NCT03282357, redacted
version v1.0, 17Jun2019

Merz Pharmaceuticals GmbH

Statistical Analysis Plan

Evaluation of the Effectiveness and Safety of Radiesse for the Correction of Nasolabial Folds

Device Pre-market in China

MRZ M900311004

Version 1.0

Date: 17-Sep-2018

Author: [redacted]

SIGNATURE PAGE

I confirm that this Statistical Analysis Plan accurately describes the planned statistical analyses to the best of my knowledge and was finalized before breaking the blind/database close.

[redacted]	18-Sep-2018	[redacted]
Author (print name)	Date (dd-mmm-yyyy)	S
[redacted]	17-SEP-2018	[redacted]
Peer Reviewer (print name)	Date (dd-mmm-yyyy)	S
[redacted]	17-SEP-2018	[redacted]
Study Medical Expert Merz (print name)	Date (dd-mmm-yyyy)	Sig
(print name)	Date (dd-mmm-yyyy)	Signature

TABLE OF CONTENTS

1 LIST OF ABBREVIATIONS

ADE	Adverse Device Effect
AE	Adverse Event
ATC	Anatomical Therapeutic Chemical
BDRM	Blind Data Review Meeting
BMI	Body Mass Index
CI	Confidence Interval
CSR	Clinical Study Report
eCRF	electronic Case Report Form
FAS	Full Analysis Set
GAIS	Global Aesthetic Improvement Scale
MedDRA	Medical Dictionary for Regulatory Activities
NLFs	Nasolabial Folds
PPS	Per Protocol Set
RS	Randomized Set
SADE	Serious Adverse Device Effect
SAE	Serious Adverse Event
SAP	Statistical Analysis Plans
SES	Safety Evaluation Set
TEAE	Treatment Emergent Adverse Event
WHODD	World Health Organization Drug Dictionary
WSRS	Wrinkle Severity Rating Scale

2 GENERAL AND TECHNICAL ASPECTS

The objective of this statistical analysis plan is to specify the statistical analyses in more detail than stated in the clinical study protocol and to be precise enough to serve as a guideline for statistical programming and creation of tables, figures and listings.

This statistical analysis plan is based on the clinical study protocol version 1.2, dated Oct 16th, 2017.

All programs will be written using SAS version 9.4. It will be assured that all programs run with SAS version 9.2, too, without errors or warnings. A preferred font size of 10 points, minimum font size of 8 points with a unique font size for the whole document required will be used for the tables and figures in section 14, corresponding to a linesize of 111 (8 points: 140) digits and a pagesize of 42 (8 points: 52) lines for an output in A4 landscape format. For listings, a standard font size of 10 points with the linesize and pagesize as defined above will be used to produce the output in A4 format. Single SAS programs will be written for all tables and figures, and all listings, respectively. All outputs will be transferred into PDF-files using the Merz internal SAS macro LST2PDF. These PDF-files will be generated separately for the tables and figures of section 14 and the listings of section 16.2 of the appendix of the clinical study report (CSR). Each PDF-file will include the corresponding table of contents, preceding the content of the file.

The standard TFLs v.6.0 will be applied.

3 CLINICAL STUDY DESIGN AND OBJECTIVES

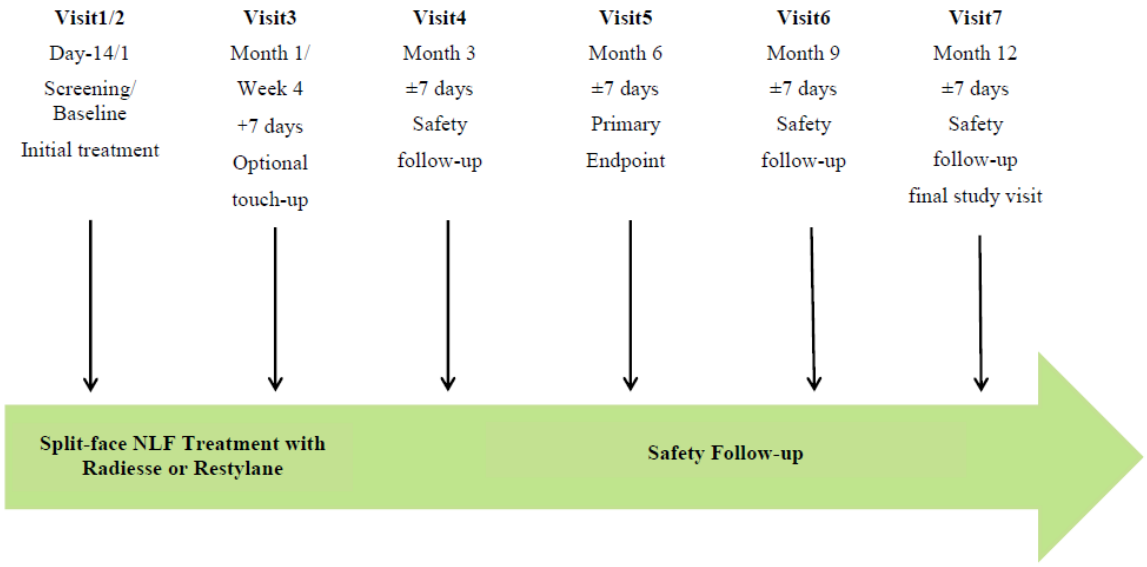
3.1 Clinical Study Design

This is a 12-month, multicenter, non-inferiority, split-face, active comparator, randomized, blinded (evaluator and subject) study evaluating the effectiveness and safety of Radiesse versus Restylane for the correction of nasolabial folds (NLFs) in adult subjects.
Approximately 174 Chinese subjects will be enrolled into this study. Subjects will be enrolled at approximately 4 investigational sites in China.

During the baseline visit for each subject, one NLF will be randomized to be treated with Radiesse and the other with Restylane. Subjects will have the option of a touch-up injection in one or both NLFs 4 weeks after the initial injection with the same randomized treatment.

The end of the study is the date of the last visit of the last subject.

Figure 1: Study Flow Chart



3.2 Clinical Study Objectives

The objectives of this study are to evaluate the effectiveness and safety of Radiesse for the correction of NLFs.

4 DETERMINATION OF SAMPLE SIZE

The estimate of the sample size is based on the following assumptions:

- Estimate of treatment effects (i.e. proportion of responders) on Test (Radiesse) and Control (Restylane) sides are at least equal, i.e. $P_{test} >= P_{control}$ (or $D = P_{test} - P_{control} >= 0$).
- Non-inferiority margin, $\Delta = 10\%$ (i.e., maximum negligible difference between the test and control proportions).
- Type I error, $\alpha = 0.025$.
- Type II error $\beta = 0.2$ (i.e. statistical power, $1-\beta = 0.8$ or 80%).
- Proportion of pairs with discordant responses $\eta = 0.175$ (i.e. 17.5%).

Using the above assumptions for sample size calculation in nQuery yields 138 effectiveness evaluable subjects needed to carry out the non-inferiority hypothesis associated with the primary effectiveness endpoint. Accounting for a 20% attrition (attributable to dropout of subject and major protocol deviations), overall 20% more subjects will be enrolled resulting in a total of approximately 174 subjects that will be randomized into this trial.

Since this is a multi-center, competitive enrollment study, the target number of subjects to be enrolled per investigational site is 44. In order to ensure that adequate subjects are enrolled per investigational site for determination of treatment effect homogeneity (or lack thereof), a minimum of approximately 22 and maximum of approximately 66 subjects will be enrolled

per investigational site. The specified proposed minimum and maximum number of subjects to be enrolled per investigational site minimizes the probability of any one site having a preponderance of treatment effect.

5 ANALYSIS SETS

The following analysis sets will be defined for the statistical analysis of this clinical study:

Randomized Set (RS)

The randomized set is defined as all subjects randomized into the study.

Safety Evaluation Set (SES)

The Safety Evaluation Set (SES) is defined as all subjects who receive an investigational product.

Full Analysis Set (FAS)

The Full Analysis Set (FAS) is defined as all subjects randomized and who receive an investigational product with at least one post-baseline effectiveness assessment (i.e. WSRS, blind evaluator's GAIS or subject's GAIS assessment).

Per Protocol Set (PPS)

The Per Protocol Set (PPS) is defined as the subset of the FAS without any major protocol deviations. Major protocol deviations will be finalized during the Blind Data Review Meeting (BDRM) prior to database lock.

6 VARIABLES FOR ANALYSIS

6.1 Effectiveness Variables

6.1.1 Primary Effectiveness Variable

The primary effectiveness variable is the treatment success at Month 6. Treatment success is defined as a ≥ 1-point improvement in NLF from baseline as assessed by a blinded evaluator on a 5-point Wrinkle Severity Rating Scale (WSRS).

6.1.2 Secondary Effectiveness Variable

The secondary effectiveness variable is the "improvement" in the blinded evaluator's Global Aesthetic Improvement Scale (GAIS) score for the NLF at Month 6. An "improvement" on the GAIS will be classified as "improved", "much improved", or "very much improved".

[redacted]

6.2 Safety Variables

Adverse events (AEs) including serious adverse events (SAEs), AEs related to treatment as assessed by the investigator, and AEs leading to discontinuation during the study. Other safety variables include vital signs and clinical laboratory assessments.

6.3 Other Variables

Other variables include

[redacted]

- Demographics
- Medical History
- Concomitant Medications/Treatments

7 STATISTICAL ANALYSIS METHODS

The effectiveness analyses on the WSRS and Blinded evaluator's GAIS will be carried out on the PPS and for sensitivity of their analyses on the FAS. All safety analyses will use the SES. [redacted]

In the efficacy tables, mean and standard deviations will be reported with two decimal places. In all other tables, mean, standard deviation, and median will be reported to one decimal place greater than the data were collected; for derived data an adequate number of decimals has to be chosen. Percentages will be calculated using the denominator of all subjects in a specified population or exposed to a specific treatment. The denominator will be specified in a footnote to the tables for clarification if necessary. Percentages will be reported to one decimal place.

7.1 Effectiveness Endpoints

For categorical variables, frequencies of raw values and (where applicable) frequencies of changes from baseline will be shown by treatment and visit together with means and standard deviations.

Response variables are shown by visit as number and percentage of subjects with response (=proportion) at that visit per treatment together with the difference of the (paired) treatment response proportions and its corresponding 95% confidence interval computed according to Newcombe's recommended method[1] [Newcombe, 1998, method 10]. In addition, cross tables are provided showing number (and percent) of subjects with response/non-response on the Radiesse treated side vs. response/non-response on the Restylane treated side by visit.

For WSRS raw values and changes from baseline as well as for blinded evaluator's GAIS, cross tables are prepared showing frequencies of values observed on the Radiesse treated side vs. values observed on the Restylane treated side by visit.

7.1.1 Primary Effectiveness Endpoint

The primary effectiveness endpoint is the difference in proportions of responding NLF sides (NLF sides with treatment success) in the Radiesse and Restylane groups at Month 6. The difference in proportions (D) is defined as the difference between the proportion of responders on the side treated with Radiesse (P_{test}) and the proportion of responders on the side treated with Restylane ($P_{control}$), i.e.:

$$D = P_{test} - P_{control}$$

As a result thereof, the null and the alternative hypothesis notations will be as follows:

$$H_0 : D \leq -\Delta \text{ (null) versus } H_1 : D > -\Delta \text{ (alternative)}$$

To test the non-inferiority of test in comparison to control, a two-sided 95% confidence interval (CI) will be constructed around D. The non-inferiority margin Δ, the clinically negligible difference, is 0.10 (or 10.0%). H_0 will be rejected if the lower bound of the 2-sided 95% CI lies above the non-inferiority margin of –Δ (-10%), i.e. a conclusion of non-inferiority of Radiesse to Restylane. The two-sided 95% CI for the differences between proportions (paired) will be constructed using the Newcombe's recommended method[1] [Newcombe, 1998, method 10]. The SAS code is provided as below:

First, obtain the Wilson score confidence intervals for the responding proportions in two groups, denoted as l_2, u_2, and l_3, u_3 respectively, as well as the frequencies: e (response with both Radiesse and Restylane), f (response with Radiesse but no response with Restylane), g (response with Restylane but no response with Radiesse) and h (no response with both Radiesse and Restylane) based on the following SAS code:

```
proc freq data=a;
      by treatment;
      tables response/binomial(wilson);
run;
```

Second, obtain the two-sided 95% CI for the difference between proportions using Newcombe's method based on the following formulas and SAS code:

Formulas:

$$\mathrm{L} = \hat{\theta} - \delta, U = \hat{\theta} + \varepsilon$$

where

$$\delta = \sqrt{dl_2^2 - 2\hat{\varphi} dl_2 du_3 + du_3^2}, \varepsilon = \sqrt{du_2^2 - 2\hat{\varphi} du_2 dl_3 + dl_3^2},$$

$$dl_2 = (e+f)/n - l_2, du_2 = u_2 - (e+f)/n,$$

$$dl_3 = (e+g)/n - l_3, du_3 = u_3 - (e+g)/n,$$

$$n = e + f + g + h,$$

and if $eh > fg$, $\hat{\varphi} = \max(eh - fg - n/2, 0)$; if $eh \leq fg$,

$\hat{\varphi} = (eh - fg)/\sqrt{[(e+f)(g+h)(e+g)(f+h)]}$, but $\hat{\varphi} = 0$ if this denominator is 0.

SAS code:

```
n=e+f+g+h;
theta=e/(e+f)-g/(g+h);
dl2=(e+f)/n-l2;
du2=u2-(e+f)n;
dl3=(e+g)/n-l3;
du3=u3-(e+g)/n;
if e*h>f*g then phi=max(e*h-f*g-n/2,0);
else if (e+f)*(g+h)*(e+g)*(f+h)=0 then phi=0;
else phi=(e*h-f*g)/sqrt((e+f)*(g+h)*(e+g)*(f+h));
delta=sqrt(dl2**2-2*phi*dl2*du3);
epsilon=sqrt(du2**2-2*phi*du2*du3+dl3**2);
L=theta-delta;
U=theta+epsilon;
```

		Response with Restylane		
		yes	no	total
Response with Radiesse	yes	*e*	*f*	
	no	*g*	*h*	
	total			*n*

The confirmatory analysis will be based on the PPS using observed values only.

7.1.2 Secondary Effectiveness Endpoint

The secondary effectiveness endpoint is the difference between the proportions (π_{test} - $\pi_{control}$) with an improvement at Month 6 in blinded evaluator's GAIS on the side treated with Radiesse (π_{test}) and on the side treated with Restylane ($\pi_{control}$).

The hypothesis on the secondary effectiveness endpoint and the corresponding analysis will be similar to that specified for the primary effectiveness analysis, however, using a less conservative margin of $\Delta = 0.15$ (or 15%). This analysis will be of exploratory nature without any inferential considerations.

7.1.3 [redacted]

[redacted]

7.2 Safety Endpoints

All safety analyses will be performed on the SES.

The assessment of safety will be based mainly on the frequency of adverse events.

Only treatment-emergent AEs (TEAEs) will be analyzed, which are defined as AEs with onset or worsening after the first administration of Restylane or Radiesse.

TEAEs will be summarized for each treatment group (as applicable) by the incidence of at least one event, the number of events, and the incidence using Medical Dictionary for Regulatory Activities (MedDRA) version xxx preferred terms within the system organ classes (SOCs). The detailed information on AE including the outcome will be listed.

In case of missing intensity or missing causal relationship of an AE the worst case principle will be applied, i.e. a missing intensity will be set to "severe" and a missing causal relationship will be set to "related".

7.2.1 Primary Safety Endpoint(s)

Not applicable.

7.2.2 Secondary Safety Endpoints

Not applicable.

7.2.3 Other Safety Endpoints

7.2.3.1 All AEs

AEs will be summarized overall, by worst intensity and by worst causal relationship, each by SOC and PT. In addition, the frequency of subjects with TEAEs are provided by PT only. The number and percentage of subjects with at least one non-serious TEAE with incidence ≥ 5% will be displayed by SOC and PT.

7.2.3.2 Serious Adverse Events (SAEs)

Serious TEAEs will be summarized by SOC and PT.

7.2.3.3 Related Adverse Events

An AE is considered to be "related" to study treatment if "relationship to study device" or "relationship to procedure" is ticked as "yes" or is missing (i.e., the relationship cannot be ruled out). Related TEAEs will be summarized by SOC and PT.

7.2.3.4 AEs Leading to Discontinuation

AEs leading to discontinuation are AEs that are recorded as leading to study termination in electronic Case Report Form (eCRF), and will be summarized by SOC and PT.

7.2.3.5 Vital signs

Vital signs will be analyzed using descriptive summary statistics including the number of non-missing observations, mean and standard deviation, median, upper and lower quartiles, minimum and maximum for values and changes from baseline.

7.2.3.6 Clinical Laboratory Assessments

Variables from laboratory assessment with continuous outcomes will be analyzed using descriptive summary statistics including the number of non-missing observations, mean and standard deviation, median, upper and lower quartiles, minimum and maximum for values and changes from baseline. Categorical data (e.g. clinical significance) will be analyzed using shift tables.

7.3 [redacted]

[redacted]

7.3.1 Disposition and Withdrawals

The assignment of subjects to analysis populations, completion of study/ premature termination of study and protocol deviations/other reasons for exclusion from analysis sets will be summarized in frequency tables. Detailed information will be provided in listings.

7.3.2 Demographic and Other Baseline characteristics

Demographic data and other baseline characteristics will be summarized for the SES, FAS and PPS. If the 'randomized subjects set' is not equal to the SES, demographic data and other baseline characteristics will be summarized for the 'randomized subjects set', too. Continuous data will be analyzed using descriptive summary statistics (n, mean, standard deviation, median, lower and upper quartiles, minimum and maximum). Categorical data will be analyzed using frequency tables. No homogeneity tests will be performed.

The following demographic data will be summarized for this study: sex, age (calculated relatively to consent date), age category, ethnicity, race, Fitzpatrick skin type, height, weight and Body Mass Index (BMI).

Detailed information will be provided in listings.

7.3.3 Medical History

Medical history and concomitant diseases will be summarized based on MedDRA system organ class and preferred term levels for the SES.

7.3.4 Prior/Concomitant Medications/Treatments

Prior and concomitant medications/treatments will be listed and summarized by Anatomical Therapeutic Chemical (ATC) class and World Health Organization Drug Dictionary (WHODD) generic name. Frequencies of concomitant medications/treatments will be given

based on different ATC code levels as well as by generic name for the SES. Study Product Exposure

The injected volumes of Radiesse and Restylane at the baseline visit and the 4-week/Month 1 optional touch-up visit, as well as the total volume administered will be summarized, using mean, standard deviation, median, minimum and maximum based on SES.

7.4 Special Statistical/Analytical Issues

7.4.1 Discontinuations and Missing Data

Effectiveness:

Besides the analyses based on observed data, additional sensitivity analyses for primary and secondary effectiveness endpoints will be carried out with missing data imputed using single value imputation methods. Imputations will be performed until Month 6.

Missing effectiveness data will be imputed as follows:

Worst Case Scenario (WCS):

- For the WCS method, missing post-baseline data for Radiesse treatment group will be imputed as non-responders, while missing post-baseline data for Restylane treatment group will be imputed as responders. As a result, the WCS is considered a conservative imputation method because missing values from the control are imputed as responders while missing values from the treatment are imputed as non-responders. The WCS will be used for sensitivity analysis.As this method is an extreme conservative method, an additional, probably more realistic but still conservative imputation method will be performed based on pattern (reason) of missingness:

- *Pattern Based Scenario (PBS):*At the BDRM, the reasons for subjects' withdrawals before Month 1, Month 3 or Month 6 assessment of the WSRS and/or GAIS assessment (with the consequence that these values are missing) are reviewed and assessed as "probably missing at random" or as "definitely or possibly missing not at random". Hereby "Lost to follow-up" will be assessed as "probably missing at random", while "AE related to treatment" will be assessed as "definitely or possibly missing not at random". For subjects with an assessment of "probably missing at random", no imputation will be performed. For subjects with an assessment of "definitely or possibly missing not at random", it will be checked whether the withdrawal is related to only one of the two treatments (this is the case if e.g. the AE leading to WD is located on one face side only). If so, the subject is treated as "non-responder" for this treatment and as "responder" for the other. If not, the subject is treated as "non-responder" for both treatments.

Safety:

Each medication/therapy will be allocated unambiguously either to previous medications or to concomitant medications.

- If stop date is before start of treatment: Previous medication
- If the stop date is at or after start of treatment or ongoing is checked: Concomitant medication
- Otherwise, if the stop date is missing or partially given:
  - Partial stop date available: Previous therapy if the latest possible stop date is before start of treatment; Concomitant therapy if the latest possible stop date is at or after start of treatment
  - If the stop date is completely missing: Concomitant therapy

Each finding on the medical history page will be allocated unambiguously either to medical history or to concomitant diseases.

- If the stop date is before start of treatment: Medical history
- If the stop date is at or after start of treatment or ongoing (even if it refers to a cut-off point before start of treatment) is ticked: Concomitant disease
- Otherwise, if the stop date is missing or partially given:
  - Partial stop date available: Medical history if the latest possible stop date is before start of treatment; Concomitant disease if the latest possible stop date is at or after start of treatment
  - Stop date is completely missing for findings coded as “Surgical and medical procedures”: Medical history
  - Stop date is completely missing for other findings: Concomitant disease

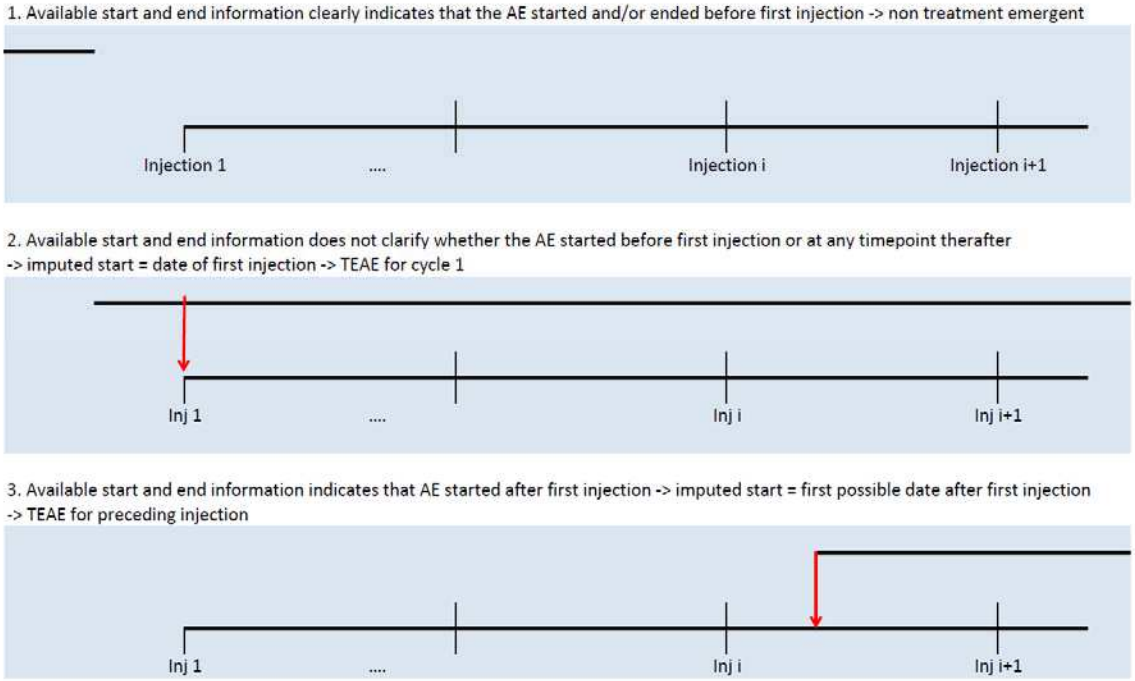
For AEs, imputation of missing start or stop dates will follow the steps below:

- If the start date (and time, if applicable) is partially missing but implies start before start of treatment or after date of final examination, the AE will be considered as Non-Treatment Emergent Adverse Event (TEAE).
  - If the start date (and time, if applicable) is partially missing and possibly after start of treatment and before or after date of final examination, will be considered as TEAE.

AEs with completely missing onset dates will be considered treatment emergent; no estimation of the dates will be performed.

For the imputed analysis start date of adverse events (variable ASTDT) in the ADaM domain ADAE the following imputation rules related to any treatment period will be used if applicable:

Figure 1: Possible start/end dates of AEs



1a) If the start date is missing completely and the end date and time* does imply end before first injection then no imputation of start date and time will be performed.

1b) If the start date is partially missing and the partially missing start date and/or the end date and time* do imply start and/or end before first injection then set missing start day to 1 / missing start month to 1, if applicable and set start time (if not available) to 23:59, if not after end date/time (else set start time to end time) (e.g. start date of AE: 2008-11 end date of AE: 2009-02-01, end time 22:00 and first injection: 2008-12-06 T17:35 → imputed start date: 2008-11-01, start time: 23:59). Set imputation flags accordingly.

2a) If the start date is missing completely and end date and time* do not imply end before first injection then set imputed start date to date of first injection and set start time (if not available) to 23:59, if not after end date/time (else set start time to end time). (e.g. start date of AE: missing, end date of AE: 2008-12-31, end time 23:00, first injection date: 2008-12-06T17:35 → imputed start date: 2008-12-06, start time: 23:59). Set imputation flags accordingly.

2b) If the start date is partially missing and the non-missing part is the same as in first injection date and the end date and time* do not imply end before first injection, then set the imputed start date to the first injection date and set start time (if not available) to 23:59, if not after end date/time (else set start time to end time) (e.g. start date of AE: 2008-12, end date of AE: 2008-12-31, end time 23:00, first injection date: 2008-12-06T17:35 → imputed start date: 2008-12-06, start time: 23:59). Set imputation flags accordingly.

3) If the start date is partially missing and imputation rule 1b and 2b do not fit, then set missing day to 1 / missing month to 1, if applicable and set start time (if not available) to 23:59, if not after end date/time (else set start time to end time) (e.g. start date of AE: 2009-02 end date of AE: 2009-02-01, end time 22:00 and first injection: 2008-12-06 T17:35 → imputed start date: 2009-02-01, start time: 22:00). Set imputation flags accordingly.

End date or end time will not be imputed.

* The end date and time may be available completely, partly or may be missing.

7.4.2 *Interim Analyses*

In the mock shells it is specified for each table/listing whether it is to be produced for the interim report and/or amendment.
Statistical analyses will be performed when all subjects reach the 6 Month post initial injection time point. Primary effectiveness variable, secondary effectiveness variables and AE data will be analyzed using the snapshot of data for the first 6 months. A Statistical and a Clinical Study Report will be generated for the aforementioned analysis. They will be amended by a description and discussion of the analyses performed when all subjects have completed the study. For further details on which tables are produced for the Clinical Study Report and which for the amendment, please refer to the TFLs.

No examination for early termination is planned.

7.4.3 *Data Monitoring Committee*

There will be no data monitoring committee for this study.

7.4.4 *Multiple Comparisons/Multiplicity*

Only the analysis of the primary effectiveness endpoint based on PPS using observed values is of confirmatory nature. Further analyses on the primary effectiveness endpoint are performed for sensitivity reasons. Analyses of all other endpoints/variables is of exploratory nature.

7.4.5 *Examination of Subgroups*

No subgroup examination will be performed for this study.

8 CHANGES IN THE PLANNED ANALYSES

As the efficacy tables on the WSRS and Blind evaluator's GAIS cover [redacted] the paragraph

"The primary effectiveness and secondary effectiveness analyses will be carried out on the PPS and for sensitivity of their analyses on the FAS. [redacted]."

was replaced by

"The effectiveness analyses on the WSRS and Blinded evaluator's GAIS will be carried out on the PPS and for sensitivity of their analyses on the FAS. All safety analyses will use the SES. [redacted]"

9 REFERENCES

1.Altman D, Machin D, Brayant TN, Gardner MJ (2000) „Statistics with confidence", 2nd edition, BMj Books.

Appendix 1: List of Tables, Figures, and Listings