[redacted]

Boehringer Ingelheim

NON-ELECTRONIC SIGNATURE STATEMENT

The signatories with no access to the Electronic Document Management System (EDMS) sign the Statistical Analysis Plan

BI Trial No.:	1237-0087
Title:	EVELUT®: Assessment of dyspnea and other symptoms as patient reported outcomes (PRO) in patients with chronic obstructive pulmonary disease (COPD), symptomatic on LABA/ICS maintenance therapy (now) treated with Spiolto® Respimat® (tiotropium/olodaterol) in comparison to open or fixed triple combination treatment in routine clinical practice.
Date of statistical analysis plan:	11-AUG-2020
Version:	FINAL 1.0

Review and approval Clinical Monitor	Date, Signature 17-Aug-2020 DocuSigned by: [redacted]
Review and approval Statistician	Date, Signature 18-Aug-2020 DocuSigned by: [redacted]
Review and approval Oversight Statistician (BI)	Date, Signature 17-Aug-2020 DocuSigned by: [redacted]

[redacted]

Trial Statistical Analysis Plan

BI Trial No.:	1237-0087
Title:	EVELUT®: Assessment of dyspnea and other symptoms as patient reported outcomes (PRO) in patients with chronic obstructive pulmonary disease (COPD), symptomatic on LABA/ICS maintenance therapy (now) treated with Spiolto® Respimat® (tiotropium/olodaterol) in comparison to open or fixed triple combination treatment in routine clinical practice.
Investigational Product(s):	Spiolto® Respimat® 2.5 microgram/2.5 microgram per puff inhalation solution
Responsible trial statistician(s):	[redacted] Phone: [redacted] Fax: [redacted]
Date of statistical analysis plan:	11 August 2020
Version:	Final 1.0

1. TABLE OF CONTENTS

LIST OF TABLES

2. LIST OF ABBREVIATIONS

Term	Definition / description
ADR	Adverse Drug Reaction
AE	Adverse event
ATT	Average treatment effect in the treated
BI	Boehringer Ingelheim Pharma GmbH & Co. KG
CAT™	COPD Assessment Test
CDF	Cumulative distribution function
CI	Confidence Interval
COPD	Chronic obstructive pulmonary disease
CTP	Clinical trial protocol
FEV1	Forced expiratory volume in one second
GERD	Gastroesophageal reflux
GP	General Practitioner
GOLD	Global Initiative for Chronic Obstructive Lung Disease
ICH	International Conference on Harmonisation
ICS	Inhaled corticosteroids
IPTW	Inverse probability of treatment weighting
IRB	Institutional Review Board
LABA	Long-acting beta2-adrenoceptor agonist
LAMA	Long-acting anticholinergic bronchodilator
Max	Maximum
MedDRA	Medical Dictionary for Drug Regulatory Activities
Min	Minimum
mMRC	Modified Medical Research Council Scale
MS	Matched set
NCI	National Cancer Institute
NIS	Non-interventional study
PGE	Physician’s global Evaluation
PP	Per-protocol set
PS	Propensity score
PV	Protocol violation

Term	Definition / description
SAS	Statistical analysis software
SmPC	Summary of Product Characteristics
TEAE	Treatment-emergent AE
TS	Treated set
TSAP	Trial Statistical Analysis Plan

3. INTRODUCTION

As per ICH E9[1], 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 TSAP assumes familiarity with the Clinical Trial Protocol (CTP), including Protocol Amendments. In particular, the TSAP is based on the planned analysis specification as written in CTP Section 9.7 “Data Analysis”. Therefore, TSAP 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 higher will be used for all analyses.

4. CHANGES IN THE PLANNED ANALYSIS OF THE STUDY

None.

5. ENDPOINT(S)

5.1 PRIMARY ENDPOINT(S)

Primary outcomes are:

- Difference between mMRC score at baseline (visit 1) and mMRC score after end of observation (ca. 12 weeks of treatment, Visit 2)
- Difference between CAT™ score at baseline (visit 1) and CAT™ score after end after end of observation (ca. 12 weeks of treatment, Visit 2)
- For descriptive analysis the CAT™ will additionaly be stratified in low (0-10 points), medium (11-20 points), high (21-30 points) and very high (31-40 points)

The primary outcomes are reported in a descriptive manner, will be compared in an explorative manner and are not safety outcomes.

5.2 SECONDARY ENDPOINT(S)

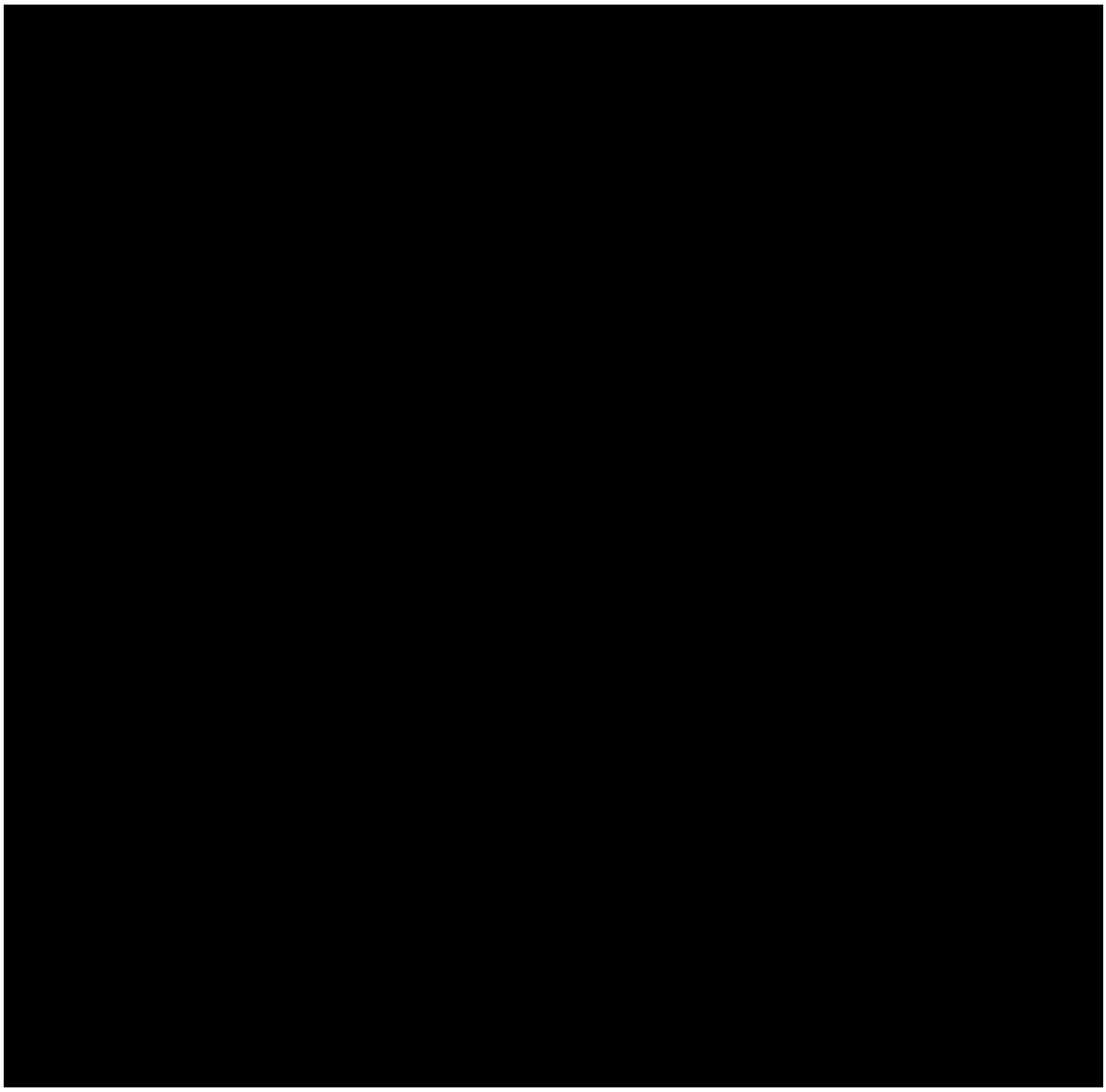
Secondary endpoints to be determined and compared in an exploratory manner include:

- Patients' general condition according to the Physician's Global Evaluation (PGE) score at baseline and end of the observation period
- Patient satisfaction with inhaler and therapy at end of observation period according to a seven-point ordinal scale (ranging from very dissatisfied to very satisfied as documented in non-interventional BI studies (BI 1237-0042, 1237-0043, 1237-0044, 1237-0045, 1237-0065, 1237-0072)
- Proportion of responders with $\Delta_{mMRC} \geq 1$ and the proportion of responders with $\Delta_{CAT} \geq 2$

The secondary outcomes are not safety outcomes.

Comparative analyses on secondary endpoints will be performed as for the main analysis but only using the primary analytical approach (propensity score matching).

[redacted]



6. GENERAL ANALYSIS DEFINITIONS

6.1 TREATMENT(S)

In the presented study, treatment with Spiolto® Respimat® will be according to product information. Patients will be treated with Spiolto® Respimat® or triple therapy and be observed for approximately 12 weeks.

Other LAMA, LABA and/or ICS therapy will be performed according to label as described in the respective SmPCs.

6.2 IMPORTANT PROTOCOL DEVIATIONS

Table 6.2: 1 defines the different categories of important protocol violations (PVs). The final column describes which PVs will be used to exclude subjects from the different patient analysis sets[2].

Table 6.2: 1 Important protocol violations

Category/Code		Description	Requirements	Excluded from
A		**Entrance criteria not met**		
	A1.1	Inclusion criterion 1 (Diagnosis of COPD)	Not met as specified in the protocol	PP, MS
	A1.2	Inclusion criterion 2 (Symptomatic (with regard to dyspnea (mMRC Dyspnea score ≥1) AND with regard to symptoms (CAT Score ≥10) at the same time))	Not met as specified in the protocol	PP, MS
	A1.3	Inclusion criterion 3 (Patients on LABA/ICS maintenance therapy who are switched to Spiolto® Respimat® in the new reusable inhaler or a free/fixed triple combination of LABA + LAMA + ICS at Visit 1 at the discretion of the treating physician.)	Not met as specified in the protocol	PP, MS
	A1.4	Inclusion criterion 4 (Adults who are contractually capable and mentally able to understand and follow the instructions of the study personnel)	Not met as specified in the protocol	PP, MS
	A1.5	Inclusion criterion 5 (Male or female aged ≥40 years of age)	Not met as specified in the protocol	PP, MS
	A1.6	Inclusion criterion 6 (Written informed consent prior to study participation)	Not met as specified in the protocol	PP, MS
	A1.7	Inclusion criterion 7 (The patient is willing and able to follow the procedures outlined in the protocol)	Not met as specified in the protocol	PP, MS
	A2.1	Exclusion criterion 1 (Patients with contraindications acc. to SmPC)	Met as specified in the protocol	PP, MS
	A2.2	Exclusion criterion 2 (Patients not on LABA/ICS maintenance treatment at visit 1, e.g., mono or dual bronchodilation only, ICS only, or a triple combination of LAMA + LABA + ICS (either as a	Met as specified in the protocol	PP, MS

Category/Code		Description	Requirements	Excluded from
		fixed combination product or as separate components))		
	A2.3	Exclusion criterion 4 (Pregnant and/or lactating females)	Met as specified in the protocol	PP, MS
	A2.4	Exclusion criterion 5 (Acute exacerbation of COPD (within 4 weeks prior to Visit 1))	Met as specified in the protocol	PP, MS
	A2.5	Exclusion criterion 6 (Frequently exacerbating COPD patients, i. e. patients with ≥2 moderate exacerbations within the last 12 months or ≥1 exacerbation leading to hospitalization within the last 12 months[1])	Met as specified in the protocol	PP, MS
	A2.6	Exclusion criterion 7 (Acute respiratory failure (pH <7,35 and/ or respiratory rate >30/min within 3 months prior to Visit 1))	Met as specified in the protocol	PP, MS
	A2.7	Exclusion criterion 8 (History or current diagnosis of asthma)	Met as specified in the protocol	PP, MS
	A2.8	Exclusion criterion 9 (History or current diagnosis of asthma-COPD overlap)	Met as specified in the protocol	PP, MS
	A2.9	Exclusion criterion 10 (History or current diagnosis of allergic rhinitis within the last 5 years)	Met as specified in the protocol	PP, MS
	A2.10	Exclusion criterion 11 (History or current diagnosis of lung cancer within the last 5 years)	Met as specified in the protocol	PP, MS
	A2.11	Exclusion criterion 12 (Participation in a parallel interventional clinical trial)	Met as specified in the protocol	PP, MS
B		**Informed consent**		
	B1	Informed consent not available/not done (Inclusion criterion 1/Exclusion criterion 3)	IC 01 not met as specified in the protocol or informed consent date missing	All

[1] Definitions of exacerbation severity according to German Pulmonary Society Guideline (DGP-Leitlinie 2019).

6.3 PATIENT SETS ANALYZED

Treated set (TS): All patients with informed consent who have received at least one dose of Spiolto® Respimat® or triple combination (free or fixed LAMA + LABA+ ICS).

Per-protocol set (PP): All patients with informed consent who met all inclusion criteria as well as no exclusion criteria and have received at least one dose of Spiolto® Respimat® or triple combination (free or fixed LAMA + LABA+ ICS).

Matched set (MS): Subset of PP excluding all patients who could not be matched in the propensity score matching.

Table 6.3: 1 Patient sets analyzed

	Patient set		
Class of endpoint	TS	PP	MS
Primary	X	X	X
Secondary endpoints	X		X
Safety endpoints	X		
Demographic/baseline	X		X

If the number of patients excluded from the PP is small (<10% of the TS), then analyses on this set may not be conducted.

[redacted]

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

Multiple imputation will be used in the estimation of propensity scores and in multivariable regression analyses. In descriptive analyses, the fraction of missing observations will be reported. Every effort will be made to collect complete data at the specified time points. Any removal from the analysis will be documented, stating the site and patient number as well as the reason for removal.

6.7 BASELINE, TIME WINDOWS AND CALCULATED VISITS

Baseline visit (Visit 1) will include historical and demographic data as well as registration and initial examination.

Treatment with Spiolto® Respimat® will be documented at visit 1 and visit 2, after approximately 12 weeks of treatment.

The mMRC breathlessness scale as well as the CAT™ questionnaire (health and functional status) is completed by the patient at visit 1 and (after approximately 12 weeks of treatment) visit 2.

In addition, the PGE is completed by the physician at visit 1 and visit 2 as well as satisfaction survey at visit 1 (regarding the preceding therapy) and visit 2 (regarding the new therapy).

7. PLANNED ANALYSIS

The analyses of the primary and secondary endpoints will be performed on the matched set (MS) using the primary analytical approach (propensity score matching).

The primary will further be analysed using propensity score weighting and multivariable regression modeling. These analyses will be performed on the treated set (as-treated analysis) and, additionally, on the per-protocol set (PP).

Additionally, the primary and secondary endpoints will be analysed on the treated set (TS) based on data as collected (i.e. without applying any analytical approach).

The propensity score will be calculated based on a logistic regression model taking the choice of treatment as dependent variable and baseline characteristics as independent variables into account. Further details are given in section 7.3.

7.1 DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS

Only descriptive statistics are planned for this section of the report. All baseline analyses will be done for the treated set and the matched set.

For categorical variables summary tabulations of the number and percentage within each category (with a category for missing data) of the parameter will be presented. For continuous variables number of values, mean, standard deviation, minimum, 25th percentile, median, 75th percentile, maximum and number of missing values will be presented. Incidence rates and 95% CI will be given when appropriate. Known right-skewed variables will be log-transformed for computation of means.

7.2 CONCOMITANT DISEASES AND MEDICATION

Only descriptive statistics are planned for this section of the report. All analyses will be done for the treated set.

7.3 PRIMARY ENDPOINTS

The primary outcomes are reported in a descriptive manner, will be compared in an explorative manner and are not safety outcomes. For the three primary outcomes the comparative treatment effect will be estimated along with 95% confidence intervals (CI).

For absolute changes in total mMRC and CAT^{TM} score between baseline and end of observation, summary statistics will be provided. Furthermore linear regression will be used to analyse the change in mMRC and CAT^{TM} score. The model will adjust for clustering (matched patients) using generalized estimating equations (GEEs) and take the respective baseline variable as covariable into account.

In an additional analysis it will be investigated if the therapy effect is heterogeneous with respect to the GOLD spirometric classification and the GOLD patient group, respectively. The stratified estimators with their corresponding 95% confidence interval will be shown. If

appropriate, a p-value of the interaction of treatment and GOLD may be provided. If some of the strata are too small they will be omitted from the analysis.**mMRC and CAT™ score**

The mMRC Dyspnea Scale consists only of one question on symptom severity and yields a score of either 0, 1, 2, 3 or 4.

The total CAT™ score ranges from 0 to 40. A higher score is indicative of worse status. According to the manual missing values will not be imputed, no total score will be calculated if at least one item is missing. However a single imputation will be done as sensitivity analysis, by imputing the response for the missing question as mean of the other questions of the respective patient.
The total CAT™ score and the corresponding subscales are calculated as follows, in case that answers to all 8 items are available:

$$\text{Total CAT}^{\text{TM}}\ \text{score} = \sum_{i=1}^{8} \text{item}_i$$

The estimation of the treatment effect is subject to potential confounding. Therefore, beside the analysis based on data as collected adjusted analyses are required. Multiple analytical approaches will be applied to allow an assessment of the sensitivity of the results to these approaches:

- Propensity score matching (primary analysis; requires data to be discarded from the analysis)
- Propensity score weighting (uses the complete data set)
- Multivariable regression modeling (uses the complete data set)

Ten multiple imputations will be applied to estimate the propensity score model and the multivariable regression models. The average of the propensity scores will be used to analyze the treatment effect.

Propensity score matching:

Estimation of propensity scores will be realized via multivariable logistic regression, modelling treatment dependent on several matching variables.

Variables used for propensity score estimation are:

- age at registration [continuous variable]
- age at registration squared [continuous variable]
- sex [male / female]
- baseline CAT™ [continuous variable]
- baseline mMRC [mMRC-grade 0 / mMRC-grade I / mMRC-grade II / mMRC-grade III / mMRC-grade IV]
- exacerbation history [yes: at least 1 exacerbation requiring hospitalization or at least 2 moderate exacerbations / no: no exacerbations or only mild exacerbations]

- pack-years of smoking [continuous variable]
- FEV1 [no finding / < 30% / 30% - 49% / 50% - 79% / ≥ 80%]
- Logarithm of eosinophil levels [continuous variable]
- Specialization of attending physician [GP/ pulmonologist/ internal specialist]

If the propensity score model can be fitted without convergence problems, no further variable selection will be performed. In case of convergence problems, variables causing failed convergence will be eliminated from the model.

Density curves of the propensity score distribution stratified by treatment group will be shown to assess overlap.

Matching based on the propensity score is realized using a (greedy) nearest-neighbor matching on the logit of the propensity score with a caliper of 0.2 standard deviations of the logit of the PS.

Matching of patients treated with Spiolto® Respimat® to patients treated with triple combination (free or fixed LAMA + LABA + ICS) will be implemented using PROC PSMATCH with a random order of treated patients to be matched. It will be started with a 1:1 matching, but if the balance can be kept a 1_{Triple}:$2_{Spiolto}$ or even 1_{Triple}:$3_{Spiolto}$ matching will be performed.

Balance of propensity score matched samples will be assessed using standardized differences. They will be applied to compare the means and occurrence rates of continuous and dichotomous variables as well as the mean of squares of continuous varibales, i.e. the varicance between the treatment groups, respectively. Standardized differences are not influenced by sample size and allow for the comparison of the relative balance of variables measured in different units. They can be derived for continuous and binary variables. Although there is no universally agreed upon criterion as to what threshold of the standardized differences can be used to indicate important imbalance, a standard difference that is less than 0.1 will be taken to indicate a negligible difference in the mean or prevalence of a covariate between treatment groups.
Furthermore side-by-side boxplots and empirical cumulative distribution functions (CDFs) will be used to compare the distribution of continuous baseline covariates between treated and control subjects in the weighted sample. These methods allow to assess whether the variability of a continuous baseline variable differs betweenthe treatment groups and whether the tails of the distribution of the variable differ between the treatment groups.

<u>Propensity score weighting (IPTW):</u>

Patients will be weighted for the analysis of the average treatment effect in the treated (subjects using the new reusable Respimat® inhaler) using inverse probabaility of treatment weighting defined as

$$w_{ATT}=Z+e(1-Z)/(1-e)$$

where Z denotes the treatment assignment, i.e. Z=1 if using the new reusable Respimat® inhaler and Z=0 if using triple combination,

e denotes the propensity score.

Hence the treated subjects receive a weight of one and the sample "triple combination" is used as reference population to which the treated and the "untreated" (subjects using the new reusable Respimat® inhaler) are standardized.

Multivariable regression modeling:

In context of these multivariable logistic or linear regression analyses, first univariate logistic or linear regressions will be performed. Afterwards, all covariates which were significant at the p-level of 0.2 in the univariate analysis will be entered into a backward selection multivariable logistic or linear regression . While the treatment variable will be kept in the model, backwards elemination will be used to remove covariates until p-values of all remaining covariates are <0.05 (based on Type III analysis of effects).
The analysis will be done for all ten datasets resulting from multiple imputation, respectively. However the results will be shown in an aggregated way.

The variables used for regression are the same as for propensity score matching:

- age at registration [continuous variable]
- age at registration squared [continuous variable]
- sex [male / female]
- baseline CAT™ [continuous variable]
- baseline mMRC [mMRC-grade 0 / mMRC-grade I / mMRC-grade II / mMRC-grade III / mMRC-grade IV]
- exacerbation history [yes: at least 1 exacerbation requiring hospitalization or at least 2 moderate exacerbations / no: no exacerbations or only mild exacerbations]
- pack-years of smoking [continuous variable]
- FEV1 [no finding / < 30% / 30% - 49% / 50% - 79% / ≥ 80%]
- Logarithm of eosinophil levels [continuous variable]
- Specialization of attending physician [GP/ pulmonologist/ internal specialist]

For each categorical variable the class with the highest frequency will be chosen as reference value, respectively.

7.4 SECONDARY ENDPOINT(S)

Comparative analyses on secondary endpoints will be performed as for the main analysis but only the primary analytical approach (propensity score matching) will be employed. All secondary endpoint analyses will be done on the matched set. Additionally, the analysis based on data as collected (i.e. without applying any analytical approach) will be done using the treated set.

For general condition of patients according to the Physician's Global Evaluation (PGE) and patient's satisfaction with Spiolto® Respimat® as well as the proportion of responders ($\Delta_{mMRC} \geq 1$, $\Delta_{CAT} \geq 2$ respectively), the number and percentage of patients within each category will be displayed.

[redacted]

7.6 EXTENT OF EXPOSURE

Not applicable.

7.7 SAFETY ANALYSIS

The analysis of adverse events will be descriptive and conducted according to Boehringer Ingelheim standards. The main focus will be on treatment emergent events, i.e. events occurring after start of treatment (events after end of treatment are not documented). Non treatment-emergent events will be assigned to "screening" and only be displayed in listings. All analyses will be based on the treated set.

The frequency and severity of drug-related adverse events or events with fatal outcome (i.e., subset of serious adverse events) will be tabulated according to MedDRA-SOC and PT. Additionally adverse events (drug-related or serious ones) leading to treatment discontinuation will be presented. Moreover, the causality of events with fatal outcome will be displayed.

Unless otherwise specified, the analyses of drug-related adverse events and events with fatal outcome will be descriptive in nature. All analyses will be based on the number of patients with AEs and not on the number of events.

An overall summary of drug-related adverse events and events with fatal outcome will be presented.

The frequency of patients with drug-related adverse events or events with fatal outcome will be summarised by primary system organ class and preferred term. Separate tables will be provided for patients with serious adverse events.

The system organ classes will be sorted alphabetically, preferred terms will be sorted by frequency (within system organ class).

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

- All AE attributes are identical (PT, NCI-CTC grade, action taken, therapy required, seriousness, reason for seriousness, relationship, outcome**)**

- The occurrences were time-overlapping or time-adjacent (time-adjacency of 2 occurrences is given if the second occurrence started on the same day or on the day after the end of the first occurrence). For classification into TEAE or Non-TEAE, the first documented start of event will be used.

For further details on summarization of AE data, please refer to [3] and [4].

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	*001-MCS-50-413*: "Handling of Protocol Violations in Clinical Trials and Projects", current version; group: Study Conduct; IDEA for CON.
3	*001-MCG-156_RD-01:* "Handling of missing and incomplete AE dates", current version; IDEA for CON. *001-MCG-156_RD-01:* "Handling of missing and incomplete AE dates", version 5; IDEA for CON.
4	*001-MCG-156*: "Handling and summarization of adverse event data for clinical trial reports and integrated summaries", current version; IDEA for CON.

[redacted]

10. HISTORY TABLE

Table 10: 1 History table

Version	Date (DD-MMM-YY)	Author	Sections changed	Brief description of change
Draft v0.1	**12-JUN-2019**	[redacted]	None	This is the first Draft-Version of TSAP without any modification
Draft v0.2	**09-SEP-2019**		2	Abbreviations completed
			5.1	Categorization of CAT™ added
			[redacted]	
			7.3	Revised
			7.4	Change in wording to explain issue more clearly
			[redacted]	
			7.7	Definition of TEAE added
Draft v0.3	**18-SEP-2019**	[redacted]	7.3	Revised
Draft v0.4	**08-OCT-2019**		7.3	Clarification of analysis for GOLD groups
Draft v0.5	**19-FEB-2020**		7.3	Number of primary outcomes corrected to three Clarification of multiple regression
Draft v0.6	**26-MAY-2020**		6.2	Header changed: deviation instead of violation
			6.3	Analysis of secondary endpoints for PP deleted Statement when PP analysis can be skipped added
			[redacted]	
			7.	Analysis of primary and secondary endpoints adapted
			7.3	p-values deleted Analysis on "data as collected" added Multivariable regression modeling adapted
			7.4	Analysis sets for secondary endpoints adapted Subgroup analyses deleted
Draft v0.7	**17-JUL-2020**	[redacted]	7.3	Definition of exacerbation history added