

**Official Title:** A PHASE II, MULTICENTER, RANDOMIZED, DOUBLE-BLIND STUDY TO EVALUATE THE EFFICACY AND SAFETY OF RO5520985 (VANUCIZUMAB) PLUS FOLFOX VERSUS BEVACIZUMAB PLUS FOLFOX IN PATIENTS WITH PREVIOUSLY UNTREATED METASTATIC COLORECTAL CANCER

**NCT Number:** NCT02141295

**Document Date:** SAP Version 1.1: 09-Sep-2016

## STATISTICAL ANALYSIS PLAN

**TITLE:** A PHASE II, MULTICENTER, RANDOMIZED, DOUBLE-BLIND STUDY TO EVALUATE THE EFFICACY AND SAFETY OF RO5520985 PLUS FOLFOX VERSUS BEVACIZUMAB PLUS FOLFOX IN PATIENTS WITH PREVIOUSLY UNTREATED METASTATIC COLORECTAL CANCER

**PROTOCOL NUMBER:** BP29262

**STUDY DRUG:** RO5520985

**VERSION NUMBER:** 1.1

**IND NUMBER:** 120910

**EUDRACT NUMBER:** 2013-005108-32

**SPONSOR:** F. Hoffmann-La Roche Ltd

**PLAN PREPARED BY:** [REDACTED]

**DATE FINAL:** August 18, 2016

**DATE(S) AMENDED** *09 September 2016*

### CONFIDENTIAL

This is a F.Hoffman –La Roche Ltd document that contains confidential information. Nothing herein is to be disclosed without written consent from F.Hoffman –La Roche Ltd.

09 September 2016: List of prognostic variable for subgroup analysis has been expanded on clinical science request.

## TABLE OF CONTENTS

1.	BACKGROUND .....	5
2.	STUDY DESIGN .....	5
2.1	Protocol Synopsis .....	7
2.2	Outcome Measures .....	7
2.2.1	Primary Efficacy Outcome Measures .....	7
2.2.2	Secondary Efficacy Outcome Measures .....	8
2.2.3	Exploratory Efficacy Outcome Measures .....	8
2.2.4	Pharmacokinetic Efficacy Outcome Measures .....	9
2.2.5	Safety Outcome Measures .....	9
2.3	Determination of Sample Size .....	9
2.4	Analysis Timing .....	10
3.	STUDY CONDUCT .....	11
3.1	Randomization Issues .....	11
3.2	Data Monitoring .....	11
4.	STATISTICAL METHODS .....	11
4.1	Analysis Populations .....	11
4.1.1	ITT Population .....	11
4.1.2	Pharmacokinetic-Evaluable Population .....	11
4.1.3	Safety Population .....	12
4.2	Analysis of Study Conduct .....	12
4.3	Analysis of Treatment Group Comparability .....	12
4.4	Efficacy Analysis .....	13
4.4.1	Primary Efficacy Endpoint .....	13
4.4.2	Secondary Efficacy Endpoints .....	14
4.4.3	Sensitivity Analyses .....	14
4.4.3.1	Unstratified Analysis for Progression-Free Survival	14
4.4.3.2	PFS including death outside the 30 days window .....	14
4.4.3.3	PFS on study .....	15
4.4.3.4	Censoring pattern evaluation .....	15

4.4.3.5	Time to surgery evaluation .....	15
4.4.3.6	Prognostic factors assessment .....	15
4.4.3.7	Non confirmed ORR.....	16
4.4.4	Subgroup Analyses .....	16
4.4.4.1	Demographic and baseline prognostic characteristics .....	16
4.4.4.2	Ang2 baseline biomarker assessment.....	16
4.5	Pharmacokinetic and Pharmacodynamic Analyses .....	17
4.5.1	Pharmacokinetic analysis .....	17
4.5.2	PK/PD Modeling .....	18
4.5.3	Tumor Growth Kinetics .....	18
4.6	Safety Analyses .....	18
4.6.1	Exposure of Study Medication .....	18
4.6.2	Adverse Events .....	18
4.6.3	Laboratory Data .....	19
4.6.4	Vital Signs.....	19
4.6.5	ECG Data Analysis .....	19
4.6.6	Concomitant Medications .....	19
4.7	Biomarker .....	20
4.8	Missing Data .....	20
4.9	Interim Analyses .....	20
5.	REFERENCES .....	21

## LIST OF TABLES

Table 1	Analysis Timings .....	10
---------	------------------------	----

## LIST OF APPENDICES

Appendix 1	Protocol Synopsis .....	22
Appendix 2	Schedule of Assessments .....	23
Appendix 3	{Additional Items to Be Appended} .....	24

## **1. BACKGROUND**

Study BP29262 is a phase II multicenter, randomized, parallel groups, double-blind study of RO5520985 to evaluate the efficacy and safety of RO5520985 plus mFOLFOX-6 versus bevacizumab plus mFOLFOX-6 in patients with previously untreated mCRC.

The target population consists of adult patients with metastatic colorectal cancer who have not previously been treated with chemotherapy for metastatic disease and who are not candidates for potentially curative resection.

The study consists of two parts: Part I (Safety Run-in) open label single arm and Part II (Randomization Phase) randomized, parallel groups, double-blind study. A safety evaluation will be performed before starting Part II of the study. Upon completion of the safety run in and selection of the recommended Phase II dose (RP2D), a total of 190 patients will be enrolled in Part II of the study.

The primary objective of Study BP29262 is to estimate the efficacy of RO5520985 in combination with oxaliplatin, folinic acid, and 5 fluorouracil (mFOLFOX-6) vs. bevacizumab in combination with mFOLFOX-6, as measured by progression-free survival (PFS).

During Part II of the study an interim analysis for futility is planned after approximately 30 progression-free (PFS) events. Administrative interim analysis, based on efficacy, will also be performed during Part II of the study after approximately 30 events and again after 50 events.

A safety Internal Monitoring Committee (sIMC) will conduct a safety analysis prior to initiation of Part II and an efficacy Internal Data Monitoring Committee (eIMC) will oversee the futility and administrative interim analyses during Part II. The sIMC and the eIMC will operate according to pre-specified agreements. The efficacy and the safety IMC Agreements define the roles and responsibilities of the IMCs including membership, scope, timing of meetings, and communication plan. Details of the interim analysis are described in the interim analysis plan.

This SAP applies to Part II of the study only and describes the final analysis plan for the primary and secondary endpoint.

## **2. STUDY DESIGN**

Upon completion of the safety run-in part with selection of the RP2D of RO5520985 in combination with mFOLFOX-6 by the IMC, eligible patients will be randomized in a ratio of 1:1 to receive either mFOLFOX-6 +RO5520985 (experimental arm) or mFOLFOX-6 +

bevacizumab (control arm). Patients will prospectively be stratified by region (United States vs. Rest of World [RoW]) and number of metastatic sites (1 vs. >1).

Patients randomized to the experimental arm will receive the fixed dose of RO5520985 IV on Day 1 of each Cycle on a Q2W schedule as recommended by the IMC based on Part I safety data review. Patients randomized to the control arm will receive bevacizumab at a dose of 5 mg/kg IV on Day 1 of each Cycle on a Q2W schedule. The weight at screening will be used for dose calculation for each patient. Recalculation of dosage is not required if weight changes.

Study treatment will be given in cycles repeated every 14 days and will consist of induction and maintenance therapy.

Induction therapy: up to 8 cycles of mFOLFOX-6 plus either bevacizumab or RO5520985. Patients may switch to maintenance therapy earlier if oxaliplatin cannot be tolerated.

Maintenance therapy: following induction therapy, oxaliplatin administration will be discontinued and patients receive 5-FU/folinic acid plus either RO5520985 (experimental arm) or bevacizumab (control arm) as maintenance therapy for a maximum period of 24 months (calculated from start of maintenance therapy), if treatment is not stopped earlier due to disease progression, unacceptable toxicity, investigator decision or consent withdrawal. If patients stop chemotherapy either in part or in whole, then they should continue on RO5520985 treatment or bevacizumab for a maximum of 24 months. Details of the induction and maintenance therapy can be found in section 3.1.1.2 of study protocol.

Tumor assessments should be performed every 8 weeks until documented disease progression. Should a patient discontinue from treatment for any reason other than progression, an End of Study CT/MRI is to be performed only if it has not been done  $\leq$  28 days prior. Patients with a global deterioration of health status requiring discontinuation of treatment without objective evidence of disease progression at that time should be reported as “symptomatic deterioration”. Every effort should be made to document objective progression even after discontinuation of treatment.

If a patient discontinues therapy for reasons other than progression, the patient will be followed by regular CT assessments until documentation of progressive disease, initiation of another anticancer therapy, withdrawal of consent, or death

All patients who discontinue from the treatment phase will be followed for subsequent anticancer therapy and survival (e.g., by phone call) approximately every 3 months after end of study visit (EoS) Visit until death, loss to follow-up, or study termination by Roche, whichever occurs first unless the patient requests to be withdrawn from study survival follow-up..

## **2.1 PROTOCOL SYNOPSIS**

The Protocol Synopsis is in Appendix 1. For additional details, see the Schedule of Assessments in Appendix 2.

## **2.2 OUTCOME MEASURES**

For those patients who discontinue study treatment before progression, all assessments conducted before start of new therapy (including surgery with curative or de-bulking intent or radiotherapy) will be accounted for the evaluation even if they were performed after stopping study treatment.

### **2.2.1 Primary Efficacy Outcome Measures**

- The primary efficacy endpoint of this study is PFS on study. PFS on study is defined as the time between randomization and the date of first documented disease progression or death from any cause on study, whichever occurs first. Progression will be based on tumor assessment made by the Investigator according to RECIST 1.1 criteria (Appendix 8 of the protocol). Death on study is defined as death from any cause within 30 days of the last study treatment. Patients without an event on study will be censored at the date of the last tumor assessment when the patient was known to be progression free either during follow up or during study treatment. Patients without any post baseline assessments or with all post baseline assessments having unknown result/response but known to be alive at the clinical cut off for the analysis will be censored at the date of randomization plus 1 day.
- In addition, a more generic definition of PFS will be employed, similar to the primary PFS analysis, with the difference that deaths regardless of whether or not the event occurred within 30 days of the last study treatment will be taken into account. As sensitivity analysis PFS is defined as the time between randomization and the date of first documented disease progression or death from any cause, whichever occurs first. Progression will be based on tumor assessments made by the Investigator according to RECIST 1.1 criteria. Patients without an event on study will be censored at the date of the last tumor assessment when the patient was known to be progression free either during follow up or during study treatment. Patients without any post baseline assessments or with all post baseline assessments having unknown result/response but known to be alive at the clinical cut-off for the analysis will be censored at the date of randomization plus 1 day.
- For sensitivity analysis a more restrictive definition of PFS on study will be employed, same as for the primary PFS analysis with the difference that only assessments or death that occurred within 30 days of the last study treatment will be taken in account. As sensitivity analysis PFS on study is defined as the time between randomization and the date of first documented disease progression or death from any cause, whichever occurs first. Progression will be based on tumor assessments made by the Investigator according to RECIST 1.1 criteria. Patients without an event



on study will be censored at the date of the last tumor assessment when the patient was known to be progression free either during study treatment or during follow up but within the 30 days from last study treatment. Patients without any post baseline assessments or with all post baseline assessments having unknown result/response but known to be alive at the clinical cut-off for the analysis will be censored at the date of randomization plus 1 day.

### **2.2.2 Secondary Efficacy Outcome Measures**

- Objective Response Rate (ORR). Objective response rate (ORR) is determined as the rate of patients with an objective tumor response (complete [CR] or partial response [PR]). Objective response (OR) is defined as a complete or partial response as determined by the Investigator using RECIST v1.1 on two consecutive occasions at least 4 weeks apart. Patients without a post-baseline tumor assessment will be regarded as non-responders.
- Duration of Objective Response (OR). For patients with an OR, duration of OR is defined as the time from the initial response (CR or PR) to disease progression, or death from any cause on study.- This will only be calculated for patients who have a best overall response of CR or PR. Methods for handling censoring and for analysis are the same as those described for PFS on study.
- For sensitivity analysis a more generic definition of Duration of Objective Response (OR). will be employed, same as for the above OR analysis with the difference that deaths will be included regardless of whether or not the event occurred within 30 days of the last study treatment. As sensitivity analysis OR is defined as the time from the initial response (CR or PR) to the date of first documented disease progression or death from any cause, whichever occurs first. Progression will be based on tumor assessment made by the Investigator according to RECIST 1.1 criteria. This will only be calculated for patients who have a best overall response of CR or PR. Methods for handling censoring and for analysis are the same as those described for PFS.
- Overall Survival (OS). OS is defined as the time from randomization until death from any cause. All deaths will be included, regardless whether they occur on study or following treatment discontinuation. For patients who have not died, OS will be censored at the last date known to be alive. Patients without any post baseline information will be censored at the time of randomization.
- Depth of response. Depth of response defined as best percentage change from baseline in tumor shrinkage will also be assessed.

### **2.2.3 Exploratory Efficacy Outcome Measures**

- Levels of circulating targets by comparing baseline vs. on-treatment values of Ang-2 and VEGF-A and their differential expression/regulation during RO5520985 versus bevacizumab therapy.

- Changes of Placenta Growth Factor (PIGF) from baseline during treatment as marker of unspecific endothelial stress response and as potential response towards RO5520985
- Assessment of tissue and soluble blood markers potentially related to predicting clinical benefit or toxicity of RO5520985 vs. bevacizumab
- An assessment of tumor growth kinetics will be made by comparing post-treatment scans with one or more pre-treatment scans not older than 12 weeks prior to C1D1. The two pre-treatment scans consist of a pre-study scan and the study baseline scan. If not available, tumor growth kinetics will be applied by comparing post-treatment scans with the baseline scan. OS prediction for comparison between the 2 arms will be performed.
- Analysis of additional CT parameters including necrotic volume/total tumor volume and intensity to assess their potential value as early response prediction markers.

#### **2.2.4 Pharmacokinetic Efficacy Outcome Measures**

- The PK profile of RO5520985 will be characterized with the plasma concentration-time data following IV administration of RO5520985, and may include the following parameters: C<sub>max</sub>, C<sub>min</sub>, T<sub>max</sub>, t<sub>1/2</sub>, AUC, AUC<sub>τ</sub>, CL, V<sub>ss</sub>, accumulation ratio (RA), as applicable
- The PK profiles of oxaliplatin (free and total) and 5-FU will be characterized with the plasma concentration-time data following IV administration of FOLFOX, and will include the following parameters: C<sub>max</sub>, C<sub>min</sub>, T<sub>max</sub>, t<sub>1/2</sub>, AUC (when applicable)
- Additional PK parameters may be evaluated as appropriate

#### **2.2.5 Safety Outcome Measures**

- Exposure to study drug medication
- Adverse events and adverse events of special interest
- Laboratory data
- Vital sign
- ECG
- Concomitant medication
- 

### **2.3 DETERMINATION OF SAMPLE SIZE**

Part II of the study will enroll approximately 190 patients, and the primary analysis will be performed after approximately 80 investigator assessed PFS events. The emphasis of the efficacy analysis will be on estimation of the magnitude of treatment effect rather than hypothesis testing. This trial is hypothesis-generating and is designed to be able to detect a meaningful benefit of the combination therapy of RO5520985 plus mFOLFOX-6 versus bevacizumab plus mFOLFOX-6. Based on the sample size of 80 events observed in the two treatment arms combined, there is an 80% power to detect a HR of 0.574 at a one-sided significance level of 0.05.

Initially 140 patients were planned to be enrolled in the study. The sample size was increased to 190 with a subsequent amendment in October 2015. At the cut-off date of 10 August 2015, the number of observed PFS events (14 events in 10 months) was significantly lower than expected. This seemed to indicate that the median progression free survival (PFS) for the control arm of bevacizumab plus FOLFOX-6 may be longer than the assumed 10 months originally planned for sample size calculation and study duration.

Additionally, as of 10 August 2015, a drop-out rate of approximately 20% of enrolled patients for reasons other than PD/death had been observed. It is of note that Adverse Events accounted for a drop out of 8% of enrolled patients, whereas 13% were other reasons, predominantly physician decisions, but also cases of secondary resectability. On 05 June 2015, the SIMC (safety internal monitoring committee) performed a blinded safety analysis of study BP29262 and came to the conclusion that the overall observed toxicity was acceptable and balanced between both treatment arms, including the incidence of drop-outs due to AE.

To mitigate the potential risk of not achieving the required number of PFS events (n=80) for the primary analysis, approximately 50 additional patients were additionally enrolled into study BP29262.

## 2.4 ANALYSIS TIMING

**Table 1 Analysis Timings**

Analysis	Timing of Analysis	Percent Information
First interim for futility	30 events	37.5%
Second interim administrative	50 events	62.5%
Final	80 events	100%

The primary analysis will occur when approximately 80 PFS events have been observed. The clinical cut off is defined as the date at which Roche is informed of the 80<sup>th</sup> event. At the time of PFS analysis OS data will not be sufficiently mature. OS data analysis will be performed only at the end of the study. The study will formally end once the survival follow-up is complete or the last patient has completed the EoS Visit or is withdrawn from the study prior to that time (whichever occurs last), but may be prematurely terminated by the Sponsor.

The study will not be stopped for evidence of efficacy. Futility and administrative interim analysis are described in the Interim Analysis Plan.

### **3. STUDY CONDUCT**

#### **3.1 RANDOMIZATION ISSUES**

In part II patients will be randomized in a double-blinded fashion in a ratio of 1:1 to receive mFOLFOX-6 + RO5520985 (experimental arm) or mFOLFOX-6 + bevacizumab (control arm). Randomization will be performed using a permuted block randomization and patients will prospectively be stratified by region (United States vs. RoW) and number of metastatic sites (1 vs. > 1). Randomization will be implemented via IxRS system.

#### **3.2 DATA MONITORING**

A safety Internal Monitoring Committee (sIMC) will conduct a safety analysis prior to initiation of Part II and an efficacy Internal Data Monitoring Committee (eIMC) will oversee the futility and administrative interim analyses during Part II. The sIMC and the eIMC will consist of employees of Roche and will operate according to pre-specified agreements. The efficacy and the safety IMC agreements define the roles and responsibilities of the IMCs including membership, scope, timing of meetings, and communication plan. Details of the interim analysis are described in the interim analysis plan.

### **4. STATISTICAL METHODS**

The analyses outlined in this Statistical Analysis Plan will supersede those specified in the protocol,

For all analysis the clinical cutoff date will be applied for all data and not the snap shot date i.e. data occurring after the clinical cutoff date will not be included in the analysis.

#### **4.1 ANALYSIS POPULATIONS**

##### **4.1.1 ITT Population**

The ITT population It will include all patients who were randomized (Part II only) and received any amount of study treatment (5-FU/folic acid, oxaliplatin, bevacizumab, or RO5520985). Patients will be assigned to the treatment arm to which they were randomized.

##### **4.1.2 Pharmacokinetic-Evaluable Population**

PK analyses will include all patients who received any amount of study treatment (5-FU/folinic acid, oxaliplatin, bevacizumab, or RO5520985), the same as the safety analysis population.

Patients may be excluded from the PK analysis population if they significantly violate the eligibility criteria, deviate significantly from the protocol or if data are unavailable or incomplete which may influence the pharmacokinetic analysis. Excluded cases will be

documented together with the reason for exclusion. All decisions on exclusions from the analysis will be made prior to database closure.

#### **4.1.3 Safety Population**

The safety analysis population will include all patients who receive any amount of study treatment (5-FU/folic acid, oxaliplatin, bevacizumab, or RO5520985). Patients will be assigned to treatment arm based on the study treatment they actually received: patients randomized to bevacizumab will be included in the RO5520985 arm if they received at least one dose of RO5520985.

Patients randomized to RO5520985 or bevacizumab who received only 5-FU/folic acid or oxaliplatin but no RO5520985 nor bevacizumab will not be included in the overall summaries by treatment arm but will be listed separately.

#### **4.2 ANALYSIS OF STUDY CONDUCT**

Descriptive statistics will be used in evaluating the conduct of the study. Enrollment, study treatment administration, and discontinuations from the study will be summarized by part and treatment arm for all ITT randomized patients. The summaries will include the following:

Total number of cycles and doses of study treatment administered (mFOLFOX-6 , RO5520985 or bevacizumab).

Reasons for discontinuation of stud treatment (mFOLFOX-6 , RO5520985 or bevacizumab).

A summary of the first non-protocol specified cancer therapy by treatment arm (including surgery with curative or debulking intent, or radiotherapy) after study discontinuation will be provided.

In addition study enrollment will be summarized by treatment arm for all randomized patients.

A listing of major protocol deviations will be produced.

#### **4.3 ANALYSIS OF TREATMENT GROUP COMPARABILITY**

The following demographic and baseline disease characteristics and cancer history will be summarized by treatment arm for all ITT patients:

Demographic and baseline disease characteristics:

sex, age, weight, height, BMI, time since first diagnosis, ECOG at baseline, number of metastatic sites (1, >1), adjuvant therapy (Y/N), sum of all lesions at baseline, sum of target lesion at baseline, primary tumor location (left/right side as per clinical science

classification), primary tumor location (all terms), primary tumor in place (Y/N), tumor staging, tumor histology grade and type, peritoneal disease (Y/N), RAS and BRAF mutation (Y/N/UNK), plasma Ang2 Dx baseline value.

In addition to the above prior cancer treatment, including surgery and radiotherapy, and medical history will also be summarized by treatment arm.

The baseline value of any variable will be defined as the last available value prior to the first administration of study drug.

#### **4.4 EFFICACY ANALYSIS**

This Phase II trial is designed to make a preliminary evaluation of the efficacy and safety of RO5520985 plus mFOLFOX-6 versus bevacizumab plus mFOLFOX-6 in patients with previously untreated mCRC. The emphasis of the efficacy analysis will be on estimation of the magnitude of treatment effect rather than hypothesis testing.

##### **4.4.1 Primary Efficacy Endpoint Progression-Free Survival by Investigator Assessment**

The primary endpoints are defined in section 2.2.1.

Kaplan-Meier methods will be used to estimate median PFS for each treatment arm and the 95% CIs for median PFS will be computed using the Brookmeyer and Crowley method.

The stratified Cox proportional hazard model will be used to estimate the hazard ratio (i.e., the magnitude of the treatment effect) and the corresponding 95% confidence interval. The stratification factors are number of metastatic sites (1 vs. > 1) and country/region (USA vs rest of the world).

P-values of comparing the two treatment arms from stratified and unstratified log-rank test will also be presented; however, given that the focus of the study is on estimation of treatment P values should not be interpreted as confirmatory testing. Treatment comparison will be based on a two sided level 0.05 stratified log rank test on all ITT patients.

Kaplan-Meier curve will be provided for a visual description of the difference over time between the two treatment arms. IVRS stratification information data will be used for the analysis.

The reasons for PFS censoring will be summarized.

For the primary PFS (death on study) the treatment comparison based on a one sided level 0.05 stratified log rank test on all ITT patients will also be reported.

#### **4.4.2            Secondary Efficacy Endpoints**

Secondary efficacy endpoints include OS, ORR confirmed, duration of objective response, depth of response and 6 months and 12 months PFS rate as defined in section 2.2.2. All the key secondary endpoints will be tested at two sided 0.05 levels and will not have adjustment for multiplicity. The p values from these analyses should be interpreted accordingly and should not be interpreted as confirmatory testing.

##### **Objective Response Rate (ORR)**

An estimate of the objective response rate and 95% confidence intervals (Casella 1986) will be calculated for each treatment arm. The difference in objective response rate between the two treatment arms and its confidence intervals (Wald, 95% CI) will be calculated as well.

##### **Duration of Objective Response (OR)**

Methods for handling censoring and for analysis are the same as those described for PFS. Kaplan-Meier curve will be provided for a visual description of the difference in duration of response over time between the two treatment arms. Descriptive summaries of duration of response by treatment arm will be produced. These descriptive summaries will consist of the Kaplan-Meier estimates of median duration of response and the unstratified HR and its 95% CI computed using the Brookmeyer and Crowley method. No adjustments will be made to account for the non random nature of this comparison

##### **Overall Survival (OS)**

Analysis methods are the same as those described for the PFS endpoint.

#### **4.4.3            Sensitivity Analyses**

##### **4.4.3.1        Unstratified Analysis for Progression-Free Survival**

As the primary analysis of progression-free survival is stratified, the unstratified analysis to compare the difference in progression-free survival between the two treatment arms, using a two-sided unstratified log-rank test at the 0.05 alpha level, will serve as a sensitivity analysis to check the robustness of the results.

The hazard ratio  $\lambda_{RO} / \lambda_{Bev}$  will be estimated using an unstratified Cox regression model.

##### **4.4.3.2        PFS including death outside the 30 days window**

As sensitivity analysis the analysis described in Section 4.4.1 will be applied to the more generic definition of PFS that also includes deaths occurring more than 30 days after last study treatment as defined in section 2.2.1.

#### **4.4.3.3 PFS on study**

As sensitivity analysis the analysis described in Section 4.4.1 will be applied to the more restrictive definition of PFS on study that includes only assessments occurring within 30 days after last study treatment as defined in section 2.2.1

#### **4.4.3.4 Censoring pattern evaluation**

Censoring reasons will be summarized and time to censoring will be evaluated across the two arms. Deaths occurring more than 30 days after last study treatment as described in section 2.2.1 will be considered as a PFS event.

Kaplan-Meier curve will be provided for a visual description of the difference in censoring over time between the two treatment arms. Descriptive summaries of median censoring time by treatment arm will be produced. These descriptive summaries will consist of the Kaplan-Meier estimates of median time to censoring and the unstratified HR and its 95% CI computed using the Brookmeyer and Crowley method.

#### **4.4.3.5 Time to surgery evaluation**

A summary of time to surgery with debulking or curative intent (mean median and range) will be provided.

#### **4.4.3.6 Prognostic factors assessment**

In order to further assess the effect of important prognostic variables on PFS, including Ang2 VEGF biomarker (see also section 4.4.4.2), a non-stratified Cox proportional hazard model will be fitted and methods like shrinkage for covariate selection may be applied to the following terms for best model selection:

- Treatment arm
- Age as continuous variable
- RAS mutation status as categorical (mutant, vs wild type),
- Primary tumor location as categorical (left vs right),
- Primary tumor in place (yes vs no)
- Time since first diagnosis as continuous variable in days,
- ECOG at baseline as categorical variable,
- Number of metastatic sites as categorical variable (1, >1)
- Biomarker Ang2 levels at baseline as categorical variable (high vs low) (dichotomized at median)
- Biomarker VegfA levels at baseline as categorical variable (high vs low, median cut off)
- Treatment by biomarker group

This analysis will only be applied to the generic definition of PFS (including all death irrespective of the 30 days window) as described in section 2.2.1



The effect of the biomarker as continuous variable may also be explored instead of the categorical biomarker high/low classification.

Missing values will be imputed as described in section 4.8

#### **4.4.3.7 Non confirmed ORR**

As sensitivity analysis not confirmed responses will also be summarized by treatment arm.

#### **4.4.4 Subgroup Analyses**

##### **4.4.4.1 Demographic and baseline prognostic characteristics**

The effects of demographic and baseline prognostic characteristics on PFS will be explored for all ITT patients. It is expected that accrual in the subgroups defined by these baseline characteristics will not be large enough for definitive treatment comparisons to be made among these subgroups.

- Race (white , non white)
- Region (USA , rest of the world)
- Number of metastatic sites (1 or >1)
- Baseline ECOG performance status
- Tumor location (left , right)
- RAS mutation status (mutant, vs wild type)
- Primary tumor location (left ,right)
- Primary tumor in place (yes, no)
- Age as categorical ( )
- Baseline target lesion size, mm (median) ( $\geq$  median, < median)
- Adjuvant treatment (Yes vs. No) Baseline Serum CEA ( $\leq 5$  vs.  $>5$  mg/dl)
- Age ( $\leq 65$  vs.  $>65$  years)
- Baseline Ang2 (high vs low)

Descriptive summaries of PFS will be produced for each level of the categorical variables listed above for each treatment arm. These descriptive summaries will consist of the Kaplan-Meier estimates of median time to event and the unstratified HR and its 95% CI. Number of patients and events in each subgroup will be summarized.

This analysis will only be applied to the generic definition of PFS as described in section 2.2.1

##### **4.4.4.2 Ang2 baseline biomarker assessment**

In addition to the analysis already specified in the ITT population, and in order to guide decisions on total ang2 plasma levels as a potential biomarker to predict treatment

response to RO5520985, a subgroup analysis based on patients' baseline ang2 plasma concentration (diagnostic assay) will also be performed. Patients will be grouped as Ang2 "high" or "low" based on their baseline total ang2 plasma concentrations levels with the median as the pre-specified cutoff point.

The subgroup analysis will be performed in the safety population..

The generic definition of PFS will be used for this analysis.

In each subgroup, the primary efficacy endpoint of PFS (including death irrespective of the 30 day time window period as described in section 2.2.1) will be compared between the two treatment arms (RO5520985 plus mFOLFOX6 vs. bevacizumab plus mFOLFOX6) and within each treatment arm between subgroups (high Ang2 vs low Ang2).

Kaplan-Meier methods will be used to estimate median PFS for each treatment arm or biomarker subgroup and the 95% CIs for median PFS will be computed using the Brookmeyer and Crowley method. The stratified Cox proportional hazard model will be used to estimate the hazard ratio (i.e., the magnitude of the treatment within each subgroup or between subgroup within each treatment) and the corresponding 95% confidence interval. The stratification factors are number of metastatic sites (1 vs.> 1) and country/region (USA vs rest of the world).

The reasons for PFS censoring with each group and treatment will be summarized.

Assessment of the effect of different cut off from the median on the unstratified PFS HR will also be conducted.

Biomarker data will also be plotted against other known important factors and stratification factors by treatment arm.

## **4.5 PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES**

### **4.5.1 Pharmacokinetic analysis**

All pharmacokinetic parameters will be presented by listings and descriptive summary statistics separately by group or cohorts. Individual and mean plasma RO5520985, oxaliplatin (free and total), and 5-FU concentration versus time data will be tabulated and plotted. The plasma pharmacokinetics of RO5520985, oxaliplatin, and 5-FU may be summarized by estimating total exposure (area under the curve [AUC]), maximum concentration, total clearance, volume of distribution at steady-state, and terminal half-life (when applicable). These parameters will be tabulated and summarized (arithmetic mean, standard deviation, geometric mean, coefficient of variation, median, minimum, and maximum). Interpatient variability and drug accumulation will be evaluated.

Additional PK analyses will be conducted as appropriate.

#### **4.5.2 PK/PD Modeling**

Exploratory graphical analyses of exposure-efficacy relationships may be produced for selected efficacy, PD and/or safety measurements if feasible. A PK/PD modeling approach may be considered in order to further explore the exposure-response relationship of selected response variables.

#### **4.5.3 Tumor Growth Kinetics**

An exploratory assessment of tumor growth kinetic will be made by comparing post treatment scans with pretreatment scans. The tumor growth kinetic methodology will be applied to predict overall survival in the two arms.

When available a pre-study scan (not older than 12 weeks prior to C1D1) and a study baseline scan will allow estimation of tumor growth rate before start of treatment and this will then be compared to the growth/shrinkage rate after start of treatment.

If only a limited number of patients had pre-study scan available, the tumor growth kinetic will be assessed by comparing post treatment scans to the baseline scan.

Data will be explored using linear and/or exponential models, as appropriate, using non linear mixed effect modelling software. Further details are provided in the appendix 3.

### **4.6 SAFETY ANALYSES**

Safety analyses will be performed for all patients in the safety analysis population. All safety parameters will be analyzed using descriptive statistics, summarized and presented in tables.

#### **4.6.1 Exposure of Study Medication**

Summaries of treatment administration will include the following:

- Total number of infusions for RO5520985 /Bevacizumab
- Number of infusions not administered or dose reduction for RO5520985 /Bevacizumab
- Total number of cycles received for mFOLFOX6
- Number of cycles not administered or dose reduction for each of the mFOLFOX6 regimen components.

#### **4.6.2 Adverse Events**

The original terms recorded on the eCRF by the Investigator for adverse events will be standardized by the Sponsor; for classification purposes, preferred terms will be assigned by the Sponsor to the original terms entered on the eCRF, using the most up-to-date version of the Medical Dictionary for Regulatory Activities (MedDRA) terminology

for adverse events and diseases. AEs will be graded according to the most up-to-date version of NCI Common Terminology Criteria for Adverse Events (NCI-CTCAE).

Adverse events will be listed and summarized by MedDRA system organ class and preferred term. Toxicity grade, seriousness and relationship to study treatment will be presented, as well as summaries of deaths, AEs leading to death and premature withdrawal from study treatment...

Glossary of adverse events, medication and procedures and

#### **4.6.3 Laboratory Data**

All clinical laboratory data will be stored on the database in the units in which they were reported. Patients' listings and summary statistics at each assessment time will be presented using the International System of Units (SI units; Système International d'Unités). Laboratory data not reported in SI units will be converted to SI units before processing. Laboratory test values will be presented by individual listings with flagging of values outside the normal ranges.

For laboratory data, summary tables of change from baseline over time based on SI (Standard International) units will be displayed. Shifts in toxicity grade from baseline to the worst grade observed during treatment will be presented for selected laboratory parameters.

Additional figures/tables/listings may be produced as deemed appropriate.

For standard reference ranges and transformation of data and definition of laboratory abnormalities refer to section 6.5.2 of the protocol.

#### **4.6.4 Vital Signs**

Vital signs data will be presented by individual listings with flagging of values outside the normal ranges and flagging of marked abnormalities. In addition, tabular summaries will be used, as appropriate. Additional figures/tables/listings may be produced as deemed appropriate.

#### **4.6.5 ECG Data Analysis**

ECG data will be presented by individual listings with flagging of values outside the normal ranges and flagging of marked abnormalities. In addition, tabular summaries will be used, as appropriate.

#### **4.6.6 Concomitant Medications**

The original terms recorded on the patients' eCRF by the Investigator for concomitant medications will be standardized by the sponsor. For classification purposes, preferred terms will be assigned by the Sponsor to the original terms entered on the eCRF, using the most up-to-date version of the d the International Non-proprietary Name (INN) Drug

Terms and Procedures Dictionary for treatments and surgical and medical procedures. Concomitant medications will be presented in listings and summary tables by treatment arm. .

#### **4.7 BIOMARKER**

Exploratory biomarker parameters may be presented by listings and descriptive summary statistics separately by treatment arms.

All analyses of PD and exploratory biomarkers will be based on the safety analysis population, in other words, any patients who received any amount of study treatment (5-FU/folic acid, oxaliplatin, bevacizumab, or RO5520985).

For the analysis of PD biomarkers the primary evaluation will be based on the observed change from baseline. Both actual values and estimated parameters will be presented in summary tables and graphically.

To assess predictability of a biomarker, the association between clinical outcome and the biomarker level or changes thereof will be explored. Details of the biomarker analysis will be provided within the biomarker analysis plan.

Analysis of Ang 2 biomarker data at baseline are provided in section 4.4.4.2.

Exploratory PKPD analysis may be performed for selected biomarker if appropriate.

#### **4.8 MISSING DATA**

In case of incomplete RECIST assessment's or death dates if the day is missing, but the month and year are available, then the day will be imputed as the first day of the month. No imputation will be done for missing month or year. For biomarker analysis values below the limit of quantitation (BLQ) will be imputed as half of the BLQ value. Details of the BLQ values employed for each biomarker are stored in the output specification document within the DAP module 2. For the categorization of patients in high or low biomarker subgroup any patient for which the biomarker value at baseline is missing because outside the upper range of quantitation of the assay will be considered as been in the high subgroup.

#### **4.9 INTERIM ANALYSES**

No stopping for efficacy is planned. For interim analysis details please refer to the interim analysis plan.

5. **REFERENCES**

# Appendix 1 Protocol Synopsis

## **Appendix 2 Schedule of Assessments**



## Appendix 3 Tumor growth kinetic analysis

### Tumor kinetics (TK) model

A non-linear mixed-effect model describing the time dynamics of tumor size (sum of lesion diameter) as a function of drug exposure will be developed on the basis of the data collected in each arm of the current study.

Adopting a structure which has already been considered in previous analysis of anti-angiogenic drug effect in mCRC patients [1,2], the following differential equation could be considered to describe such a time dynamics:

$$\frac{dy}{dt} = k_G y - \text{exposure} \cdot k_s \cdot y \cdot e^{-\lambda t} \quad (1)$$

with  $y$ , the tumor size,  $y(0) = y_0$ , the tumor size at baseline,  $k_G$ , the tumor growth rate,  $k_s$  the tumor shrinkage rate and  $\text{exposure}$ , the drug exposure level. The tumor shrinkage rate is expected to decrease over time (in this equation, exponentially, at a rate  $\lambda$ ).

Various alternatives models will be considered to describe drug exposure ranging from daily dose (K-PD approach) to individual concentration at any time  $t$ , as predicted from a population PK model.

This model will be used to extract patient-level tumor kinetics metrics such as time to growth (TTG) and relative change in tumor size (TSR) from baseline at first post-treatment visit ( $TTG_i$  and  $TSR_i$  respectively) already identified as relevant predictors of survival in mCRC (see [1,2] for further details).

### Validation of the TK model

In order to validate the TK model (1), it is proposed to leverage historical data from the phase III study NO16966 with bevacizumab/FOLFOX4. [3].

A tumor kinetics model (similar to the one described in the previous section) will be fitted to the NO16966 data collected in the 350 patients treated with combination therapy with FOLFOX-4+Bev. From this model, the individual estimates of tumor kinetics metrics will be derived.

A Kolmogorov-Smirnov test (or equivalent) will be used to compare the distributions of  $TTG_i$  and  $TSR_i$  in patients treated with chemotherapy+Bev in studies BP29262 and NO16966.

To sustain this comparison, it will assumed that the effects of FOLFOX-4 and mFOLFOX6 on the tumor size dynamics are not markedly different [4].

### Prediction of overall survival

As observed in multiple clinical trials with anti-angiogenic treatments, changes in tumor size correlates strongly with OS [5]. Since exposure is explicitly influencing tumor shrinkage in model (1), the drug effect is carried over in the survival model throughout the estimates of individual tumor kinetics (e.g.  $TTG_i$ ). Hence, in a survival model

applied to mCRC patients, tumor kinetics metrics can be used as surrogate marker of the drug effect.

In order to predict the overall survival in patients treated with mFOLFOX6+Van or Bev, it is proposed again to rely on the NO16966 study where 351 patients were allocated to FOLFOX-4+placebo and 350 to FOLFOX-4+bevacizumab and for which survival data is available.

A tumor kinetics model (similar to the one described in the previous section) will be fitted to the tumor size data collected in these 701 patients. This model will be used to extract individual estimates of tumor kinetics metrics.

Then, a parametric survival model will be developed to describe the distribution of survival time in study NO16966. The following tumor kinetics covariates will be tested for inclusion in the model:  $y_0$ , TTG and TSR. In addition, baseline characteristics will also be tested for inclusion in the model, such as ECOG performance status, number of metastatic sites, alkaline phosphatase level and presence or absence of prior adjuvant therapy, as they may impact the death hazard in this patient population.

Various distributions will be considered to fit the time to survival model, including (but not limited to) exponential, log-normal and Weibull. The most suitable distribution is selected on the basis of penalized log-likelihood value (the smaller, the better). There is no clinical insight involved in the process of selecting the best parametric model.

If deemed suitable for prediction purpose, the survival model developed on the basis of the NO16966 data will be used for predicting OS in patients whose tumor kinetics correspond to the one observed in the current (BP29262) study.

To estimate the distribution of the difference in median OS, 1000 replicates of independent phase III studies (each of size  $N=1000$ ) of mFOLFOX6+Bev vs. mFOLFOX6+Van (allocation ratio 1:1) were simulated. Each replicate will contribute a median OS time in each treatment group.

For each replicate (*i.e.* each simulated study)

- model parameters (for TK and OS models respectively) will be sampled from the model parameter uncertainty distribution (standard error),
- individual tumor kinetics metrics ( $TTG_i$  and  $TSR_i$ ) will be derived from individual tumor kinetics profiles simulated using individual model parameter estimates sampled from log-normal distributions (representing inter-individual variability in tumor kinetics),
- individual survival times will be drawn from the parametric survival model probability density function,
- baseline characteristics (e.g.  $Age_i$ ,  $ECOG_i$ , ...) will be generated using non-parametric bootstrap from the baseline characteristics observed in study BP29262, in each treatment arm respectively.

The distribution of the difference in median OS ( $\Delta$ OS) will be displayed and the proportion of simulated studies with  $\Delta$ OS >5 month will be calculated. This quantity will correspond to the predictive power of a phase III study (with N=1000).

### **Validation of the OS model**

In order to validate the OS model (1), it is proposed to leverage literature data, where 1L mCRC patients were treated with either FOLFOX-4 or mFOLFOX6, given in combination with either Bev or Cetuximab or Panitumumab..

The Kaplan-Meier curves observed in the CRYSTAL [6,7], OPUS [8], FIRE-3 [9], TRIBE-3 [10], PRIME [11], PEAK [12], HORIZON-3 [13] and PACCE [14] will be overlaid to the simulated curves (described in the previous section) and the one observed in study NO16966 in patients treated with FOLFOX-4+Bev.

The differences observed between the study-specific curves and the simulated ones will be assessed by overlaying the KM simulated from the model to the ones reported in published trials.

### **References**

- [1] Claret et al. 2009 doi:10.1200/JCO.2008.21.0807
- [2] Claret et al. 2013 doi:10.1200/JCO.2012.45.0973
- [3] Saltz et al. 2008 doi:10.1200/JCO.2007.14.9930
- [4] Ohhara et al. 2015 doi:10.2147/OTT.S77190
- [5] Bruno et al. 2014 doi:10.1038/clpt.2014.4
- [6] Van Cutsem et al. 2009 doi:10.1056/NEJMoa0805019
- [7] Van Cutsem et al. 2011 doi:10.1200/JCO.2010.33.5091
- [8] Bokemeyer et al. 2011 doi:10.1093/annonc/mdq632
- [9] Heinemann et al. 2014 doi:10.1016/S1470-2045(14)70330-4
- [10] Loupakis et al. 2014 doi:10.1056/NEJMoa1403108
- [11] Douillard et al. 2010 doi:10.1200/JCO.2009.27.4860
- [12] Schwartzberg et al. 2014 doi:10.1200/JCO.2013.53.2473
- [13] Schmoll et al. 2012 doi:10.1200/JCO.2012.42.5355
- [14] Hecht et al. 2008 doi:10.1200/JCO.2008.19.8135