NOVARTIS

Clinical Development

AMN107 (Nilotinib)

CAMN107DDE06 / NCT00756509

An open-label, multi-center, single-arm study to evaluate the efficacy of nilotinib in adult patients with metastatic or unresectable gastrointestinal stromal tumors in first line treatment

Statistical Analysis Plan (SAP) of Follow-up phase

Author:	CRO Biometrician
Document type:	SAP Documentation (previously RAP Module 3) of Follow up Phase
Document status:	Final Version 1.2
Release date:	17 MAR 2025
Number of pages:	18

Document History – Changes compared to previous version of SAP (Follow up Phase)

Date	Reason for update	Outcome for update	Section and title impacted (Current)
26 JUL 2022	Final Version 1.0	Description of statistical analysis of the follow up phase of the study	N.A.
03 FEB 2025	Final Version 1.1	Renaming of document according to current requirements & addition to the analysis sets	Title page & 2.2
17 MAR 2025	Final Version 1.2	Definition of the visit display added	2.1.1

Table of contents

List of abbreviations

AE	Adverse event
ATC	Anatomical Therapeutic Classification
bid	bis in diem/twice a day
CR	Complete response
CRF	Case Report Form
CRO	Contract Research Organization
CSR	Clinical Study report
CTC	Common Toxicity Criteria
CTCAE	Common Terminology Criteria for Adverse Events
ECG	Electrocardiogram
ENR	Enrolled Patient population
ITT	Intention to treat
MedDRA	Medical Dictionary for Drug Regulatory Affairs
nsAE	non-serious adverse event
ORR	Objective response rate
OS	Overall Survival
[redacted]	[redacted]
PFS	Progression-Free Survival
PK/ PD	Pharmacokinetics/ Pharmacodynamic
PPS	Per-Protocol Set
PR	Partial response/ remission
RAP	Report and Analysis Process
RDI	Relative dose intensity
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	Serious adverse event
SAF	Safety Set
SD	Stable disease
TEAE	treatment-emergent adverse events
TFL	Tables, Figures, Listings
VAP	Validation and Analysis Plan
WHO-DRL	World Health Organization – drug reference list

1 Introduction

This document describes how the final analyses of the follow-up phase of the study will be implemented.

Based on the tables/figures/listings (TFL) resulting from this RAP M3 the Clinical Study Report (CSR) will be written. In addition, this RAP M3 describes which Patient Data Listings will be generated.

It is based on the Study Protocol version 06 (Amendment 5) dated 26-NOV-2018.

Primary analysis of the core study was already analyzed as indicated in the study protocol and SAP and includes all data until 6 months after LPFV. CSR for the core study is available (CAMN107DDE06_CSR_final1.0_20120330).

Primary analysis was based on

- CAMN107DDE06_RAP_M3_final1.0_20111118
- CAMN107DDE06_RAP_M7_final1.0_20111118.

1.1 Study design

This was a multicenter, single-arm, phase II trial evaluating the efficacy of Nilotinib in adult patients with unresectable or metastatic gastrointestinal stromal tumors.

The individual treatment phase lasted up to 6 months with monthly visits. Patients who profited from the treatment (SD, PR or CR) were offered to continue on the study drug during a follow up phase.

Following the decision to close-out the Novartis-Sponsored studies CAMN107G2301 and CAMN107DDE05, the enrollment of the study CAMN107DDE06 was re-opened in order to ensure continued access to nilotinib to the patients currently in the CAMN107G2301 trial and CAMN107DDE05 trial in Germany and benefiting from the nilotinib treatment. Patients who moved from CAMN107G2301 or CAMN107DDE05 studies to CAMN107DDE06 study entered directly in the follow-up phase of study CAMN107DDE06.

The study was designed as an open-label, single-arm trial, hence there is no randomization.

No interim analysis was planned.

1.2 Study objectives and endpoints

- The **Primary objective** relates to data from baseline to month 6 and was analyzed within the core analysis (summarized in the core study CSR).
- Additionally to the primary endpoint of the core study, the proportion of patients with a best overall response of CR or PR or SD will be derived using all available tumor assessments, i.e. including assessments beyond Week 24/Month 6 for patients continuing to receive treatment in the follow-up phase.

Secondary objectives:

- To evaluate (best) overall response by/ based on
  - end of follow-up
  - local assessment by end of follow-up by mutation status
  - First occurrence (CR or PR)
- Objective status based on local assessment by visit for follow-up phase
- To evaluate the objective tumor response rate based on RECIST criteria (complete response (CR) and partial response PR)
- To evaluate time to overall response (PR or CR)
- To evaluate duration of response
  - Time to overall response (CR or PR) according to RECIST
- To evaluate progression free survival (PFS)
  - under treatment
  - according to RECIST
- To evaluate overall survival (OS)
- To evaluate the safety and tolerability of Nilotinib as measured by rate and severity of adverse events

2 Statistical methods

2.1 Data analysis general information

Primary efficacy and safety analyses have been conducted on all patient data at the time all patients who were still receiving study drug had completed at least 6 months of treatment. Analysis based on 6 months data was performed by Novartis and is summarized in the core CSR dated 30.03.2012.

Additional data for any patients continuing to receive study drug past this time, as allowed by the protocol, were further summarized once these patients completed the study. The data of the follow-up phase will be analyzed by [redacted]. Analysis will be carried out using the SAS (Statistical Analysis System) software, version 9.2 or higher, SAS Institute Inc., Cary, North Carolina, USA.

All statistical analyses were performed by, or under supervision of, Novartis Pharma, Nuremberg, Germany. The protocol did not envisage data analyses carried out independently by the investigator; if performed, they were recommended to be submitted to Novartis before publication or presentation.

Data were summarized with respect to demographic and baseline characteristics, efficacy observations and measurements, and safety observations and measurements. The patient sets for analysis, variables, and methods for analysis were selected for their relevance for testing the hypotheses of the study and are described in more detail below.

In general, **categorical data** will be presented by absolute and relative frequencies. Percentages will always be calculated based on the total number N of the corresponding analysis population, unless otherwise stated. In general, missing values are considered for calculation of percentages if not otherwise specified.

For the analyses of the dosage and adverse events, the number m of events is displayed in addition to the number n of patients. In this case the percentages for number of events will always be calculated based on the total number of events and percentages for number of patients will be calculated based on the total number of patients.

For **continuous data**, mean, standard deviation, minimum, median, and maximum will be presented.

The data from each center are intended to be pooled with data from other centers conducted under this protocol so that an adequate number of patients will be available for analysis.

2.1.1 General definitions

The term “study treatment” is used for Nilotinib 400 mg bid.

“Study start” is defined as date of informed consent.

Baseline/ Screening is labelled as ‘Visit 1’. “Baseline value” is defined as all data documented at Visit 1. The visit with the first intake of treatment (day 1) is labelled as ‘Visit 2’.

The “follow-up phase” starts at visit FUP 1.

“Last contact” is defined as date of study completion follow-up form or, if not applicable, date of study completion form for the core study.

Assignment of the dosing administration records to the visits will be done using the visit dates. If entries of the follow up phase do not correspond to visit dates, they will be assigned to the nearest visit according to the suggested visit schedule. Therefore, the time periods between the visits will be divided into halves. Visit dates up to and including the date separating the time periods will be assigned to the previous visit. Visit dates later than the date line will be assigned to the later visit. If more than one assessment could be assigned to one visit, only the closest one will be chosen.

In all per visit displays, visits will not be sorted by visit number, but by the week planned for this visit. This procedure is based on the fact that the visit schedule was changed during the field phase as part of Protocol Amendment 2 (CSP V3.0, release date 26-October-2009) and therefore the numbering of the visits up to visit 207 does not correspond to the order of the

assessments in all cases, as was identified during dry run review. Sorting by the week planned for this visit, results in a clear sequence except for week 36, which may occur twice. In this case (one has been identified), further sorting will be based on the visit name (first FUP 1, then FUP 2).

This results in the following sequence of visits that was confirmed to correspond to the sequence of the examinations:

Week	Visit number
28	201
36 (FUP 1)	205
36 (FUP 2)	202
44	203
48	206
52	204
60	207
72	208
84	209
96	210
108	211
120	212
etc.	

2.2 Analysis sets

The **Enrolled Patient Population (ENR)** includes all patients who signed an informed consent regardless whether they received study treatment or not, except screening failures.

The **Intent-to-treat (ITT)** comprises all patients from the ENR who received at least one dose of study drug and have at least one post-baseline assessment of the primary efficacy variable (tumor assessment according to RECIST). Patients without any post-baseline assessment of tumor will be included if they are defined as progressive disease based on clinical evaluation.

The **Intent-to-treat follow-up (ITT_F)** comprises all patients from the ITT who entered the follow-up phase.

The **Safety Set (SAF)** consists of all patients who received at least one dose of study drug and had at least one post-baseline safety assessment. Of note: the statement that a patient had no adverse events (on the Adverse Event CRF) constitutes a safety assessment. Patients who have

received at least one dose of study drug but who have no post-treatment safety data of any kind would be excluded from the safety set.

The **Safety follow-up set (SAF_F)** consists of all patients from the SAF who entered the follow-up phase.

Roll-over patients (RO) are those patients who moved from the studies [CAMN107G2301] or [CAMN107DDE05]. They skipped the core study phase and entered directly into the follow up. They are not included in above analysis sets but form a separate analysis set.

Roll-over patients are included in the **Safety roll-over analysis set (SAF_RO)**, who received at least one dose of study drug in the follow-up phase.

2.2.1 Subgroup of interest

No subgroup analyses are planned.

2.3 Patient disposition, demographics and other baseline characteristics

2.3.1 Patient disposition

Patient disposition for patients in Follow-up will be displayed by frequencies of patients completed and discontinued study by center for the ITT_F. In addition, patient disposition will be displayed for roll-over patients.

The frequency of patients who discontinued the follow-up phase of the study prematurely is displayed and the reason for discontinuation is analyzed according to CRF-given categories. These frequencies will be calculated based on the ITT_F population.

Furthermore, the number of patients by visit will be displayed by frequency tables for the SAF_F.

Frequencies of major and minor protocol violations in the follow-up phase will be reported for the ITT_F.

2.3.2 Demographic and other baseline data

Demographic and other baseline data (including disease characteristics) was analyzed within the core analysis and is summarized in the core study CSR.

Prior medical conditions and relevant current medical conditions (active problem at first administration of study drug) have been summarized descriptively in the core study CSR and will not be repeated in this evaluation. In the analysis of the follow-up phase demographic data (and vital signs at Screening) will be displayed for the ITT_F and roll-over patients, if data is available.

Concomitant COVID-19 diseases during follow-up phase will be listed.

In addition, listings for demographic and other baseline data as well as for medical history will be provided for the ITT – patients who entered the FUP will be flagged.

2.4 Treatments (study treatment, rescue medication, concomitant therapies, compliance)

2.4.1 Study treatment / compliance

All analyses of treatment (study drug, concomitant therapies) during follow-up phase will be performed for the Safety set.

Duration (days) of exposure to study medication will be summarized using descriptive statistics for the safety set. Dosage averages will be calculated including and excluding zero doses for periods of temporary interruption of treatment regardless of whether this was due to safety reasons or patients' non-compliance. Mean daily dose levels will be summarized descriptively. Dose intensity will be summarized descriptively. The number of dose adjustments and patients with dose adjustments will be presented by reason for dose adjustment using frequency distributions.

In the same way, the number of dose reductions (including temporary dose interruption) will be presented.

2.4.2 Prior, concomitant and post therapies

Concomitant medications including vaccination against COV-19 and significant non-drug therapies beginning with visit FU 1 will be coded according to the WHO Drug Reference List and summarized by ATC class and preferred term using frequency distributions.

2.5 Analysis of the primary objective

Primary objective was already analyzed within the core analysis and is summarized in the core study CSR.

2.5.1 Primary endpoint

The primary efficacy variable was defined as the proportion of patients with a best overall response of CR or PR or SD by Week 24/Month 6 based on local assessment according to RECIST (Version 1.0).

Additionally, the proportion of patients with a best overall response of CR or PR or SD will be derived using all available tumor assessments, i.e. including assessments beyond Week 24/Month 6 for patients continuing to receive treatment in the follow-up phase.

The best overall response for each patient will be determined from the sequence of overall lesion responses according to the following rules (see Post-text supplement 3 to the study protocol):

- CR = at least two determinations of CR at least 4 weeks apart before progression.
- PR = at least two determinations of PR or better at least 4 weeks apart before progression (and not qualifying for a CR).
- SD = at least one SD assessment >6 weeks after start of treatment (and not qualifying for CR or PR).
- PD = progression or death due to underlying cancer lower or equal to 12 weeks after start of treatment (and not qualifying for CR, PR or SD). Patients with symptoms of rapidly progressing disease without radiologic evidence will be classified as progression only when clear evidence of clinical deterioration is available and patient discontinued due to 'Disease progression'.
- UNK = all other cases (i.e. not qualifying for confirmed CR or PR and without SD after more than 6 weeks or early progression within the first 12 weeks)

The proportion of patients in whom a best overall response of CR, PR or SD was observed will be presented with the one-sided exact 90% (80% two-sided) confidence interval for the ITT_F analysis set. The confidence interval will be computed using the exact method of Clopper and Pearson.

If progression has not been documented and one or more (non-)target lesions have not been assessed or have been assessed using a less sensitive method than baseline, the response status was to be determined as UNK according to RECIST.

Although planned in the protocol, the radiologic assessment according to Choi criteria was not performed as per protocol amendment 3. Thus, the supportive analysis of the primary objective as well as secondary analyses based on radiologic assessments according to Choi criteria were not performed.

2.6 Analysis of secondary efficacy objective(s)

2.6.1 Secondary endpoints

Secondary efficacy variables for the core and follow up phase will be analyzed exploratively using the ITT_F.

If applicable, secondary objectives will be analyzed using the Kaplan-Meier method. Results will be presented by Kaplan- Meier statistics as well as the Kaplan-Meier curve. For convenience, any tests will be performed at a two-sided significance level of 5%. The appropriate two-sided 95% confidence intervals will be displayed. Adjustments for multiplicity were neither planned in the protocol nor performed.

2.6.2 Statistical hypothesis, model, and method of analysis

Objective response rate (ORR) was defined as the proportion with a best overall response of CR or PR by Month 6. The ORR was presented with exact 95% confidence intervals (Clopper-

Pearson method). ORR was already analyzed within the core analysis and is summarized in the core study CSR.

Disease control rate (DCR) was defined as the proportion of patients with a best overall response of CR or PR or SD by Month 6. The DCR was presented together with the exact 95% confidence interval. DCR was already analyzed and is summarizes in the core study CSR.

In the analysis of the follow-up phase, absolute and relative frequencies of patients with clinical benefit (CR/PR/SD) was observed and will be presented together with the 95% confidence interval for the ITT_F, taking into account all available assessments (including those beyond Month 6). In the same way frequencies for each category (CR/PR/SD/PD/UNK) will be presented.

The following objectives were already analyzed within the core analysis and are summarized in the core study CSR. In addition, they will also be analyzed for patients who entered follow up phase (ITT_F):

Time to response is defined as the time from start of treatment to the first objective tumor response (PR or CR) observed. Patients who did not achieve a confirmed PR or CR will be censored at last adequate tumor assessment date when they did not progress (including deaths not due to underlying disease). Time to response will be summarized by time intervals and explored graphically by presenting the Kaplan-Meier curve.

Duration of response is defined as the time from onset of response (CR/PR) to objective tumor progression or death from any cause. Patients not experiencing progression or death will be censored with the date of their last adequate tumor assessment. Duration of response will be explored graphically by presenting the Kaplan-Meier curve.

Progression-free survival (PFS) is defined as the time from first study drug administration to objective tumor progression or death from any cause. If a patient has not had an event, PFS is censored at the date of last adequate tumor assessment. PFS will be explored graphically by presenting the Kaplan-Meier curve.

Overall survival (OS) is defined as the time from first study drug administration to death due to any cause. If a patient is not known to have died, survival will be censored at the date of last contact. OS will be explored graphically by presenting the Kaplan-Meier curve.

2.6.3 Handling of missing values/censoring/discontinuations

If progression has not been documented and one or more (non-)target lesions have not been assessed or have been assessed using a less sensitive method than baseline, the response status will be determined as UNK according to RECIST. Subjects with a best overall response of UNK will be included in the denominator but counted as non-responders.

2.7 Safety analyses

Safety of the core phase was already analyzed and is included in the core phase's CSR.

Adverse events that occurred during FU phase will be listed cumulatively to the events in the core phase. Laboratory data and other safety data that were evaluated during FU phase will be listed cumulatively (update).

Adverse events of roll-over patients, who skip the core study phase and enter directly into the safety follow up, will be listed separately.

2.7.1 Adverse events (AEs)

All adverse events (AE) recorded during the study will be listed cumulatively. Adverse events were coded by primary system organ class and preferred term according to the Medical Dictionary for Regulatory Activities (MedDRA). In case of missing information regarding seriousness, the AE will be defined as non-serious (nsAE). In case of missing relationship, the AE will be defined as not related.

In the data listings of adverse events, the severity of an AE, the relationship to study drug, whether or not it is a serious AE and the action taken will be indicated. Adverse events occurred during follow-up will be flagged.

2.7.1.1 Adverse events of special interest / grouping of AEs

No (S)AEs of special interest are defined.

2.7.2 Deaths

Deaths will be listed by patient.

2.7.3 Laboratory data

All laboratory values will be converted into SI units and the severity grade calculated using appropriate toxicity criteria (mild, moderate, severe). Laboratory data will be listed cumulatively.

2.7.4 Other safety data

2.7.4.1 ECG and cardiac imaging data

ECG results will be listed.

2.7.4.2 Vital signs

Vital signs, including weight, were presented in the Core analysis. In the follow-up analysis Vital signs will be listed.

2.8 Pharmacokinetic endpoints

According to protocol amendment 5 the pharmacokinetic analysis was removed from the protocol as no new results are expected. Thus, no pharmacokinetic endpoints for the follow up phase are analyzed.

2.9 PD and PK/PD analyses

Not applicable.

2.10 Patient-reported outcomes

Not applicable.

2.11 Biomarkers

[redacted]

No additional analysis of biomarker will be performed for FU phase.

[redacted]

2.12 Other Exploratory analyses

Not applicable.

2.13 Interim analysis

Not applicable. Interim analyses were neither planned nor performed.

3 Sample size calculation

The study followed an exact binomial single-stage design (A'Hern 2001). The study required 39 evaluable subjects to decide whether the proportion responding, $\boldsymbol{p}$, was less than or equal to $\boldsymbol{p_0}$ = 50% or greater than or equal to $\boldsymbol{p_1}$ = 70%.

If the number of responses was 24 or more, the hypothesis that p is less than or equal to $\boldsymbol{p_0}$ = 50% was to be rejected with a target error rate α of 10% and an actual error rate of 10%. If the number of responses was 23 or less, the hypothesis that p is greater than or equal to $\boldsymbol{p_1}$ = 70% was to be rejected with a target error rate β of 10% and an actual error rate of 9.4%. The design was estimated using the procedure for single-stage phase II trials of NCSS Trial and PASS 2002. Power for analysis of secondary variables was not assessed.

4 Change to protocol specified analyses

Analyses of Study treatment / compliance will be performed using the Safety Set instead of the ITT Population as the duration and dose of study medication intake are safety concerns.

CTCAE grading does not exist for adverse events and laboratory values. Severity categories: mild, moderate and severe will be displayed.

5 Appendix

5.1 Imputation rules

5.1.1 Study drug

The relative dose intensity is calculated as follows:
RDI = DDI/SDI x 100 [%],
where **SDI** (Standard Dose Intensity) = planned dose for total treatment time/total time to complete therapy and **DDI** (Delivered Dose Intensity) = actually delivered dose/total time to complete therapy.

DDI is understood as the mean daily dose (including zero doses for periods of dose interruption).

5.1.2 AE date imputation

In case of missing or partial start date or end date of an AE, the start date should be shifted to the earliest possible date, but not earlier than the first intake of study drug while the end date should be shifted forward as far as possible.

5.1.3 Concomitant medication date imputation

In case of missing or partial start date or end date of a concomitant medication, the start date should be shifted to the earliest possible date, but not earlier than the first intake of study drug while the end date should be shifted forward as far as possible.

5.1.3.1 Prior therapies date imputation

Not applicable.

5.1.3.2 Post therapies date imputation

In case of missing or partial start date or end date of a post therapy, the start date should be shifted to the earliest possible date, but not earlier than the first intake of study drug while the end date should be shifted forward as far as possible. The date of first intake of study drug is the date of visit 2 (day 1).

5.1.3.3 Other imputations

Not applicable.

5.2 AEs coding/grading

Adverse events will be assessed by the severity of mild, moderate, severe.

5.3 Laboratory parameters derivations

See 2.8.3.

5.4 Statistical models

5.4.1 Primary analysis

No statistical hypothesis or model was underlying the analysis.

5.4.2 Key secondary analysis

Not applicable.

5.5 Rule of exclusion criteria of analysis sets

Protocol deviation specifications are given in full detail in the Data-Review-Meeting Minutes and the VAP Module 2.

Considering the current situation with the Covid-19 pandemic, a new PD category has to be implemented: 7 = Other. The first PD in this category is defined as O_01.

Using the string 'PD due to COVID-19' for the description of the deviation itself, all relevant PDs can be clearly identified to be listed in the CSR.

To fulfill the requirements of the Novartis Coronavirus SWAT the following subcategories need to be implemented:

Handling of manual PDs due to COVID-19

1st subcategory:

	PD Description
1	Missed visit due to COVID-19
2	Visit not at study site due to COVID-19
3	Assessment / procedure changed due to COVID-19
4	Drug supply method changed due to COVID-19
5	Treatment not given due to COVID-19
6	Discontinuation due to COVID-19
7	Other

2nd subcategory:

	Predefined relationship options	**When to use**

a	COVID-19 health status related	i.e. patients' infection did lead to this PD
b	COVID-19 situation: Site issue	e.g. site closed, personnel not available
c	COVID-19 situation: Lockdown/ Quarantine of patient	e.g. site is active but patient not allowed to come
d	COVID-19 situation: Patient concern	e.g. site is active, patient could come but refused to come / do assessment
e	COVID-19 situation: Drug supply issue	e.g. drug was delivered to home
f	COVID-19 situation: Other	e.g. situation not already covered by the information above

6 Reference

Not applicable.

NOVARTIS

Clinical Development

AMN107 (nilotinib)

CAMN107DDE06

An open-label, multi-center, single-arm study to evaluate the efficacy of nilotinib in adult patients with metastatic or unresectable gastrointestinal stromal tumors in first line treatment

RAP Module 3 – Detailed Statistical Methodology

Author: [redacted] Statistician

Document type: RAP Documentation

Document status: Final 1.0

Release date: 18-Nov-2011

9.7 Statistical methods planned in the protocol and determination of sample size

9.7.1 Statistical and analytical plans

Primary efficacy and safety analyses were conducted on all patient data at the time all patients who were still receiving study drug had completed at least 6 months of treatment. Additional data for any patients continuing to receive study drug past this time, as allowed by the protocol, were further summarized once these patients completed the study. All statistical analyses were performed by, or under supervision of, Novartis Pharma, Nuremberg, Germany. The protocol did not envisage data analyses carried out independently by the investigator; if performed, they were recommended to be submitted to Novartis before publication or presentation. Data were summarized with respect to demographic and baseline characteristics, efficacy observations and measurements, and safety observations and measurements. The patient sets for analysis, variables, and methods for analysis were selected for their relevance for testing the hypotheses of the study and are described in more detail below.

9.7.1.1 Patient sets for analysis

The **intent-to-treat set (ITT)** consisted of all patients who received at least one dose of study drug and had at least one post-baseline assessment of the primary efficacy variable (tumor assessment according to RECIST 1.0). Patients without any post-baseline assessment of tumor were included if they were defined as progressive disease based on clinical evaluation.

The **per-protocol set (PP)** consisted of all patients of the intent-to-treat population who did not show any major protocol deviation. The per-protocol set was identified before database lock.

The **safety set** consisted of all patients who received at least one dose of study drug and had at least one post-baseline safety assessment. Of note, the statement that a patient had no adverse events (on the Adverse Event CRF) constituted a safety assessment. Patients who had received at least one dose of study drug but who had no post-treatment safety data of any kind were to be excluded from the safety set.

9.7.1.2 Patients and treatments

Demographic and other baseline data

Demographic and other baseline data (including disease characteristics) were summarized for the ITT set. Categorical data were presented as frequencies and percentages. For continuous data, mean, standard deviation, minimum, median, and maximum were presented.

Medical history was coded using MedDRA. Frequencies were presented by system organ class and MedDRA preferred term. Separate tables were provided for past medical condition and current medical condition.

Treatments (study drug, concomitant therapies)

All analyses of treatment (study drug, concomitant therapies) were performed for the Safety set. Duration (days) of exposure to study medication was summarized using descriptive statistics for the safety set. Dosage averages were calculated including and excluding zero doses for

periods of temporary interruption of treatment regardless of whether this was due to safety reasons or patients' non-compliance. Mean daily dose levels were summarized descriptively. The number of dose adjustments and patients with dose adjustments was presented by reason for dose adjustment using frequency distributions. Permanent treatment discontinuations were analyzed by frequencies.

Concomitant medications and significant non-drug therapies prior to and after the start of the study drug were coded the WHO-DRL and were summarized by drug class and preferred term. Separate tables were provided for prior therapies (ending before start of study drug) and concomitant therapies. Prior antineoplastic medications were coded using the WHO-DRL and were summarized by drug class and preferred term using frequency distributions. Prior antineoplastic surgeries and radiotherapies were not coded. The number of prior antineoplastic surgeries/patients with antineoplastic surgeries was summarized by reason for surgery using frequency distributions. Prior antineoplastic radiotherapies were summarized analogously by setting.

9.7.1.3 Efficacy evaluation

Primary efficacy evaluation

The primary efficacy variable was defined as the proportion of patients with a best overall response of CR or PR or SD by Week 24/Month 6 based on local assessment according to RECIST (Version 1.0). Additionally, the proportion of patients with a best overall response of CR or PR or SD was derived using all available tumor assessments, i.e. including assessments beyond Week 24/Month 6 for patients continuing to receive treatment in the follow-up phase.

The best overall response for each patient was determined from the sequence of overall lesion responses according to the following rules (see Post-text supplement 3 to the study protocol):

- CR = at least two determinations of CR at least 4 weeks apart before progression.
- PR = at least two determinations of PR or better at least 4 weeks apart before progression (and not qualifying for a CR).
- SD = at least one SD assessment >6 weeks after start of treatment (and not qualifying for CR or PR).
- PD = progression or death due to underlying cancer lower or equal to 12 weeks after start of treatment (and not qualifying for CR, PR or SD). Patients with symptoms of rapidly progressing disease without radiologic evidence will be classified as progression only when clear evidence of clinical deterioration is available and patient discontinued due to 'Disease progression'.
- UNK = all other cases (i.e. not qualifying for confirmed CR or PR and without SD after more than 6 weeks or early progression within the first 12 weeks)

The proportion of patients in whom a best overall response of CR, PR or SD was observed was presented with the one-sided exact 90% (80% two-sided) confidence interval for the ITT (eligible) set. The confidence interval was computed using the exact method of Clopper and Pearson. The trial was designed to test the null hypothesis that the proportion of patients with a best overall response of CR, PR or SD, p, is less than or equal to p_0 = 50% against the alternative hypothesis that the proportion is greater than or equal to p_1 = 70%. If the number of responses

was 24 or more, the hypothesis that *p* is less than or equal to 50% was to be rejected with a target error rate α of 10% and an actual error rate of 10%. If the number of responses was 23 or less, the hypothesis that *p* is greater than or equal to 70% was to be rejected with a target error rate β of 10% and an actual error rate of 9.4%.

If progression has not been documented and one or more (non-)target lesions have not been assessed or have been assessed using a less sensitive method than baseline, the response status was to be determined as UNK according to RECIST.

The proportion of patients in whom a CR, PR or SD was observed was additionally presented with the appropriate confidence interval for the per-protocol set. [redacted]

Although planned in the protocol, the radiologic assessment according to Choi criteria was not performed as per protocol amendment 3. Thus, the supportive analysis of the primary objective as well as secondary analyses based on radiologic assessments according to Choi criteria were not performed.

Secondary efficacy evaluation

Secondary efficacy variables were analyzed exploratively for the ITT set. Any tests were performed at a two-sided significance level 5%. Adjustments for multiplicity were neither planned in the protocol nor performed. For convenience, the appropriate two-sided 95% confidence intervals were computed. For all safety analyses, the safety analyzable population was used.

Objective response rate (ORR) was defined as the proportion with a best overall response of CR or PR by Month 6. The ORR was presented with exact 95% confidence intervals (Clopper-Pearson method).

Disease control rate (DCR) was defined as the proportion of patients with a best overall response of CR or PR or SD by Month 6. The DCR was presented together with the exact 95% confidence interval.

Time to response is defined as the time from start of treatment to the first objective tumor response (PR or CR) observed. Patients who did not achieve a confirmed PR or CR will be censored at last adequate tumor assessment date when they did not progress (including deaths not due to underlying disease). Time to response will be explored graphically by presenting the Kaplan-Meier curve.

Duration of response is defined as the time from onset of response (CR/PR) to objective tumor progression or death from any cause. Patients not experiencing progression or death will be censored with the date of their last adequate tumor assessment. Duration of response will be explored graphically by presenting the Kaplan-Meier curve.

Progression-free survival (PFS) is defined as the time from first study drug administration to objective tumor progression or death from any cause. If a patient has not had an event, PFS is censored at the date of last adequate tumor assessment. PFS will be explored graphically by presenting the Kaplan-Meier curve.

Overall survival (OS) is defined as the time from first study drug administration to death from any cause. If a patient is not known to have died, survival will be censored at date of last contact. OS will be explored graphically by presenting the Kaplan-Meier curve.

Although planned in the protocol, the analysis of objective response rate, time to response, duration of response, and progression-free survival according to Choi criteria was not performed. Per amendment no. 3, the assessment of radiologic images according to Choi criteria was deleted without replacement.

Pharmacokinetic evaluations (change / add PD, PK/PD, Biomarkers, as needed)

Pharmacokinetics

Pharmacokinetic assessments were analyzed separately from the clinical data and reported in an appendix to the Clinical Study Report..

Biomarkers

[redacted]

Safety evaluation

Safety was evaluated using assessment of adverse events and laboratory data. The assessment of safety was based mainly on the frequency of adverse events and on the number of laboratory values that are new or worsening based on the common toxicity criteria (CTCAE) grade. Other safety data (e.g. vital signs) were considered as appropriate. Any statistical tests performed to explore the data were used only to highlight any interesting comparisons that may warrant further consideration.

All adverse events recorded during the study were summarized. Adverse events were coded by primary system organ class and preferred term according to the Medical Dictionary for Regulatory Activities (MedDRA). The incidence of treatment-emergent adverse events (new or worsening from baseline) were summarized by the number and percentage of patients in each primary system organ class and preferred term. For summaries by severity (based on CTCAE grades) of event, the most severe occurrence for a particular preferred term was used for a given patient. In the data listings of adverse events, the severity of an AE, whether or not an AE is study drug related, and whether or not it is a serious AE was indicated. Deaths, all adverse events, serious adverse events, AE leading to study drug discontinuation, AE causing dose adjustment or interruption, and AE requiring additional therapy were listed.

All laboratory values were converted into SI units and the severity grade calculated using appropriate CTCAE criteria. Laboratory data were summarized by presenting summary statistics of raw data and changes from baseline values, by presenting worsening tables and shift tables using CTCAE grades as well as normal ranges (baseline to most extreme post-baseline value).

Data from other tests (e.g., electrocardiogram or vital signs) were summarized descriptively and listed as appropriate, and any other information collected was listed as appropriate.

Interim analyses

Not applicable. Interim analyses were neither planned nor performed.

Other topics

No other topics were studied.

9.7.2 Determination of sample size

The study followed an exact binomial single-stage design (A'Hern 2001). The study required 39 evaluable subjects to decide whether the proportion responding, $\boldsymbol{p}$, was less than or equal to $\boldsymbol{p_0}$ = 50% or greater than or equal to $\boldsymbol{p_1}$ = 70%. If the number of responses was 24 or more, the hypothesis that p is less than or equal to 50% was to be rejected with a target error rate α of 10% and an actual error rate of 10%. If the number of responses was 23 or less, the hypothesis that p is greater than or equal to 70% was to be rejected with a target error rate β of 10% and an actual error rate of 9.4%. The design was estimated using the procedure for single-stage phase II trials of NCSS Trial and PASS 2002. Power for analysis of secondary variables was not assessed.