STATISTICAL ANALYSIS PLAN

STUDY NUMBER: AIO-STO-0217

CONFIDENTIAL

Ipilimumab or FOLFOX in combination with Nivolumab and Trastuzumab in previously untreated HER2 positive locally advanced or metastatic EsophagoGastric Adenocarcinoma

The randomized phase 2 INTEGA trial.

Sponsor: Wolfgang Hiegl
AIO-Studien-gGmbH
Kuno-Fischer-Straße 8, 14057 Berlin, Germany
Phone: +49 (0)30 8145344-54
E-Mail: Wolfgang.hiegl@aio-studien-ggmbh.de

Author: Marcos Marin-Galiano
M.A.R.C.O. GmbH & Co. KG
Schirmerstraße 71
D-40211 Duesseldorf
Phone: +49 (0)221/ 97793-578
Email: marcos.marin@marco-institut.de

Version: 1.0

Date: 6th July, 2021

The undersigned have discussed the topics of this document and agree on the content.

Wolfgang Hiegl / AIO-Studien-gGmbH	**Date**
Marcos Marin-Galiano Statistician / M.A.R.C.O.	**Date**

Table of contents

Listing of tables

Listing of abbreviations and definition of terms

AE	Adverse event
CPS	Combined positive score
CT	Computerized tomography
CTC	Circulating Tumor cells
CTCAE	Common Terminology Criteria for Adverse Events
ctDNA	Circulating Tumor DNA
d1	Day 1
DRM	Data Review Meeting
DRP	Data Review Plan
ECG	Electrocardiogram
EORTC	European Organization of Research and Treatment of Cancer
EOT	End of treatment
FOLFOX	FOLinic acid 5-Fluorouracil OXaliplatin
FU	Follow-up
GC	Gastric carcinoma
GEJ	Gastroesophageal junction
HER	Human epidermal growth factor receptor
IC	Immune cell
ICF	Informed consent form
IDMC	Independent Data Monitoring Committee
IQR	Inter-quartile range
ISH	In-Situ Hybridization
ITT	Intention-to-Treat
i.v.	intravenous
kg	kilogram
LLT	Lowest level term
LLOQ	Lower limit of quantification
LV	Leucovorin
Max	Maximum
MedDRA	Medical dictionary for regulatory activities
mg	milligram
Min	Minumum
MRI	Magnetic resonance imaging
N	Number
NCI	National Cancer Institute
Med	Median
ORR	Overall response rate
OS	Overall survival
PD-L1	Programmed Death Receptor Ligand 1
PFS	Progression free survival
PT	Preferred term
Q1	Lower quartile
Q3	Upper quartile
RECIST	Response Evaluation Criteria in Solid Tumors
RR	Response rate
SAE	Serious adverse event
SAP	Statistical analysis plan

SD	Standard deviation
SFU	Safety follow-up
SOC	System organ class
t	Time
TPS	Tumor proportion score
TRA	Tumor response assessment
WHO	World health organization

1. Background and scope

As a general rule:
All text copied from the protocol is set in italics type in this statistical analysis plan.

This statistical analysis plan (SAP) is based on:

- Trial protocol (Version 5.0, 04-MAR-2019)

This SAP covers:

- More detailed planning of the final statistical analyses as outlined in the trial protocol

Responsibilities:

- Statistics: M.A.R.C.O. GmbH & Co. KG

2. Study objectives

Primary objective:
The primary objective is to determine the clinical performance of ipilimumab or FOLFOX in combination with nivolumab and trastuzumab in patients with previously untreated HER2 positive locally advanced or metastatic esophagogastric adenocarcinoma in terms of overall survival.

Secondary objectives:
The main secondary objective is to determine safety and tolerability, according to NCI CTC AE v4.03 and to the obtained data on vital signs, clinical parameters and feasibility of the regimen.
Further secondary objectives are to determine efficacy in terms of progression free survival and objective response rate acc. to RECIST v1.1 of the experimental regimen.
In addition immune response signatures (e.g. TiL repertoire and NGS immunoprofiling of immunoglobuline and T-cell receptor rearrangements), changes in HER2 and PD-L1 status in CTC and ctDNA and PD-L1 status in biopsy will be correlated with efficacy. Imaging will be centrally review and ORR and PFS will be determined according to modified RECIST.

3. Study design

Number of centers:	multicenter
Randomized:	Yes
Blinded:	No
Design:	Randomized, open labelled, multicenter phase II trial
Placebo controlled:	No
Strata:	• Prior surgery of the primary tumour yes vs. no • HER2 status IHC 3+ vs. IHC 2+ and ISH amplified
Treatments:	
Arm A:	<u>Week 1-12</u> Trastuzumab 6mg/kg d1 every 3 weeks (loading dose 8mg/kg) Nivolumab 1mg/kg i.v. d1 every 3 weeks Ipilimumab 3mg/kg i.v. d1 every 3 weeks <u>Week 13 till EOT</u> Trastuzumab 4mg/kg d1 every 2 weeks Nivolumab 240mg i.v. d1 every 2 weeks
Arm B:	Trastuzumab 4mg/kg d1 every 2 weeks (loading dose 6mg/kg) Nivolumab 240mg i.v. d1 every 2 weeks mFOLFOX6 every 2 weeks Oxaliplatin at a dose of 85 mg/m^2 IV over two hours (day 1) 5-Flurouracil 400 mg/m^2 IV bolus (day 1) LV at a dose of 400 mg/m^2 iv over two hours (day 1) 5-Flurouracil at a dose of 2400 mg/m^2 IV over 46 hours (day 1-3)
Planned sample size	Overall, 82 patients to be included and randomized into the two experimental arms (41 in each experimental treatment group). The rate of drop-outs is estimated to be 15%. Hence, the total number of patients to be recruited is 97.

A schedule of the study visits and assessments are provided in Table 1.

Table 1 Schedule of visits and assessments

Study Schedule Visit	Inclusion	During treatment		End of Treatment	Safety follow up	Follow up
Study week (wks)	Baseline (within 4 wks prior to cycle 1)	Cycle 1, day1 afterwards every 2/3 wks (+3/-2 d)	Day 12 (+/-3 d) until wk 13 only in arm A		After 30 and 100 d [7] (±7 d)	every 3 months (±28 d)
Informed consent	X					
Eligibility criteria	X					
Medical history, demographics	X					
Physical examination[1]	X	X	X	X	X	X[6]
Vital signs[2]	X	X	X	X		
Oxygen saturation	X	X	X	X	X	
Performance status (ECOG)	X	X	X	X	X	X
Echocardiography and ECG	X	X[8]		X		
Obtain tumor tissue	X					
Blood draw translat. research	X	X[4]		X		X[10]
Laboratory determinations[3]	X	X	X	X	X	
Treatment		X				
Tumor markers (CA 72-4, optional CEA, CA 19-9)	X	X[5]		X		X[6]
Quality of life assessment (EORTC QLQ C30 and STO 22)	X	X[5]		X		(x)[5]
Tumor assessment (CT/MRI)	X	X[5]		X		X[6]
Concomitant medication	X	X		X	X	
Further treatment						X
AE monitoring/ assessment of toxicity	X					X[9]
Survival	X					

1: including heart, chest, abdomen, skin, lymph nodes, neurological exam, inspection of accessible mucosa and weight, height (only baseline)
2: blood pressure, heart rate, respiratory rate, body temperature
3: hematology panel (hemoglobin, platelets, WBC with neutrophils, lymphocytes), chemistry panel (sodium, potassium, calcium, magnesium, serum creatinine, urea, alkaline phosphatase, AST, ALT, total bilirubin, glucose, lipase, amylase) screening at baseline, day 1 prior every treatment cycle, (in arm A until week 13 on day 12 of every cycle), EOT and 30 days safety follow up; coagulation (INR, aPTT), LDH, albumine, and total protein, screening at baseline and EOT; HIV, hepatitis b/c screening only baseline; serum pregnancy test in women of childbearing potential screening at baseline and every 4 weeks, EOT and 30 days safety follow up,,and free T3/T4 and TSH at baseline, every 6 weeks, EOT and 30 days safety follow up
4: blood draw baseline and cycle 2 and 5 and end of treatment
5: every 8 weeks (±7 days) for 12 months after randomization in parallel to tumor assessments afterwards every 3 months until radiologic disease progression and/or EOT, whichever is longer.
6: only in case of no progressive disease during or after 1st line treatment
7: safety follow up should be on site
8: echocardiography every 3 months
9. protracted toxicity
10. at progression (if not reason for EOT)

4. Patient population

The main inclusion criteria are defined as:

- Inoperable, advanced or metastatic esophagogastric adenocarcinoma
- HER2-positive disease defined as either IHC 3+ or IHC 2+, the latter in combination with ISH+, as assessed locally on a primary or metastatic tumour
- Previously untreated with systemic treatment (including HER2 inhibitors) given as primary therapy for advanced or metastatic disease
- ECOG performance status score of 0 or 1
- Males and females, ≥ 18 years of age
- Written informed consent

5. Primary, secondary and exploratory endpoints

5.1. Primary endpoint and hypotheses

The primary endpoint is:

- *Overall Survival including milestone rate @ 12 months*

OS *will be determined as time from the randomization date to the date of death. A patient who has not died will be censored at last known date alive.*

The survival curve for the standard treatment (chemotherapy and trastuzumab) was estimated based on historical data. Bang, Van Cutsem et al. 2010 observed an overall survival of 55% at 12 months after this standard therapy. The hazard rate λ is assumed to be constant over time for the standard treatment. In this case, the survival is exponential, namely (time *t* measured in years):

$$S_H(t) = 100e^{-\lambda t},\ with\ \lambda = -\ln\left(\frac{55}{100}\right) = 0.6.$$

Each of the experimental arms would be considered promising, if the true 12-month-OS rate amounts to 70 %, i.e.

$$\mathrm{S_{A/B}(t)} = 100\mathrm{e}^{-\lambda \mathrm{t}},\ \text{with}\ \lambda = -\ln\left(\frac{70}{100}\right) = 0.36$$

This translates into a hazard ratio of 0.6 compared to the standard OSR@12 of 55% for chemotherapy and trastuzumab:

$$\gamma_0 = \frac{0.36}{0.6} = 0.6$$

For the primary endpoint, the following hypotheses are tested in the confirmatory sense at significance level α=0.05 (one-sided) for both experimental treatment arms:

$$H_0: R \geq \gamma_0 \quad vs. \quad H_1: R < \gamma_0 = 0.6$$

where:

R hazard ratio of unknown hazard function of new treatment (A or B) and known hazard function for historical data

γ_0 bound for R below which the treatment arm is considered promising

5.2. Secondary endpoints

The secondary endpoints will include:

- *Safety and tolerability (acc. to NCI CTC AE v4.03 and to the obtained data on vital signs, clinical parameters and feasibility of the regimen)*
- *Progression Free Survival (PFS) according to RECIST v1.1 Response Rate (RR) according to RECIST v1.1*
- *Quality of life (EORTC QLQ-C30 and STO-22)*
- *Translational research (correlation of immune response signatures, changes in HER2 and PD-L1 and HER signaling status in tissue, CTC and ctDNA with efficacy)*
- *central imaging review and determination of ORR and PFS according to modified RECIST)*

5.3. Exploratory endpoints

Not applicable

6. General conventions

6.1. Layout of tables

Table 2 Layout of tables

Issue	Specification
Basic layout of end-of-text tables and listings (header, footnote)	3 horizontal lines per page: one above column headers, one below column headers, one below the body of the Table or Listing (footnotes below this last line) Footnotes for all tables, figures, listings: • `Source: xxx` (not for listings) • `Program: xxx.SAS, Generation date: xxMON20Ix` • `Page X of Y (`***note: display in lower right corner of page in the same line as program and generation date***`)`
Page margins	Landscape: left / right: 2.5 cm top / bottom: 3 cm
Font and font size for end-of-text tables and listings	Arial 9 for Tables/Graphs Arial 8 for Listings Titles in bold and using Arial 10 for Tables and Graphs / Arial 9 for listings.
Text (Title, footnotes, organizational variables and column headers, contents of table)	Case-sensitive, the first letter should be capitalized
Subject or Patient as label?	Patient
Patient identification numbers	Use of patient number
Order in patient data listings	Patients sorted by: • Treatment group • Center • Patient Number
Labels and order of treatment groups	• Trast/Nivo/Ipil (A) • Trast/Nivo/Fol (B)
Labels and order of visits	• Screening For arm A: • Cycle 1 start (week 1) • Cycle 2 start (week 4) • Cycle 3 start (week 7) • etc. (new cycle every three weeks) • Cycle 1, Day 12

Issue	Specification
	• Cycle 2, Day 12 • Cycle 3, Day 12 • Cycle 4, Day 12 <u>For arm B:</u> • Cycle 1 start (week 1) • Cycle 2 start (week 3) • Cycle 3 start (week 5) • etc. (new cycle every two weekes) • TRA 1 • TRA 2 • TRA 3 • … • EOT • SFU30 • SFU60 • SFU100 • FU 1 • FU 2 • FU 3 • …
Labels for descriptive statistics for continuous variables	N, Mean, SD, Minimum, Q1, Median, Q3, Maximum
Categorical Variables	Display of all possible categories (even if a category is not present in the data)
In case a category does not occur	Display absolute frequency 0 (instead of "-")
Display of percentages in tables:	Aligned by decimal point, one decimal place: • 20.3% • 9.3% • 0.0% • 100.0%
Display of ranges	• xx – yy
Display of units	Presentation case-sensitive and in square brackets, e.g. Concentration [mg/mL]

The following short and long decodes will be used for stratification, treatment, visit, and subgroup variables.

Table 3 Labels for variable levels

Variable	Level	Short label	Long label
Surgery	1	Yes	Yes
	2	No	No
HER2 status	1	IHC 3+	IHC 3+
	2	IHC 2+/ISH	IHC 2+ and ISH amplified
ECOG	0	Fully active	Fully active
	1	Restricted	Restricted in physically strenuous activity
Treatment group	1	Trast/Nivo/Ipil (A)	Trastuzumab/Nivolumab/Ipilimumab
	2	Trast/Nivo/Fol (B)	Trastuzumab/Nivolumab/Folfox
Visit	-28 to -1	Screening	Screening
For arm A:	1	Cycle 1 start	Start of treatment cycle 1 (week 1)
	2	Cycle 2 start	Start of treatment cycle 2 (week 4)
	3	Cycle 3 start	Start of treatment cycle 3 (week 7)
	…	…	etc. (new cycle every three weeks)
	51	Cycle 1, Day 12	Day 12 in cycle 1 until week 13 only in arm A
	52	Cycle 2, day 12	Day 12 in cycle 2 until week 13 only in arm A
	53	Cycle 3, Day 12	Day 12 in cycle 3 until week 13 only in arm A
	54	Cycle 4, Day 12	Day 12 in cycle 4 until week 13 only in arm A
For arm B:	101	Cycle 1 start	Start of treatment cycle 1 (week 1)
	102	Cycle 2 start	Start of treatment cycle 2 (week 3)
	103	Cycle 3 start	Start of treatment cycle 3 (week 5)
	…	…	etc. (new cycle every two weeks)
	160	TRA 1	Tumor response assessment 1
	161	TRA 2	Tumor response assessment 2
	162	TRA 3	Tumor response assessment 3
	…	…	…
	999	EOT	End of treatment
	1001	SFU30	Safety follow-up after 30 days
	1002	SFU60	Safety follow-up after 60 days
	1003	SFU100	Safety follow-up after 100 days
	2001	FU 1	Follow-up 1
	2002	FU 2	Follow-up 2
	2003	FU 3	Follow-up 3
	…	…	…
Gender	1	Male	Male patients
	2	Female	Female patients

6.2. General calculation rules

Data will be listed with the observed number of decimal places. Summary statistics will be presented to the same number of decimal places as the observed data, apart from the means and medians (to one more decimal place), and standard deviations (SD) (two more decimal places). Calculated data and summary statistics of calculated data will be presented with an appropriate number of decimal places (at least three significant digits).

The complete set of summary statistics will be given in case that the number of patients is at least three. For one or two patients, summary statistics will only comprise minimum and maximum.

Data from unscheduled assessments will only be listed and not presented in summary statistics. In case of unscheduled assessments prior to first dosing, the last measured value prior to dosing will be used as baseline value for presentation of summary statistics, if applicable.

The age of the patients will be calculated as: “date of informed consent” minus “birth date”, if applicable.

Relative study times and dates will be calculated with respect to the start time and date of the first administration of study drug.

The relative day of onset of an AE and the duration of an AE will be expressed in days as follows:
The relative day for onset of an AE is calculated as “day of onset” minus “day of start of treatment”. The duration of an AE is calculated as “stop date” minus “start date” plus 1.

If parts of the start/stop date of an AE are missing (day, month, or year), the relative day for onset and/or the duration of the AE are not calculated.

In order to determine if an AE or the worsening of an AE is a treatment-emergent event, the following rules will be used as a worst case scenario:

- If the date (and time) of onset or change is completely known, the AE is considered as treatment-emergent, provided the day (and time) of onset or change is after or on the same day (and time) as first study treatment.
- If the time of onset or change is missing, but the day, month and year is known, the AE is considered as treatment-emergent, provided the day of onset or change is after or on the same day as first study treatment, unless other information (e.g. stop date/time) suggest otherwise.
- If the time and day of onset or change is missing, but the month and year is known, the AE is considered as treatment-emergent, provided the month of onset or change is after or in the same month as first study treatment, unless other information (e.g. stop date) suggest otherwise.

- If the time, day and month of onset or change is missing, but the year is known, the AE is considered as treatment-emergent, provided the year of onset or change is after or in the same year as first study treatment, unless other information (e.g. stop date) suggest otherwise.
- If the time and date of onset or change is completely missing, the AE is considered as treatment-emergent, unless other information (e.g. stop date) suggests otherwise.

Baseline is defined as the last observation within 4 weeks prior to cycle 1.

Calculation of absolute changes from baseline:
- post-dose value - baseline value

Frequencies / percentages for qualitative data:
- Number of observations (frequency)
- Percent
  For time-independent variables (e. g. baseline characteristics, subgroups), percentages will be calculated as “frequencies” divided by the number of patients available (i. e. this may include missing data in the denominator).
  For time-dependent variables, e. g. percentages will be presented by time-point during study course, percentages will be calculated as “frequencies” divided by the number of non-missing data.

6.3. Descriptive statistics

Basic descriptive statistics for continuous data:
- Number of observations (N)
- Arithmetic mean (Mean)
- Standard deviation (SD)
- Minimum
- Lower quartile (Q1)
- Median
- Upper quartile (Q3)
- Maximum

7. Analysis sets and data review

7.1. Analysis sets

Analysis set 'Valid for safety'
All patients receiving at least one dose of study treatment will be evaluable for safety and included in the safety population.
It also includes patients who may not have received the full volume of injection.
All kinds of safety analyses will be based on this analysis set.

Analysis set 'Intention-to-treat (ITT)'
The Intention-to-treat (ITT) population will include all randomized patients in the study (signed ICF and confirmation of eligibility).

Analysis set 'Per-protocol (PP)'
The per-protocol analysis set includes all patients who

- are part of the ITT set
- are without any major protocol violations that would compromise the interpretation of efficacy data
- did receive study therapy for at least 6 weeks.

All primary and secondary efficacy analyses will be performed for both analysis sets (ITT and PP) where the ITT analysis is considered as the primary one.

7.2. Protocol violations/data review meeting

Possible protocol violations will be collected by the Data Management of the sponsor who will also prepare the data review meeting. At the data review meeting, possible protocol violations will be reviewed, categorized by severity and analysis sets will be defined. The final data will contain a dataset detailing the protocol violations and analysis set flags and comments in case of exclusion from analyses sets will be added to the demographic data.

8. Data analysis

8.1. Statistical software

Statistical analysis will be performed using SAS version 9.4 (or higher) on a Windows 10 personal computer.

8.2. Definitions, derived variables and handling of missing data

8.2.1. Derived variables

EORTC QLQ-C30

The EORTC QLQ-C30 contains the following scales:

A: Global health status / QoL
A1: Global health status / QoL — Items: 29, 30

B: Functional Scales
B1: Physical functioning — Items: 1, 2, 3, 4, 5
B2: Role functioning — Items: 6, 7
B3: Emotional functioning — Items: 21, 22, 23, 24
B4: Cognitive functioning — Items: 20, 25
B5: Social functioning — Items: 26, 27

C: Sympton scales / items
C1: Fatigue — Items: 10, 12, 18
C2: Nausea and vomiting — Items: 14, 15
C3: Pain — Items: 9, 19
C4: Dyspnoea — Item: 8
C5: Insomnia — Item: 11
C6: Appetite loss — Item: 13
C7: Constipation — Item: 16
C8: Diarrhoea — Item: 17
C9: Financial difficulties — Item: 28

It at least half of the items are not missing, the scoring of the scale is derived using the following algorithm.

First, the average of the no-missing items contributing to the scale will be calculated (raw score) as follows:

Raw score = RS = $(I_1 + I_2 + \ldots + I_n) / n$

Second, a linear transformation will be used for the standardization of the raw score, so that scores range from 0 to 100. A higher score represents a higher ('better') level of functioning, or a higher ('worse') level of symptoms:

Functional scales: S = (1 - (RS - 1) / range)*100
Symptom scales / items: S = ((RS-1) / range)*100
Global health status / Qol: S = ((RS-1) / range)*100

Range is the difference between the maximum possible value of RS and the minimum possible value. The QLQ-C30 has been designed so that all items in any scale take the same range of values. Therefore, the range of RS equals the range of the item values.

The overall summary score is computed as the mean of the subscale scores B1 to B5 and C1 to C8 (13 scores in total). The summary score will only be computed if all of the required scale scores are available.

STO-22
The items of this questionnaire are shown as items 31 to 52 on the questionnaire page on the CRF.

The STO-22 contains the following scales:

D: Functional scales	
D1: Body image	Item: 49
E: Symptom scales	
E1: Dysphagia	Items: 31, 32, 33
E2: Pain	Items: 34, 35, 36, 37
E3: Reflux symptoms	Items: 38, 39, 40
E4: Eating restrictions	Items: 41, 42, 43, 46
E5: Anxiety	Item: 47, 48, 50
E6: Dry mouth	Item: 44
E7: Taste	Item: 45
E8: Hair Loss	Item: 51, 52

The scores are calculated in the same way as the scales of the QLQ-C30. An overall summary score will not be derived.

8.2.2. Handling of missing values

Data from patients who prematurely terminate the trial will be used to the maximum possible extent.

All analyses will be based on available data. No procedures will be applied for replacement of missing values.

The primary endpoint OS *will be determined as time from the randomization date to the date of death. A* patient *who has not died will be censored at last known date alive.*

In case of secondary resection or local ablation and consecutive complete remission, patients will be counted as not evaluable in terms of overall response thereafter, but will be further evaluated for progression free survival until progression or death.

8.3. Disposition

8.3.1. Disposition summary

An overview of all patients who were randomized, treated and who completed the study will be provided by treatment group and overall (“trial profile”).

Reasons for all post-randomized discontinuations will be listed.

Furthermore, an overview of all patients who are in the different analysis sets will be provided by treatment group.

Frequencies of patients with post-randomized discontinuation and their reasons will be provided as well as frequencies of reasons for end of treatment.

8.3.2. Patient disposition

Individual patient profiles will be displayed in an Excel table (refer to template in Section 12).

8.4. Demographic data and baseline characteristics

Demographic data and baseline characteristics will be summarized for each of the three analysis sets.

Demographic and baseline characteristics
The following demographic and baseline characteristics will be summarized descriptively by treatment group:

- *Gender and age*
- Height and weight
- *ECOG performance status*
- Prior surgery
- Locally determined HER2 status
- Prior perioperative chemotherapy
- Location (AEG Type I-III, stomach)
- central tumor assessment (MSS or MSI, Infection with Epstein Barr Virus, centrally determined HER2 status, tumor proportion score (TPS), combined positive score (CPS), immune cell (IC) score

Summary statistics (number of patients, mean, SD, minimum, Q1, median, Q3 and maximum) will be presented for quantitative data (e.g. age). For qualitative data (e.g. gender), frequency tables including percentages will be provided.

Medical history
Medical history entries are only listed.

Concomitant medication
Concomitant medication entries are only listed.

8.5. Analysis of primary endpoint

The analyses of the primary endpoint OS will be conducted on the ITT and the PP analysis sets.

8.5.1. Definition of covariates/subgroups

Subgroup analyses will be performed for all patients with a CPS > 1 and all patients with a CPS > 5.

8.5.2. Confirmatory analysis of the primary endpoint

The primary endpoint as well as the null-hypothesis to be tested are defined in Section 5.1.

Each of the experimental arms would be considered promising, if the true 12-month-OS rate amounts to 70 %. This translates into a hazard ratio of 0.6 compared to the standard OSR@12 of 55% for chemotherapy and trastuzumab.

Hypothesis testing will be performed with a one-sided logrank test (against a constant hazard derived from the historical data on the standard therapy [...]) at a one-sided significance level of 0.05.

For each patient $i = 1, \dots, n$, let time-to-event data be given by:

- Δ_i: Censoring indicator.

$$\Delta_i = \begin{cases} 0, if\ patient\ i\ is\ censored \\ 1, if\ patient\ i\ had\ an\ event \end{cases}$$

- T_i: Right-censored survival time

The number of events observed in the experimental treatment arm is

$$O = \sum_{i=1}^{n} \Delta_i.$$

Let $\Lambda_H(t) = \int_0^t \lambda_H(u)du$ denote the cumulative hazard function of the historical control, and let E_H denote the expected number of events given a hazard ratio R of 1:

$$E_H = \sum_{i=1}^{n} \Lambda_H(\mathrm{T}_i).$$

Likewise, $\gamma_0 E_H$ may be interpreted as the expected number of events if the hypothesis $H_0: R = \gamma_0$ holds true.

The one-sample log-rank test of $H_0: R \geq \gamma_0$ at one-sided significance level α is defined by

$$Z = \frac{O - \gamma_0 E_H}{\sqrt{\gamma_0 E_H}} \leq \Phi^{-1}(\alpha)$$

H_0 will be rejected if $Z \leq \Phi^{-1}(\alpha)$,

where Φ denotes the standard normal distribution function.

8.5.3. Explorative analysis of primary and co-primary endpoints

Not applicable.

8.5.4. Interim analyses

Not applicable.

8.6. Analysis of secondary efficacy endpoints

Analysis of the secondary endpoints concerning safety and tolerability is covered in Section 8.11.

All secondary efficacy analyses, excluding safety and tolerability, will be conducted on the ITT analysis set.

8.6.1. Progression Free Survival (PFS)

Time from the *randomization date to the date of first observed disease progression (investigator assessment according to RECIST 1.1) or death from any cause. Clinical deterioration in the absence of unequivocal evidence of progression (per RECIST 1.1) is not considered progression for purposes of determining PFS.* Patients *who die without a reported prior progression will be considered to have progressed on the date of their death.* Patients *who did not progress or die will be censored on the date of their last evaluable tumor assessment.* Patients *who did not have any on study tumor assessments and did not die will be censored on the date they were registered.* Patients *who started any subsequent anti-cancer therapy without a prior reported progression will be censored at the last evaluable tumor assessment prior to or on the date of initiation of the subsequent anti-cancer therapy.*

The progression-free survival curves will be estimated for each treatment group separately by use of the Kaplan-Meier method, and will be graphically displayed.

Subgroup analyses will be performed for all patients with a CPS > 1 and all patients with a CPS > 5.

8.6.2. Response rate (RR)

Response rate will be assessed according to RECIST v1.1. Overall response rate will be defined as the proportion of randomized patients *with best response of complete or partial response.*

The response rate will be summarized descriptively by time point and treatment group by generating frequency tables (including outcomes for CR, PR, SD). For these frequency tables, subgroup analyses will be performed for all patients with a CPS > 1 and all patients with a CPS > 5.

The time to deterioration after response is measured as the time from first occurrence of any response according to RECIST v1.1 until the first deterioration from response. The time to deterioration is analysed using the Kaplan-Meier method. Patients with no deterioration will be censored at the time of last determination of response. Patients with no positive response will not be included into this analysis.

Spider plots will be generated for the percentage change in sum of diameters of target tumor lesions, and waterfall plots will be produced for the percentage change in sum of diameters of target tumor lesions for the best response for the two treatment groups.

8.6.3. Quality of life assessments (EORTC QLQ-C30 and STO-22)

Quality of life will be assessed using the EORTC QLQ C30 questionnaire and the module STO22 at baseline, during treatment every 8 weeks during the first 12 months, afterwards every 3 months and at the end of treatment.

The scores of the EORTC QLQ-C30 and STO-22 will be summarized by subscale, treatment group and time point.

8.7. Descriptive efficacy analyses

Time to event distributions will be estimated using Kaplan-Meier techniques. Median OS along with 95% CI will be constructed based on a log-log transformed CI for the survivor function. Rates at fixed timepoints (e.g. milestone rate of OS at 12 months) will be derived from the Kaplan-Meier estimate and corresponding confidence interval (CI) *will be derived based on Greenwood formula for variance derivation and on log-log transformation applied on the survivor function.*

Kaplan-Meier survival curves will be graphically displayed. Median, Q1 and Q3 of survival times for both the OS and PFS including 95% confidence intervals and mean survival times including standard error will be presented. Monthly survival rates will be presented using the life table estimate.

8.8. Pharmacokinetic data analyses

Not applicable.

8.9. Pharmacodynamic data analyses

Not applicable.

8.10. Evaluation of relationships between endpoints

Not applicable.

8.11. Safety analysis

All safety analyses will be conducted on the 'Valid for safety' analysis set.

8.11.1. Extent of exposure

Frequency tables for the number of cycles, mean and median number of cycles are presented by treatment group overall and by single treatments (5FU/FS, Oxaliplatin, Trastuzumab, Nivolumab, Ipilimumab). Dose will be summarized per cycle and overall, also as percent of planned dose to assess treatment compliance.
Furthermore, frequencies of overall dose modifications and particular dose delays and reduced doses will be summarized per treatment arm and separated for chemotherapy (FOLFOX), trastuzumab and immunotherapy (Nivo+/-Ipi).

8.11.2. Primary safety analysis

Not applicable.

8.11.3. Secondary safety analysis

Not applicable.

8.11.4. Adverse events

Adverse events (AEs) will be coded by Dr. Notghi Contract Research GmbH according to MedDRA (latest version available).

An overview over all AEs (original term) will be generated, including onset relative to start of treatment and start of last cycle, duration, intensity, seriousness and relationship to study treatments, action taken and outcome.

A summary table counting number of patients with at least 1 AE, number of patients with treatment-emergent AEs, number of patients with CTCAE grade 1, 2, 3, 4 and 5 and also with CTCAE grade ≥ 3 by relationship, number of patients with at least one serious adverse event (SAE), number of patients with at least one treatment related SAE and AEs leading to

treatment withdrawal or interruption will be provided. This table will also be presented for number of AEs.

Treatment emergent events are AEs not present at baseline, or AEs that worsened after start of treatment even if they were present at baseline.

The incidence of treatment emergent AEs and SAEs will be summarized by treatment group using MedDRA terms grouped by PTs, by SOCs, also by CTCAE grades and by relationship to study treatment. A listing will be provided linking the original and coded terms.

The occurrence of one AE in the same patient more than once will be counted only once. The descriptive statistics will include the number and percentage of patients who experienced treatment emergent AEs.

Deaths, SAEs and certain other significant AEs will be listed separately (if applicable).

8.11.5. Clinical laboratory

Complete listings for laboratory values will be generated. Laboratory values outside the normal range will be marked.
Separate listings will be presented for abnormal values assessed as clinically relevant.

Results from laboratory tests (hematology, chemistry, coagulation, hormones, and tumor markersl) will be summarised by treatment group and time point using descriptive statistics including changes from baseline.

8.11.6. Vital signs

Vital sign values (blood pressure, heart rate, respiratory rate, body temperature, oxygenation saturation and left ventricular ejection fraction) will be summarised by treatment group and time point using descriptive statistics including changes from baseline.

8.11.7. ECG

Data listings will be provided for ECG results.

8.11.8. Physical examination

Data listings will be provided for results of *physical examination*.

8.11.9. Other assessments

<u>Translational research</u>

The following translational research is currently planned, but may be adapted taking into account new research data

- *Tumor-infiltrating lymphocytes (TiL) repertoire determination from tumor*

- *Liquid biopsy next-generation sequencing (NGS) immunoprofiling (TCRβ & IgH) before treatment initiation and before second cycle to determine response predictive immune signature (diversification pattern as read-out for ongoing immune activation, TiL clone expansion in peripheral blood)*
- *In addition FFPE will be centrally tested for PD-L1, HER2 (IHC and ISH), MSI, EBV and HER signaling alterations (amplifications and/or mutations in e.g. EGFR, HER2, HER3, PIK3CA) and correlated with clinical efficacy.*
- *CTC will be evaluated for changes in HER2 and PD-L1 status*
- *ctDNA will be evaluated for HER signaling alterations (amplifications and/or mutations in e.g. EGFR, HER2, HER3, PIK3CA)*
- *Central imaging review and determination of ORR and PFS according to modified RECIST (refer to appendix D* of the protocol*).*

Thus, the tumor block for TiL analysis, HER2, PD-L1 and HER signaling assessment will be obtained at baseline. Blood will be collected prior to first treatment and beginning of cycle 2 and 5 and at progression and/or end of treatment. In addition, imaging will be retrospectively collected.

Evaluation of translational research results is beyond the scope of this statistical analysis plan.

9. Changes to the protocol

In addition to protocol, a per-protocol analysis (PP) set was defined. Efficacy analyses are now to be performed on both ITT and PP sets. In this context, generation of a data review plan and performance of a data review meeting was addressed.

Furthermore, Waterfall plots for best response and spider plots for percentage change in sum of diameters of target tumor lesions were added.

10. Contents of statistical tables and figures

For analyses which will be provided for different analysis sets, sub-outputs will be created using letters a), b), c). This will only be the case, if analysis sets differ from each other. If two (or more) analysis sets are identical, one of the identical analysis sets will be mentioned in a subtitle.

Number and titles of Tables, Figures and Listings are subject to change if the need arises during analysis.

Table / Figure	No.	14 Tables and figures referred to but not included in the text
		14.1 Demographic data and general study information
Table	14.1-1	Trial profile
Table	14.1-2	Reasons for premature termination of the study

Table / Figure	No.	14 Tables and figures referred to but not included in the text
Table	14.1-3	Reasons for end of treatment
Table	14.1-4	Frequency of patients with protocol violations
Table	14.1-5	Reasons for restriction of validity
Table	14.1-6	Descriptive statistics for demographic data
Table	14.1-7	Frequencies of demographic data
		14.2 Analysis of primary efficacy endpoint
		14.2.1 All patients
Table	14.2.1-1	Kaplan-Meier estimates of overall survival
Figure	14.2.1-2	Kaplan-Meier curves for overall survival
Table	14.2.1-3	Monthly overall survival rates using the life table estimate
Table	14.2.1-4	Median and 12-month overall survival
Table	14.2.1-5	Log-rank test results
		14.2.2 Subgroups analysis
Table	14.2.2-1	Kaplan-Meier estimates of overall survival – Patients with CPS > 1
Figure	14.2.2-2	Kaplan-Meier curves for overall survival – Patients with CPS > 1
Table	14.2.2-3	Monthly overall survival rates using the life table estimate – Patients with CPS > 1
Table	14.2.2-4	Median and 12-month overall survival – Patients with CPS > 1
Table	14.2.2-5	Log-rank test results – Patients with CPS > 1
Table	14.2.2-6	Kaplan-Meier estimates of overall survival – Patients with CPS > 5
Figure	14.2.2-7	Kaplan-Meier curves for overall survival – Patients with CPS > 5
Table	14.2.2-8	Monthly overall survival rates using the life table estimate – Patients with CPS > 5
Table	14.2.2-9	Median and 12-month overall survival – Patients with CPS > 5
Table	14.2.2-10	Log-rank test results – Patients with CPS > 5
		14.3 Analysis of secondary endpoints
		14.3.1 All patients
Table	14.3.1-1	Kaplan-Meier estimates of progression free survival
Figure	14.3.1-2	Kaplan-Meier curves for progression free survival
Table	14.3.1-3	Monthly progression free survival rates using the life table estimate
Table	14.3.1-4	Median and 12-month progression free survival
Table	14.3.1-5	Frequencies of response rates by RECIST criteria
Table	14.3.1-6	Kaplan-Meier estimates of time to deterioration after response
Figure	14.3.1-7	Kaplan-Meier curves for time to deterioration after response
Table	14.3.1-8	Median time to deterioration after response
Table	14.3.1-9	Descriptive statistics for best response (percentage change in sum of diameters of target tumor lesions)
Figure	14.3.1-10	Waterfall plots for best response (percentage change in sum of diameters of target tumor lesions)
Figure	14.3.1-11	Spider plots for percentage change in sum of diameters of target tumor lesions
Table	14.3.1-12	Descriptive statistics of EORTC questionnaire

Table / Figure	No.	14 Tables and figures referred to but not included in the text
Table	14.3.1-13	Descriptive statistics of STO-22 questionnaire
		14.3.2 Subgroup analysis
Table	14.3.2-1	Kaplan-Meier estimates of progression free survival – Patients with CPS > 1
Figure	14.3.2-2	Kaplan-Meier curves for progression free survival – Patients with CPS > 1
Table	14.3.2-3	Monthly progression free survival rates using the life table estimate – Patients with CPS > 1
Table	14.3.2-4	Median and 12-month progression free survival – Patients with CPS > 1
Table	14.3.2-5	Frequencies of response rates by RECIST criteria – Patients with CPS > 1
Table	14.3.2-6	Kaplan-Meier estimates of progression free survival – Patients with CPS > 5
Figure	14.3.2-7	Kaplan-Meier curves for progression free survival – Patients with CPS > 5
Table	14.3.2-8	Monthly progression free survival rates using the life table estimate – Patients with CPS > 5
Table	14.3.2-9	Median and 12-month progression free survival – Patients with CPS > 5
Table	14.3.2-10	Frequencies of response rates by RECIST criteria – Patients with CPS > 5
		14.4 Safety data
		14.4.1 Extent of exposure
Table	14.4.1-1	Frequencies, mean and median number of cycles by treatment group
Table	14.4.1-2	Descriptive statistics for dose and percent of planned dose
Table	14.4.1-3	Frequencies of overall dose modifications, dose delays and reduced doses
		14.4.2 Display of adverse events
Listing	14.4.2-1	Overview of all adverse events
Table	14.4.2-2	Summary of all adverse events
Table	14.4.2-3	Incidences of treatment emergent adverse events
Table	14.4.2-4	Incidences of treatment emergent serious adverse events
Table	14.4.2-5	Incidences of treatment emergent adverse events by intensity
Table	14.4.2-6	Incidences of treatment emergent adverse events by CTCAE grade
Table	14.4.2-7	Incidences of treatment emergent adverse events by relationship to study medication
Table	14.4.2-8	Incidences of treatment emergent serious adverse events by relationship to study medication
		14.4.3 Listings of deaths, other serious and significant adverse events
Listing	14.4.3-1	Listing of deaths
Listing	14.4.3-2	Listing of serious adverse events (death excluded)
Listing	14.4.3-3	Listing of severe adverse events (death and serious adverse events

Table / Figure	No.	14 Tables and figures referred to but not included in the text
		excluded)
Listing	14.4.3-4	Listing of adverse events leading to study discontinuation
		14.4.4 Laboratory values
Table	14.4.4-1	Summary statistics for hematology variables including changes from baseline
Table	14.4.4-2	Summary statistics for clinical chemistry variables including changes from baseline
Table	14.4.4-3	Summary statistics for coagulation variables including changes from baseline
Table	14.4.4-4	Summary statistics for hormones including changes from baseline
Table	14.4.4-5	Summary statistics for tumor makers including changes from baseline
		14.4.5 Vital signs
Table	14.4.5-1	Summary statistics for vital signs including changes from baseline

11. Contents of appendices' listings

16 Appendices

16.1.9 Documentation of statistical methods

16.1.9-1 Log-rank test for Table 14.2.1-5
16.1.9-2 Log-rank test for Table 14.2.2-5
16.1.9-3 Log-rank test for Table 14.2.2-10
16.1.9-4 Calculated values of duration of response

16.2 Patient data listings

16.2.1 Discontinued patients

Listing 16.2.1-1 Trial completion information

16.2.2 Protocol deviations

Listing 16.2.2-1 Protocol deviations

16.2.3 Patients excluded from the efficacy analysis

Listing 16.2.3-1 Analysis sets with reason for exclusion

16.2.4 Demographic data

Listing 16.2.4-1 Demographic data
Listing 16.2.4-2 Medical history
Listing 16.2.4-3.1 Description of inclusion criteria
Listing 16.2.4-3.2 Inclusion criteria – answers
Listing 16.2.4-3.3 Description of exclusion criteria
Listing 16.2.4-3.4 Exclusion criteria – answers
Listing 16.2.4-4 Eligibility of patients
Listing 16.2.4-5 Written informed consent

16.2.5 Compliance and drug concentration data

Listing 16.2.5.1 Course of study dates – regular visits
Listing 16.2.5.2 Course of study dates - unscheduled visits
Listing 16.2.5.3 Administration of study medication

16.2.6 Individual efficacy response data

Listing 16.2.6-1 Overall survival
Listing 16.2.6-2.1 Progression free survival
Listing 16.2.6-2.2 Response rate
Listing 16.2.6-2.3 Quality of life assessment

16.2.7 Adverse event Listings (each patient)

Listing 16.2.7-1.1 Adverse events – details
Listing 16.2.7-1.2 Adverse events – coding
Listing 16.2.7-1.3 Serious adverse events

16.2.8 Listing of individual laboratory measurements by patient

Listing 16.2.8-1 Laboratory normal ranges

Listing 16.2.8-2.1	Laboratory evaluations – sampling
Listing 16.2.8-2.2	Laboratory evaluations – hematology
Listing 16.2.8-2.3	Laboratory evaluations – chemistry
Listing 16.2.8-2.4	Laboratory evaluations – coagulation
Listing 16.2.8-2.5	Laboratory evaluations – hormones
Listing 16.2.8-2.6	Laboratory evaluations – tumor markers
Listing 16.2.8-2.7	Laboratory comments

16.2.9 Vital signs and other safety data

Listing 16.2.9-1	Vital signs
Listing 16.2.9-2	Electrocardiogram
Listing 16.2.9-3	Physical examination

16.2.10 Other patient data

Listing 16.2.10-1	Further cancer treatment
Listing 16.2.10-2	Other concomitant medication
Listing 16.2.10-3	Comments

12. Other outputs

Excel template for individual patient disposition:

Patient_disposition_template.xlsx

14. References

Bang, Y. J., E. Van Cutsem, A. Feyereislova, H. C. Chung, L. Shen, A. Sawaki, F. Lordick, A. Ohtsu, Y. Omuro, T. Satoh, G. Aprile, E. Kulikov, J. Hill, M. Lehle, J. Ruschoff and Y. K. Kang (2010). "Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial." Lancet **376**(9742): 687-697.