NOVARTIS

Clinical Development

MCS110

CMCS110X2201

A Phase II randomized, double-blind (Parts A, B, and C), placebo controlled (Parts A and B) study followed by open label dosing to assess safety, tolerability and effect on tumor size of MCS110 in patients with pigmented villonodular synovitis (PVNS)

TSc RAP Module 3: Detailed Statistical Methodology

Personal Data

Document type:	RAP Module 3: Detailed Statistical Methodology
Document status:	Final/Amendment 1
Release date:	23-Feb-2017

Table of contents

List of tables

1 Introduction to RAP documentation

1.1 Scope

The RAP documents contain detailed information to aid the production of Statistics & Programming input into the Clinical Study Report (CSR) for trial ***"CMCS110X2201".***

1.2 Changes to RAP documentation (M3)

Refer to corresponding guidance's and NIBR RAP Addendum template for detailed information on the requirements of documenting changes to RAP documentation.

For the statistical methodology (M3), any major changes occurring before database lock to the statistical methodology should be reflected in the RAP M3 documentation via version control (new document version to be approved by the trial team as the original module.

Major changes include, but are not limited to, changes in protocol that affect study design and statistical methodology.

Minor changes to the RAP M3 documentation can be captured e.g. by a study note to file / note in RAP Addendum or within the CSR itself. Minor changes include, but are not limited to, change in statistical model. Corrections of typographical errors or modification of spelling (from English to American, for example) do not need to be incorporated into the RAP M3 documentation.

2 Study objectives and design

Major study protocol amendments have occurred which require new versions of the RAP documents to be created.

2.1 Study objectives

Primary objectives Part A

- To assess efficacy of a single i.v. dose of MCS110 in reducing the size of PVNS tumors compared to placebo in patients awaiting surgery over 4 weeks evaluated by volume of PVNS lesion by MRI.
- To assess safety and tolerability of MCS110 in this population.

Secondary objectives Part A

- To characterize the pharmacokinetics of a single dose of MCS110 in PVNS patients.
- To assess pharmacodynamic effects of MCS110 on:
  - total circulating M-CSF levels in PVNS patients.
  - the soluble form of M-CSFR (sCSF-1R).
  - serum C-terminal Type I Collagen peptide (CTX-I), a biomarker of bone resorption.
  - circulating CD14+ monocytes and CD14+CD16+ monocytes in blood by FACS.
- To assess the immunogenicity of MCS110 in this population.
- To assess the duration of clinical response, if any, by joint range of motion.
- To assess the degree of functional recovery at time-points according to Assessment Schedule.

Primary objectives Part B

- To assess efficacy of multiple i.v. doses of MCS110 in reducing the volume of PVNS or GCTTS tumors evaluated by MRI after 8 weeks post last dose.
- To assess safety and tolerability of multiple i.v. doses of MCS110 in this population.

Secondary objectives Part B

- To characterize the pharmacokinetics of multiple doses of MCS110 in PVNS or GCTTS patients.
- To assess pharmacodynamic effects of MCS110 on:
  - total circulating M-CSF levels in PVNS or GCTTS patients.
  - the soluble form of M-CSFR (sCSF-1R).
  - serum C-terminal Type I Collagen peptide (CTX-I), a biomarker of bone resorption.

- To assess the immunogenicity of MCS110 in this population.
- To assess the duration of clinical response, if any, by joint range of motion.
- To assess the degree of functional recovery (questionnaires EQ5D, mHAQ, and joint specific questionnaire KOOS [Knee and Osteoarthritis Outcome Scale], SPADI [Shoulder Pain and Disability Index] and PREE [Patient Rated Elbow Evaluation]).
- To assess joint pain using a visual analog scale (VAS).
- To assess time to surgery.
- To assess time to relapse (based on MRI).

Primary objectives Part C

- To assess efficacy of multiple doses of 3 mg/kg, 5 mg/kg and 10 mg/kg MCS100 in reducing the size of PVNS or GCTTS tumors after 8 weeks following the last dose evaluated by volume of PVNS or GCTTS lesion by MRI
- To assess safety and tolerability of MCS110 in this population

Secondary objectives Part C

- To characterize the PK of multiple doses of MCS110 in PVNS or GCTTS patients.
- To assess pharmacodynamic effects of MCS110 on:
  - Total circulating M-CSF levels in PVNS or GCTTS patients.
  - Serum C-terminal Type I Collaged peptide (CTX-I), a biomarker of bone resoprtion.
- To assess the immnogenicity of MCS110 in this population.
- To assess the duration of clinical response, if any, by joint range of motion.
- To assess the degree of functional recovery using questionnaires.
- To assess joint pain using a visual analog scale (VAS).
- To assess time to surgery
- To assess time to relapse (based on MRI).

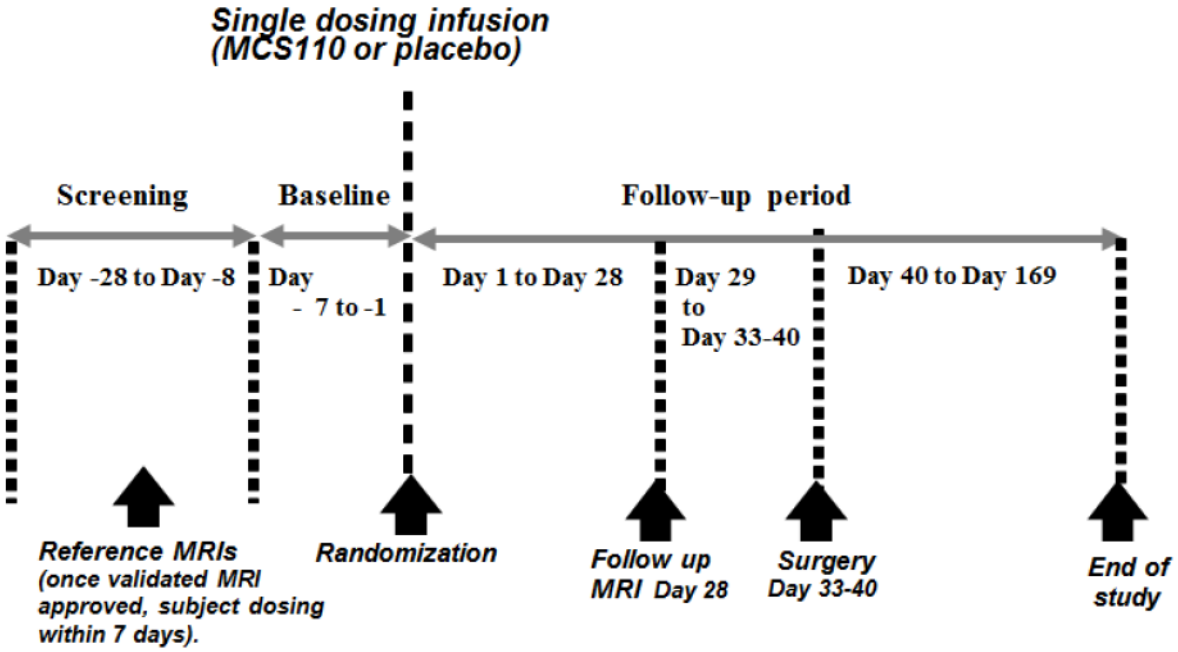
2.2 Study design and treatment - Part A

The study is a randomized, double-blind, placebo-controlled, parallel-arm, single dose exploratory PoC clinical trial with 2:1 randomization of MCS110 to placebo. Enrolment will proceed until up to 12 evaluable patients have been recruited, where "evaluable" is defined as completing the protocol through surgery (approximately Day 33). Patients diagnosed with PVNS, who are awaiting surgery will be treated with a single i.v. dose of 10 mg/kg of MCS110. Patients will only be eligible for the study, if they have at least one tumor site that can be measured by MRI. MRI of PVNS provides unique features depicting the extent of the tumor, synovial proliferation, joint effusion, bone erosion and deposits of hemosiderin.

The primary outcome of tumor shrinkage will be determined by MRI at 4 weeks, followed by surgery to remove residual tumor and to provide histological evidence of tumor cellularity. Patients will be monitored over a period of approximately 6 months to evaluate safety and tolerability of MCS110 and for relapse.

Subjects who meet the eligibility criteria at screening will be admitted to baseline evaluations. All baseline safety evaluation results must be available prior to dosing. Safety assessments will include physical examinations, ECGs, vital signs, standard clinical laboratory evaluations (hematology, blood chemistry, urinalysis) adverse event and serious adverse event monitoring. A full list of assessments can be seen in the list of assessments in the protocol.

For all outputs treatments will be labelled as 'MCS110' or 'Placebo'.



2.3 Study design and treatment - Part B

The study CMCS110X2201 Part B is a randomized, double-blind, placebo-controlled, parallel-arm, exploratory PoC clinical trial with 2:1 randomization of MCS110 to placebo for the first dose followed by open-label multiple doses. The schematic figure below shows the steps a patient takes through the study in Part B. Part B first dose will be administered double-blind, placebo-controlled. All patients will receive active treatment after the first dose and the study will continue as an open label study.

Day1 blinding will be maintained until week 12, which corresponds to the fourth dose for the potential Day 1 MCS110 patients. After patient reaches week 12 (Visit 7), Day1 treatment received by the patient will be un-blinded in order to allow Day1 placebo patients to receive a fourth dose of MCS110.

If patient received MCS110 at Day1, then the patient will follow Assessment Schedule Part B1 after Visit 7.

If patient received placebo at Day1, then the patient will follow Assessment Schedule Part B2 after Visit 7.

In the multiple dose Part B patients receive up to 4 doses of 10 mg/kg MCS110 administered once every 4 weeks.

The primary outcome of Part B of the study will be tumor volume after multiple doses of MCS110 monitored by MRI. The primary outcome will be assessed 8 weeks after the last dose of MCS110, a time point, where free drug is expected to decline rapidly, based on modelling and simulations of free MCS110 concentrations.

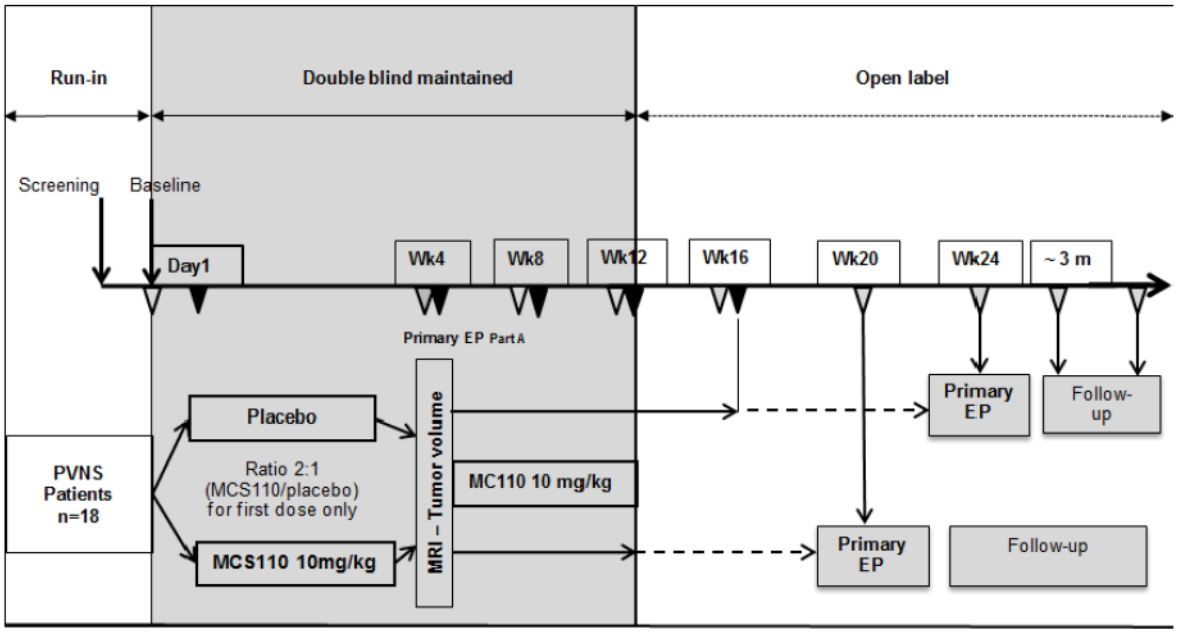
Subjects who meet the eligibility criteria at screening will be admitted to baseline evaluations. All baseline safety evaluation results must be available prior to dosing.

Safety assessments will include physical examinations, ECGs, vital signs, standard clinical laboratory evaluations (hematology, blood chemistry, urinalysis) adverse event and serious adverse event monitoring. A full list of assessments can be seen in the list of assessments in the protocol.

In order to monitor the duration of the tumor size reducing effect of MCS110, patients will have follow-up visits every 3 months for up to 2 years to document changes in growth of the residual tumor or to detect recurring tumor by MRI. The time intervals of 3 months are according to the standard of care for patients, who had undergone surgery.

In case of relapse or regrowth of residual PVNS tumor during the follow-up period, patients will be offered to start another treatment course once the previous MCS110 treatment was longer than 6 months ago. At any time during the study, investigator and/or patient can decide to undergo surgery if patient's PVNS conditions require such a treatment.

Total enrolment in the study (Part A and Part B) will proceed until up to 18 evaluable patients have been recruited, where Part B "evaluable" patient is defined as completing the protocol until at least 2 doses of MCS110 were administered and a follow-up MRI was performed. Patients will only be eligible for the study, if they have at least one tumor site that can be measured by MRI.



MRI – Tumor volume measurement

Dosing day – 1hr i.v. infusion

Primary Endpoint – Tumor volume reduction measured by MRIMRI Time-point: last MCS110 dose + 8 weeks

Follow-up on relapse every 3 months until up to 2 years based on MRI and symptoms

Double-blind: only at Day 1 – Day 1 Double blind maintained until week 12

2.4 Study design and treatment - Part C

The study CMCS110X2201 Part C is a randomized, double-blind, parallel-arm adaptive dose exploration study part. Approximately 15 patients, including adults and adolescents (≥12 years old), will be randomized to receive 3 mg/kg, 5 mg/kg or 10 mg/kg MCS110 in a ratio of approximately 2:2:1 and a potential further extension of the efficient lower dose arm. Tumor volume reduction after 3 doses, assessed by MRI at week 12, will guide the decision on whether to continue with the initial dosing (3 mg/kg or 5mg/kg) or increase for the following three doses to 10 mg/kg to allow the patient to benefit from the drug.

Conditions of dose level change after 3 doses

Patients enrolled in 3 mg/kg and 5 mg/kg arms: 3 additional MCS110 doses of the same dose will be administered, if MCS110 is well tolerated and if a reduction in tumor volume of ≥45% can be demonstrated after the 3rd MCS110 dose evaluated by MRI taken just before the 4th dose. This MRI is reflecting potential efficacy of 3 doses. If the drug is well tolerated and the tumor volume reduction is <45%, then three additional MCS110 doses of 10 mg/kg will be administered.

Process of dosing after 3 doses

All patients within each arm will receive 3 doses of the same initial dose at Day 1, week 4 and week 8.

At week 12, an MRI and safety examinations will be taken before the planned 4th dose. The unblinded statistician(s) and an unblinded Novartis medical person (unblinded safety reviewer) will review the tumor volume reduction data from baseline to week 12 MRI and patient's safety data. The unblinded safety reviewer will inform the site's unblinded pharmacist on the dose level to administer at week 12:

- Patients enrolled in 3 mg/kg and 5 mg/kg arms can receive same initial dose or increase to 10 mg/kg.
- Patients enrolled in 10 mg/kg arm: can either stop after 4 doses or receive up to 6 doses of 10 mg/kg under conditions described in the Protocol Section 5.5.5.2 (Dosing extension to 6 consecutive doses once every 4 weeks).

Corporate Confidential Information

The primary outome of Part C of the study will be the tumor volume after multiple doses of MCS110 determined by MRI which will be assessed 8 weeks after the last dose of MCS110.

Eligibility criteria and safety assessments will be similar as in Part B. Patients will also be followed up to 2 years every 4 months and they will be offered another treatment course with 10 mg/kg of MCS110 in case of relapse, where relapse is diagnosed at least 6 months after the last MCS110 dose received during the first treatment course.

Efficacy assessments will include tumor volume reduction, Corporate Confidential Information joint specific questionnaires, VAS and time to relapse.

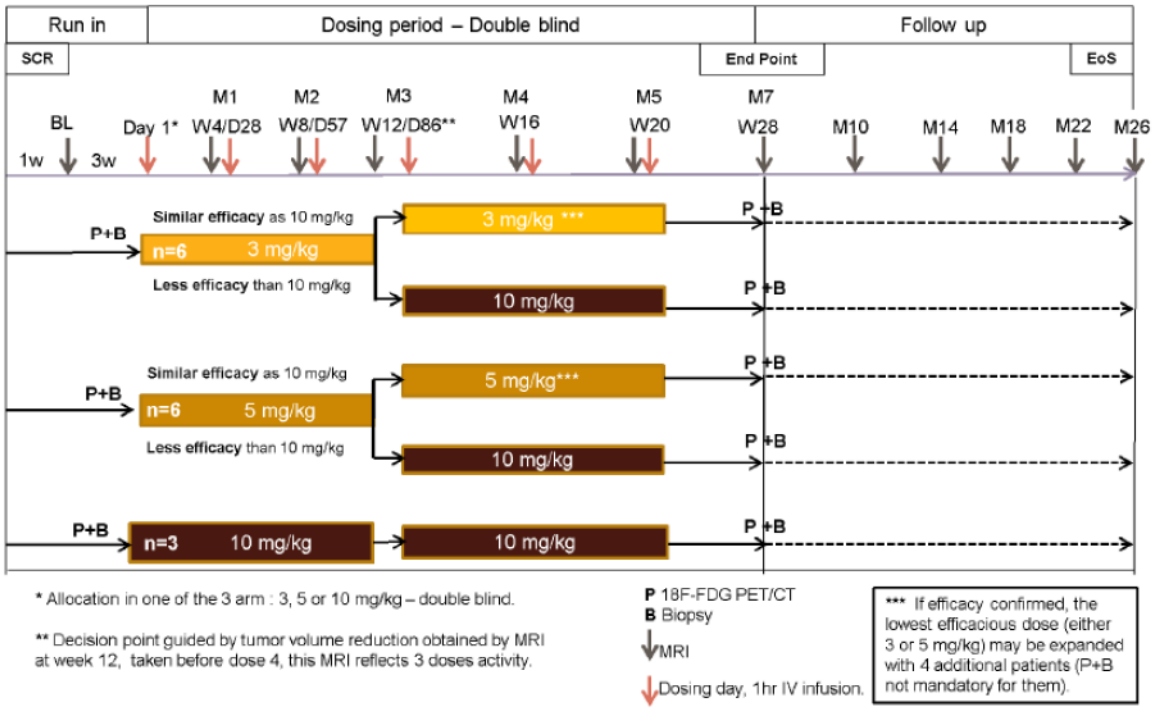


Figure 2-1: Study design Part C – adults and adolescents

Conditions of lower dose arm recruitment extension

If a lower dose is identified with at least the same efficacy as 10 mg/kg, defined as tumor volume reduction evaluated by MRI of at least 45%, additional patients may be enrolled in this lower dose group to extend the arm up to approximately 10 patients. Only one dose level, either 3 or 5 mg/kg may be extended and the additional patients will be allocated to the selected dose level in an open label fashion with the same assessment schedule applied,

Corporate Confidential Information

The planned enrollment number in Part C is approximately 15 patients. If a lower dose is identified to be, at least, as efficacious as 10 mg/kg, additional patients may be enrolled in open label to extend the efficacious 3 mg/kg or 5 mg/kg arm to approximately 10 patients.

3 First interpretable results (FIR)

First interpretable results (FIR) will be provided for this trial for each interim analysis and final analysis.

Corporate Confidential Information

5 Statistical methods: Analysis sets

Analysis sets will be defined the same for parts A,B and C.

For subjects for which the actual treatment received does not match the randomized treatment the treatment actually received will be used for the analysis.

All patients that received the study drug dose and with no protocol deviations with relevant impact on safety will be included in the safety analysis set.

All patients with at least one available valid (i.e. not flagged for exclusion) PK concentration measurement, who received the study drug and experienced no protocol deviations with relevant impact on PK data will be included in the PK analysis set.

All patients with any evaluable PD data who received any study drug (or at least 2 consecutive MCS110 doses for Part B) and experienced no protocol deviations with relevant impact on PD data will be included in the PD data analysis set.

The analysis sets and protocol deviation codes are related as follows:

Table 5-1 Protocol deviation severity codes and analysis sets

Protocol deviation severity code		Safety analysis set	PK analysis set	PD analysis set
Code	Text			
5	Exclude subject from all safety analysis	-	-	-
20	Exclude subject from PK analysis set	+	-	+
22	Exclude subject from PD analysis set	+	+	-
23	Exclude subject from PK and PD analysis set	+	-	-
49	Report relevant protocol deviation – include subject in all analysis sets	+	+	+

+ = include in analysis set, - = exclude from analysis set, NA = not applicable

6 Statistical methods for Pharmacokinetic (PK) parameters

PK analysis set will be used for the PK related summary tables described in this section.

PK data will be listed and summarized by Part A and Part B separately.

Noncompartmental pharmacokinetic analysis will be performed on MCS110 concentration-time profiles and the following PK parameters characterizing the disposition of MCS110 will be generated using WinNonlin Phoenix (Version 6.2, Pharsight, Mountain View, CA):

Part A: AUClast, AUCinf, Tmax, Cmax, CL, Vss, and MRT.

Part B: AUCτ (τ=dosing interval in Part B), Racc (accumulation ratio), Tmax, Cmax, T1/2.

Part C: AUCτ (τ=dosing interval in Part C), Racc (accumulation ratio), Tmax, Cmax, T1/2.

Additional PK parameters may be determined and compartmental PK modeling may be performed where appropriate.

Biofluid concentrations will be expressed in mass per volume units. All concentrations below the limit of quantification or missing data will be labeled as such in the concentration data listings. Concentrations below the limit of quantification will be treated as zero in summary statistics and will be omitted for the calculation of pharmacokinetic parameters.

Descriptive statistics of pharmacokinetic parameters will include mean, SD, and CV, min and max. When a geometric mean will be presented it will be stated as such, missing values will be set to zero and excluded for the purposes of calculating the geometric mean. Since T_{max} is generally evaluated by a nonparametric method, median values and ranges will be given for this parameter.

7 Statistical methods for Pharmacodynamic (PD) parameters

7.1 Data post-surgery will not be used for primary analysis - Primary Endpoint

Part A, B and C pooled for weeks 1 to 4 (referred as 'Part A*' hereafter)

The tumor size (volume) as assessed by MRI technique up to and including week 4 will be the primary measure of efficacy. This data will be listed as collected.

Consolidated tumor volume will be calculated as the geometric mean from two readers. Summary statistics of the consolidated tumor volume will be presented by treatment and visit time. The tumor volume measured at screening is treated as baseline. The percent change and absolute change from baseline of the consolidated tumor volume will be summarized in the same fashion. Waterfall plot on percent change from baseline after 4 weeks will be produced. The tumor size at baseline and after 4 weeks will be log transformed. The log ratio to baseline of the tumor size (change from baseline of log transformed volume) will be the primary measure of efficacy for the single dosing part of the study (Part A*).

The analysis of the primary endpoint will quantify the accumulation of evidence about the treatment effect as patients complete their 4 week treatment. The evidence will be summarized in the posterior distributions of the treatment group means and their differences. These posterior distributions will be updated after certain patients complete the 4 week treatment.

For the analysis, the distributional model for the observed data (log ratio) will be the normal model with unknown parameters μ and σ, with μ being the parameter of primary interest (treatment mean) and σ representing the sampling error. This model will be applied to each MCS110 dose group separately. Little prior information is available for these parameters and therefore non-informative prior distributions will be placed on them. For μ a normal prior distribution centered at 0 (no treatment effect) will be used with large standard deviation, and for σ a uniform (0,2) prior distribution.

The model is formally described below, LR being the log ratio of tumor volume:

Likelihood

$LR_i \sim N(\mu, \sigma)$

with

$\mu = \mu_p$ for observations from placebo group

$\mu = \mu_m$ for observations from MCS110 group

Prior distributions on treatment means μ's

$\mu_p \sim N(0, 0.3)$

$\mu_m \sim N(0, 10)$ * i.e. Neutral prior or

$\mu_m \sim N(0, 0.3)$ * i.e. Skeptical prior

Prior distributions on sigma

$\sigma \sim$ Uniform (0,2)

Decision Rule

The PoC criterion is two-fold, as follows:

- Prob(LR(treated) < LR(placebo)) > 0.90; i.e. 90% confidence in superiority to placebo
- Prob(LR(treated) - LR(placebo) < log(0.6)) > 0.50; i.e. 50% confidence in the treated group ratio being 40% below the placebo group ratio

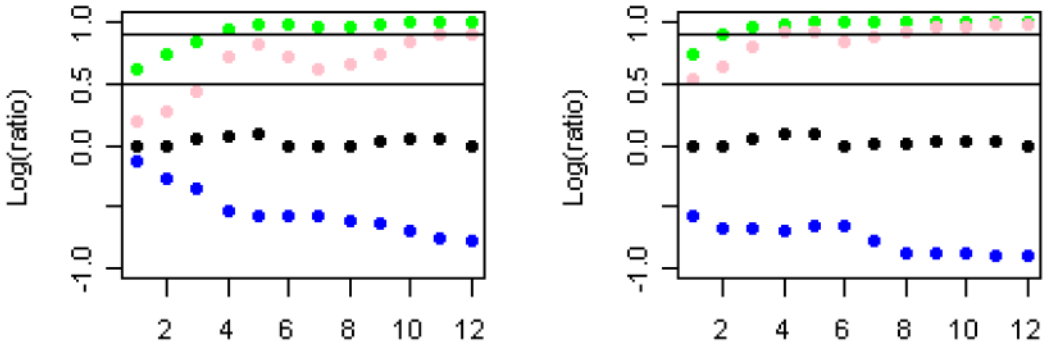
At least one placebo patient and two MCS110 patients must be included to declare PoC. Both neutral and skeptical priors will be used to assess the PoC criterion. The skeptical prior is used to shift prior beliefs toward no effect of MCS110 i.e. stronger evidence from the data in favor of MCS110 will be needed to reach PoC. The skeptical analysis will only be run as a sensitivity analysis.

Once the PoC criterion is reached, the clinical trial team (CTT) may decide to continue recruitment to collect more data on MCS110.

If after 6 patients have completed the 4 week treatment, the PoC criterion is not reached, the probability to reach PoC with the addition of 6 new patients will be assessed, and if this probability is less than 10% the trial may be stopped for futility.

For illustration, an example of a trial path simulation with skeptical and neutral priors on mean treatment effect in MCS110 group is shown below (Figure 7-1):

Figure 7-1 Example of a sample path



The left plot displays the skeptical analysis and the right plot the neutral analysis.

The blue dots represent the posterior mean in the MCS110 group and the black dots represent the posterior means in the placebo group as more patients contribute data.

The green dots show the probability of MCS110 being superior to placebo, and the pink dot the probability to be better by 40%.

In this example, the PoC criterion is reached after 3 patients when neutral prior (second graph) is used, and after 4 patients when skeptical prior (first graph) is used.

The posterior probability of LR(treated), LR(placebo) and their difference will be estimated using MCMC techniques using at least 40000 simulations discarding a sufficient number of the initial samples to insure convergence of the markov chain (based on trace plot visual inspections).

In the event that there are any complete responders a square root ratio analysis will be performed instead. In this analysis everything else will be the same as above other than using square root ratio instead of log ratios.

The same model as above will also be applied to compare the efficacy of the 3 mg/kg and 5 mg/kg to the 10 mg/kg dose (instead of Placebo).

Part B and C (referenced as Part B* hereafter)

Part B includes all data from multiple dosing Parts. In Part B, treatment group 1 = MCS110 throughout the trial, treatment group 2 = Placebo for the first dose and MCS110 from the second dose. In Part C, treatment arm 1 = MCS110 3 mg/kg, treatment arm 2 = MCS110 5 mg/kg, treatment arm 3 = MCS110 10 mg/kg, for patients in treatment arms 1 and 2 that switch to 10 mg/kg after 3 doses, they will be included in their original treatment group for the first 3 doses and will added in a separate group after that (4= MCS110 3 mg/kg for 3 doses and 10 mg/kg afterwards and 5= MCS110 5 mg/kg for 3 doses and 10 mg/kg afterwards), no re-calculation of baseline will occur for the people who change treatments.

Tumor volumes will be listed at each MRI assessment timepoint till the end of the study by treatment groups.

Part B*: MCS110 data only, by including data starting from the 1st dose of MCS110 in all treatment groups of Parts B and C. More specifically, for Part B treatment group 2, the measurement right before patients receiving the first dose of MCS110 (2nd dose from screening) will be treated as the new baseline and assessment time-points will be adjusted accordingly. Summary statistics of the consolidated tumor volume for the raw values, percent change and absolute change from baseline.

Arithmetic mean (+/-) SE on percent change from baseline vs time plot will be produced for Part B*. Waterfall plot of the maximum reduction until 8 Weeks after last dose will be produced in regards of number of doses received.

Overall Time to complete tumor ablation will be summarised up to the primary endpoint (8 weeks after the last MCS110 dose), along with a summary of overall time to ≥50%, and ≥80% ablation. The number and frequency of patients reaching each level will also be included in the summary.

Overall Time to complete ablation is the time (in weeks) - from first MCS110 dose to the first time when there is no tumor detected. Overall time to ≥50%, and ≥80% ablation is the time from first MCS110 dose to reach maximum ablation of [50%, 80%) and [80%, complete) respectively.

Part C

Tumor volumes will be listed and descriptively summarized at each MRI assessment time by treatment arm (treatment arm 1 = MCS110 3 mg/kg, treatment arm 2 = MCS110 5 mg/kg, treatment arm 3 = MCS110 10 mg/kg). For patients in treatment arms 1 and 2 that switch to 10 mg/kg after 3 doses, they will be treated as a separate group throughout the study (4= MCS110 3 mg/kg for 3 doses and 10 mg/kg afterwards and 5= MCS110 5 mg/kg for 3 doses and 10 mg/kg afterwards) This approach will be applied for all analyses of Part C where the treatment group is used. In addition, incidences of patients with partial response (PR), stable disease (SD) and progressive disease (PD) will be listed by treatment arm.

All summary tables (tumor volumes and response) will also be provided by age group (<18 years and ≥18 years).

7.1.1 Handling of missing values/censoring/discontinuations

Patients who withdraw early before receiving the second/third dose and a follow-up MRI may be replaced. Missing data will not be imputed and left missing for statistical analyses.

7.1.2 Supportive analyses

Part A, Part B and Part C pooled for weeks 1 to 4 (Part A*)

As a supportive analysis, log ratio to baseline of tumor volume at week 4 may be subjected to an Analysis of Covariance (ANCOVA) model. The model will include log transformed tumor volume at baseline and a factor for treatment. Point estimates and 95% confidence interval will be calculated for MCS110 and Placebo separately, and for the difference of means between MCS110 and Placebo. Results will then be transformed back to the original scale.

Part B*:

To study the time course of tumor volume reduction from baseline, a repeated measures ANCOVA model on log ratio to baseline will be fit by resetting the tumor volume at week 4 as baseline for patients in treatment group 2 and the visit number will be adjusted accordingly. The model will include log tumor volume at baseline and visit as fixed effects. Standard fit statistics will be used to determine the best variance-covariance structure. Point estimates and 95% confidence interval will be calculated at each visit. Baseline adjusted tumor volume (+/- SE) over time will be plotted.

Part C:

As a supportive analysis, log ratio to baseline of tumor volume at week 4 may be subjected to an Analysis of Covariance (ANCOVA) model. The model will include log transformed tumor volume at baseline and a factor for treatment. Point estimates and 95% confidence interval will be calculated for each treatment arm and for the difference of means between treatment groups. Results will then be transformed back to the original scale.

The statistical analysis will also be provided by age group (<18 years and ≥18 years).

Parts A, B and C :

A dose-time-response model (Lange and Schmidli 2015) will be used to analyze tumor response data (log ratio to baseline of tumor volume) from Parts A, B, and C. The following equation will be used for the latent concentration C(t) of a patient who receives K doses with dose D_k received at time t_k, k=1,..,K :

$$C(t) = \sum_{k=1}^{q} \frac{D_k \theta_1}{\theta_1 - \theta_2} \left\{ exp\left[-\theta_2 \left(t - \sum_{j=0}^{k-1} \delta_j \right) \right] - exp\left[-\theta_1 \left(t - \sum_{j=0}^{k-1} \delta_j \right) \right] \right\}$$

where $\delta_0 = t_1$, $\delta_j = t_{j+1} - t_j$ for j=1,…, K-1 and $q \in \{1, \dots, K-1\}$ is such that $t_q < t \leq t_{q+1}$ and q=K if $t > t_k$.

The above latent equation will be substituted in the equation below which will be used to model the dose response:

$$y_{ij} = a + \frac{\theta_4 C(t)}{\theta_3 C(t)} + \epsilon_{ij}$$

where $\epsilon_{ij} \sim N(O, \Sigma)$ i.i.d are the residuals. The observations will be considered to be independent between patient but dependent within a patient.

Preliminary results of the Placebo group have shown that a Placebo time-changing response may have been observed. To adjust for that the α term may be replaced with $\sum_{r=5}^{R} \theta_r \gamma_r(t)$, where θ_i are unknown real-valued parameters and the $\gamma_i(t)$ are arbitrary real valued known functions (which will be decided at that time).

The model will include tumor at baseline as a covariate (baseline will be centered) but other demographic/baseline variables may also be added. A random effect may also be considered if the data permits.

Only post-baseline values will be considered for the analysis.

Time-dose-response models describe and predict drug effect for complex dosing regimens. The mechanistic non-linear regression model above will allow for both inference on model parameters and prediction of the time-changing tumor response for the dosing regimens used in this study. Mechanistic models are based on prior knowledge and provide more reliable predictions. To this end, a Bayesian approach (using prior information) with weakly informative prior will be used. Point estimates and 95% confidence intervals of the model parameters, i.e., the expected response after multiple doses, will be generated using the Emax model. The Emax model will estimate tumor reduction from baseline at each dose and timepoint. Dose-time-response curves based on the model will be generated and graphically presented. The pooled dataset should result in approximately 19 patients on 10mg/kg, 6 patients each on 3 and 5 mg/kg (up to 10 if a lower efficacious dose is identified), and 5 patients (2 in Part A and 3 as first dose in Part B) on placebo.

7.2 Secondary Endpoint

7.2.1 Bone marker serum C-terminal Type I Collagen peptide (CTX-I) ; Monocytes (hematology); total circulating M-CSF levels; and the soluble form of M-CSFR

To be treated as with all biomarkers. See section 9.

7.2.2 The Joint Range of Motion

The Joint Range of Motion is only assessed for the affected joint and the impairment is recorded.

- Range of motion typically is measured with the neutral-zero method.
- For each plane assessed in the joint, the maximal angle of extension-neutral in degrees or the maximal angle of neutral –flexion in degrees is recorded.
- For the shoulder joint we have the additional planes of abduction to adduction and exterior flexion to interior flexion.

We will present Part A* (i.e., **Part A,Part B and Part C pooled for weeks 1 to 4) in one set of outputs and further sets of outputs for Part B* and Part C as:**

Spaghetti plots will be provided by joint and plane of motion. For example, all knee patients' extension-neutral in degrees vs. time will be plotted in one graph, and their neutral-flexion will be plotted in a separate graph.

In addition all data will be listed showing the baseline, raw value, absolute and percent change from baseline.

7.2.3 Knee and Osteoarthritis Outcome Scale (KOOS) - Parts B and C only

KOOS will only be assessed on the patients whose tumor is in knees.

This measurement was added in after the protocol amendment v03, hence the analysis plan below only applies to the data collected afterwards.

KOOS consists of 5 subscales; Pain, other Symptoms, Function in daily living (ADL), Function in sport and recreation (Sport/Rec) and knee related Quality of life (QOL). The previous week is the time period considered when answering the questions. Standardized answer options are given (5 Likert boxes) and each question is assigned a score from 0 to 4.

A normalized score (100 indicating no symptoms and 0 indicating extreme symptoms) is calculated for each subscale.

Assign the following scores to the boxes:

None	Mild	Moderate	Severe	Extreme
□	□	□	□	□
0	1	2	3	4

Each subscale score is calculated independently. Calculate the mean score of the individual items of each subscale and divide by 4 (the highest possible score for a single answer option). The normalized score for each subscale is

- PAIN 100 - (Mean Score (P1- P9)x100/4), where P1-P9 is questions 1 to 9 for PAIN subscale
- SYMPTOMS 100 - (Mean Score (S1- S7)x100/4), where S1-S7 are questions 1 to for SYMPTOMS subscale
- ADL 100 - (Mean Score (A1- A17)x100/4), where A1-A17 are questions 1 to 17 for ADL subscale
- SPORT/REC 100 - (Mean Score (SP1- SP5)x100/4), where SP1-SP5 are questions 1 to 5 for SPORT/REC subscale
- QOL 100 - (Mean Score (Q1- Q4)x100/4), where Q1-Q4 are questions 1 to 4 for QOL subscale

Missing data:

As long as at least 50% of the subscale questions are answered for each subscale, a mean score can be calculated of the data that is present. If more than 50% of the subscale items are omitted, the response is considered invalid and no subscale score should be calculated. For the subscale Pain, this means that at least 5 items must be answered; for Symptoms, at least 4 items; for ADL, 9 items; for Sport/Rec, 3 items; and for QOL, at least 2 items must be answered in order to calculate a subscale score. Subscale scores are independent and can be reported for any number of the individual subscales, i.e. if a particular subscale is not considered valid (for example, the subscale Sport/Rec 2 weeks after total knee replacement), the results from the other subscale can be reported at this time-point.

A total score has not been validated and is not recommended.

Part B

Summary statistics presenting each of KOOS sub scores and absolute change from baseline in each KOOS sub score will be provided with MCS110 arms pooled (Part B*). Panelwise graphs per sub-score will be used, where mean (+/-SE) will be plotted over time with MCS110 arms pooled (Part B*). All KOOS data will be listed by treatment group (defined in Part B).

Part C

Summary statistics presenting each of KOOS sub scores and absolute change from baseline in each KOOS sub score will be provided. Panelwise graphs per sub-score will be used, where mean (+/-SE) will be plotted over time by treatment group. All KOOS data will be listed by treatment group.

7.2.4 Patient Rated Elbow Evaluation (PREE) – Parts B and C only

PREE will only be assessed on patients whose tumor is in elbows.

This measurement was added in after the protocol amendment v03, hence the analysis plan below only applies to the data collected afterwards.

The PREE is a 20-item questionnaire designed to measure elbow pain and disability in activities of daily living. The PREE allows patients to rate their levels of elbow pain and disability from 0 to 10, and consists of 2 subscales:

1) PAIN subscale (0 = no pain, 10 = worst ever)

Pain - 5 items

2) FUNCTION subscale (0 = no difficulty, 10 = unable to do)

Specific activities - 11 items

Usual activities - 4 items

In addition to the individual subscale scores, a total score out of 100 can be computed by equally weighting the PAIN subscale score (sum of five items) and the FUNCTION subscale score (sum of fifteen items, divided by 3).

Computing the Subscales

Pain Score = Sum of the 5 pain items (out of 50). Best Score = 0, Worst Score = 50

Function Score = Sum of the 15 function items, Divided by 3 (out of 50) Best Score = 0, Worst Score = 50

Computing the Total Score

Total Score = Sum of pain + function scores. Best Score = 0, Worst Score = 100

If there is an item missing, the item will be replaced with the mean score of the subscale.

Part B

Summary statistics and graphical displays of total PREE score and absolute change from baseline in PREE score will be provided with MCS110 arms pooled (Part B*). All PREE data will be listed by treatment group (Part B).

Part C

Summary statistics and graphical displays of total PREE score and absolute change from baseline in PREE score will be provided by treatment group All PREE data will be listed by treatment group.

7.2.5 Shoulder Pain and Disability Index (SPADI) - Parts B and C only

SPADI will only be assessed on the patients whose tumor is in shoulders.

This measurement was added in after the protocol amendment v03, hence the analysis plan below only applies to the data collected afterwards.

The Shoulder Pain and Disability Index (SPADI) is a self-administered questionnaire that consists of two dimensions, one for pain and the other for functional activities. The pain dimension consists of five questions regarding the severity of an individual's pain. Functional activities are assessed with eight questions designed to measure the degree of difficulty an individual has with various activities of daily living that require upper-extremity use.

Total pain score: sum of 5 questions / 50 x 100%

(Note: If a person does not answer all questions divided by the total possible score, e.g. if 1 question missed divide by 40)

Total disability score: sum of 8 questions / 80 x 100 %

(Note: If a person does not answer all questions divided by the total possible score, e.g. if 1 question missed divide by 70)

Total Spadi score: sum of all 13 questions/ 130 x 100 %

(Note: If a person does not answer all questions divided by the total possible score, e.g. if 1 question missed divide by 120)

Part B

Summary statistics and graphical displays of overall SPADI score and absolute change from baseline in SPADI score will be provided with MCS110 arms pooled. All SPADI data will be listed by treatment group.

Part C

Summary statistics and graphical displays of overall SPADI score and absolute change from baseline in SPADI score will be provided by treatment group. All SPADI data will be listed by treatment group.

7.2.6 Joint Pain measured on a Visual Analogue Scale (VAS) - Part B* and C only

During the study, patient's pain will be assessed with a 100 mm visual analog scale (VAS).

The pain VAS will be completed by the patient. The patient will be asked to place a line perpendicular to the VAS line at the point that represents his/her pain intensity. Using a ruler,

the score is determined by measuring the distance (mm) on the 10-cm line between the "no pain" anchor and the patient's mark, providing a score from 0-100.

Part B

Summary statistics and graphical displays of overall VAS pain scores and absolute change from baseline mean(+/-SE) will be provided over time with MCS110 arms pooled. All VAS data will be listed by treatment group.

Part C

Summary statistics and graphical displays of overall VAS pain scores and absolute change from baseline mean(+/-SE) will be provided over time by treatment group. All VAS data will be listed by treatment group.

7.2.7 Time to surgery – listed for Parts B and C only

Patients with their time to surgery will be listed by Part and treatment groups as defined in Section 7.1. Time to surgery is the time from first dose to surgery.

7.2.8 Time to relapse – listed by Part

Patients who experience recurring PVNS tumor / tumor growth of residual tumor and their time, location, etc. to relapse /re-growth will be listed by Part and treatment group.

7.3.3 Bone Mineral Density by DXA- Parts B and C only

For exploratory purpose in Parts B and C summary statistics and graphical displays for the change from baseline DXA will be provided by Part. In addition, the data will be listed by Part showing the baseline, raw value and absolute change from baseline. All DXA data will be listed by Part.

7.3.3.1 Medical Outcome Short Form (36) Health Survey (SF-36®) – Part A only

SF-36 was collected in Part A only and removed from Part B. The SF-36 measures the impact of disease on overall quality of life and consists of eight subscales (physical function, pain, general and mental health, vitality, social function, physical and emotional health) which can be aggregated to derive a physical-component summary score and a mental-component summary score. Scores are normally determined with the use of norm-based methods which standardize scores based on an assessment of the general U.S. population free of chronic conditions. Composite scores range from 0 to 100, with higher scores indicative of better health. The data will be analyzed at Quality Metrics using their software to derive the 8 subscales and the physical-component summary score and a mental-component summary score.

Raw data from Part A for SF-36 will be listed only.

7.3.3.2 Analysis of mHAQ data - Parts B and C only

SF-36 has been replaced with mHAQ for part B as the team felt that SF-36 would not be evaluable due to the small number of patients in the study.

The modified Health assessment questionnaire (mHAQ) is a patient reported outcome (PRO) which is usually self-administered by the patient. The mHAQ is a modification of the HAQ. It was developed to include questions concerning perceived patient satisfaction regarding the same activities of daily living, along with perceived change in degree of difficulty. The number of activities of daily living was reduced from 20 (HAQ) to 8 (mHAQ).

The eight activities measured by the mHAQ are:

1. dressing and grooming
2. arising
3. eating
4. walking
5. hygiene
6. reach
7. grip
8. common daily activities

Patients are asked to rate these daily activities on a scale ranging from 1 to 4 with:

1. without difficulty
2. with some difficulty
3. with much difficulty

4. unable to do

The mHAQ is calculated as the average of the single scores. To do that the following scorin g is applied:

- without difficulty = 0
- with some difficulty = 1
- with much difficulty = 2
- unable to do = 3

Values <0.3 are considered normal.

Summary statistics and graphical displays of overall mHAQ score will be provided. All mHAQ data will be listed by Part and treatment group.

7.3.3.3 Analysis of EQ5D data

EQ5D is collected in Parts A, B and C. EQ5D is a generic, self-completed, and easy-to-use questionnaire which provides a health profile and a global rating of perceived health using a visual analogue scale (VAS). EQ5D has two parts consisting of a self-reported description using a five-dimensional classification (health profile - $EQ5D_{profile}$), and a self-rated global valuation of perceived health using a visual analogue scale or ''thermometer'' (perceived health - $EQ5D_{vas}$) (perfect health = 100; worst possible health = 0). Each of the five dimensions of the $EQ5D_{profile}$ - mobility, self-care, performance of usual activities, pain/discomfort, and anxiety/ depression - is divided into three levels of difficulty: ''no problem,'' ''some problem,'' or ''extreme problem.'' Then a health utility score $EQ5D_{utility}$ can be derived from the five dimensions- by summing the results after a coefficient has been applied (see details below). The $EQ5D_{vas}$ score will be reported as collected. In addition, there is a general question to assess the general state of health.

Data handling

Missing data will be excluded from the above analysis.

The EQ5D utility score $EQ5D_{utility}$ is derived as following. Let 1 corresponds to ''no problem,'' 2 for ''some problem,'' and 3 for ''extreme problem.'' The utility score of 1.000 corresponds to full health state (1, 1, 1, 1, 1). The following coefficients are used to compute for utility score:

Table 7-1 Coefficients for calculating utility score

Dimension	No problem (1)	Some problem (2)	Extreme problem (3)
Mobility	0	0.069	0.314
Self-care	0	0.104	0.214
Usual activity	0	0.036	0.094
Pain / discomfort	0	0.123	0.386
Anxiety / depression	0	0.071	0.236
		Constant = 0.081	Constant = 0.269

For example, for a health state of mobility = 1, self-care = 1, usual activity = 2, pain/discomfort = 2, and anxiety/depression = 3, the utility score is 1.0 - 0.081 - 0.036 -

0.123 - 0.236 - 0.269 = 0.255. The worst health state (3, 3, 3, 3, 3) corresponds to utility score -0.594.

Endpoints

- Absolute change from baseline of $EQ5D_{vas}$ score at each visit
- Absolute change from baseline of $EQ5D_{utility}$ score at each visit

Data for Part A and B pooled for weeks 1-4 change from baseline of $EQ5D_{vas}$ score at Week 2 and 4 will be plotted as an individual overlay plot with a mean (+/-SE) line overlaid.

Data for Part B change from baseline of $EQ5D_{vas}$ score at each visit will be plotted as an individual overlay plot with a mean (+/-SE) line overlaid.

Data for Part A, B and C pooled for weeks 1-4 change from baseline of $EQ5D_{utility}$ score at Week 2 and 4 will be plotted as an individual overlay plot with a mean (+/-SE) line overlaid.

Data for Part B change from baseline of $EQ5D_{utility}$ score at each visit will be plotted as an individual overlay plot with a mean (+/-SE) line overlaid.

Data for Part C change from baseline of $EQ5D_{utility}$ score at each visit will be plotted as an individual overlay plot with a mean (+/-SE) line overlaid.

All data for EQ5D will be listed by Part and treatment group.

8 Statistical methods for safety and tolerability data

All safety data will be listed and summarized by Part unless otherwise stated.

Disposition

Frequency counts of the number of patients completing/discontinuing from the study will be provided together with frequency count of the reason for discontinuation

Demographics

A summary of demographics will be produced containing Frequency counts of Race, Ethnicity and Sex together with summary statistics for Baseline Age, Height, Weight and BMI, and other related baseline values.

Immunogenicity

All Anti-drug antibodies will be listed by treatment, patient, visit/time, antibody Y/N, titer value. Summary statistics will be provided by treatment and visit.

Vital signs

All vital signs data will be listed by treatment, patient, and visit/time and if ranges are available abnormalities (and relevant orthostatic changes) will be flagged. Summary statistics will be provided by treatment and visit/time.

In Part C all summary tables will also be provided by age group (<18 years and ≥18 years).

ECG evaluations

All ECG data will be listed by treatment, patient and visit/time, abnormalities will be flagged. Summary statistics will be provided by treatment and visit/time.

In Part C all summary tables will also be provided by age group (<18 years and ≥18 years).

Standard clinical laboratory evaluations

All laboratory data will be listed by treatment, patient, and visit/time and abnormalities will be flagged. Summary statistics will be provided by treatment and visit/time.

In Part C all summary tables will also be provided by age group (<18 years and ≥18 years).

Additional time-profiles will be provided graphically for circulating monocytes, Creatine Kinase and ALT, AST (liver enzymes). This will be presented for Part A*, Part B* and Part C.

Special clinical laboratory evaluations

All laboratory data will be listed by treatment, patient, and visit/time and if ranges are available abnormalities will be flagged. Summary statistics will be provided by treatment and visit/time.

Adverse events

All information obtained on adverse events will be displayed by treatment and patient. The number and percentage of subjects with adverse events will be tabulated by body system and preferred term with a breakdown by treatment. An adverse event starting in one period and continuing into the next period is counted only in the onset period. A subject with multiple adverse events within a body system /within a treatment period/ is only counted once towards the total of this body system.

Frequency counts will be produced for adverse events by maximum severity and the same table produced for adverse events that are related to study drug. An overall summary of the number of adverse events categorised by serious, related, leading to discontinuation and death. Summary statistics will be produced for the duration of adverse events by preferred term.

Additional listing of adverse events of special interest will also be provided. Adverse Events of special interest will be identified by medical review of the preferred terms occuring in the data in addition to periorbital edema already identified.

In Part C all summary tables will also be provided by age group (<18 years and ≥18 years).

Concomitant medications / Significant non-drug therapies

All concomitant therapies will be listed by treatment group and patient.

10 Reference list

Eisenhauer EA, Therasse P, Boggerts J et al (2009) New response evaluation criteria in solid tumors: Revised RECIST guideline (version 1.1). Eur J Cancer 45: 228-247.

Lange, MR, Schmidli, H. (2015) Analysis of clinical trials with biologics using dose-time-response models. Statistic Med. doi: 10.1002/sim.6551, 2015.

Shankar LK, Hoffman JM, Bacharach S, et al (2006) Consensus recommendations for the use of 18F-FDG PET as an indicator of therapeutic response in patients in National Cancer Institute Trials. JNM 47:1059-1066.

Stacchiotti S, Crippa F, Messina A, et al (2013) Response to imatinib in villonodular pigmented synovits (PVNS) resistant to nilotinib. Clin Sarcoma Res 13 (3):8Steinbach LS, Neumann CH, Stoller DW, et al (1989) MRI of the knee in diffuse pigmented villonodular synovitis. Clinical Imaging 13: 305-316.

Ware, J.E. Jr (1993)*: SF-36 Health Survey. Manual & Interpretation Guide*. The Health Institute, New England Medical Center. Boston, Massachusetts.

Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y. Development of a shoulder pain and disability index. Arthritis Care Res. 1991 Dec;4(4):143-9.

MacDermid JC: Outcome evaluation in patients with elbow pathology: issues in instrument development and evaluation. *J Hand Ther*. 2001;14:105-114.