Trial Statistical Analysis Plan

c02157629-01

BI Trial No.:	1200.161
Title:	LUX-Head & Neck 3 A randomised, open-label, phase III study to evaluate the efficacy and safety of oral afatinib (BIBW 2992) versus intravenous methotrexate in patients with recurrent and/or metastatic head and neck squamous cell carcinoma who have progressed after platinum-based therapy Including Protocol Amendment 1 [c02089607-03] Including Protocol Amendment 2 [c09086960]
Investigational Product(s):	Afatinib (BIBW 2992)
Responsible trial statistician(s):	[redacted] Phone: +[redacted] Fax: +[redacted]
Date of statistical analysis plan:	1 APR 2016 SIGNED
Version:	Final

1. TABLE OF CONTENTS

LIST OF TABLES

2. LIST OF ABBREVIATIONS

Term	Definition / description
AE	Adverse event
BRPM	Blinded report planning meeting
CTC	Common Terminology Criteria
CTP	Clinical Trial Protocol
CTR	Clinical Trial Report
DM&SM	Boehringer Ingelheim Data Management and Statistics Manual
DRA	Drug Regulatory Affairs
DMG	Dictionary Maintenance Group
EMEA	European Agency for the Evaluation of Medicinal Products
ICH	International Conference on Harmonisation
LOCF	Last observation carried forward
MedDRA	Medical Dictionary for Regulatory Activities
MQRM	Medical Quality Review Meeting
O*C	Oracle Clinical
PK	Pharmacokinetics
PPS	Per protocol set
PSTAT	Project Statistician
PT	Preferred term
PV	Protocol violation
Q1	Lower quartile
Q3	Upper quartile
SA	Statistical analysis
SD	Standard deviation
SMQ	Standardised MedDRA query
SOC	System organ class
TCM	Trial Clinical Monitor
TOC	Table of contents
TMW	Trial Medical Writer
TSAP	Trial statistical analysis plan

3. INTRODUCTION

As per ICH E9, 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 Section 7.6 of the CTP "Statistical Methods and Determination of Sample Size". 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, randomisation.

SAS® Version 9.2 or later versions will be used for all analyses.

4. CHANGES IN THE PLANNED ANALYSIS OF THE STUDY

The following changes were made to the statistical methods specified in the CTP and amendments:

The CTP describes that the effect of afatinib on PFS compared with methotrexate will be tested at the one-sided, 0.025 significance level. This is identical to the effect of afatinib being tested at the more commonly used two-sided 0.05 significance level if the treatment effect is in favor of afatinib. In order to aid in the interpretation of this study, two-sided p-values will be presented in the analysis results.

The CTP specifies two subgroup variables for PFS and OS: baseline ECOG performance score (0 or 1) and prior use of EGFR-targeted antibody for recurrent and/or metastatic head and neck squamous cell carcinoma (HNSCC) (Yes or No). The TSAP pre-specifies additional subgroup variables, detailed in Section 6.4. Subgroup analysis will also be performed for objective response.

5. ENDPOINTS

Please refer to Section 5.1 and 7.3 in the CTP for more details on the endpoints for this study.

5.1 PRIMARY ENDPOINT

The primary endpoint in this study is progression-free survival (PFS), defined as the time from the date of randomisation to the date of disease progression evaluated according to RECIST 1.1 or to the date of death from any cause, whichever occurs first.

The disease progression status and date of progression will be assessed at the investigational sites, as well as by an independent central review board, comprising of three radiologists (two primary readers and one adjudicator as needed). As stated in the CTP and Section 7.4 below, the primary analysis of PFS will be based on independent central review.

5.2 SECONDARY ENDPOINTS

5.2.1 Key secondary endpoint

Not applicable.

5.2.2 Secondary endpoints

Secondary efficacy endpoints include:

- Overall survival (OS), defined as the time from the date of randomization to the date of death (regardless of the cause of death)
- Objective response
- Health-related quality of life (HRQoL)

Refer to Section 5.1 of the CTP for more details.

6. GENERAL ANALYSIS DEFINITIONS

6.1 TREATMENT

In the clinical trial report, all randomized patients will be classified into either 'Afatinib 40' or 'MTX 40'. For efficacy analyses, patients will be analysed as randomised. For safety analyses, patients will be analysed as treated, according to the initial treatment taken.

Even though this trial is open-label, according to BI SOP [1], report planning will be done in a 'blinded' manner. Therefore, before database lock, all patients will be randomly assigned a dummy treatment code, DummyRed or DummyBlue, and displayed as such in all programmed tables, listings and figures (TLFs).

For safety reporting purpose, the following treatment periods are defined:

- Screening: from informed consent to the date of first dose of study medication
- On-treatment: date of first dose of study medication to last dose
- Post-treatment: date of last dose + 1 day to the 28th day after last dose
- Post-study: from 29th day after last dose onwards

In summarizing the safety data, all adverse events with onset date during the on- and post-treatment periods will be considered as 'treatment-emergent'. Refer to Section 7.8.1 for more details.

6.2 IMPORTANT PROTOCOL VIOLATIONS

Patients with potentially important protocol violations will be identified based on Table 6.2: 1 and documented. No per protocol set analysis is planned for this study. The list of PVs in Table 6.2: 1 is considered a 'working' list, which is expected to be updated throughout the trial and finalized prior to the primary analysis of PFS.

Table 6.2: 1 Important protocol violations

Category / Code		Description	Comment/Example	Efficacy / Safety
A		Entrance criteria not met		
	A1	Inclusion criteria not met		
	A1.1	Not histologically or cytologically confirmed recurrent and/or metastatic SCC in oral cavity, oropharynx, hypopharynx or larynx	Inclusion criterion 1	E
	A1.2	Not progressed after first-line platinum-based therapy for R/M disease, or not received minimum dose of such therapy	Inclusion criterion 2	E
	A1.3*	No measurable disease according to RECIST 1.1 at screening	Inclusion criterion 3 e.g., all non-nodal target lesions have longest diameters < 10mm per CT, or MRI or caliper measurement at baseline; lymph nodes < 15mm	E
	A1.4*	ECOG score >= 2 at baseline	Inclusion criterion 4	E/S
	A1.5*	Age < 18 years at baseline	Inclusion criterion 5	S
	A1.6*	Baseline imaging performed too early	Baseline imaging is 21 days prior to randomization.	E
	A2	Important exclusion criteria met		
	A2.1	PD within 3 months after completion of curative treatment	Exclusion criterion 1	E
	A2.2	Malignancy other than SCC in oral cavity, oropharynx, hypopharynx or larynx, or brain metastasis.	Exclusion criterion 2, 10, 11	E
	A2.3	>1 chemo regimen given for R/M disease	Exclusion criterion 3	E
	A2.4*	Prior treatment with EGFR-targeted small molecules	Exclusion criterion 4	E
	A2.5*	Treatment of investigational drug or anti-cancer therapy, or major surgical procedure less than four weeks prior to randomisation;	Exclusion criteria 5, 9	S
	A2.6*	Pre-existing condition at screening that may cause additional risk from investigational drug or exacerbate expected adverse events	Exclusion criteria 6, 7, 12-17, 19-21	S
	A2.7*	Abnormal screening values for ANC, platelet count, bilirubin, AST, ALT or calculated creatinine clearance	Exclusion criterion 18	S
B		Informed consent		
	B1*	Informed consent not available		S
	B2*	Informed consent signed and dated after first study related activity	Informed consent after screening visit date	S
	B3	Informed consent not in compliance with ICH-GCP or local law	Inclusion criterion 6	S
C		Trial medication & randomisation		

Category / Code		Description	Comment/Example	Efficacy / Safety
	C1	Incorrect medication dose taken	Starting dose is not 40 mg for afatinib or 40 mg/m2 for methotrexate; dose was not paused and reduced according to protocol; dose escalation not done properly.	E/S
	C2*	Randomisation not followed	Randomized to afatinib but received MTX or vice versa; stratification factor wrong	E
	C3*	Non-compliance with trial medication	Compliance rate < 80%	E
D		Concomitant medication		
	D1*	Other concurrent anti-cancer therapy use	Other anti-cancer therapy use prior to discontinuation of trial medication	E/S
	D2	Use of prohibited concomitant medications	Exclusion criterion 8	S

* IPVs will be identified programmatically.

6.3 PATIENT SETS ANALYSED

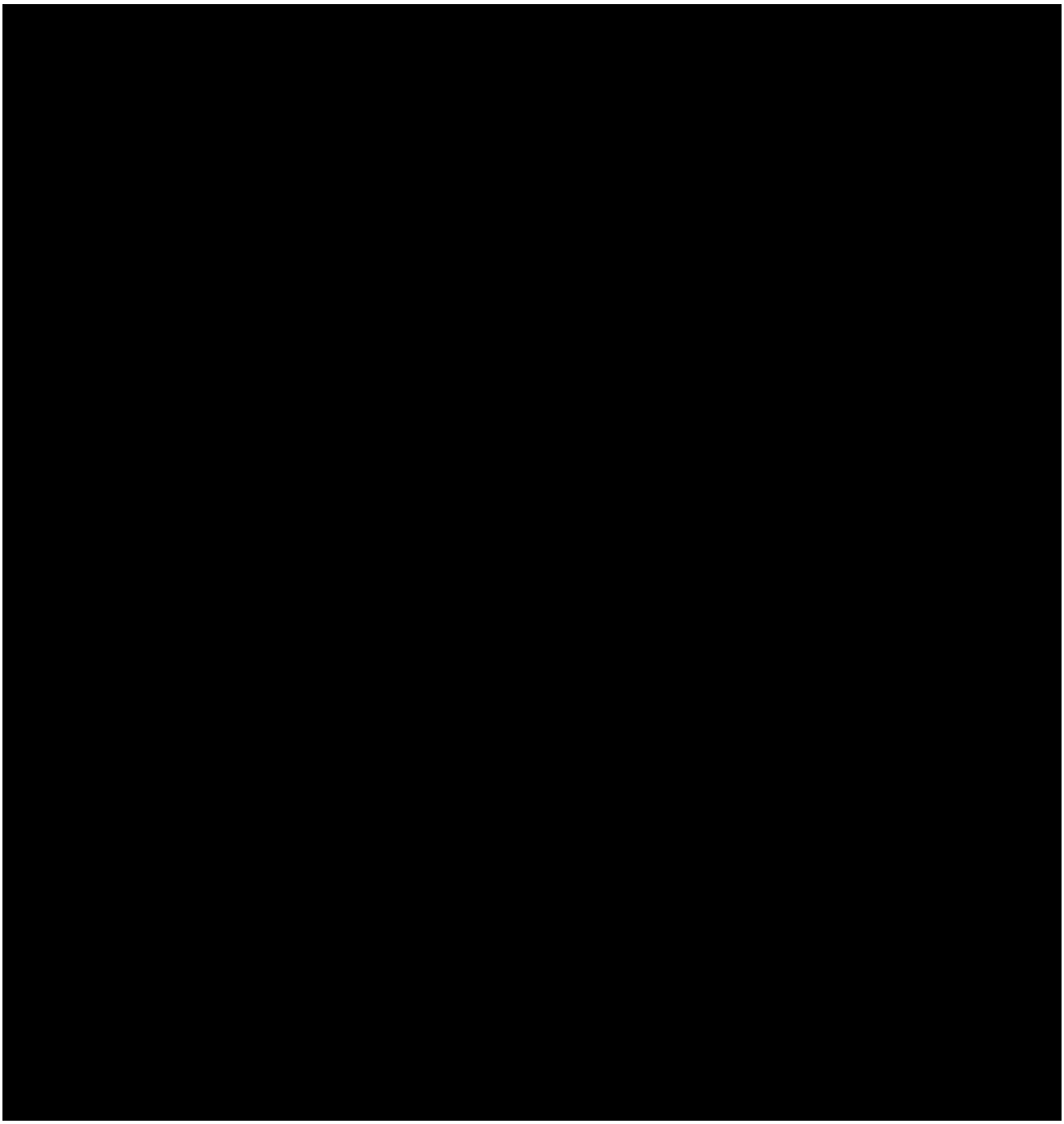
The following two patient sets are defined for this study:

- Randomised set, which includes all randomised patients, whether treated or not.
- Treated set, which includes all patients who were documented to have taken at least one dose of study medication.

The following table shows which patient set will be used for each planned analysis.

Table 6.3: 1 Patient sets analyzed

Endpoints	Treated set	Randomised set
Primary endpoint		As randomised
Secondary endpoints		As randomised
Safety endpoints	As treated	
Demographic/baseline characteristics		As randomised



6.5 POOLING OF CENTRES

This section is not applicable because centre is not included in any of the statistical models.

6.6 HANDLING OF MISSING DATA AND OUTLIERS

In general, missing efficacy data will not be imputed and all reasonable efforts will be taken during the study conduct to obtain such data. Patients lost to follow up, with unknown vital status, missing tumour imaging data etc. will be censored for time to event analyses. Detailed rules for censoring are provided in Table 7.4: 1.

Missing or incomplete AE dates are imputed according to BI standards (see "Handling of missing and incomplete AE dates")[2].

Handling of missing PK data will be performed according to BI standards (see "Standards and processes for analyses performed within Clinical Pharmacokinetics/Pharmacodynamics")[11].

6.7 BASELINE, TIME WINDOWS AND CALCULATED VISITS

In general, baseline values will be the measurements taken on the date of the first dose of study medication. If this value is not available, the measurement at screening visit will be used. If the screening lab is repeated, then the second one should be used. For tumour imaging data, the last imaging done before randomisation will be used as baseline (typically, on or before the screening visit).

Visit will be labelled according to the flow chart in the protocol: Visit 1 (for screening), Visit 2 (randomisation and start of treatment), ..., EOT (End of treatment), FUV (28 days after EOT), OP1(Observation Period 1), OPx. OPx will be scheduled for every 4 weeks after FUV. Treatment is expected to start as soon as patient is randomised, if not on the same date, start within 4 days of randomisation. Visit 2 is used for both randomisation and treatment start, in case the two are not on the same date, the visit number will not be recalculated.

For safety displays where 'Start day' and 'End day' are needed, they will be calculated relative to the date of the first dose of study medication. The date of first dose will be 'Day 1'; the days prior to first dose will be Day -1, Day -2 etc. Day 0 will not be used.

Tumour imaging will be performed at screening (within 21 days prior to start of treatment, i.e. Visit 2), every 6 weeks after Visit 2 until week 24, and every 8 weeks thereafter. Tumour imaging data will be displayed as Screening, Week 6, 12, 18, 24, 32, 40, ..., etc. The number of weeks will be calculated using their relative day (from randomisation) with appropriate time windows: ± 3 weeks for weeks 6 – 24; and ±4 weeks for Week 32 and onwards.

7. PLANNED ANALYSIS

The primary analysis for the primary endpoint, PFS, will take place after the number of required events have occurred based on independent review. Final analyses will be done after final database lock.

For display purpose, in the End-of-Text tables, the set of summary statistics is: N / Mean / StD / Min / Median / Max for continuous variables. 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.

An independent data monitoring committee (DMC) will oversee the trial conduct and review safety data according to DMC SAP [3].

7.1 DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS

Only descriptive statistics are planned for this section of the report. Standard summary tables will be created based on the randomised set. All subgroup variables defined in Section 6.4 will be included.

7.2 CONCOMITANT DISEASES AND MEDICATION

Only descriptive statistics are planned for this section of the report, based on the randomised set. Concomitant disease will be coded according to MedDRA dictionary (current version at the time of analysis), presenting the system organ class (SOC) and preferred term (PT) levels. Concomitant medication will be coded based on the WHO-DD (current version at the time of analysis), displaying the ATC3 class and PT levels. PTs that belong to different ATC3 classes will be shown under all applicable ATC3 classes.

7.3 TREATMENT COMPLIANCE

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

For afatinib, treatment compliance will be summarized based on two measurements:

$$1)\ (\%) = \frac{\text{Total number of tablets taken} + \text{Days missed dose due to AE}}{\text{Total number of days on treatment (drug stop - start date} + 1)} \times 100\%$$

$$2)\ (\%) = \frac{\text{Total number of tablets taken}}{\text{Total number of days on treatment (drug stop - start date} + 1)} \times 100\%$$

The first measure does not consider missing doses due to AE as non-compliance because treatment interruptions are recommended by protocol in the presence of certain AEs in order for patients to recover from the adverse events. The second measure is intended for treatment

adherence, which considers non-adherence as long as a patient is not taking the medication regardless of the cause.

Similarly, for methotrexate, treatment compliance will be summarized based on the number of skipped infusions:

$$1) \quad (\%) = \frac{\text{Total number of infusions} + \text{infusions skipped due to AE}}{\text{Total number of infusions expected}} \times 100\%$$

$$2) \quad (\%) = \frac{\text{Total number of infusions}}{\text{Total number of infusions expected}} \times 100\%$$

where Total number of infusions expected = (date of last infusion – first infusion)/7 + 1.

For both treatment groups, the first measure will be used in determining compliance to study medication as protocol violation, code C3 in Table 6.2: 1.

7.4 PRIMARY ENDPOINT

Primary analysis

The primary analysis of PFS will be performed based on the results from independent review.

Stratified log-rank test (two-sided, 0.05 significance level) will be used to test the effect of afatinib on PFS compared with methotrexate. Stratified Cox proportional hazards model will be used to estimate the hazard ratio (afatinib vs. methotrexate) and 95% confidence intervals (CI).

Stratification will be done on the two stratification factors used in randomisation: baseline ECOG performance score (0 or 1) and use of prior EGFR-targeted antibody in the R/M setting (Yes or No). These two variables are entered on baseline CRFs and via the IVRS system. In case they differ, the values entered on the CRFs will be used in analysis.

Kaplan-Meier curves will be calculated for each treatment group, separately, and the estimates of PFS probabilities from the curves and 95% CI (using the Greenwood standard error estimate) will be tabulated at 6-week intervals up to week 24. For each time point, a comparison of afatinib vs. methotrexate will be done using a z-test (the difference of Kaplan-Meier estimates between afatinib and methotrexate divided by its standard error follows approximately the standard normal distribution under the null hypothesis of equality between the two groups).

For the stratified Cox model, the assumption of proportional hazards within each stratum and across strata will be checked descriptively.

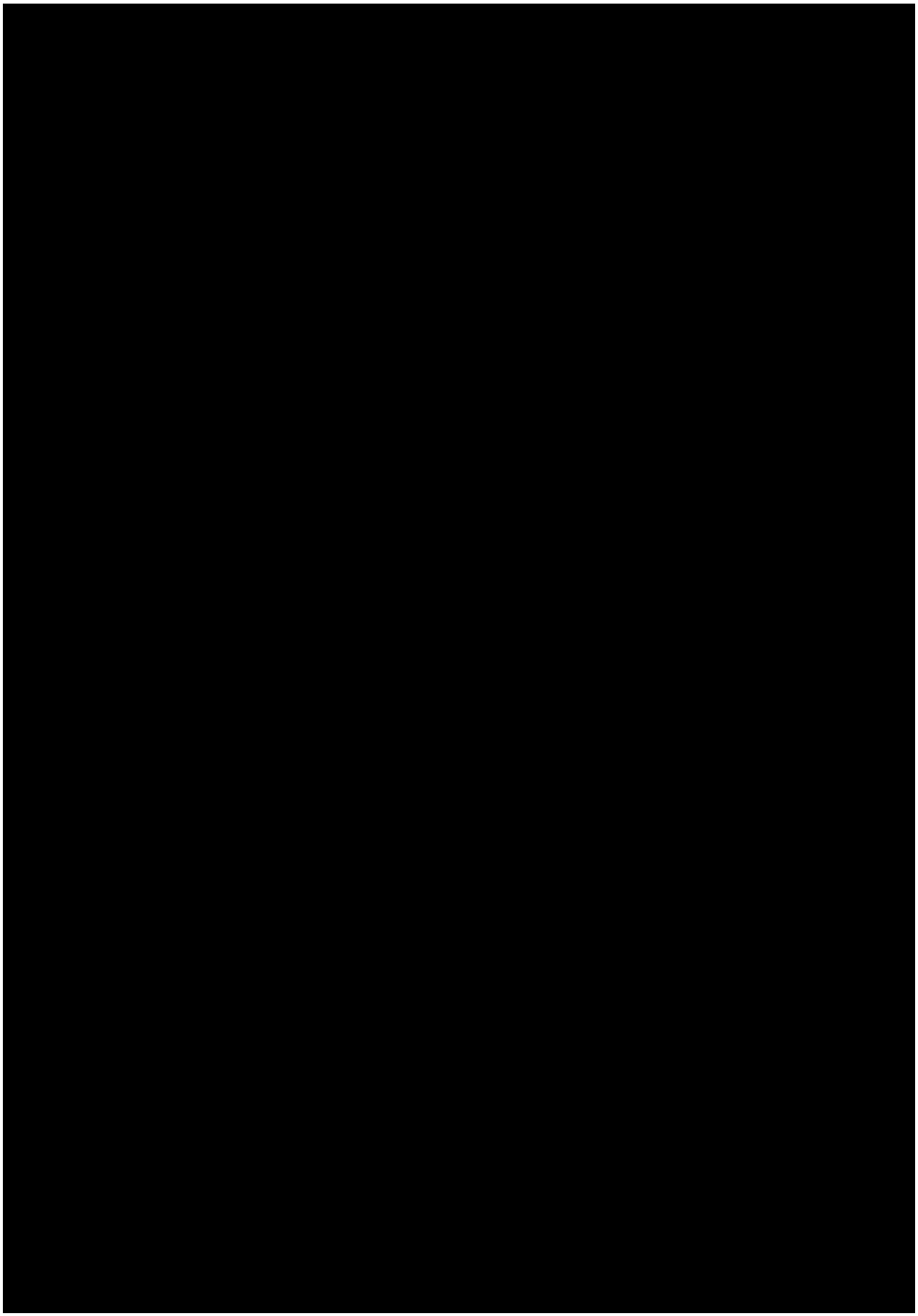
By definition, a patient with known date of progression or death will be considered as having had a PFS event on the earliest date of progression or death. According to the FDA Guidance recommendations [4], we also consider the following two important factors in determining

whether or not a patient has had a PFS event and the date of such event: 1) whether there have been missing tumour imaging scans and the timing of such missing scans, 2) whether a patient has had taken other anti-cancer therapy prior to progression or death. The rules are described in Table 7.3.1.1: 1 of the CTP. The table is copied here as Table 7.4: 1 for completeness and ease of review. A summary table of the reasons for censoring will be produced for each treatment group.

Table 7.4: 1 Determination of PFS events and censoring for the primary analysis

Situation	Date of event or censoring	Outcome
Without post-baseline radiological assessments		
1. Alive	Date of randomisation	Censored
2a. Death prior to the second scheduled image	Date of death	Event
2b. Death beyond the second scheduled image	Date of randomisation	Censored
With post-baseline radiological assessments BUT no other anti-cancer therapy		
3. Alive and no progression	Date of last radiological assessment of measured lesions	Censored
4a. Death but no progression, zero or one missed image prior to death	Date of death	Event
4b. Death but no progression, two or more missed images prior to death	Date of last radiological assessment of measured lesions	Censored
5a. Progression, zero or one missed image prior to progression	Date of radiological assessment of progression	Event
5b. Progression, two or more missed images prior to progression	Date of last radiological assessment of measured lesions prior to progression	Censored
With post-baseline radiological assessments AND other anti-cancer therapy		
6. New anti-cancer therapy started before progression (or death)	Date of last radiological assessment before other new anti-cancer therapy	Censored
7a. Progression before new anti-cancer therapy, zero or one missed image prior to progression	Date of radiological assessment of progression	Event
7b. Progression before new anti-cancer therapy, two or more missed image prior to progression	Date of last radiological assessment of measured lesions prior to progression	Censored

Note: Within rule 6, regardless of whether the patient had PD or not, as long as a patient started new anti-cancer therapy, the patient will be censored on the date of last radiological assessment prior to the new anti-cancer therapy.



[redacted]

7.5 SECONDARY ENDPOINTS

7.5.1 Key secondary endpoint

Not applicable.

7.5.2 Secondary endpoints

Overall survival

OS is calculated as the time form date of randomisation to death, regardless of cause of death. Patients for whom there is no evidence of death at the time of analysis will be censored on the date that they were last known to have been alive.

Similar analyses to PFS will also be done for OS by stratified log-rank test (nominal two-sided p-value). Stratified Cox PH model will be used to calculate hazard ratio and 95% CI. Proportional hazard assumption will be checked within and across strata. Kaplan-Meier curves will be calculated for each treatment group, and the estimates of OS probabilities from the curves and 95% CI (using the Greenwood standard error estimate) will be tabulated at 3-month intervals. For each time point, a comparison of afatinib vs. methotrexate will be done using a z-test. Because sample size is not powered for OS, the analyses for OS are exploratory.

The effect of non-study anti-cancer therapy on OS is a known issue in trials evaluating overall survival. In order to explore the potential impact on OS, the following additional summaries will be produced:

- Summarize the type of non-study anti-cancer therapy after discontinuing trial medication by treatment group. This includes a graphical display of patient numbers by treatment group for the number of subsequent lines of therapy.
- A summary by scheduled tumour assessments of the following subgroups of patients:
    - Number of patients alive at the beginning of interval
        - Patients receiving study treatment at anytime during the interval
        - Patients who started new anti-cancer treatment during the interval
- A summary of the cumulative proportion of deaths at each scheduled tumour assessment time point and if necessary, an analysis based on rank preserving structural failure time models [5, 6].
- If necessary, an analysis based on inverse probability of censoring weighted methods and on inverse probability of treatment weighted methods [7, 8].

Objective response

Objective response is defined as the best overall response of complete response (CR) or partial response (PR) according to RECIST 1.1 based on independent review. Objective

response is a binary variable. Patients who do not show CR or PR will be considered as non-responders, irrespective of protocol violation or missing data.

Objective response rate will be presented for each treatment as the proportion of responders, with the exact 95% Clopper-Pearson intervals. For descriptive purpose, the Cochran-Mantel-Haenszel test will be used to test the null hypothesis of equal response rate between afatinib and methotrexate while controlling for the two baseline stratification factors. Stratified logistic regression model will also be used to calculate the odds ratio of response (afatinib vs. methotrexate).

For patients with objective response, the time to response measured from randomisation to the date of first response will be summarised by the planned imaging time points. Duration of response measured from the date of first response until progression or death will be explored among the responders. Kaplan-Meier curves for the two treatment groups will be produced for the duration of response. The end time, progression or death, will be handled the same way as in the primary PFS analysis.

Same subgroup variables in Section 6.4 will be explored for response rate. The logistic regression model with treatment group as the only factor will be run for each subgroup category to calculate the odds ratio within the category. A forest plot will be provided showing the odds ratio (afatinib vs. methotrexate) and the corresponding 95% CI for each subgroup category.

Per RECIST 1.1, confirmation of response (response shown on a subsequent image taken at least 4 weeks afterwards) is not required for Phase III studies that have PFS as the primary endpoint; hence the above summaries will be produced without the requirement for confirmation. However, for completeness and comparability with other trial data, the above summaries and analyses will be repeated with the requirement of confirmation, these will be considered secondary in nature. Subgroup analyses will not be performed for confirmed response.

All the above analyses for objective response will be repeated based on the investigator assessment data. The difference between investigator and independent review data will be summarized.

Health-related Quality of Life (HRQoL)

The HRQoL questionnaires, EORTC QLQ-C30 and QLQ-H&N35, measured at baseline, during treatment, first follow-up visit after stopping study drug will be included in analyses. The main analysis of HRQoL questionnaires will focus on the following 3 scales:

- Pain scale from H&N35: hn31, pain in the mouth; hn32, pain in the jaw; hn33, soreness in the mouth; hn34, painful throat.
- Swallowing scale from H&N35: hn35, problems swallowing liquid; hn36, problems swallowing pureed food; hn37, problems swallowing solid food; hn38, choked when swallowing.

- Global health status/QoL scale from C30: c29, overall health; c30, overall quality of life.

Scoring of the symptom scales/items will follow the EORTC scoring manual and a linear transformation of the scores to a 0-100 point scale, see reference in Section 7.3.2.5 of the CTP.

For each of the 3 scales, the following 3 analyses will be performed:

Distribution of patients improved, stable or worsened

Improvement will be defined as a score that improves from baseline by at least 10 points (on the 0-100 point scale) at anytime during the study. If a patient has not improved, worsening will be defined as a 10-point worsening at anytime during the study. Patients who have neither improved nor worsened will be considered as stable.

The number and percentage of patients falling into each of these three categories will be summarised by treatment group. Similar to the objective response endpoint, stratified logistic regression model will be used to compare the distribution of improved vs. not improved (i.e. stable and worsened) between the two treatment groups.

Change in scores over time

Changes in scores over time will be assessed using longitudinal mixed-effects growth curve models with the average profile over time for each endpoint described by a piecewise linear model adjusted for the fixed effects baseline ECOG performance score and prior use of EGFR-targeted antibody for R/M HNSCC. The models will allow the slope to change at 6, 12, 18 and 24 weeks. The model will also include two random effects of intercept and slope (time from randomisation). The area under the estimated growth curve (AUC) up to the median follow-up time will be calculated as a summary measure for each treatment group; this will be divided by the median follow-up time so that it can be interpreted as the mean HRQOL score up to the median follow-up time. The treatment effect will be estimated as the average difference between the treatment group mean scores, together with a 95% confidence interval and associated p-value based on a t-statistic with degrees of freedom calculated using the Kenward-Roger method.

Time to deterioration

Time to deterioration is defined as the time from randomisation to the first 10-point worsening on the 0-100 point scale. Patients with no deterioration (including those with disease progression) will be censored at the last available HRQOL assessment date. Patients with no post-baseline assessments will be censored on the day of randomisation. One exception is that patients who die within 4 weeks of drug discontinuation or randomisation with no evidence of deterioration will be considered to have deteriorated at the time of death.

Time to deterioration will be analysed and summarised using stratified log-rank test and stratified Cox proportional hazard models, similar to PFS.

In addition to the above mentioned 3 scales, all single items and scales (functional and symptom) from both questionnaires will be analysed in a similar fashion to summarise the impact of therapy over the entire profile of the measures, and to examine the consistency of component items with the composite measures. Results will be presented as forest plots (odds ratio for improvement, mean score differences from longitudinal model, hazard ratios for time to deterioration) and summary tables.

7.7 EXTENT OF EXPOSURE

Extent of exposure will be explored based on total treatment time, calculated from start to end of treatment, including off-drug periods due to non-compliance or toxicity prior to permanent discontinuation. For methotrexate, time between the weekly infusions will be counted toward total treatment time. Standard descriptive summaries, by treatment group, of these data will be provided for all treated patients (the treated set), according to the initial treatment taken.

Kaplan-Meier curves will be calculated to show the time from treatment start to permanent discontinuation, the number of patients at risk (exposed) during the study treatment period.

Dose escalation and reductions are allowed in all patients, details in Section 4 of the CTP. Exposure will also be summarised by dosing level:

- Treatment time by each dose level (50, 40, 30 and 20). Dose level unit is mg for afatinib and mg/m^2 for methotrexate.
- Number and proportion of patients on each dose level over time.
- Time to first dose reduction, Kaplan-Meier plot of time from first dose of study medication to the first dose reduction
- For afatinib, duration of off-drug periods prior to dose reduction.

7.8 SAFETY ANALYSIS

All safety analyses will be performed on the treated set.

7.8.1 Adverse events

Unless otherwise specified, the analyses of adverse events (AEs) will be descriptive in nature. All analyses of AEs will be based on the number of patients with AEs and NOT on the number of AEs.

Furthermore, for analysis of AE attributes such as duration, severity, etc, multiple AE occurrences collected on the CRF will be collapsed into AE episodes provided that all of the following applies:

- The same MedDRA lowest level term was reported for the occurrences

- The occurrences were time-overlapping or time-adjacent (time-adjacent if the second occurrence started on the same day or on the day after the end of the first occurrence)

- Treatment did not change between the onset of the occurrences OR treatment changed between the onset of the occurrences, but no deterioration was observed for the later occurrence

For further details on summarization of AE data, please refer to the guideline 'Handling and summarization of adverse event data for clinical trial reports and integrated summaries' [9].

The analysis of adverse events will be based on treatment-emergent adverse events, i.e. all adverse events with an onset date between first drug intake and 28 days after last drug intake. All adverse events occurring before first drug intake will be assigned to ‘screening’ and all adverse events occurring after last drug intake + 28 days will be assigned to ‘post-study’. For details on the treatment definition, see Section 6.1.

An overall summary of adverse events will be presented. This summary will exclude the rows ‘Severe AEs’, and ‘Other significant AEs’ but will include additional rows for ‘AEs leading to dose reduction’, ‘AEs leading to permanent discontinuation of study drug’, and ‘AEs by highest CTCAE grade’.

The frequency of patients with adverse events will be summarised by treatment, primary system organ class (SOC) and preferred term (PT) of MedDRA (using the version that is current at the time of analysis). All tables will be sorted by SOC according to the standard sort order specified by the European Agency for the Evaluation of Medicinal Products (EMEA); PTs will be sorted by frequency (within SOC). Frequency tables will be created for the following events:

- All AEs
- Drug-related AEs
- AEs leading to dose reduction
- AEs leading to treatment discontinuation
- Drug-related AEs leading to treatment discontinuation
- Serious AEs
- Drug-related serious AEs
- Fatal AEs

These tables will be further categorised by highest CTCAE grade, including all grades, Grade 1-5.

In addition, preferred terms have been grouped to capture similar events, e.g. skin reactions that include rash, acne etc. Six grouped terms (rash/acne, stomatitis, ocular effects, lip effects, nail effects and fatigue) have been formulated for afatinib. These grouped terms are carefully reviewed by the Team Member Drug Safety for afatinib for each MedDRA version and maintained at the substance data base. The above frequency tables will be repeated using these grouped terms and remaining MedDRA PTs (i.e. the original PTs under these grouped terms will not be displayed). The tables will be sorted by descending frequency.

Additional AE tables will be produced for AEs of special interest (diarrhoea, the grouped term of rash/acne, stomatitis, renal insufficiency, hepatic impairment, and hematological events), providing further details on highest CTCAE grade, action taken with study drug and time to first onset of AE etc.

In recent trials, it has been observed that patients treated with an EGFR TKI who develop rash events tend to have better response and progression free survival. In order to explore this effect in this study, afatinib-treated patients will be grouped by the occurrence of CTCAE grade >= 2 rash/acne events and provide descriptive statistics for objective response and progression free survival. This analysis is exploratory in nature and not for comparison with methotrexate, therefore will be limited to afatinib patients only.

7.8.2 Laboratory data

The analyses of laboratory data will be descriptive in nature and will be based on BI standards [10]. Similar to AEs, lab values reported between drug start and 28 days after drug discontinuation will be analysed. CTCAE grades will be applied to applicable laboratory parameters according to NCI CTCAE v3.0.

Descriptive statistics of all laboratory values by visit will be provided including changes from baseline. Possible clinically significant abnormalities will be summarised by treatment group. For those lab parameters that have a CTCAE grading, possible clinically significant abnormalities are defined as post-baseline laboratory values with a CTCAE grade ≥ 2 that have had an increase of ≥1 grade from baseline. For those parameters for which no CTCAE grade has been defined, standard BI project definitions will be used to decide on clinical significance. Further frequency tables will show the transition of CTCAE grade from baseline to worst value and from baseline to last value on treatment.

The focus of the laboratory data analysis will be on the following laboratory parameters:

- Low values: haemoglobin, white blood cells, neutrophils, lymphocytes, platelets, potassium, sodium, GFR
- High values: creatinine, asparate aminotransferase (AST), alanine aminotransferase (ALT), bilirubin, alkaline phosphatase

7.8.3 Vital signs

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

A summary table of left ventricular ejection fraction (LVEF) will be produced for the final and minimum on-treatment values along with the corresponding percentage change from baseline. The number of patients with a significant LVEF event will also be presented. A significant event is defined as a decrease of ≥20% from baseline that is also below the lower limit of normal of the investigational site (50% will be used if the lower limit of normal is missing).

7.8.4 ECG

Not applicable as ECG measurement data are not collected in this study.

7.8.5 Others

Pharmacokinetic analyses (i.e. of afatinib pre- and post-dose plasma concentrations) will be performed according to Section 7.3.6 of the CTP.

8. REFERENCES

1 *001-MCS-50-409:* "Report Planning and Database Lock", current version, IDEA for CON.

2 *001-MCG-156_RD-01*: "Handling of missing and incomplete AE dates", current version; IDEA for CON.

3 *"Statistical Analysis Plan (SAP) for the external Data Monitoring Committee (DMC) of Boehringer Ingelheim Trial 1200.161"*, current version, BIRDS.

4 *"Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics",* US Department of health and Human Services, FDA, CDER, CBER. May 2007.

5 P Korhonen and J Palmgren. Effect modification in a randomized trial under non-ignorable non-compliance: an application to the alpha-tocopherol beta-carotene study, *Applied Statistics* (2002) 51, Part 1: 115-133 [R10-4999]

6 P Korhonen, N Laird and J Palmgren. Correcting for non-compliance in randomized trials: An application to the ATBC study, *Statistics in Medicine* (1999). 18, 2879-2897 [R10-5000]

7 J Robins and D Finkelstein. Correcting for Noncompliance and Dependent Censoring in an AIDS Clinical Trial with Inverse Probability of Censoring Weighted (IPCW) Log-Rank Tests, *Biometrics* (Sept 2000) 56, 779-788 [R10-4487]

8 J Robins. Information recovery and bias adjustment in proportional hazards regression analyses of randomized trials using surrogate markers, *Proceedings of the Biopharmaceutical Section, American Statistical Association* (1993) 24-33 [R10-4484]

9 *001-MCG-156*: "Handling and summarisation of adverse event data for clinical trial reports and integrated summaries", current version; IDEA for CON.

10 *001-MCG-157*: "Display and Analysis of Laboratory Data", current version, IDEA for CON.

11 *001-MCS-36-472*: “Standards and processes for analyses performed within Clinical Pharmacokinetics/Pharmacodynamics”, current version, IDEA for CON.

10. HISTORY TABLE

Table 10: 1 History table

Version	Date (DD-MMM-YY)	Author	Sections changed	Brief description of change
Final	**01-APR-16**	[redacted]	None	This is the final TSAP without any modification.