-baseline through to the completion of the follow-up evaluation will be summarized. Data for the pre-treatment period will be included in data listings.

7.1. Adverse Events

7.1.1. Adverse Event Dictionary

Clinical and laboratory AEs will be coded using the current version of MedDRA. System organ class (SOC), high-level group term (HLGT), high-level term (HLT), preferred term (PT), and lower-level term (LLT) will be provided in the AE dataset.

7.1.2. Adverse Event Severity

Adverse events are graded by the investigator as Grade 1, 2, 3, 4, or 5 according to the Common Terminology Criteria for Adverse Events (CTCAE) Version 4.03. The severity grade of events for which the investigator did not record severity will be categorized as "missing" for tabular summaries and data listings. The missing category will be listed last in data presentation.

7.1.3. Relationship of Adverse Event to Study Treatment

Related AEs are those for which the investigator selected "Related" on the AE case report form (CRF) to the question of related to study treatment. Relatedness will always default to the investigator's choice, not that of the medical monitor. Events for which the investigator did not record relationships to study treatment will be considered related to study treatment for summary purposes. However, by-subject data listings will show the relationship as missing from that captured on the CRF.

7.1.4. Serious Adverse Events

Serious adverse events (SAEs) will be identified and captured as SAEs if AEs met the definitions specified in the study protocol as determined by the investigator. Serious adverse events captured and stored in the clinical database will be reconciled with the SAE database from the Gilead Drug Safety and Public Health Department before database finalization.

7.1.5. Treatment-Emergent Adverse Events

7.1.5.1. Definition of Treatment-Emergent

Treatment-emergent AEs (TEAEs) are events in a given study period that meet any of the following criteria:

- Any AEs with onset date of on or after andecaliximab/placebo start date and no later than 30 days after permanent discontinuation of all study treatment (andecaliximab /placebo and chemotherapy) or
- Any AEs with onset date of on or after the andecaliximab/placebo start date and no later than 55 days after permanent discontinuation of andecaliximab/placebo or
- AEs leading to discontinuation of andecaliximab/placebo

7.1.5.2. Incomplete Dates

All AEs with partial onset dates will be identified and the partial dates will be imputed for TEAE determination. The imputation rules are as follows:

- If day and month are missing but year is available, then the imputed day and month will be 01Jan or the first dosing date if they have the same year, whichever is later.
- If day is missing but the month and year are available, then the imputed day will be the first day of the month or the first dosing date if they have the same month and year, whichever is later.
- AEs with completely missing onset dates and stop dates, or with the onset date missing and a stop date later than the first dosing date of andecaliximab/placebo, will be considered to be treatment emergent. In addition, an AE with the onset date missing and an incomplete stop date with the same or later month and year (or year alone if month is not recorded) as the first dosing date of study drug will be considered treatment emergent.

7.1.6. Summaries of Adverse Events and Deaths

Treatment-emergent AEs will be summarized based on the Safety Analysis Set.

7.1.6.1. Summaries of AE Incidence by Severity

The number and percentage of subjects who experienced at least 1 TEAE will be provided and summarized by SOC, PT, maximum severity, and treatment group. For other AEs described below, summaries will be provided by SOC, PT, and treatment group:

- any TEAE
- any Grade 3 or higher TEAE

- any TE SAE
- any TEAE leading to andecaliximab/placebo dose interruption
- any TEAE leading to 5-FU bolus dose modification or interruption
- any TEAE leading to 5-FU infusion dose modification or interruption
- any TEAE leading to oxaliplatin dose modification or interruption
- any TEAE leading to andecaliximab/placebo discontinuation
- any TEAE leading to 5-FU bolus discontinuation
- any TEAE leading to 5-FU infusion discontinuation
- any TEAE leading to oxaliplatin discontinuation
- any TEAE leading to death
  - AE leading to death disease related
  - AE leading to death not disease related

For TEAEs, Grade 3 or higher TEAEs, and SAEs, the summary of AEs will be further categorized into related to andecaliximab/placebo, related to 5-FU bolus, related to 5-FU infusion, and related to oxaliplatin.

A brief, high-level summary of AEs described above will be provided by treatment group and by the number and percentage of subjects who experienced the above AEs.

Multiple events will be counted only once per subject in each summary. Adverse events will be summarized and listed first in alphabetic order of SOC (if applicable), and then by PT in descending order of total frequency within each SOC. For summaries by severity grade, the most severe grade will be used for those AEs that occurred more than once in an individual subject during the study.

In addition to the above summary tables, TEAEs, Grade 3 or higher TEAEs, SAEs, TEAEs leading to death, andecaliximab /placebo-related AEs, and TEAEs leading to discontinuation of andecaliximab /placebo will be summarized by PT only in descending order of total frequency.

Death will be summarized by whether it is considered as treatment emergent and the cause of the death (AE, due to PD, or other reasons).

Data listings will be provided as follows:

- AEs
- AEs with Grade 3 or higher
- SAEs
- AEs leading to andecaliximab/placebo interruption
- AEs leading to 5-FU bolus modification or interruption
- AEs leading to 5-FU infusion modification or interruption
- AEs leading to oxaliplatin modification or interruption
- AEs leading to andecaliximab/placebo discontinuation
- AEs leading to 5-FU bolus discontinuation
- AEs leading to 5-FU infusion discontinuation
- AEs leading to oxaliplatin discontinuation
- AE leading to Death
- All Deaths

7.2. Laboratory Evaluations

Laboratory data collected during the study, scheduled and unscheduled, will be analyzed and summarized using both quantitative and qualitative methods. The analysis will be based on values reported in conventional units. When values are below the LOQ, they will be listed as such, and the closest imputed value will be used for the purpose of calculating summary statistics as specified in Section 3.7. Hemolyzed test results will not be included in the analysis, but they will be listed in by-subject laboratory listings.

A by-subject listing for laboratory test results will be provided by subject ID number and time point in chronological order for hematology, serum chemistry, and urinalysis separately. Values falling outside of the relevant reference range and/or having a severity grade of 1 or higher on the CTCAE severity grade will be flagged in the data listings, as appropriate.

7.2.1. Summaries of Laboratory Results

A baseline laboratory value will be defined as the last measurement obtained on or prior to the date/time of first dose of andecaliximab/placebo. In the case of multiple values in an analysis window, data will be selected for analysis as described in Section 3.8.3.

Descriptive statistics (n, mean, StD, median, Q1, Q3, minimum, and maximum) will be provided by treatment arm for selected lab test parameters (Appendix 1) as follows:

- Baseline values
- Values at each post-baseline time point
- Change from baseline at each post-baseline time point

Change from baseline to a post baseline visit will be defined as the visit value minus the baseline value. The mean, median, Q1, Q3, minimum, and maximum will be displayed to the reported number of digits; StD to the reported number of digits plus 1. The mean (+/- SE) of change and percent change from baseline over time will also be plotted for each treatment arm.

Categorical laboratory data will be summarized using the number and percentage of subjects with the given response at baseline and each scheduled post-baseline time point by treatment arm.

7.2.2. Graded Laboratory Values

CTCAE Version 4.03 will be used for assigning toxicity grades (0 to 4) to laboratory results for analysis. Grade 0 includes all values that do not meet the criteria for an abnormality of at least Grade 1. Some laboratory tests have criteria for both increased and decreased levels; analyses for each direction (ie, increased, decreased) will be presented separately.

7.2.2.1. Treatment-Emergent Laboratory Abnormalities

Treatment-emergent laboratory abnormalities are defined as values that increase at least 1 toxicity grade from baseline at any post-baseline time point, up to 30 days after the last dose of all study treatment, or 55 days after the last dose of andecaliximab /placebo for subjects who permanently discontinued all study treatments, or all available data in the database snapshot for subjects who were still on treatment at the time of analysis. If the relevant baseline laboratory value is missing, then any abnormality of at least Grade 1 observed within the time frame specified above will be considered treatment emergent.

7.2.2.2. Summaries of Laboratory Abnormalities

The summaries (number and percentage of subjects) for treatment-emergent laboratory abnormalities will be provided by lab test for each treatment arm; subjects will be summarized by the most severe post-baseline abnormality grade for given lab tests.

A by-subject listing of treatment-emergent Grade 3 or higher laboratory abnormalities will be provided by subject ID number and time point in chronological order. This listing will include all test results that were collected throughout the study for the lab test of interest, with all applicable severity grades displayed.

7.2.3. Shifts Relative to the Baseline Value

Shift tables will be presented by showing change in CTCAE grade from baseline to select time points and to the worst post-baseline grade.

7.3. Vital Signs and Weights

Descriptive statistics will be provided by treatment arm for vital signs (temperature, pulse, systolic blood pressure, diastolic blood pressure, and respiratory rate) and weights as follows:

- Baseline value
- Values at each post-baseline visit (including unscheduled and scheduled)
- Change from baseline at each post-baseline visit (including unscheduled and scheduled)

A baseline value will be defined as the last available value collected on or prior to the date/time of first dose of study drug. Change from baseline to a post-baseline visit will be defined as the post-baseline value minus the baseline value.

In the case of multiple values in an analysis window, data will be selected for analysis as described in Section 3.8.3. No formal statistical testing is planned.

A by-subject listing of vital signs will be provided by subject ID number in ascending order and by visit in chronological order. In the same manner, a by-subject listing of body weight will be provided separately.

7.4. Prior and Concomitant Medications

7.4.1. Prior Medications

Prior medications are defined as any medications begun before a subject took the first study drug.

Prior medications will be summarized by Anatomical Therapeutic Chemical (ATC) drug class Level 2 and preferred name using the number and percentage of subjects for each treatment group and overall. A subject reporting the same medication more than once will be counted only once when calculating the number and percentage of subjects who received that medication. The summary will be ordered alphabetically by ATC medical class and then by preferred term in order of descending overall frequency within each ATC medical class. For drugs with the same frequency, sorting will be done alphabetically.

For the purposes of analysis, any medication with a start date prior to the first dosing date of study drug will be included in the prior medication summary regardless of when the stop date is. If a partial start date is entered the medication will be considered prior unless the month and year (if day is missing) or year (if day and month are missing) of the start date are after the first dosing date. Medications with a completely missing start date will be included in the prior medication summary, unless otherwise specified.

All prior and concomitant medications (other than per-protocol study drugs) will be provided in a by-subject listing sorted by subject ID number and administration date in chronological order.

7.4.2. Concomitant Medications

Concomitant medications are defined as medications taken while a subject took study drug. Concomitant medications will be summarized by ATC drug class Level 2 and preferred name using the number and percentage of subjects for each treatment group. A subject reporting the same medication more than once will be counted only once when calculating the number and percentage of subjects who received that medication. The summary will be ordered alphabetically by ATC medical class and then by preferred term in descending overall frequency within each ATC medical class. For drugs with the same frequency, sorting will be done alphabetically.

For the purposes of analysis, any medications started prior to or on the first dosing date of study drug and continued to take after the first dosing date, or started after the first dosing date but no later than last dosing date of study drug will be considered concomitant medications. Therefore, medications with a stop date that is on or prior to the date of first dosing date of study drug or a start date that is after the last dosing date of study drug will be excluded from the concomitant medication summary. If a partial stop date is entered, any medication with the month and year (if day is missing) or year (if day and month are missing) prior to the date of first study drug administration will be excluded from the concomitant medication summary. If a partial start date is entered, any medication with the month and year (if day is missing) or year (if day and month are missing) after the study drug stop date will be excluded from the concomitant medication summary. Medications with completely missing start and stop dates will be included in the concomitant medication summary.

Summaries will be based on the Safety Analysis Set. No formal statistical testing is planned.

7.5. Electrocardiogram (ECG) Results

Electrocardiogram (ECG) analysis results are intended to identify meaningful changes in the QT interval. If potential abnormalities of interest are identified, further analyses may be conducted. Summaries of investigator assessment of ECG readings will be provided for the Safety Analysis Set for each scheduled and unscheduled time point. No formal statistical testing is planned.

7.5.1. Investigator Electrocardiogram Assessment

A shift table of the investigators' assessment of ECG results at each visit compared with baseline values will be presented by treatment arm using the following categories: normal; abnormal, not clinically significant; abnormal, clinically significant; or missing. The number and percentage of subjects in each cross-classification group of the shift table will be presented. Subjects with a missing value at baseline or post-baseline will not be included in the denominator for percentage calculation. No formal statistical testing is planned.

A by-subject listing for ECG assessment results will be provided by subject ID number and time point in chronological order.

7.5.2. Corrected QT Intervals

The QT interval (measured in millisecond [msec]) is a measure of the time between the start of the Q wave and the end of the T wave in the heart's electrical cycle. The QT interval represents electrical depolarization and repolarization of the ventricles. The QT interval is affected by heart rate and a number of methods have been proposed to correct QT for heart rate.

Corrected QT (QTc) intervals will be derived using Fridericia's correction (QTcF) as follows:

$$QTcF = \frac{QT}{\sqrt[3]{RR}}$$

where QT is measured in msec; RR = 60/Heart Rate (beats per min [bpm]) and is measured in seconds

The maximum postdose QTcF interval values obtained during the study will be summarized within the following categories:

- > 450 msec
- > 480 msec
- > 500 msec

The maximum postdose change in QTcF interval values obtained during the study will also be summarized within the following categories:

- > 30 msec
- > 60 msec

QTcF and uncorrected QT values at each visit and change from baseline at each visit will be summarized by treatment arm using descriptive statistics.

7.5.3. PR and QRS Intervals

The PR interval (measured in msec) is a measure of the time between the start of the P wave (the onset of atrial depolarization) and the beginning of the QRS complex (the onset of ventricular depolarization). The QRS interval measures the duration of the QRS complex. The maximum ventricular rate (VR) and PR and QRS intervals observed during the study will be categorized. The number and percentage of subjects having values in the following ranges will be presented by treatment arm:

- VR > 100 bpm
- PR interval > 200 msec
- QRS interval > 110 msec

In addition, VR, PR interval, RR, and QRS interval values at each visit and change from baseline at each visit will be summarized by treatment arm using descriptive statistics.

7.6. Other Safety Measures

A data listing will be provided for the results of all pregnancy tests conducted in the study. Post-treatment anti-cancer therapies will be listed for applicable subjects.

8. PHARMACOKINETIC (PK) ANALYSES

8.1. PK Sample Collection

PK plasma samples will be collected for andecaliximab PK at 30(± 15) minutes after the end of infusion on Day 1 of Cycle 1, prior to dosing and 30(± 15) minutes after the end of infusion on Day 1 of Cycles 2, 3, 5, 7, and every 3 cycles thereafter, at the EOT and EOS visits.

CCI

8.2. Estimation of PK Parameters

PK parameters will be estimated using Phoenix WinNonlin® software using standard noncompartmental methods. The linear/log trapezoidal rule will be used in conjunction with the appropriate noncompartmental model, with input values for dose level, dosing time, plasma concentration, and corresponding real-time values, based on drug dosing times whenever possible.

All predose sample times before time-zero will be converted to 0.

For area under the curve (AUC), samples BLQ of the bioanalytical assays occurring prior to the achievement of the first quantifiable concentration will be assigned a concentration value of 0 to prevent overestimation of the initial AUC. Samples that are BLQ at all other time points will be treated as missing data in WinNonlin. The nominal time point for a key event or dosing interval (τ) may be used to permit direct calculation of AUC over specific time intervals. The appropriateness of this approach will be assessed by the PK scientist on a profile-by-profile basis.

Pharmacokinetic parameters such as AUC_{tau}, λ_z and $t_{1/2}$ are dependent on an accurate estimation of the terminal elimination phase of drug. The appropriateness of calculating these parameters will be evaluated upon inspection of PK data on a profile-by-profile basis by the PK scientist.

8.3. PK Parameters

PK parameters will be generated for all subjects in the PK substudy analysis set.

CCI

CCI

8.4. Statistical Analysis Methods

Individual subject concentration data for andecaliximab, oxaliplatin, and 5-FU will be listed and summarized using descriptive statistics by treatment. Summary statistics (n, mean, StD, coefficient of variation [%CV], median, min, max, Q1, and Q3) will be presented for individual subject concentration data by time point.

Individual concentration data listings and summaries will include all subjects in PK analysis set. The sample size for each time point will be based on the number of subjects with nonmissing concentration data at that time point. The number of subjects with concentration BLQ will be presented for each time point. For summary statistics, BLQ values will be treated as 0 at predose and one-half of the LLOQ for postdose time points.

Individual subject PK parameters (PK substudy analysis set) for andecaliximab, oxaliplatin, and 5-FU will be listed and summarized using descriptive statistics by treatment. Summary statistics (n, mean, StD, coefficient of variation [%CV], median, min, max, Q1, and Q3) will be presented for individual subject PK parameters by treatment. Moreover, the geometric mean, 90% CI, and the mean and StD of the natural log-transformed values will be presented for individual subject PK parameter data.

Individual PK parameter data listings and summaries will include all subjects in PK substudy analysis set. The sample size for each PK parameter will be based on the number of subjects with nonmissing data for that PK parameter.

The following tables will be provided for each analyte by treatment:

- Individual subject concentration data and summary statistics (PK analysis set)
- CCI

The following figures may be provided for each analyte by treatment (PK analysis set):

- Mean (± StD) concentration data versus time (on linear and semilogarithmic scales)
- Median (Q1, Q3) concentration data versus time (on linear and semilogarithmic scales)
- Individual subject concentration versus time (on linear and semilogarithmic scales)

Individual, mean, and median postdose concentration values that are ≤ LLOQ will not be displayed in the figures and remaining points connected.

The following listings will be provided

- PK sampling details by subject, including procedures, differences in scheduled and actual draw times, and sample age will be provided in listings (PK analysis set).
- CCI [redacted]

9. IMMUNOGENICITY ANALYSES

9.1. ADA Sample Collection

Serum samples for anti-andecaliximab antibody will be collected prior to dosing on Day 1 of Cycles 1, 2, 3, 5, 7, and every 3 cycles thereafter, EOT, EOS, and 30-day Safety Follow-Up visits.

9.2. Statistical Analysis Methods

Immunogenicity of andecaliximab will be evaluated based upon the positive ADA rate. The number and percentage of subjects exhibiting positive ADA status, defined as ADA presence in serum confirmed in validated assay and reported by bioanalytical laboratory, at each specified time point will be summarized. The number and percentage of subjects exhibiting positive ADA status at any post-andecaliximab time point will also be summarized.

The time to ADA onset will also be assessed, ant it is defined as the time interval from the date of first dose of andecaliximab to the date positive ADA status is first achieved. The analyses of time to ADA onset will be based on the subjects in the immunogenicity analysis set who have negative ADA status at baseline and achieve positive ADA status at post baseline. Summary statistics (n, mean, StD, coefficient of variation [%CV], median, min, max, Q1, and Q3) will be presented.

The impact of ADA on andecaliximab PK may be evaluated as appropriate. Individual subject andecaliximab predose concentration data may be summarized using descriptive statistics by ADA status (positive or negative) of the same time point. Summary statistics (n, mean, StD, coefficient of variation [%CV], median, min, max, Q1, and Q3) will be presented by ADA status. The number of subjects with concentration BLQ will be presented for each time point. For summary statistics, BLQ values will be treated as 0 at predose. Individual subject andecaliximab predose concentration data versus ADA results may be plotted (on linear and semilogarithmic scales). PK plasma samples meeting below criteria will be included in this analysis:

- Collected on Cycle 3 Day 1 or later
- Collected 288 – 384 hours (12 to 16 days) post most recent dose

A by-subject listing for ADA status at each time point and the titer for subjects with positive ADA status will be provided by subject ID number and time point in chronological order.

10. BIOMARKER ANALYSES

Biomarker analyses will be described in the biomarker analysis plan (BAP)

11. REFERENCES

Hung HM, Wang SJ, O'Neill R. Statistical considerations for testing multiple endpoints in group sequential or adaptive clinical trials. J Biopharm Stat 2007;17 (6):1201-10.

Koch GG, Carr GJ, Amara IA, Stokes ME, Uryniak TJ. Categorical Data Analysis. Chapter 13 in Berry, D.A. (ed.). Statistical Methodology in the Pharmaceutical Sciences. New York: Marcel Dekker, Inc., 1989:pp. 414-21.

Lan KKG, Hu P, Proschan M. A Conditional Power Approach to the Evaluation of Predictive Power. Statistics in biopharmaceutical research 2012;1 (2):131-6.

Tamhane AC, Mehta CR, Liu L. Testing a primary and a secondary endpoint in a group sequential design. Biometrics 2010;66 (4):1174-84.

12. SOFTWARE

SAS® (SAS Institute Inc., Cary, NC) is to be used for all programming of tables, listings, and figures.

WinNonlin® (Pharsight Corporation, Mountain View, CA) is to be used for all PK analyses.

13. SAP REVISION

Revision Date (DD MMM YYYY)	Section	Summary of Revision	Reason for Revision
27 SEP 2018	2.4, 3.4	Primary (OS) and secondary (PFS and ORR) efficacy endpoints will be examined in subgroups defined by baseline tumor PD-L1 status (positive vs. negative). This is a new analysis which was not specified in the protocol (Section 2.4) and the analysis is described in Section 3.4.	Add a new subgroup analysis.

14. APPENDICES

Appendix 1. List of Laboratory Tests for Safety Summary

Serum Chemistry	Hematology	Coagulation
Sodium	White Blood Cell Count	PT/INR
Potassium	Hemoglobin	aPTT
Chloride	Hematocrit	
Glucose	Platelet Count	
BUN	ANC	
Creatinine[a]	Neutrophils	
ALT	Lymphocytes	
AST	Monocytes	
Alkaline phosphatase		
Total bilirubin		
Total protein		
Albumin		
Calcium		
Magnesium		
Phosphate		
Lipase		
Bicarbonate		

a Estimated creatinine clearance (CL_{cr})/glomerular filtration rate will be calculated based on the Cockroft-Gault formula using actual body weight: CL_{cr} (mL/min) = (140 – age [years])* weight (kg) / (serum creatinine [mg/dL]*72). If the subject is female, multiply the quantity by 0.85.

Appendix 2. List of Participating Country and Geographic Region

Country		Region
1	Australia	Asia Pacific
2	Belgium	Europe
3	Colombia	Latin America
4	Czech Republic	Europe
5	France	Europe
6	Germany	Europe
7	Hungary	Europe
8	Italy	Europe
9	Peru	Latin America
10	Poland	Europe
11	Romania	Europe
12	Spain	Europe
13	Turkey	Europe
14	United Kingdom	Europe
15	United States	North America
16	Chile	Latin America

Appendix 3. Details for Testing Efficacy Endpoints

This appendix provides details for testing multiple efficacy endpoints at interim and final analyses.

The final analysis will occur when 286 overall survival (OS) events have been observed. Two formal interim analyses are planned: the futility interim analysis and the efficacy interim analysis will be performed when approximately 33.3% and 66.7% of OS events have occurred, respectively.

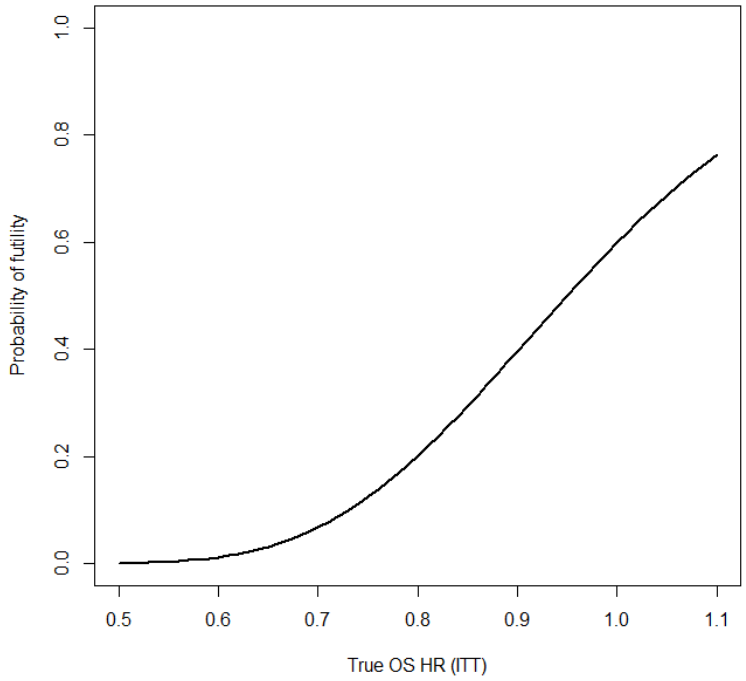
Futility Interim Analysis

The end point of the futility interim analysis is OS. The futility analysis will occur approximately after 95 OS events (33.3% of the information relative to the final analysis) have been observed in the ITT population. The analysis will perform a stratified log-rank test for OS in the ITT population. A non-binding futility rule will be implemented. Based on the analysis results, the IDMC may recommend terminating the study for lack of efficacy if the predictive power (PP) on OS in ITT population is < 14%. Otherwise, the study will be continued.

Predictive power is defined as a weighted average of the conditional power; the weighting function is determined by the sampling distribution of the observed hazard ratio based on the data at the interim analysis {Lan 2012}. When exactly 95 OS events are observed at the interim, this futility rule of PP <14% corresponds to observing an OS hazard ratio (HR) > 0.95 in the ITT population.

Appendix Figure 1 presents the operating characteristics of futility rule at the interim of 95 OS events with futility boundary of 0.95. When the true hazard ratio is 0.7, the probability of terminating the study early is 7%. On the other hand, when the true hazard ratio is 1, indicating lack of efficacy, the probability of terminating the study early is 60%.

Appendix Figure 1. Probability of Early Stop due to Futility



Efficacy Interim and Final Analysis

In the efficacy interim and final analysis, the primary and secondary endpoints will be tested sequentially in the following gatekeeping order: the primary OS endpoint, then the secondary PFS endpoint, and finally the secondary ORR endpoint. Specifically:

- At the efficacy interim analysis, the OS event point will be tested at one-sided alpha level of 0.006. If the OS endpoint is not rejected, the study will continue to final analysis. If the OS endpoint is rejected, the IDMC would recommend early stop for efficacy. Furthermore, the PFS endpoint will be tested at one-sided alpha level of 0.016. Only if the PFS endpoint is rejected, the ORR endpoint will be tested at one-sided alpha level of 0.016.
- At the final analysis, the OS event point will be tested at one-sided alpha level of 0.023. If the OS endpoint is rejected, the PFS endpoint will be tested at one-sided alpha level of 0.016. If the PFS endpoint is also rejected, the ORR endpoint will be tested at one-sided alpha level of 0.016.

This gatekeeping testing strategy is summarized in the table below. Overall, this testing strategy employs O'Brien-Fleming boundary for the primary OS endpoint and the Pocock type boundary for the secondary PFS/ORR endpoints. The testing strategy controls the overall one-sided family-wise type I error to be at 0.025, equivalent to two sided 0.05 by appropriately adjusting for multiplicity in the efficacy interim and final analyses.

Appendix Table 1. Gate-keeping Decision Boundaries for Efficacy Analysis

Efficacy Analysis	Event (%)	One-sided Decision Boundary		
		OS	PFS (After OS boundary is crossed)	ORR (After PFS boundary is crossed)
Interim	191 (66.7%)	0.006	0.016	0.016
Final	286 (100%)	0.023	0.016	0.016

In the following scenarios we will provide detailed justification for this testing strategy. During efficacy interim and final analysis, type I error could occur in the following non-overlapping scenarios:

1) Reject H_{01} when H_{01} is true

2) Reject H_{01} and H_{02} when H_{11} and H_{02} is true

3) Reject H_{01}, H_{02} and H_{03} when H_{11}, H_{12}, and H_{03} is true

where H_{ij} (i=0: null, i=1: alternative) are the hypothesis for the jth endpoint (1 = primary OS, 2 = secondary PFS, 3 = secondary ORR).

Scenario 1

The type I error rate is controlled by the O'Brien-Fleming (OF) boundaries on the primary OS endpoint.

Scenario 2

We herein justify that the use of 1-sided significance level of 0.016 for the testing of secondary PFS endpoint at both the interim and final analyses will control the Type I error rate at 0.025 one-sided. We follow the methods in Hung et al (2007) {Hung 2007} and Tamhane et al (2010) {Tamhane 2010}.

Per Hung et al {Hung 2007}, the one-sided type I error rate in Scenario 2 in the hierarchical testing strategy is:

$$\alpha_2 = \Pr(T_{11} > C_{11}, T_{21} > C_{21} \mid H_{02}) + \Pr(T_{11} \le C_{11}, T_{12} > C_{12}, T_{22} > C_{22} \mid H_{02}) \qquad (1)$$

where T_{jk} and C_{jk} are the test statistic and critical value, respectively, for the *j*th endpoint (1 = primary OS, 2 = secondary PFS) and the *k*th analysis (1=interim, 2=final). Accordingly, (1) can also be written as:

$$\alpha_2 = \Pr(Z_{11} > z_{11}, Z_{21} > z_{21} \mid H_{02}) + \Pr(Z_{11} \le z_{11}, Z_{12} > z_{12}, Z_{22} > z_{22} \mid H_{02}) \qquad (2)$$

where Z_{jk} and z_{jk} are the corresponding test statistics and critical values expressed in the setting of the standard normal distribution (ie, after standardization). Equation (2) can be evaluated analytically and we prove in the following that in the scenario 2, under the proposed the testing strategy, the one-sided type I error rate is controlled at 0.025.

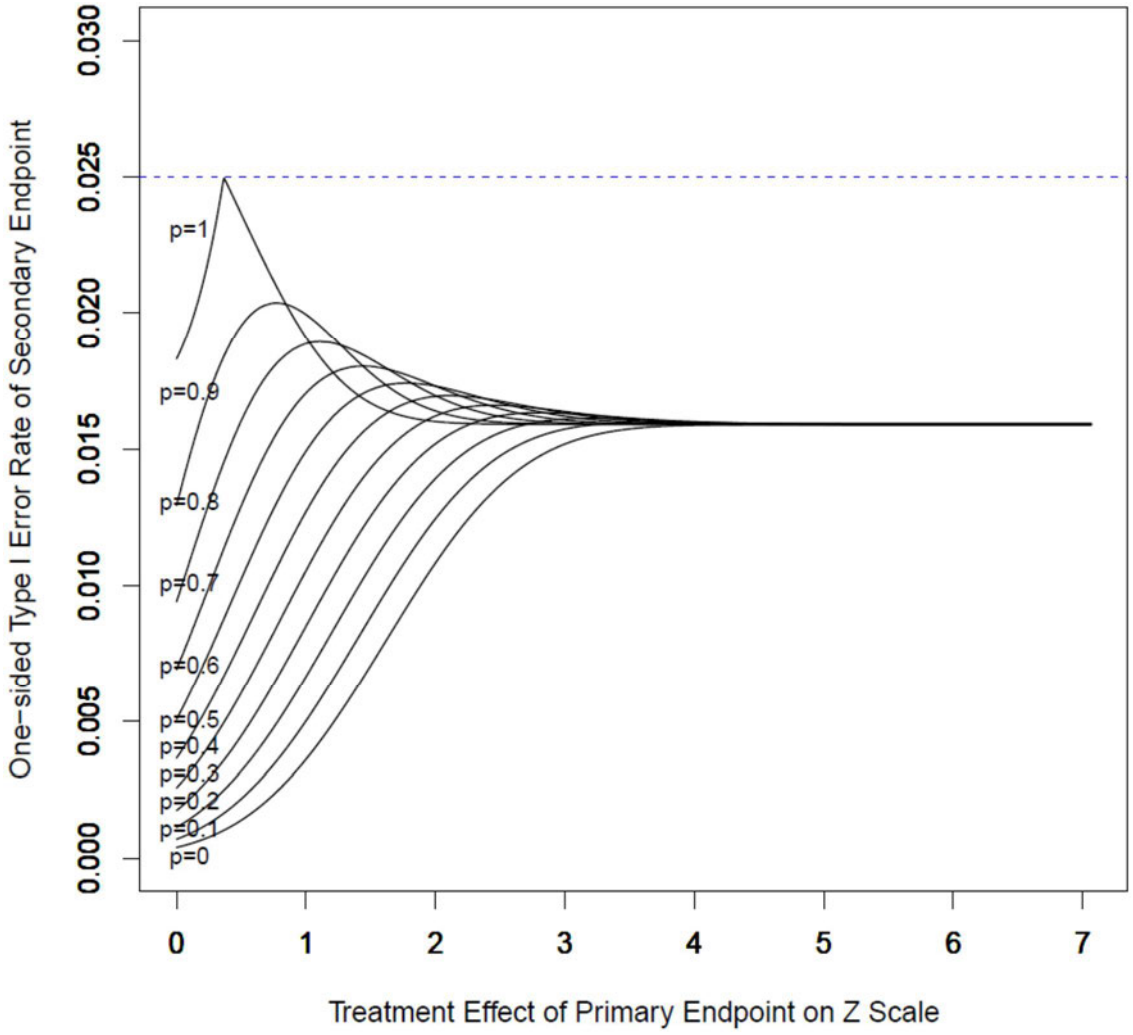
It can be derived that $\tilde{Z} = (Z_{11}, Z_{21}, Z_{12}, Z_{22})$ follows the multivariate normal distribution below:

$$\begin{pmatrix} Z_{11} \\ Z_{21} \\ Z_{12} \\ Z_{22} \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mu_{11} \\ \mu_{21} \\ \mu_{12} \\ \mu_{22} \end{pmatrix}, \begin{pmatrix} 1 & \rho & \sqrt{t} & \rho\sqrt{t} \\ \rho & 1 & \rho\sqrt{t} & \sqrt{t} \\ \sqrt{t} & \rho\sqrt{t} & 1 & \rho \\ \rho\sqrt{t} & \sqrt{t} & \rho & 1 \end{pmatrix} \right) \qquad (3)$$

where (μ_{11}, μ_{12}) and (μ_{21}, μ_{22}) are the normalized mean treatment effects on the primary and secondary endpoints, respectively, for data up to the interim and final analysis. The mean treatment effects between the interim and final analysis can be related by the information fraction t: $\mu_{j1} = \mu_{j2}/\sqrt{t}$ where $j = 1, 2$. Under the assumption of Scenario 2, the mean treatment effect on the primary OS endpoint (μ_{11}, μ_{12}) is positive representing the alternative hypothesis, and the mean treatment effect on secondary PFS endpoint (μ_{21}, μ_{22}) is set to (0, 0) under the null hypothesis. ρ denotes the correlation between the primary and the secondary test statistics at each analysis stage, ranging from 0 to 1, assuming a non-negative correlation. $t = 0.667$ denotes the planned information fraction at the interim efficacy analysis. The one-sided OF boundaries for the primary endpoint are given by $z_{11} = z_{0.006}$ and $z_{12} = z_{0.023}$. Thus, equation (2) can be evaluated using the pmvnorm function in the R package mvtnorm based on the mean and covariance structure specified in equation (3).

Appendix Figure 2 shows the analytically calculated type I error rate in the second scenario, which is the error rate for the secondary endpoint. It demonstrates that the use of a 1-sided significance level of 0.016 (equivalent to 2-sided level 0.032) for the testing of secondary endpoints at both the interim and final analysis controls the overall type I error rate at the 1-sided level 0.025 regardless of the correlation between the primary and secondary test statistics and the magnitude of treatment effect of the primary endpoint.

Appendix Figure 2. Type I Error Rate of Secondary Endpoint in a Group Sequential Design with OF Boundaries for Primary Endpoint and 1-sided Significance Level of 0.016 for the Secondary Endpoints with 1 Interim and 1 Final Analysis



Scenario 3

The one-sided type I error rate of Scenario 3 is bounded by the following inequality:

$$\begin{aligned} &P(Reject\ H_{01}, Reject\ H_{02}, Reject\ H_{03} | H_{11}, H_{12}, H_{03}) \\ &\leq P(Reject\ H_{01}, Reject\ H_{03} | H_{11}, H_{12}, H_{03}) \\ &= P(Reject\ H_{01}, Reject\ H_{03} | H_{11}, H_{03}) \end{aligned}$$

Applying the same argument in Scenario 2 (with PFS replaced by ORR), the type I error rate can be protected by using one-sided significance level of 0.016 for the testing of secondary ORR endpoint at both the interim and final analyses.

Appendix 4. Formula and Sample Code to Compute Predictive Power

The predictive power at the futility interim analysis will be evaluated as

$$\Phi\left(\sqrt{\frac{d_1}{d-d_1}}\cdot Z_{\frac{\alpha}{2}}-\sqrt{\frac{d\cdot d_1}{4\cdot(d-d_1)}}\cdot\log\hat{\lambda}_1\right)$$

Where d_1 is the number OS events observed by the time of futility interim analysis; $d = 286$ is the number of OS events at the final analysis; $\hat{\lambda}_1$ is the observed log HR (treatment arm vs. placebo) at futility interim; $\alpha = 0.05$ is the Type I error of the final analysis; $\Phi(\cdot)$ is the probability function of standard normal distribution; Z is the inverse of Φ.

Sample SAS code:

```
/*count number of death events observed by futility interim analysis
d1*/
proc sql;
      select Count(distinct(a.subjid)) into: d1
      from eff_strat_1 as a
      where a.cnsr_os=0; /* cnsr_os=0 denotes death events*/

quit;

/*compute hazard ratio adjusted by stratification factor*/
ods output HazardRatios=os strat_1;
proc phreg data=eff_strat_1;
  class cohort;
  model os*cnsr_os(1)=cohort; /* cnsr_os=1 denotes censored*/
  strata strata;
  hazardratio cohort/diff=ref alpha=0.05;
run;

/*compute predictive power*/
data HR_PP_1;
      set os_strat_1;
      /*assume HazardRatio is treatment arm vs. placebo. Otherwise,
invert the Hazard ratio for PP calculation*/
      Z=sqrt(&d1/(286-&d1))*quantile("normal", 0.025)-
sqrt(286*&d1/4/(286-&d1))*log(HazardRatio);
      PP=cdf("normal", z);
run;
```

Appendix 5. EORTC Scoring Guideline

Scoring the EORTC QLQ-C30 version 3.0

Table 1: Scoring the QLQ-C30 version 3.0

	Scale	Number of items	Item range*	Version 3.0 Item numbers	Function scales
Global health status / QoL					
Global health status/QoL (revised)†	QL2	2	6	29, 30	
Functional scales					
Physical functioning (revised)†	PF2	5	3	1 to 5	F
Role functioning (revised)†	RF2	2	3	6, 7	F
Emotional functioning	EF	4	3	21 to 24	F
Cognitive functioning	CF	2	3	20, 25	F
Social functioning	SF	2	3	26, 27	F
Symptom scales / items					
Fatigue	FA	3	3	10, 12, 18	
Nausea and vomiting	NV	2	3	14, 15	
Pain	PA	2	3	9, 19	
Dyspnoea	DY	1	3	8	
Insomnia	SL	1	3	11	
Appetite loss	AP	1	3	13	
Constipation	CO	1	3	16	
Diarrhoea	DI	1	3	17	
Financial difficulties	FI	1	3	28	

* *Item range* is the difference between the possible maximum and the minimum response to individual items; most items take values from 1 to 4, giving *range* = 3.
† (revised) scales are those that have been changed since version 1.0, and their short names are indicated in this manual by a suffix "2" – for example, PF2.

For all scales, the *RawScore*, *RS*, is the mean of the component items:

$$RawScore = RS = (I_1 + I_2 + ... + I_n)/n$$

Then for **Functional scales**:

$$Score = \left\{1 - \frac{(RS-1)}{range}\right\} \times 100$$

and for **Symptom scales / items** and **Global health status / QoL:**

$$Score = \{(RS-1)/range\} \times 100$$

Examples:

Emotional functioning

$$RawScore = (Q_{21} + Q_{22} + Q_{23} + Q_{24})/4$$
$$\text{EF } Score = \{1 - (RawScore - 1)/3\} \times 100$$

Fatigue

$$RawScore = (Q_{10} + Q_{12} + Q_{18})/3$$
$$\text{FA } Score = \{(RawScore - 1)/3\} \times 100$$

Gastric cancer module: QLQ-STO22

Scope

The gastric cancer module is meant for use among patients with gastric cancer varying in disease stage and treatment modality (i.e. surgery, chemotherapy, radiotherapy, etc.). It should always be complemented by the QLQ-C30.

Scoring

	Scale name	Number of items	Item range	QLQ-STO22 item numbers
Functional scales				
Body image	STOBI	1	3	49
Symptom scales				
Dysphagia	STODYS	3	3	31 – 33
Pain	STOPAIN	4	3	34 – 37
Reflux symptoms	STORFX	3	3	38 – 40
Eating restrictions	STOEAT	4	3	41 to 43,46
Anxiety	STOANX	3	3	47,48,50
Dry mouth	STODM	1	3	44
Taste	STOTA	1	3	45
Body image	STOBI	1	3	49
Hair loss	STOHL	2/1	3	51,52*

Remarks

- Item 52 is an optional item and depends on the answer to item 51. Item 52 should only be answered and assessed if 'yes' has been answered to item 51.