NOVARTIS

Clinical Development

RAD001

CRAD001C24114 / NCT00767819

Multicenter, triple-arm, single-stage, phase II trial to determine the preliminary efficacy and safety of RAD001 in patients with histological evidence of progressive or metastatic bone or soft tissue sarcomas

RAP Module 3 – Detailed statistical methodology

Author: [redacted]

Document type: RAP Documentation

Document status: Final 1.0

Release date: 07-Jul-2011

Novartis approval signatures for:

RAP Module 3 – Detailed statistical methodology

[redacted] Signature Date

[redacted] Signature Date

1 Data analysis

This was a non-randomized, multicenter, triple-arm, phase II trial evaluating the treatment of RAD001 in the following three arms: in patients with progressive or metastatic bone or soft tissue sarcoma (except for GIST), in patients with progressive or metastatic GIST after failure or intolerance of treatment with imatinib or sunitinib, and in patients with progressive or metastatic alveolar soft part sarcoma (ASPS), respectively.

The study was designed to assess the activity, safety, and tolerability of the therapy with RAD001 in progressive bone or soft tissue sarcoma in the **three** arms each.

The data were analyzed by Novartis. Any data analysis carried out independently by the investigator(s) was recommended to be submitted to Novartis before publication or presentation. It was planned that the data from participating centers in this protocol were combined within each of the arms, so that an adequate number of patients was available for analysis.

Data were summarized with respect to demographic and baseline characteristics, efficacy observations and measurements, safety observations and measurements, and pharmacokinetic measurements. Categorical data were presented as absolute frequencies and percentages. For continuous data, N (the number of valid observations), N miss. (the number of missing observations), mean, standard deviation, median, minimum, and maximum were presented.

It was planned that the data of each arm were to be analyzed and reported independently of the other arm once the number of patients needed for analysis was reached. However, the methods described in this analysis plan apply for all three arms, if not otherwise indicated.

For simplification, populations for analysis, and the analysis of patient demographics/other baseline characteristics, treatment, primary and secondary efficacy variables, and safety were described without referring to a specific arm. As mentioned above, the methods of analysis were applied to each arm separately.

Changes from the analysis pre-planned in the protocol:

1. Although planned in the protocol, the central radiologic review was only incompletely performed due to a variety of reasons (no images submitted from local to central radiologist, when already the first tumor assessment resulted in an overall lesion response of ‘progressive disease’; or images were not assessable according to the central radiologist). An incomplete sequence of central assessments cannot be used to derive the primary and secondary efficacy variables. Thus, the trial will be analyzed based on local tumor assessments only.
2. The intention-to-treat (ITT) set will be used for the primary efficacy analysis rather than the per-protocol (PP) set, since this is what is recommended by RECIST criteria (version 1.0, see Therasse et al. 2000: ‘All patients included in the study must be assessed for response to treatment, even if there are major protocol treatment deviations or if they are ineligible. [...]All conclusions should be based on all eligible patients.’)

1.1 Populations for analysis

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. Patients were analyzed according to treatment received. Of note, the statement that a patient had no adverse events also constituted a safety assessment.

The **intent-to-treat (ITT) set** 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 (target/non-target lesion assessment according to RECIST). Patients without any post-baseline assessment of tumor were included if they were defined as progressive disease based on clinical evaluation.

The **per-protocol (PP) set** consisted of all patients of the intent-to-treat population who did not show any major protocol violations. As major protocol violations were considered those that may have an impact on the study outcome. Criteria that were assumed to have such an impact were defined in a Review Meeting before database lock.

1.2 Patient demographics/other baseline characteristics

Demographic and background information were summarized for the ITT set, using frequency distributions for categorical variables and descriptive statistics of mean, standard deviation, minimum, median and maximum for continuous variables. Background information included prior medication, past/current medical conditions, diagnosis and extent of cancer, ECOG performance status and tumor evaluation at baseline.

Medical history was coded using MedDRA and was presented by system organ class, and MedDRA preferred term. Separate tables were provided for past medical condition and current medical condition. Prior medications (including prior antineoplastic medications and radiotherapy) were coded according to WHO Drug Reference List (WHO-DRL). Prior medications were summarized by ATC class and preferred term.

1.3 Treatments (study drug, rescue medication, other concomitant therapies, compliance)

Study drug and concomitant therapies were analyzed for the safety set.

Study drug

Duration of exposure to study drug was defined as the time from start of treatment to the last end date documented in the Study Drug Administration CRF. Duration of exposure was summarized descriptively.

Mean daily doses of study drug 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 doses were summarized descriptively.

The number of dose changes/patients with dose changes (including temporary dose interruption) was presented by reason for dose change by frequency distribution. Permanent treatment discontinuations were analyzed by frequencies.

Concomitant medications

Concomitant medications were coded according to WHO-DRL and summarized by ATC class and preferred term using frequency tables.

1.4 Analysis of the primary objective(s)

1.4.1 Variable

The primary variable is defined as the proportion of patients in whom a best overall response of complete (CR) or partial (PR) response or stable disease (SD) was observed at 16 weeks according to RECIST (version 1.0, Therasse et al. 2000) based on local radiologic assessments.

The best overall response for each patient was determined from the sequence of overall (lesion) responses at Visit 4/Week 8 and Visit 6/Week 16, respectively, according to the following rules (see Post-text supplement 1 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 ≤ 16 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'. Furthermore, patients without radiologic evidence for progression who were prematurely discontinuing from the study for 'unsatisfactory therapeutic effect', 'new cancer therapy' or 'death' (from any reason) will be classified as 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 16 weeks)

1.4.2 Statistical hypothesis, model, and method of analysis

The best overall response rate (CR/PR/SD/PD/UNK) was presented together with the appropriate two-sided confidence interval for the ITT set. Since the one-sided alpha was set to 10%, two-sided 80% confidence intervals were presented. Exact confidence intervals were computed according to the method of Clopper and Pearson.

The absolute number of patients showing a best overall response of CR, PR or SD was determined to conclude preliminary activity or non-activity of the study drugs in this patient population according to the rules outlined in CSP section 10.2. Furthermore, the lower limit of the confidence interval was used to support the decision in favor of p_0 or p_1: if the lower limit of the confidence interval overlapped p_0, the hypothesis that p is greater than or equal to p_1 could be rejected; on the other side, if the lower limit of the confidence interval excluded p_0, the hypothesis that p is greater than or equal to p_1 could be accepted.

1.4.3 Handling of missing values/censoring/discontinuations

The derivation of the primary efficacy variable from the sequence of overall lesion responses recorded at each assessment followed the Novartis guidance 'Guidelines for Response, Duration of Overall Response, TTF, TTP, Progression-Free Survival and Overall Survival (based on RECIST)'. Subjects not qualifying for confirmed CR or PR and without SD after more than 6 weeks or early progression within the first 12 weeks were classified as 'UNK' for their best overall response.

1.4.4 Supportive analyses

The number and percentage of patients showing a best overall response of CR, PR or SD as well as the number and percentage of patients for each category (CR/PR/SD/PD/UNK) was additionally presented for the per-protocol (PP) set.

Additionally, for patients who entered the follow-up period starting after Visit 6/Week 16, the best overall response was determined using all available local tumor assessments.

1.5 Analysis of secondary objectives

1.5.1 Efficacy (secondary)

Secondary efficacy variables were analyzed for the ITT set.

Objective response rate (ORR) was defined as the proportion of patients in whom a complete (CR) or partial (PR) response was observed at Visit 6/Week 16 according to RECIST (version 1.0) based on central radiologic review. Absolute and relative frequencies were presented together with the appropriate exact confidence interval.

Duration of stable disease (CR, PR or SD) applied only to those patients whose best overall response was CR, PR or SD based on local radiologic assessments and was defined as the time from start of treatment to progression or death from underlying disease. Patients not experiencing progression or death at 16 weeks were censored with the date of their last tumor assessment. Duration of response was explored using the Kaplan-Meier method. The Kaplan-Meier estimate for median duration of response as well as the Kaplan-Meier curve was displayed. Although duration of response (CR or PR) was planned to be analyzed within the protocol, this was changed to reach a more valuable analysis since no patient reached a best overall response of CR or PR.

Progression-free survival (PFS) was defined as the time from the date of start of treatment to the date of event defined as the first documented progression or death from any cause. If a patient had not had an event, PFS was censored at the date of the last adequate tumor assessment, which was the date of Visit 6 for the core study phase and the last available tumor assessment for the follow-up phase. PFS was explored by using the Kaplan-Meier method. The Kaplan-Meier estimate for median PFS as well as the Kaplan-Meier curve was displayed.

Time to progression (TTP) was defined as the time from the date of start of treatment to the date of event defined as the first documented progression or death from underlying disease. If a patient had not had an event, TTP was censored at the date of the last adequate tumor assessment, which was the date of Visit 6 for the core phase and the last available tumor

assessment for the follow-up phase. TTP was explored by using the Kaplan-Meier method. The Kaplan-Meier estimate for median TTP as well as the Kaplan-Meier curve was displayed.

Overall survival (OS) was defined as the time from the date of start of treatment to death from any cause. If a patient was not known to have died, OS was censored at the date of the last contact, which was the date of Visit 6 for the core phase and the last available visit for the follow-up phase. OS was explored by using the Kaplan-Meier method. The Kaplan-Meier estimate for median OS as well as the Kaplan-Meier curve was displayed.

1.5.2 Safety

The assessment of safety was based mainly on the frequency of adverse events and on the number of laboratory values that fall outside of pre-determined ranges. Other safety data (e.g. electrocardiogram, vital signs, and special tests) were considered as appropriate.

1.5.2.1 Adverse events

Adverse events (AE) were summarized by presenting the number and percentage of patients having any adverse event, having an adverse event in each body system and having each individual adverse event. Furthermore, the incidence of AE was summarized by maximum severity, for AE with suspected drug relation, for serious AE, for AE leading to permanent discontinuation of study drug, and for significant AE. As significant (S)AE were defined those (S)AE who

- were suspected to be drug related,
- lead to dose adjustment or temporary interruption,
- lead to permanent discontinuation,
- required concomitant medication/non-drug therapy, and
- Deaths

Any other information collected (e.g. severity or relatedness to study medication) was listed as appropriate.

1.5.2.2 Laboratory values

Laboratory data were summarized by presenting summary statistics of raw data and changes from baseline value by visit. Although foreseen in the protocol, notable values were not analyzed since no notable ranges were defined in the protocol. Any statistical tests performed to explore the data were used only to highlight any interesting comparisons that may warrant further consideration.

Data from other tests (e.g. electrocardiogram or vital signs) were planned to be listed, notable values were recommended to be flagged (if appropriate), and any other information collected was planned to be listed as appropriate.

1.5.3 Tolerability

Not applicable.

1.5.4 Resource utilization

Not applicable.

1.6 Sample size calculation

The study followed an exact binomial single-stage design in each arm. (A'Hern 2001). Values of P_0 and P_1 followed the recommendation of the EORTC Soft Tissue and Bone Sarcoma Group's previous publication evaluating PFS in STS patients treated with 2nd line active and inactive compounds.

In arm I (patients with bone and soft tissue sarcoma except for GIST), the study required 36 evaluable subjects to decide whether the proportion responding (best overall response of CR, PR or SD), p, was less than or equal to $p_0 = 20\%$ or greater than or equal to $p_1 = 40\%$ (Table 10-1). In arm II or III (patients with GIST after failure or intolerance of 1st and 2nd line treatment with imatinib or sunitinib, and patients with progressive or metastatic alveolar soft part sarcoma (ASPS), respectively), the study required 24 evaluable subjects to decide whether the proportion responding best overall response of CR, PR or SD), p, was less than or equal to $p_0 = 20\%$ or greater than or equal to $p_1 = 40\%$.

If the number of responses in arm I was 11 or more, the null hypothesis that p is less than or equal to 20% was to be rejected with a target type I error rate of 10% and an actual error rate of 8.9%. If the number of responses in arm I was 10 or less, the alternative hypothesis that p is greater than or equal to 40% was to be rejected with a target type II error rate of 10% and an actual error rate of 9%.

If the number of responses in arm II or arm III, respectively, was 8 or more, the null hypothesis that p is less than or equal to 20% was to be rejected with a target type I error rate of 10% and an actual error rate of 8.9%. If the number of responses in arm II was 7 or less, the alternative hypothesis that p is greater than or equal to 40% was to be rejected with a target type II error of 20% and an actual error rate of 19.2%.

The designs were estimated with NCSS Trial and PASS 2002.

Table 10-1 Design features for exact binomial single-stage designs

	Arm I	Arm II	Arm III
p_0 (maximum response proportion of an inactive drug)	20%	20%	20%
p_1 (minimum response proportion of an active drug	40%	40%	40%
Type I error (one-sided test)	10%	10%	10%
Type II error (one-sided test)	10%	20%	20%
Sample size	36	24	24
Accept inactivity at the end if	≤ 10 responders	≤ 7 responders	≤ 7 responders
Reject inactivity at the end if	≥ 11 responders	≥ 8 responders	≥ 8 responders
Actual type I error level	8.9%	8.9%	8.9%
Actual type II error level	9.0%	19.2%	19.2%

The type I error is the probability of rejecting the hypothesis that p is lower or equal to p_0 when this is true. **The type II error** is the probability of rejecting the hypothesis that p is greater or equal to p_1 when this is true. Response was defined as a best overall response of CR, PR or SD.

1.7 Power for analysis of critical secondary variables

Not applicable.