Statistical and Epidemiological Analysis Plan

c09070203-02

BI Trial No.:	1199.223
Title:	A non-interventional biomarker study in patients with Non-Small Cell Lung Cancer (NSCLC) of adenocarcinoma tumour histology eligible for treatment with Vargatef® according to the approved label.
Investigational Product(s):	Vargatef® (nintedanib)
Responsible trial statistician(s):	Phone: Fax:
Date of statistical analysis plan:	28 Mar 2019 SIGNED
Version:	"Final"

1. TABLE OF CONTENTS

LIST OF TABLES

2. LIST OF ABBREVIATIONS

Include a list of all abbreviations used in the SEAP

Term	Definition / description
AE	Adverse Event
ADR	Adverse Drug Reaction
BES	Biomarker evaluation set
BM	Biomarker
CI	Confidence Interval
CNV	Copy Number Variant
CRF	Case Report Form
CTP	Clinical Trial Protocol
DNA	Deoxyribonucleic Acid
DBL	Database Lock
DBLM	Database Lock Meeting
DM&SM	Boehringer Ingelheim Data Management and Statistics Manual
eCRF	Electronic Case Report Form
EMA	European Medicines Agency
ES	Entered Set
FDR	False Discovery Rate
ICH	International Conference on Harmonisation
IHC	Immunohistochemistry
iPD	Important Protocol Deviation
MedDRA	Medical Dictionary for Regulatory Activities
NIS	Non-interventional Study
NSCLC	Non-Small Cell Lung Cancer
PD	Progression
PD-L1	Programmed cell death 1 ligand 1
PFS	Progression-Free Survival
PPS	Per Protocol Set
PT	Preferred Term
RNA	Ribonucleic Acid
RPM	Report Planning Meeting
SD	Standard Deviation
SEAP	Statistical and Epidemiological Analysis Plan
SOC	System Organ Class

Term	Definition / description
TMB	Tumor Mutational Burden
TS	Treated Set
TSFL	Time since first line
TSFLT	Time since first line therapy

3. INTRODUCTION

As per ICH E9 (1) (International Conference on Harmonisation) and the ISPE Guidelines of Good Pharmacoepidemiology Practice (2), the purpose of this document is to provide a more technical and detailed elaboration of the principal features of the analysis described in the protocol, and to include detailed procedures for executing the statistical analysis of the primary and secondary variables and other data.

This Statistical and Epidemiological Analysis Plan (SEAP) assumes familiarity with the Clinical Trial Protocol (CTP), including Protocol Amendments. In particular, the SEAP is based on the planned analysis specification as written in CTP Section 9.7 “Data Analysis”. Therefore, SEAP readers may consult the CTP for more background information on the study, e.g., on study objectives, study design and population, treatments, definition of measurements and variables, planning of sample size, randomization.

SAS® Version 9.4 (or later version) will be used for all standard analyses and R 3.2.2 (or later version) will be used for the multivariate analyses described in Section 7.6.2.

The **primary objective** of the study is to generate hypotheses whether gene expression patterns (alone or combined with clinical covariates) or genetic markers could be used to predict OS in patients treated with Vargatef® according to the approved label.

Therefore, OS and PFS will be calculated (medians, Kaplan-Meier and 1 year rates), based on the entered set (ES), for all patients and for the ES-subset of fast progressing patients on prior 1st line therapy (time since start of 1st line therapy <9mts and separately for PD as best response on 1st line therapy), irrespective of any BMs (demographics and other baseline characteristics as well as exposure will be reported descriptively for all three populations). Subgroup analyses will be performed for the overall population (ES).

Multivariate BM variable selection will be based on OS only but the resulting selection of variables will be used for PFS, too. IPF Lasso will be applied using modalities as defined in Section 7.7.2, in addition, standard Lasso for multivariate analyses will be performed separately for each of the following 5 sets of variables on the biomarker evaluation set (BES)
a) Gene expression-based (Ribonucleic acid (RNA)) pathways and cell types plus selected individual angiogenesis genes (but not for each single gene)
b) Deoxyribonucleic acid (DNA) alterations by gene plus putative Tumor mutational burden (TMB)
c) Gene-expression-based pathways and cell types, DNA and clinical variables
d) Immunohistochemistry (IHC)
e) Gene expression-based pathways and cell types, DNA, genotyping, IHC and clinical variables.
The data will be analysed in these 5 separate groups because on the one hand for IHC and genotyping the number of valid samples is expected to be smaller than for DNA and RNA data and on the other hand clinical variables as pre-specified in the CTP have a higher level of prior evidence and this should be taken into account in the analyses.

Univariate analyses will be performed for gene expression-based pathways and cell types plus selected individual angiogenesis genes (but not for each single gene), DNA alterations by gene, putative TMB, IHC, genotyping and the clinical subgroups for OS and PFS on the BES.

Further objectives of the study are:

- to investigate potential association between the covariate "time since start of first line chemotherapy until start of Vargatef® therapy" and gene-expression patterns.
- to identify potential biomarkers, tumour genomic alterations and/or gene expression patterns to characterize the survival outcome potential of patients with late progression (≥ 9 months) versus early progression(< 9 months) from start of first line chemotherapy until start of Vargatef® therapy.

Therefore, time since start of 1st line therapy (1st further objective) will be analysed with univariate and multivariate techniques (first with RNA, DNA alterations by gene plus putative TMB and clinical variables and second with RNA, DNA alterations by gene plus putative TMB, genotyping, IHC and clinical variables as above). The same will be done for PD best response on 1st line therapy. The resulting signatures will then be evaluated for their prognostic potential (2nd further objective).

Listings of the data will be shown only for relevant variables including BMs and clinical variables.

4. CHANGES IN THE PLANNED ANALYSIS OF THE STUDY

The planned number of 250 overall survival (OS) events will not be reached, therefore, in agreement with EMA (European Medicines Agency), the study will be analysed earlier. Progression-free survival (PFS) will be added as endpoint (further endpoint) based on the collected information in the electronic case report form (eCRF) to enable comparisons with other published study results.
Since patients who received immunotherapies were also recruited into this study, the cutpoint of 9 months for the biomarker (BM) time since first line therapy (TSFLT) might not be optimal anymore therefore an additional analysis of continuous TSFLT will be performed in addition and an analysis on patients with progression (PD) as best response to first-line therapy.

5. ENDPOINT(S)

5.1 PRIMARY ENDPOINT(S)

The primary endpoint of this study is Overall Survival (OS) as defined in the CTP section 9.3.2.

5.2 SECONDARY ENDPOINT(S)

5.2.1 Key secondary endpoint(s)

Not applicable

5.2.2 (Other) Secondary endpoint(s)

Not applicable

Safety:

Safety will be collected with a focus on the following variables:

- all adverse drug reactions (ADR) (serious and non-serious),
- all AEs with fatal outcome.

6. GENERAL ANALYSIS DEFINITIONS

The **primary objective** of the study is to generate hypotheses whether gene expression patterns (alone or combined with clinical covariates) or genetic markers could be used to predict OS in patients treated with Vargatef® according to the approved label.

Further objectives of the study are:

- to investigate potential association between the covariate "time since start of first line chemotherapy until start of Vargatef® therapy" and gene-expression patterns.
- to identify potential biomarkers, tumour genomic alterations and/or gene expression patterns to characterize the survival outcome potential of patients with late progression (≥ 9 months) versus early progression(< 9 months) from start of first line chemotherapy until start of Vargatef® therapy.

6.1 TREATMENT(S)

All patients receive the same treatment: Vargatef® (200 mg administered twice daily except on the day of docetaxel infusion) in combination with docetaxel 75 mg/m² every 3 weeks.

6.2 IMPORTANT PROTOCOL DEVIATIONS

Data discrepancies and deviations from the CTP will be identified for all entered patients. Listings of protocol deviations and of unresolved discrepancies will be provided to be discussed at the combined report planning and database lock meeting (RPM/DBLM). At this meeting, it will be decided whether the discrepant data can be used as they are or whether the data have to be corrected in the clinical database.

Each protocol deviation must be assessed to determine whether it is an important protocol deviation (iPD). A PD is important if it affects the rights or safety of the study patients or if it can potentially influence the primary outcome measure(s) for the respective patients in a way that is neither negligible nor in accordance with the study objectives. This last category of important PD forms the basis for the decision of whether a patient does or does not belong to an analysis set. PDs that do not influence the patient's rights and safety or the evaluability of the patients for the main study objectives are called non-important PDs. These are only considered when checking the study quality in general.
If any important PDs are identified, they are to be summarised into categories and will be captured in the RPM/DBLM minutes via an accompanying Excel spreadsheet (4) [001-MCS-40-413_RD-00]. The following table contains the categories which are considered to be important PDs in this study. If the data show other important PDs, this table will be supplemented accordingly by the time of the RPM/DBLM.

Table 6.2: 1 Protocol deviations

Category / Code	Description	Requirements	Excluded from TS[1]	Excluded from PPS[2]

Category / Code		Description	Requirements	Excluded from TS[1]	Excluded from PPS[2]
A		**Entrance criteria not met**			
	A1	Inclusion criteria not met	Inclusion criteria not met as specified in the protocol.	No	Potentially
	A1.1	Patient not treated according to the label	Label not met, e.g. patients not 2nd line chemotherapy after failing a platinum based chemotherapy in NSCLC adenocarcinoma or patients have received a too low dose	No	Potentially
	A2	Exclusion criteria not met	Exclusion criteria not met as specified in the protocol	No	No
B		**Informed consent**			
	B1	Informed consent not available/not done	Informed consent date missing	Yes	Yes
C		**Study medication**			
	C1	Medication not taken according to the label[1]	Vargatef not taken according to the label	No	Potentially

PPS – Per Protocol Set
1 patients will only be excluded from TS if they have not taken any nintedanib.
2 to mimic a patient population as it was observed in the pivotal trial patients who have received more than one line of therapy and are consequently not considered to be treated per label are excluded from the PPS. Those patients will be identified in a dedicated meeting prior to database lock (DBL).

6.3 PATIENT SETS ANALYSED

Entered Set (ES):
This patient set includes all patients who entered the study (screening failures are excluded) no matter if they have taken nintedanib or not.

Biomarker Evaluation Set (BES):
This patient set includes all patients from ES for whom a valid per protocol tissue sample (eg. sample was taken after start of 1st line therapy) was available. The patients excluded from BES will be determined in a dedicated meeting prior to DBL.

Treated Set (TS):
This patient set includes all patients who were documented to have taken at least one dose of nintedanib.

Per protocol Set (PPS):
This patient set includes all patients who have not experienced any PV which leads to the exclusion from this analysis set. The patients excluded from the PPS will be determined in a dedicated meeting prior to DBL.

Table 6.3: 1 Patient sets and analyses

Class of endpoint	Patient sets			
	ES	BES	TS	PPS
Primary and further endpoints	X	X		X
Biomarker analysis (including primary and further objectives)		X		
Safety analysis			X	
Demographic/baseline characteristics		X		

6.5 POOLING OF CENTRES

This section is not applicable because centre/country is not included in the statistical model.

6.6 HANDLING OF MISSING DATA AND OUTLIERS

All available safety data for treated patients who failed to complete all stages of the study (who withdraw or are removed from the study) will be reported.

All patient discontinuations will be documented and the reason for discontinuation will be recorded.

Censoring rules for OS are described in the CTP section 9.7.1.

Censoring rules for PFS are described in the Section 7.6.5.

It is not planned to impute missing values, with exception of missing AE start dates and times. These are imputed according to BI standards (see DM&SM “Handling of missing and incomplete AE dates” (6) [001-MCG-156_RD-01]).

6.7 BASELINE, TIME WINDOWS AND CALCULATED VISITS

The data of visit 1 will be considered as baseline data (the baseline visit is the visit from docetaxel administration up to 7 days after Vargatef® treatment initiation).

Germline blood or buccal swab samples could be taken at any time during routine clinical assessments and are all considered as baseline because of the stability of the analyte.

Follow-up visits – every 6 months (± 1 month) until death or the end of the study whichever occurs first. No time windows will be calculated.

7. PLANNED ANALYSIS

For End-Of-Text tables, the set of summary statistics is: N / Mean / SD / Min / Median / Max for continuous BMs P25 and P75 are reported additionally.

For tables that are provided for endpoints with some extreme data, median, quartiles and percentiles should be preferred to mean, SD, minimum and maximum.

Tabulations of frequencies for categorical data will include all possible categories and will display the number of observations in a category as well as the percentage (%) relative to the respective treatment group (unless otherwise specified, all patients in the respective patient set whether they have non-missing values or not). Percentages will be rounded to one decimal place. The category missing will be displayed only if there are actually missing values.

Tables describing disposition of subjects, important PDs and analysis sets will be presented categorically (number and percentage) by subgroup (see Section 6.4) for the ES. The reasons for the drop-outs will be categorized (AE / Death / Lack of efficacy / Lost to follow up / Physician's decision / Pregnancy / Non-compliant with protocol / Patient wish / other). The categorization will be discussed before DBL.

In general, all biomarker analyses will be performed on the BES and all safety analyses will be performed on the TS. OS and PFS will be calculated (medians, Kaplan-Meier and 1 year rates), based on the ES, for all patients and for the ES-subset of fast progressing patients on prior 1st line therapy (time since start of 1st line therapy <9mts and another analysis in patients with PD best response to first-line therapy), irrespective of any BMs (demographics and other baseline characteristics as well as exposure will be reported descriptively for all three populations). Subgroup analyses will be performed only in the overall population (BES).

Listings of the data will be done for all efficacy variables, demographics and for selected BM assessments. This might include a listing of mutations, CNVs and fusions by patient.

7.1 DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS

Only descriptive statistics are planned for this section of the report including demographics, oncological history, prior first-line chemotherapy, prior adjuvant anti-cancer therapies, prior neoadjuvant anti-cancer therapies, prior other anti-cancer therapies, previous radiotherapies, prior surgeries and tumour tissue sampling.

7.2 CONCOMITANT DISEASES AND MEDICATION

Only descriptive statistics are planned for this section of the report.

7.3 TREATMENT COMPLIANCE

Only descriptive statistics are planned for this section of the report.

7.4 PRIMARY ENDPOINT(S)

As described in the CTP (section 9.7.1) the primary analysis of OS is based on Kaplan-Meier estimates. The primary analysis was planned to be conducted after 250 OS events (see

protocol section 9.5 for details). However, according to section 9.2 the study will be terminated and analysed earlier.
OS (in months) will be measured from start of therapy (First administration of Vargatef®) until death. Patients not known to have died will be censored at their date of last contact. This will be the last date of the following CRF fields:

For all dates mentioned below, only non-imputed complete dates are taken into account.

AE page: AEONYMD, AEENDYMD (the raw values not the imputed ones),
Sampling pages: BLDDT, LABDT, BIONYMD, BIONYMD
Exposure page: ADMSPDT, ADMSDT,
Termination of Vargatef/Docetaxel pages: DRGYMD, DRGYMD
Patient Status pages: PROGYMD, DTH and LCONDT with PASTA = ALIVE, ALIDT with PASTA = Lost to follow-up, CONWDT
Withdrawal of biobanking consent page: CONWADT
Subsequent AC pages: C TBEGYMD, CTENDYMD and Date of first administration and Date of last administration for all compounds.

Kaplan-Meier estimates will be used for the graphical presentation of primary endpoint and their summary statistics, median and its 95% confidence intervals (CIs) and the range of OS. Breslow's method for handling ties will be used. The Brookmeyer and Crowley method (2.) will be used to produce 95% CIs for the median OS on the log-log scale.

Subgroups described in Section 6.4 will be applied to overall survival analysis as well.

Descriptive statistics for the use of subsequent anti-cancer therapies will be provided.

A sensitivity analysis on the PPS population will be performed for all patients and for all patients with TSFLT <9mts.

7.5 SECONDARY ENDPOINT(S)

7.5.1 Key secondary endpoint(s)

This section is not applicable as no key secondary endpoint has been specified in the protocol.

7.5.2 (Other) Secondary endpoint(s)

This section is not applicable as no secondary endpoint has been specified in the protocol.

7.7 STATISTICAL MODELS

The aim of the following analyses is to find prognostic BMs for OS based on categorical and dichotomized BMs. PFS is only used to verify the OS findings. Additionally, the established prognostic and predictive factor "Time since start of first line therapy" will be correlated with the BMs.
Descriptive statistics for time since first line (TSFL), mutations, copy number alterations and genotypes and all variables in Section 6.4 will be presented for BES.
If deemed necessary, demographics and other baseline characteristics for specific BM groups will be tabulated.

7.7.1 Univariate statistical model

The analyses will be carried out for all variables in Section 6.4 (categorical and/or continuous as defined above). All analyses will be performed on the BES. False Discovery Rate (FDR) corrected p-values will be presented according to Section 7.8.5.

Association of BMs and TSFL with overall survival
Univariate statistical models for analysing associations of BM including the subgroups defined in Section 6.4 with OS are described in detail the CTP section 9.7.2.1 and will be applied to all categorical BMs. Hazard ratios (HR) with 95% (Score Test) CIs will be

reported.
Kaplan-Meier curves will be produced for the graphical presentation of the BM groups where deemed necessary.
Additionally, BM (e.g. TSFL) will be considered as continuous variable in the Cox model, where possible. The HR will be evaluated at minimum, median and maximum and presented graphically.

Association of BMs with dichotomized TSFL and PD best response to 1st line therapy
Dichotomized TSFL will be used as response in a logistic regression with categorical BMs including the subgroups defined in Section 6.4 as covariate (cf. CTP section 9.7.2.1). The odds ratios with 95% (profile likelihood) CIs will be reported.
Dichotomized PD best response to prior therapy will be analysed with the same statistical model.

Association of BMs with continuous TSFL
Continuous TSFL will be used as outcome in a linear regression with BMs including the subgroups defined in Section 6.4 as covariate (see model below).

Linear regression

$$y_i = \beta_0 + \beta_B B_i,$$

where

y_i Continuous outcome,

β_0 Intercept,

β_B Effect of BM,

B_i BM.

The coefficient β_B with corresponding 95% Wald CI will be reported. The model will be presented graphically, if deemed necessary.

7.7.2 Multivariate statistical model

Multivariate statistical models analysing associations of BM-signatures with OS are outlined in the CTP section 9.7.2.2, and are described here in more detail:

The IPF Lasso (7) will use the following two modalities

a) Clinical variables, IHC and genotyping
b) Gene expression-based pathways and cell types plus selected individual angiogenesis genes (please compare Table 9.2:1 and 9.3:1 in Section 9) but not for each individual gene included in the panel and DNA alterations by gene plus putative TMB (putative refers in this case to the fact that we infer the mutational burden from a self-designed NSCLC gene panel following the description by Chalmers ZR et al. (8) instead of performing exome sequencing for TMB determination).

The modalities should contain variables with different prior information on the prognostic effect. Since, many clinical variables are known to be associated with the prognosis of patients they are in the first modality together with IHC because in particular PD-L1 is used as a predictive marker in NSCLC and the evaluated genotype which showed a potential signal

in a previous Nintedanib study (1199.243 substudy of the pivotal NSCLC study 1199.13). Since it is expected that there are many missing values for IHC the IPF Lasso will be applied twice. Once with and once without IHC, other variables with many missing variables (more or equal 50%) might be excluded, too. Note that a full design matrix is required for the penalized regression (IPF and standard Lasso); therefore certain BMs will be anyway excluded from all analyses if the number of missing is below a certain threshold (compare CTP Section 9.7).

The standard Lasso is carried out for the following sets of BMs separately

- DNA (Genomic characterization of tumour tissue)
- RNA (continuous pathways and cell types plus selected individual angiogenesis genes)
- DNA, RNA and clinical covariates
- IHC (Percentage of PD-L1 positive tumour and immune cells, PD-L1 H-Score and KI67)
- IHC, genotyping (Genomic evaluation of blood samples or buccal swab), DNA, RNA and clinical covariates

All variables can be found in Section 6.4. Continuous variables will be included as continuous except if a non-linear relationship might be present.

The BES will be used for all multivariate analyses.

The endpoint OS will be used for all multivariate analyses. The resulting selections of variables will be applied to PFS, too.

For all multivariate models the HR of all signatures will be evaluated at minimum, median and maximum of the signature.

Prognostic signatures derived by LASSO

In the following section a prognostic signature is defined as a linear combination of BMs which jointly have prognostic potential.

A prognostic signature based on a set of BMs $Z_{i1},\ldots, Z_{ip}$ is derived by fitting a LASSO (L1-penalized) regression model with penalty parameter optimized through 10-fold CV using the R function *cv.glmnet* from the package *glmnet*. The part of the resulting linear predictor including $Z_{i1},\ldots,Z_{ip}$ yields the prognostic signature.

The same procedure will be repeated for 100 random subsamples of size 0.5 × size of the original dataset, following the stability selection principles of Meinshausen and Bühlmann (2010) (R15-5998). The frequency of selection of each candidate BM over the 100 runs is computed. The average signature obtained by averaging the signatures fitted in the 100 iterations will be considered as an additional – possibly more stable – signature. This signature will consequently be used for PFS.

The only difference between IPF Lasso and standard Lasso is that the first one uses one penalty parameter for each modality. Therefore, two modality parameters need to be estimated.

Prognostic signatures derived by Random forest

A random forest is an ensemble of classification or regression trees. Predictions for new patients are based on majority voting for classification tasks or mean prediction in the regression framework. Implementation will be done e.g. by the R packages *party* or *partykit.* The random forest will be applied to only once to all variables together (IHC, genotyping (Genomic evaluation of blood samples or buccal swab), DNA, RNA and clinical covariates).

Association of BMs with dichotomized TSFL
The same multivariate procedure (IPF, standard Lasso and random forest) is repeated using TSFL as response and hence penalized logistic regressions will be used.
The resulting signature will be evaluated using a Cox regression model for OS and the regarding signature as covariate to perform a comparison with the dichotomized TSFL result and the other developed prognostic signatures.

Since the cutoff of 9mts might be no longer appropriate because of the changing treatment landscape the association with continuous time since will be performed as well and on the subgroup of patients with PD as best response to first line treatment.

Association of BMs with continuous TSFL
The same multivariate procedure (IPF, standard Lasso and random forest) is repeated using TSFL as response and hence a penalized linear regression will be used.
The resulting signature will be evaluated using a Cox regression model for OS and the regarding signature as covariate to perform a comparison with the dichotomized TSFL result and the other developed prognostic signatures.

Association of BMs with best response to first line therapy
The same multivariate procedure (IPF, standard Lasso and random forest) is repeated using Best response to first line therapy (PD vs no-PD) and hence penalized logistic regressions will be used.
The resulting signature will be evaluated using a Cox regression model for OS and the regarding signature as covariate to perform a comparison with the dichotomized TSFL result and the other developed prognostic signatures.

7.7.3 Global test (Goeman et al., 2006) (R18-3810)

The gene-set test will be used to assess if predefined groups of covariates, e.g. the set of genes as defined in Section 9.1 for the RNA pathways, is associated with a clinical endpoint. The null hypothesis of the global test is that none of the covariates in the tested group is associated with the clinical endpoint. The alternative is that at least one of the covariates has such an association.

The global test is based on regression models in which the distribution of the clinical endpoint variable is modeled as a function of the covariates. The type of regression model depends on the clinical endpoint, i.e. linear regression models are used for continuous endpoints and logistic regression models for binary endpoints, respectively. The regression modeling

approach allows to adjust the test for the confounding effect of nuisance covariates: covariates that are known to have an effect on the clinical endpoint and which are correlated with (some of) the covariates of interest.

The R package *globaltest* will be used for these analyses.

The gene set test will be first used to associate the pathways with OS and furthermore to associate the pathways with the prognostic factor TSFL.
All analyses will be based on the BES.

7.7.4 Subgroup analysis

Subgroup analysis using the clinical pre-defined subgroups defined in Section 6.4 will be performed for the OS and PFS analyses defined above (Section 7.8.1, 7.8.2, and 7.8.3). Subgroup analysis will only be performed, if at least 10 subjects per category (subgroup) and per treatment have valid results. Subgroup analyses will be performed for all patients in BES.

7.7.5 Cut-points

For gene expression data three groups based on the 33 and 66 percentile will be used as cut-point. However, if necessary (eg. because of small sample size; less than 10 subjects per category) an additional cut-point based on the median with only two groups will be used. This logic will be applied not only to gene expression data.

7.7.6 Correlation coefficients between TSFL and gene-expression

Spearman correlation coefficients between TSFL and gene-expression will be calculated (see Section 9.7.3 in the CTP).

7.7.7 Determination of False Discovery Rate corrected p-values

For a set $p_1, \cdots, p_m$ of m p-values that correspond to testing m null hypotheses, FDR corrected p-values are derived in a stepwise procedure.
The p-values are sorted in increasing order. The sorted p-values are denoted by

$$p_{(1)} \leq \cdots \leq p_{(m)}.$$

The FDR corrected p-values are defined as

$$p^*_{(i)} = \begin{cases} p_{(m)} & i = m \\ min\left(\frac{m}{i} \cdot p_{(i)}; p^*_{(i+1)}\right) & i = m-1,..,1 \end{cases}$$

Refer to Benjamini and Hochberg (P05-11198) for more details. The FDR correction will be carried out using the SAS procedure PROC MULTTEST.
In this exploratory analysis, p-values will be grouped into subsets, and FDR correction will be applied to each subset separately. The grouping of the p-values into subsets will be done in the following hierarchical order.
1) Group p-values according to the investigated efficacy endpoint.

2) Group p-values by the source/analysis method of the data (e.g. DNA, RNA, IHC, Genotyping and OS, PFS).

7.7.8 Extent of exposure

All analyses of exposure are performed for the TS including exposure duration of nintedanib and docetaxel.

Treatment exposure will be displayed descriptively as number of days on treatment (continuous and categorical by weeks 1, 2, …).

Extent of exposure is defined as the time from first administration of Vargatef® until the last administration of Vargatef® or date of DBL if patient is still ongoing at that time.
Time to first dose reduction of Vargatef® is defined as the time from first administration of Vargatef® to the first administration of the first reduced dose of Vargatef®.

7.8 SAFETY ANALYSIS

All safety analyses will be performed on the TS.

7.8.1 Adverse events /adverse drug reactions

The analyses of AEs will be descriptive in nature. All analyses will be based on the number of patients with AEs (not the number of AEs) in the TS.
AEs will be coded with the most recent version of Medical Dictionary for Regulatory Activities (MedDRA). The analyses of AEs will be descriptive in nature. Reporting of adverse events (AEs) in this study will focus on serious and non-serious adverse drug reactions (ADRs) to Vargatef® and fatal AEs. All analyses of ADRs and fatal AEs will be based on the number of subjects with ADRs and fatal AEs and not on the number of ADRs and fatal AEs.

For analysis, multiple AE occurrence data on the eCRF will be collapsed into one event provided that all of the following applies:

- All AE attributes are identical (lower level term, intensity, action taken, therapy required, seriousness, reason for seriousness, relationship, outcome)
- The occurrences were time-overlapping or time-adjacent (time-adjacency of two occurrences is given if the second occurrence started on the same day or on the day after the end of the first occurence)

For further details on summarization of AE data, please refer to (6,9).
The analysis of AEs will be based on the concept of treatment emergent AEs. That means that all AEs will be assigned to the treatment as defined in Section 6.1. An overall summary of AEs will be presented. This overall summary will comprise summary statistics for the class of other significant AEs according to ICH E3 (10**Error! Reference source not found.Error! Reference source not found.**).

The analyses of AEs will distinguish between treatment-emergent AEs and post treatment emergent AEs, where a treatment emergent AE has an onset in the analysing treatment period. The main AE analysis will be based on the **"on-treatment period"**, which starts with the date of first administration of study medication (Vargatef®) and ends 30 days after the last administration of study medication (Vargatef®) but post treatment emergent AEs will be analysed as well.

According to ICH E3 (**Error! Reference source not found.**), AEs classified as "other significant" needs to be reported and will include those non-serious AEs with "action taken=discontinued" or "action taken= reduced".

Listings of AEs will be displayed by patients. The actual dose of Vargatef® administered at the day of AE onset will be derived and included in the listings.

An overall summary of ADRs/AEs by severity, seriousness and outcome will be tabulated for all three periods separately: on-treatment period, entire study period and post-treatment period.

The frequency of patients with ADRs/AEs will be summarised by primary system organ class (SOC) and PT (mention MedDRA levels to be displayed in the tables) by severity for all three periods. The same will be done for serious ADRs/AEs and fatal AEs.

Frequency of patients with ADRs/AEs leading to dose reduction and discontinuation of Vargatef will be displayed.

Frequency of patients with ADRs/AEs by PT and severity will be summarised for all three periods.

Patient with drug exposure during pregnancy, abuse, off-label use, misuse, medication error, occupational exposure, lack of effect, and unexpected benefit will be listed.

The SOCs will be sorted alphabetically, PTs will be sorted by frequency (within SOC).

7.8.2 Laboratory data

Not applicable.

7.8.3 Vital signs

Not applicable.

7.8.4 ECG

Not applicable.

7.8.5 Others

Not applicable.

7.8.6 Interim analyses

Not applicable.

8. REFERENCES

1	*CPMP/ICH/363/96: "Statistical Principles for Clinical Trials", ICH Guideline Topic E9,Note For Guidance on Statistical Principles for Clinical Trials, current version.*
2	*ISPE Guidelines of Good Pharmacoepidemiology Practice, Pharmacoepidemiology and Drug Safety 2008; 17: 200–208*
3	Bringing the next Generation of Immuno-Oncology Biomarkers to the Clinic Alessandra Cesano * and Sarah Warren (https://www.nanostring.com/download_file/view/1737/3718)
4	*001-MCS-40-413*: "Identify and Manage Important Protocol Deviations (iPD)", current version; group: Clinical Operation; IDEA for CON.
5	Lan et al.: Association between PD-L1 expression and driver gene status in non-small-cell lung cancer: a meta-analysis, Oncotarget. 2018 Jan 26; 9(7): 7684–7699. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5800936/
6	*001-MCG-156_RD-01:* "Handling of missing and incomplete AE dates", current version; IDEA for CON.
7	Boulesteix, De Bin, Jiang, Fuchs (2017) IPF-LASSO: Integrative *LL*1-Penalized Regression with Penalty Factors for Prediction Based on Multi-Omics Data. Comp. and Math. Meth. in Medicine
8	Chalmers et al.: Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden, Genome Med. 2017; 9: 34. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5395719/
9	*001-MCG-156*: "Analysis and Presentation of adverse event data from clinical trials ", current version; IDEA for CON.
10	*CPMP/ICH/137/95: "Structure and Content of Clinical Study Reports", ICH Guideline Topic E3; Note For Guidance on Structure and Content of Clinical Study Reports, current version*
11	nCounter Advanced Analysis 2.0 Plugin for nSolver Software User Manual (https://www.nanostring.com/download_file/view/1169/3842)
R15-5998	*Meinshausen N, Bühlmann P. Stability selection. J R Stat Soc Ser B Stat Methodol 72 (4), 417 - 473 (2010)*
R18-3810	*Goeman JJ, van de Geer SA and van Houwelingen JC. Testing against a high-dimensional alternative. Journal of the Royal Statistical Society Series B Statistical Methodology, 68(3): 477–493(2006).*
P05-11198	*Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc (B) 57 (1), 289 -300 (1995)*

10. HISTORY TABLE

Table 10: 1 History table

Version	Date (DD-MMM-YY)	Author	Sections changed	Brief description of change
Final	**28-Mar-2019**		None	This is the final SEAP without any modification

APPROVAL / SIGNATURE PAGE

Document Number: c09070203 **Technical Version Number:3.0**

Document Name: 8-01-seap

Title: Statstical and epidemiological analysis plan

Signatures (obtained electronically)

Meaning of Signature	Signed by	Date Signed
Author-Statistician		28 Mar 2019 15:01 CET
Approval-Team Member Medicine		28 Mar 2019 15:10 CET
Author-Pharmacogenomics		29 Mar 2019 13:36 CET
Approval-Scientific Monitor		05 Apr 2019 12:45 CEST

(Continued) Signatures (obtained electronically)

Meaning of Signature	Signed by	Date Signed