Document Type:	Statistical Analysis Plan
Official Title:	Open-label, randomized, two-arm, controlled study to assess the efficacy, safety, and tolerability of intravitreal (IVT) aflibercept compared to laser photocoagulation in patients with retinopathy of prematurity (ROP)
NCT Number:	NCT04004208
Document Date:	10 MAY 2021

Title page

Open-label, randomized, two-arm, controlled study to assess the efficacy, safety, and tolerability of intravitreal (IVT) aflibercept compared to laser photocoagulation in patients with retinopathy of prematurity (ROP)

Aflibercept for ROP – IVT injection versus laser therapy (FIREFLEYE)

Bayer study drug	BAY 86-5321 / Aflibercept		
Study purpose:	Primarily to collect data on the efficacy and safety outcomes in patients after treatment of ROP with aflibercept versus laser surgery		
Clinical study phase:	3	**Date:**	10 MAY 2021
Study No.:	20090	**Version:**	4.0
Author:	PPD PPD		

This Statistical Analysis Plan is produced on a word-processing system and bears no signatures.

The approval of the Statistical Analysis Plan is documented in a separate Signature Document.

Table of Contents

Abbreviations

ADAs	Anti-drug antibodies
AE	Adverse event
AP-ROP	aggressive posterior ROP
CRF	Case Report Form
CSP	Clinical Study Protocol
DMC	Data Monitoring Committee
FAS	Full Analysis Set
ICROP	International Classification of Retinopathy of Prematurity
IOP	intraocular pressure
IVT	intravitreal
LLOQ	Lower Limit Of Quantification
LOCF	Last Observation Carried Forward
MedDRA	Medical Dictionary for Regulatory Activities
mFAS	Modified Full Analysis Set
PKS	Pharmacokinetic Analysis Set
PMDA	Pharmaceuticals and Medical Devices Agency
PPS	Per-protocol set
PT	Preferred term
PTP	Planned Timepoint
RC	Reading Center
ROP	Retinopathy of Prematurity
SAE	Serious adverse event
SAF	Safety Analysis Set
SAP	Statistical Analysis Plan
SD	Standard Deviation
SOC	System organ class
TEAE	treatment-emergent adverse event
TESAE	treatment-emergent serious adverse event
TLFs	Tables, listings, figures
WHO	World Health Organization

1. Introduction

The study is designed for providing evidence for efficacy, safety, and tolerability of IVT aflibercept versus laser photocoagulation in subjects diagnosed with ROP.

ROP is a proliferative vascular retinopathy caused by an abnormal development of the vascularization of the peripheral retina in prematurely born infants. It affects mainly newborns with a preterm gestational age (≤ 32 weeks) and very low birth weight (≤ 1500 g). ROP remains a major cause of childhood blindness globally.

The present Statistical Analysis Plan (SAP) reflects the final statistical analyses to be performed for study 20090 and is based on the Integrated Clinical Study Protocol (CSP), Version 2.0, dated 23 JUN 2020 and the local protocol amendment for Japan (amendment number JPN-1), dated 09 MAY 2019. The aforementioned integrated CSP version 2.0 (23 JUN 2020) was amended from version 1.0 (22 MAR 2019) for non-COVID 19-pandemic related reasons, i.e. in order to allow for additional collection of PK data beyond 4 weeks after initial aflibercept injection.

2. Study Objectives

The primary objective of this study is to assess the efficacy of aflibercept in subjects diagnosed with ROP in comparison to laser.

The secondary objectives of this study are:

- To assess the safety and tolerability of aflibercept
- To assess the treatment burden of aflibercept and laser
- To describe the systemic exposure to aflibercept

Further objectives are:

- To characterize further aspects of the effect of aflibercept in the treatment of ROP
- To further investigate the study intervention and similar drugs (ie, mode-of-action-related effects and/or safety) and to further investigate pathomechanisms deemed relevant to diseases of the eye and associated health problems

3. Study Design

This study is a prospective, multicenter, parallel-group, randomized, unmasked for study treatment Phase 3 clinical study with two treatment arms comparing IVT administered aflibercept versus laser photocoagulation.

After one or two screening/baseline visit/s, eligible subjects will be randomized with a ratio of 2:1 to treatment with either aflibercept injection or laser photocoagulation, respectively. After the initial treatment with a single injection of aflibercept per eligible eye for subjects randomized to the aflibercept arm or the initial treatment with laser photocoagulation for subjects randomized to the control arm (multiple sessions within 1 week to complete the procedure are allowed), a 23-week treatment period (including possible retreatment and rescue treatment), and a final visit at Week 24 (up to Week 27 for subjects treated after Week 21) follow.

One or both eyes can be treated according to the investigator's assessment of the study's eligibility criteria. The second eye of subjects who start the study with only one eligible eyeshould be kept under observation. Second eyes which develop ROP requiring treatment during the study should receive treatment according to the randomization assignment of the first eye. Eyes treated as randomized before or on study visit 9 (≤8 weeks from baseline) will be included in the efficacy and safety analyses as appropriate, whereas those treated after visit 9 (>8weeks from baseline) will be included in the safety analysis only.

Subjects will be stratified for randomization according to Japanese and Non-Japanese study sites as well as by ROP classification in Zone I, Zone II, or AP-ROP according to investigator assessment. If both eyes are treated at baseline, the eye with the more severe disease will be considered for stratification. To prevent imbalances, the randomization result of subjects who do not complete baseline treatment will be assigned to a subsequent subject once it is known that the baseline treatment was not completed.

If required, each treated eye can receive re-treatment (up to two times in the aflibercept arm at least 28 days after the previous injection). For subjects randomized to aflibercept, rescue treatment with laser may be performed. For subjects randomized to laser, rescue treatment with aflibercept may be performed. The criteria for the allowance of retreatment and rescue treatment are specified in the CSP.

Database release will occur when all subjects have completed the Week 24 visit or have dropped out of the study. In case of the scenario where a subject has a 30-day safety follow-up visit after Week 24 or the final PK sample results come in later and these data can not be cleaned in time for the initial data release, these additional safety data will be included in a planned re-release of the database.

Definition of treatment completion: Subject had assessments of re-treatment / rescue criteria up to and including Week 23 and received all intended treatments (retreatment and rescue treatment) as prescribed in the study protocol. A subject is considered a treatment completer if treatment was completed for at least one eye.

Definition of study completion: Subject completed Week 24 or Week 27 if indicated. A subject is considered also as a completer of the study if the study intervention is discontinued and the subject continues only on observational basis for safety evaluations.

4. General Statistical Considerations

4.1 General Principles

The statistical evaluation will be performed by using the software package SAS release 9.2 or higher (SAS Institute Inc., Cary, NC, USA). All variables will be analyzed by descriptive statistical methods. The number of data available and missing data, mean, standard deviation (SD), minimum, quartiles, median, and maximum will be calculated for metric data. Frequency tables will be generated for categorical data.

Descriptive analyses will be shown where applicable

- for all subjects

and separately

- for subjects treated in both eyes (bilateral cases)
- for subjects treated in one eye only (unilateral cases)

within the respective analysis set.

For time to event analysis, the start date will be the date of randomization (i.e. the baseline visit).

4.2 Handling of Dropouts

In case a subject qualifies for retreatment or rescue treatment and there is a decision against retreatment or rescue treatment and subjects permanently discontinue study interventions (ie, does not meet the definition of treatment completion, see Section 3), the subject should continue in the study on an observational basis for safety evaluation.

Subjects who prematurely discontinue the study (participant discontinuation/ withdrawal from study or in the case of lost to follow-up) or permanently discontinue the study treatment for any reason will not be replaced.

The number of subjects who prematurely discontinue the study or permanently discontinue the study treatment for any reason will be reported and the corresponding reasons will be shown. Potential differences between the treatment groups in the proportion of patient withdrawals or in the timing of withdrawals will be evaluated using Kaplan-Meier plots for "Time to end of study" and "Time to permanently discontinuation of study treatment". Potential dropout patterns with respect to

- ROP classification (Zone I vs. Zone II vs. AP-ROP)
- Gestational age at birth ("<=26" vs ">26" weeks) and
- Number of aflibercept administrations / laser treatments or requirement for rescue treatment from baseline to Week 24
- For subjects affected by the COVID-19 pandemic (based on reported important protocol deviations) vs subjects not affected by the COVID-19 pandemic.

will be described.

The details for the handling of missing data due to dropouts are described in Section 4.3.

4.3 Handling of Missing Data

All missing or partial data will be presented in the subject data listing as they are recorded on the Case Report Form (CRF).

Additional descriptive analyses in the presence of missing data

The number, timing, and pattern for missing values of the primary endpoint will be displayed by means of descriptive statistics and visually displayed if applicable. Data exploration will include investigation of

- potential missing data imbalance between the two treatment groups and ROP classification (Zone I vs. Zone II vs. AP-ROP)
- baseline characteristics of subjects with and without missing values
- the direction of change over time with time courses of subjects with and without missing data.

General rules

When appropriate, the following rules will be implemented so as not to exclude subjects from statistical analyses due to missing or incomplete dates:

- **Safety Variables**
  In general, missing data will not be imputed for the safety analysis. If dates of adverse events are missing or partially missing so that the determination of whether or not the event is treatment-emergent is questionable, the event will be presumed to be treatment-emergent.

- **Prior/concomitant medication data**
  For the tabulation of prior and concomitant medication, missing or partially missing start dates of the medication will be imputed by the earliest possible time point, missing or partially) missing stop dates will be imputed by the latest possible time point.

- **Date of discontinuation of study**
  Partially missing dates will be imputed by the minimum approach, i.e., the earliest possible date.

Details for the handling of missing data for the efficacy analysis are described in Section 6.2.

4.4 Interim Analyses and Data Monitoring

No interim analysis is planned. However, safety assessments will be continuously performed.

During the study, an independent Data Monitoring Committee (DMC) will perform regular safety assessments to determine if the study shows unacceptable risks for the subjects, and issue recommendations to proceed or terminate the study. Detailed information regarding the DMC procedures are explained in a separate DMC charter.

4.5 Data Rules

Definition of Screen Failure: A screen failure is any subject for whom informed consent was signed but did not complete the screening period and was therefore not randomized.

Definition of Baseline: Baseline assessment is defined as the latest, valid assessment available before first administration of study treatment (including unscheduled assessments).

Unscheduled assessments: Extra assessments (laboratory data or vital signs associated with non-protocol visits or obtained in the course of investigating or managing adverse events (AEs)) will be included in listings, but not in summaries. If more than one value is available for an unscheduled assessment, all observations will be presented in the listings.

Handling of repeated measurements at the same visit: If measurements were repeated at the same visit, the value actually used for statistical summaries and analyses will be the

- Last non-missing repeated measurement, if respective visit is before start of treatment, and
- First non-missing repeated measurement, if respective visit is after start of treatment.

Handling of Week 24/ Early Termination Visit:

The study eCRF collects the final assessment per subject in a single visit folder. In case a subject completed the study, this information refers to Week 24, otherwise the information refers to the early discontinuation of a subject and will be assigned to an early termination visit. The information is included in VISITNUM=700000. For the analysis visit (AVISITN), the information will be split into AVISITN=168 if the subject completed and into AVISITN=950 for subjects who discontinued early.

Correction of AP-ROP definition based on central reading center (RC): The RC graded AP-ROP as:

- YES, AGGRESSIVE NV WITHOUT RIDGE AND WITH PLUS
- YES, ZONE I/II WITH STAGE II/III AND PLUS,

which can be identified in the data with PARAMCD=”ROPAGGR” and OEEVALN=27 in ADOE.

For the RC grading of "AGGRESSIVE NV WITHOUT RIDGE AND WITH PLUS", the RC actually allowed less than 4 quadrants to be affected by plus disease, causing a discrepancy to the agreed criteria in the RC charter.

To correctly apply the AP-ROP definition based on the RC data, AP-ROP should only be assigned in case

1) The RC assessed AP-ROP as "YES, ZONE I/II WITH STAGE II/III AND PLUS"
2) The RC assessed AP-ROP as "YES, AGGRESSIVE NV WITHOUT RIDGE AND WITH PLUS" and reported that 4 quadrants have been affected with plus disease.

The number of quadrants with plus disease is documented in ADOE with PARAMCD=" NUMQUAAF" and OEEVALN=27. AVAL=5 reflects 4 quadrants with plus disease.

4.6 Blind/Masked Review

In this document, the analysis of an open-label study is described. Validity reviews will be also unmasked. The results of the final data assessment will be documented in the final list of important deviations, validity findings and assignment to analysis set(s).

5. Analysis Sets

5.1 Assignment of Analysis Sets

Final decisions regarding the assignment of subjects/eyes to analysis sets will be made during the review of study data and documented in the final list of important deviations, validity findings and assignment to analysis set(s).

In general, a subject is assigned to an analysis set if at least one eye meets the respective criteria. Eye level analyses include only eyes that meet the respective criteria, i.e. single eyes could be excluded from an analysis while the subject itself is valid for the analysis.

The primary and secondary efficacy variables will be analyzed using the FAS, mFAS, and the PPS. The explorative efficacy variables will be analyzed using the FAS and mFAS. In all cases, the analysis with the FAS is considered as primary and mFAS and PPS are considered as supportive. Safety variables will be analyzed based on the SAF. Pharmacokinetic related analyses will be performed based on the PKS.

Full Analysis Set (FAS)

All subjects who received any type of study treatment, and had a baseline and at least one post-baseline assessment of efficacy.

The analysis on the FAS will be performed according to the treatment assigned at baseline (as randomized).

Modified Full Analysis Set (mFAS)

All subjects with central reading center positive confirmed disease stages meeting the inclusion criteria who completed baseline treatment, had a baseline and at least one post-baseline central reading center assessment of efficacy.

The analysis on the mFAS will be performed according to the treatment assigned at baseline (as randomized).

Per Protocol Set (PPS)

All subjects in the mFAS who have no validity findings or important deviations that could affect the primary efficacy endpoint.

The analysis of the PPS will be performed according to the treatment the subject actually received (as treated; determined by the first study treatment that the subject actually received during the study).

The status of each subject, with regard to protocol deviations, will be determined by the sponsor and documented before database release.

Safety Analysis Set (SAF)

All subjects who received any type of study treatment.

The analysis of the SAF will be performed according to the treatment the subject actually received (as treated; determined by the first study treatment that the subject actually received during the study).

Pharmacokinetic Analysis Set (PKS)

All subjects who received aflibercept treatment at the baseline visit and who had at least one non-missing PK assessment following the first dose of study drug.

Immunogenicity analysis sets

Anti-drug antibody (ADA) data will be analyzed using the ADA Analysis Set (AAS) and neutralizing antibody (NAb) data will be analyzed using the NAb Analysis Set (NAS).

The AAS will include all participants who received study intervention with aflibercept and had at least 1 non-missing result in the ADA assay following the first study dose.

The NAS will include all participants who received any study intervention with aflibercept and with at least 1 non-missing result in the NAb assay. Participants who are negative in the ADA assay and those who are positive in the ADA assay but are not treatment-emergent will be set to negative in the NAb analysis set. Analysis of both immunogenicity analysis sets will be performed according to the treatment the participant actually received (as treated).

6. Statistical Methodology

6.1 Population Characteristics

6.1.1 Sample Sizes, Subject Validity Status and Subject Disposition

The following will be summarized descriptively:

- Study sample size by country and site
- Protocol deviations, screen failures and premature discontinuations
- The number of subjects with important protocol deviations
- The number of screened subjects and primary reasons for screening failures

- The total number and percentage of randomized subjects (ie, subjects who received a randomization number)
- The number and percentage of subjects who qualified as FAS, mFAS, PPS, PKS, and SAF (defined in Section 5.1) and primary reasons for exclusions
- The total number and percentage of subjects who permanently discontinued the study treatment with the reasons for discontinuation (ie, does not meet the definition of treatment completion, see Section 3). This description will be on a subject level considering a subject on treatment, as long as at least one eye is on treatment.
- The total number and percentage of subjects who discontinued the study with the reasons for discontinuation

6.1.2 Demography, Baseline Characteristics and Medical History

Demographic and baseline assessments to be summarized will include:

- Chronological age at randomization (date of randomization - date of birth + 1)
- Gestational age at birth (in weeks and days)
- Post-menstrual age at randomization (Gestational age at birth + Chronological age at randomization)
- Gender
- Race
- Body weight, head circumference and body length at birth and baseline
- APGAR score at 1, 5, 10 and 15 minutes after birth categorized as 0-3, 4-7,8-10

Baseline disease characteristics to be summarized will include:

- ROP classification (Zone I, Zone II, plus status and AP-ROP)
- ROP classification according to International Classification of Retinopathy of Prematurity stages (Table 6–1)

- O2 supplementation at baseline (yes, no)
- History of sepsis (yes, no)
- History of necrotizing enterocolitis (yes, no)
- History of intraventricular hemorrhage (yes, no)

Demographic data and baseline characteristics variables will be summarized using descriptive statistics for all four analysis sets (ie, FAS, mFAS, PPS, and SAF).

Maternal and subject medical history will be recorded and will be coded according to latest available version of Medical Dictionary for Regulatory Activities (MedDRA). Medical history data will be evaluated by frequency tables, showing number of subjects with medical history findings by primary system organ class (SOC) and preferred term (PT).

6.1.3 Exposure and Compliance to Study Treatment

Study intervention will be administered by a qualified ophthalmologist. Recorded details of aflibercept injections and of laser procedures will be described descriptively and include:

- Number of aflibercept administrations from baseline to Week 24 (see secondary efficacy endpoint described in Section 6.2.2.3) – per eye and per subject
- Number of laser treatments from baseline to Week 24 (see secondary efficacy endpoint described in Section 6.2.2.3) – per eye and per subject
- Time required to perform treatment (analysed as "Other Pre-Specified Efficacy Endpoints", see further description in Section 6.2.3.2) - per eye and subject
- Requirement for sedation or general anesthesia (analysed as "Other Pre-Specified Efficacy Endpoints", see further description in Section 6.2.3.3) - per eye and subject:
    - type of anesthesia (general, sedation, local or other)
    - need for additional respiratory assistance and type (endotracheal intubation, non-invasive or other)
    - treatment site
- Injection volume (for aflibercept treatment) - per eye
- Number of burns, wave length, spot sizes, power and duration of laser pulse (for laser treatment) - per eye
- Time to first rescue treatment – per subject

Usually a single (and mostly bilateral) aflibercept treatment session is planned with up to 2 retreatments per eye after baseline if needed, compliance with study medication will not be calculated.

6.1.4 Prior/Concomitant Medication

Prior and concomitant medication will be analyzed based on the SAF.

All medications the subject received prior to the Screening Visit or during the study as well as maternal prior medications taken during pregnancy or for breastfeeding mothers after the umbilical cord clamping (if transfer of these products into human milk is feasible) will be tabulated by substance according to the World Health Organization (WHO) Drug Global.

Prior medication is defined as any medication that was taken before start of study treatment and that has a stop date before the treatment start date.

Concomitant medication is defined as any medication that was taken after first study intervention, i.e., any medication with a stop date after the start of study intervention or which is ongoing after the end of study.

Whenever applicable in displays, a distinction is made between:

- maternal prior medications taken during pregnancy (i.e. medication taken by the mother with start day prior to date of birth of child but after conception)
- prior medication taken by the breastfeeding mother (i.e. medication taken by the mother with start date after date of birth but prior to the start of study drug and end date prior to the start of study drug)
- concomitant medication taken by the breastfeeding mother
- prior medication taken by the subject

- concomitant medication taken by the subject.

The number of subjects with O2 supplementation at each visit, as well as at the end of the study of each subject, will be presented descriptively.

6.1.5 COVID-19 pandemic-related assessments

The study started patient enrollment in September 2019 already prior to the worldwide spread of the novel Coronavirus disease (COVID- 19) caused by the severe acute respiratory syndrome Coronavirus 2 (SARS-CoV2). The WHO declared this Coronavirus disease in March 2020 a pandemic. The study was never paused since then , and is ongoing.

Apart from protocol deviations for non-pandemic related reasons, any deviations from the protocol for pandemic-related reasons were collected on an ongoing basis. The following statstical analyses are planned to assess any potential impact on the efficacy and safety evaluation and interpretation of the collected study data and the overall conduct of the study:

- A listing of all subjects affected by COVID-19 pandemic, whether due to the specific protocol deviations, an early discontinuation of the study due to COVID-19 pandemic or whether a subject reported COVID-19 as an adverse event by country and investigator.
- The number and percentage of subjects affected by COVID-19 pandemic per analysis population
- A summary for the number of subjects who did not complete a study period (Treatment/Follow-up) for COVID-19 pandemic-related reasons.
- Presentation of COVID-19 pandemic-related protocol deviations, overall and per subject, as well as by treatment arm, region and country.
- Present primary efficacy analysis for subjects with COVID-19 pandemic-related protocol deviations and for subjects without any reported COVID-19 pandemic-related protocol deviation.

6.2 Efficacy

6.2.1 Primary Efficacy Endpoint

The primary efficacy endpoint is “Proportion of patients with absence of active ROP and unfavorable structural outcomes at 24 weeks after starting study treatment”, where

- Active ROP is defined as ROP requiring treatment (according to the inclusion criterion).
- Unfavorable structural outcome is defined as
    - retinal detachment,
    - macular dragging,
    - macular fold, or
    - retrolental opacity.

One eye or both eyes of a subject will be included into the analysis to determine the primary endpoint on a subject level, if treated and meeting the inclusion criteria. Data of a second eye of subjects who started the study with only one eligible eye and that develops ROP during study will be included in the efficacy analyses only if treated before or on Visit 9 (approximately 8 weeks from baseline).

All subjects requiring rescue treatment will be counted as missing the primary endpoint.

A subject for whom both eyes have been included in the analysis is defined as responder if both eyes show the defined absence. If only one eye has been treated this fact is accounted for by the statistical model (see below).

The number and percentage of eyes with active ROP or any of the unfavorable structural outcomes (overall and by type of unfavorable structural outcome) defined for the assessment of this endpoint at Week 24 will be displayed separately by treatment group using descriptive statistics.

Primary Analysis

The primary analysis for this endpoint will be based on the investigators' assessments of ROP and analyzed for FAS and repeated for mFAS and PPS

As it is assumed that in most subjects both eyes will be treated, this will be accounted for by using the following bivariate binomial model:

$l_i \sim$ Bernoulli (p): response of subject i in left eye

$r_i \sim$ Bernoulli (p): response of subject i in right eye

l_i and r_i are correlated with correlation coefficient ρ

Bivariate probability distribution:

left eye / right eye	**Response (r_i=1)**	**No response (r_i=0)**	
Response (l_i=1)	$p^2+ \rho\ p(1-p)$	$p(1-p)(1- \rho)$	p
No response (l_i=0)	$p(1-p)(1- \rho)$	$(1-p)(1-p(1- \rho))$	1-p
	p	1-p	

Based on this model the probability for a patient to be a responder is

$$\pi = p^2+ \rho\ p(1-p)).$$

The primary efficacy endpoint will be analyzed using a Bayesian statistical model with a non-informative prior probability distribution for the response probability for a single eye (p). For the correlation coefficient ρ, an informative prior distribution allowing positive values only will be assumed. The model is based on following distribution assumptions:

$p \sim$ beta(1,1): non-informative prior for the response probability in one eye

$\rho \sim$ beta(1,1): prior for the correlation between the two eyes of one subject (allowing only a positive correlation)

Based on this model, the primary endpoint "Proportion of subjects with absence of active ROP and unfavorable structural outcomes at 24 weeks after starting study treatment," (e.g.

response on a subject level, ie $\pi = p^2 + \rho\ p(1-p)$) will be analyzed for each of the 2 treatment groups.

This will be implemented by SAS using the following code (variable and dataset names are for illustration only and may be subject to change):

```
proc mcmc data=oerespm_ seed=1 nbi=5000 nmc=50000 outpost=regOut monitor=(prespa prespl diff)
stats(alpha=0.1) = all plots=all;
  *peyea, peyel: priors for eye level reponse probability for aflibercept and laser,
            rho: correlation;
   parms peyea peyel rho;  *eye level reponse probability for aflibercept and laser;
   prior peyea ~ beta(1, 1);
   prior peyel ~ beta(1, 1);
   prior rho ~ beta(1, 1);
   peye = dtreat*peyea + (1-dtreat)*peyel;
   prespa = peyea**2 + rho*peyea*(1-peyea);
   prespl = peyel**2 + rho*peyel*(1-peyel);
   diff = prespa - prespl;
   model first  ~ binary(peye);
   model second ~ binary(first*(peye + rho*(1-peye)) + (1-first)*((1-rho)*peye));

RUN;
```

The efficacy of aflibercept and laser photocoagulation will be characterized by the resulting posterior distribution for the probability of response. Mean, median, and mode of the posterior distributions for the response probabilities of the two treatments as well as for the difference in response probabilities between the treatments will be provided in addition to 90% equal tail credible intervals. For the mode, the "half sample mode" (HSM) will be calculated based on all simulated posterior outcomes. The half sample mode will be calculated as described by Robertson T. and Cryer J.D. (1974).

Success of this study will be mainly concluded, if the posterior probability:
P(response probability for aflibercept > (response probability for laser -5%)) ≥ 95%.

This is the case if the lower limit of the one-sided 95% credible interval for the treatment difference (aflibercept – laser photocoagulation) is greater than -5%. As non-informative prior distributions are used for the analysis this success criterion corresponds to a frequentist non-inferiority test with a non-inferiority margin of 5 percentage points and a significance test of 5% (one-sided test).

The posterior distribution of both response probabilities, as well as the difference in response probabilities will be displayed using normal kernel densities.

If the success criterion is met, in a second step superiority of aflibercept over laser photocoagulation is evaluated by comparing the lower limit of the 95% one-sided credible interval with 0. As a supportive analysis, the Japanese success criterion will be analysed to demonstrate superiority in response proportion over 66% in the aflibercept group.

Handling of missing data for primary analysis

All subjects are required to have efficacy assessments 24 weeks after starting study intervention (with an allowed visit window of +/- 7 days). Missing Week 24 data for a treated eye will be imputed as described in the following as long as respective eyes not having had an unfavorable structural outcome or rescue treatment before dropping out, as in these cases the eye would always be considered as non-responding:

1. If at the last visit before dropping out the subject had no active ROP and Zone II was completely vascularized (aflibercept subjects) or laser treatment was completed (laser subjects) the respective eye will be considered as responding.
2. Otherwise, if the subject dropped out at or after Week 16, the last non-missing post-baseline ROP staging before dropping out will be carried forward and used for determining the response in this eye (last observation carried forward (LOCF) approach).
3. If the subject dropped out before Week 16, the missing information will be imputed as follows:
    a. If there was a clear documentation that the subject dropped out due to lack of efficacy, the respective eye will be considered as non-responding;
    b. otherwise, a multiple imputation approach will be used giving the same probability of success as subject's having the same treatment group and initial staging (Zone I versus II versus AP-ROP). Further details are provided in Appendix 9.1.

Sensitivity Analyses

A sensitivity analysis will be performed where the RC data will be used instead of the investigator assessments of ROP at week 24 for FAS, mFAS and PPS.

After release of the final database, one of the diagnostic criteria for AP-ROP applied by the RC was identified to be incorrect, and in deviation from the international classification of ROP (ICROP, 2005) and the agreed RC charter. As per RC charter the presence of plus disease in all 4 quadrants of the posterior retina is one prerequisite for the diagnosis of AP-ROP. Section 4.5 decribes the data rules to correctly apply the AP-ROP definition according to the charter and based on the available RC data.
A full re-run of the tables, listings and figures (TLF) documents is needed after implementing the updated AP-ROP definition.

Furthermore sensititivity analysis to evaluate the impact of missing data will be conducted:

- considering all drop-outs as non-responders unless completely vascularized (aflibercept subjects) or laser treatment was completed (laser subjects) for FAS, mFAS and PPS (worst case imputation)

As a further sensitivity analysis, the "Proportion of patients with absence of active ROP and unfavorable structural outcomes at 24 weeks after starting study treatment" according to the investigator assessment will be analyzed by the corresponding frequentist approach for FAS. Reading center results will be analysed for mFAS and PPS.
Asymptotic 90% confidence intervals for the difference in response rates will be calculated based on a 2x2 contingency table for response (Yes, No) vs. Treatment (Aflibercept injection, Laser photocoagulation). In case that both eyes of a subject are treated, the subject will only be considered as a responder if both eyes responded, if only one eye is treated, then the subject will be considered as a responder if this eye responded. Aflibercept will be considered to be non-inferior to laser with regards to the sensitivity analysis if the 90% confidence interval of the difference lies entirely above -5%. Asymptotic 90% confidence intervals will be presented in addition for for the Mantel-Haenszel weighted treatment difference calculated using normal approximation and adjusting for each individual stratification factor (baseline ROP classification and Japan vs Non-Japan). The following methodology (Koch et al, 1990) is used:

$$d = \left(\sum_h w_h (\hat{p}_{ht} - \hat{p}_{hc}) \right) / \sum_h w_h \text{, where } w_h = n_{ht} n_{hc} / (n_{ht} + n_{hc}) .$$

Then

$$\hat{\text{var}}(\hat{d}) = \left(\sum_h w_h^2 \left(\hat{p}_{hc}(1 - \hat{p}_{hc}) / (n_{hc} - 1) + \hat{p}_{ht}(1 - \hat{p}_{ht}) / (n_{ht} - 1) \right) \right) / \left(\sum_h w_h \right)^2 .$$

With this, the 90% confidence interval (CI) can be given as:

$$\hat{d} \pm z_{\alpha/2} \sqrt{\hat{\text{var}}(\hat{d})}$$

($z_{\alpha/2}$ being the lower $\alpha / 2$ quantile of the standard normal distribution).

In the formulae,

- h: number of strata
- p_{ht}: proportion of subjects with absence of active ROP and unfavorable structural outcomes at 24 weeks after starting study treatment in the aaflibercept arm, in stratum h,
- p_{hc}: proportion of subjects with absence of active ROP and unfavorable structural outcomes at 24 weeks after starting study treatment in the laser arm, in stratum h,
- n_{ht}: number of subjects in the aflibercept arm in stratum h,
- n_{hc}: number of subjects in the laser arm in stratum h.

CCI

[redacted]


CCI

6.2.2 Secondary Efficacy Endpoints

6.2.2.1 Requirement for Intervention with a Second Treatment Modality from Baseline to Week 24

A second treatment modality for ROP in this study can be either rescue treatment as defined per protocol or treatment with any other surgical or nonsurgical treatment for ROP (e.g. IVT anti-VEGF injection, ablative laser therapy, cryotherapy, or vitrectomy captured as concomitant medication or surgeries after study start).

Second treatment modalities other than rescue treatment will be identified by:

- Any concomitant medication starting after start of study treatment with standardized medication term in ("BEVACIZUMAB" "PEGAPTANIB" "RANIBIZUMAB" "AFLIBERCEPT" "BROLUCIZUMAB")
- Any ocular surgery after start of study treatment assessed by the medical experts as second treatment modality.

The same Bayesian statistical model as described for the primary analysis of the primary endpoint will be used to estimate probabilities for a requirement of intervention with a second treatment modality within the time between baseline and Week 24. As for the primary analysis, a subject might contribute with a single eye or with both eyes. Analyses will be presented for FAS, mFAS and PPS.

Handling of missing data

Missing data regarding requirements for intervention with a second treatment modality will be imputed analogously to the approach used for the primary endpoint and as described in the following as long as respective eyes did not have had a second treatment modality before dropping out:

1. If at the last visit before dropping out the subject had no active ROP and Zone II was completely vascularized (aflibercept subjects) or laser treatment was completed (laser subjects) the respective eye will be considered as not requiring intervention with a second treatment modality.
2. Otherwise, if the subject dropped out at or after Week 16, the respective eye will be considered as not requiring for intervention with a second treatment modality.
3. If the subject dropped out before Week 16, the missing information will be imputed as follows:
    a. If there was a clear documentation that the subject dropped out due to lack of efficacy, the respective eye will be considered as requiring intervention with a second treatment modality;
    b. otherwise, a multiple imputation approach will be used giving the same probability of success as subjects having the same treatment group and initial staging (Zone I versus II versus AP-ROP).

6.2.2.2 Recurrence of ROP from Baseline to Week 24

Recurrence of ROP until Week 24 aims to monitor the disease activity during the study and is defined for this study as a need for retreatment or rescue treatment in cases where the question "presence of active ROP requiring treatment" had been previously answered with "No" .

This binary endpoint will be analysed using the same Bayesian statistical model as described for the primary analysis of the primary endpoint. Analyses will be presented for FAS, mFAS and PPS.

Handling of missing data

Missing data regarding recurrence of ROP will be imputed analogously to the approach used for the primary endpoint and as described in the following as long as respective eyes did not have had a recurrence before dropping out:

1. If at the last visit before dropping out the subject had no active ROP and Zone II was completely vascularized (aflibercept subjects) or laser treatment was completed (laser subjects) the respective eye will be considered as not having a recurrence.
2. Otherwise, if the subject dropped out at or after Week 16, the respective eye will be considered as not having a recurrence.
3. If the subject dropped out before Week 16, the missing information will be imputed as follows:
    a. If there was a clear documentation that the subject dropped out due to lack of efficacy, the respective eye will be considered as having a recurrence;
    b. otherwise, a multiple imputation approach will be used giving the same probability of success as subjects having the same treatment group and initial staging (Zone I versus II versus AP-ROP).

6.2.2.3 Number of Aflibercept Administrations/ Laser Treatments from Baseline to Week 24

The number of aflibercept treatments and the number of laser treatments from baseline to Week 24 will be described descriptively per treatment arm in frequency tables.

In case multiple sessions of laser treatment are necessary within 1 week from baseline, they will be counted as a single treatment. The same approach will be applied for laser treatment as re-treatment or rescue treatment.

Accounting for eventual reduced follow-up times due to dropping-out of subjects, the number of aflibercept treatments and the number of laser treatments from baseline to Week 24 will be also displayed descriptively only for treatment completers (see Section 3 for definition).

6.2.2.4 To explore new Retinopathy of Prematurity Activity Scale proposed by the International Neonatal Consortium

Based on the International Classification of Retinopathy of Prematurity (ICROP) classification, an activity scale and severity classification into three subcategories of ROP

(mild, moderate, severe) was derived (Smith et al, 2018). Classifications are shown in Table 6–1.

Table 6–1: ROP activity scale and ROP subcategories derived from the ICROP classification

Zone	Stage	ROP Activity Scale	Category
none	none	0	mild
III	1	1	
III	2	2	
II	1	3	
III	1+	4	
III	3	5	
III	2+	6	
II	2	7	
II	3	8	moderate
III	3+	9	
I	1	10	
II	1+	11	
I	2	12	
II	2+	13	severe
II	3+	14	
I	1+	15	
I	3	16	
I	2+	17	
I	3+	18	
Any zone	AP-ROP	19	
Any zone	4a	20	
Any zone	4b	21	
Any zone	5	22	

The distribution and changes from baseline over time of the ROP Activity Scale and the three subcategories as assessed by the central reading center will be described descriptively per eye by summary statistics (inclunding quartiles), frequency and shift tables. The number and percentage of subjects with at least a two step decrease by visit will be presented. Analyses will be presented for FAS, mFAS and PPS.

6.2.3 Other Pre-Specified Efficacy Endpoints

6.2.3.1 Evaluation of Visual Function at Week 24

Visual function per eye will be presented by descriptive statistics for the following variables:

- abnormalities of the retina or optic nerve as well as unfavorable ocular structural outcomes
- fixation (central, steady and maintained)
- fixing and following a 5-cm toy

- in case visual function cannot be assessed with fixation, visual evoked potentials can be used for instance to evaluate if the visual function is normal or abnormal
- cycloplegic refraction (sphere, cylinder, axis, pseudophakia or intraocular vision correction)
- ocular motility tests (if motility test done: assessment of ocular palsy with affected nerves, assessment of strabismus, and nystagmus).

6.2.3.2 Time Required to Perform Treatment

The start and stop times for any aflibercept administrations or laser procedures (irrespective of initial treatment, retreatment or rescue treatment) will be captured. The required time to perform aflibercept administrations or laser treatments will be calculated by "stop time - start time" per eye and subject, respectively and summarized with descriptive statistics by treatment arm.

The total time of administrations for each study intervention will be analyzed per eye and per subject, i.e. the sum over all initial and retreatment and rescue treatment administrations.

6.2.3.3 Requirements During Study Intervention

Frequencies of the following during all aflibercept administrations or laser treatments (including also retreatment or rescue treatments) will be described descriptively per eye and per subject:

- type of anesthesia (general, sedation, local or other)
- need for additional respiratory assistance and type (endotracheal intubation, non-invasive or other)
- treatment site

6.2.3.4 Requirement for Treatment with more than one Aflibercept Injection

The number of eyes, as well as the number of subjects, with more than one aflibercept injection will be described descriptively by frequency tables.

6.2.3.5 Time to Intervention with a second Treatment Modality for ROP or Development of Unfavorable Structural Outcome

The time to either

- the first intervention with a second treatment modality for ROP (see definition of secondary endpoint "requirement for intervention with a second treatment modality from baseline to Week 24") for at least one eye per subject

Or

- the development of unfavorable structural outcome defined as retinal detachment, macular dragging, macular fold, or retrolental opacity based on the investigator's assessment of ROP for at least one eye per subject

will be displayed descriptively by treatment group on subject level (for all subjects and separately for bilateral cases and for unilateral cases) and descriptive statistics using Kaplan-Meier estimates. Descriptive statistics include

- number of subjects with event
- number of subjects censored

- median, 25th and 75th percentiles
- as well as rates of subjects without any event at Week 6, 12 and 24 (42, 84 and 168 days) with corresponding 95% confidence intervalls).

Partially missing event dates will be imputed as the minimum possible date on or after start of study intervention.

Handling of censored data:

Subjects without an event will be censored at the date of last visit if they prematurely discontinued from the study. All censored data will be treated as censored completely at random.

6.2.3.6 Time to Recurrence of ROP

Recurrence of ROP until Week 24 on subject level is considered as an event for this endpoint and is defined as follows:

For at least one eye of a subject a need for retreatment or rescue treatment in cases where the question "presence of active ROP requiring treatment" had been previously answered with "No" for the same eye.

The time to the first event will be analysed on subject level analougously to the approach described in Section 6.2.3.5.

6.2.3.7 Regression of Plus Disease, Regression of Pre-Retinal Vascularized Ridge and Progression of Retinal Vascularization beyond the Ridge from Baseline to Week 24

Plus disease describes an increase in severity of ROP where blood vessels are particularly enlarged and twisted. The frequency of regression of plus disease, i.e. that plus disease is present at baseline, but not at Week 24, will be described descriptively per eye.

There are 6 stages (stages 1, 2, 3, 4a, 4b, 5) indicating the severity of the disease. For the regression of the pre-retinal vascularized ridge from baseline to Week 24, the decrease/improvement of the stage is analysed. A reduction of at least one stage as well as the amount of change will be described descriptively.

There are three zones indicating the extent to which blood vessel growth (vascularization) has been completed. A progression of retinal vascularization beyond the ridge from baseline to Week 24, i.e. if the normal vascularization has grown towards the periphery is indicated by a higher zone classification (improvement of the disease). An increase of at least one zone as well as the amount of change will be described descriptively.

The primary analysis for this endpoint will be based on the investigators' assessments while the reading center assessment will be presented in addition.

6.2.3.8 Progression to Stage 4 or 5 ROP from Baseline to Week 24

The frequency of a progression to stage 4 or 5 at any time until Week 24 based on the investigators' assessments will be described per eye by visit and treatment group.

6.2.3.9 Completion of Vascularization of the Peripheral Retina to within one Disc Diameter of the Ora Serrata at Week 24

Whether the vascularization is complete or not at Week 24 based on the investigators' assessments will be described per eye by frequency tables.

6.2.3.10 Time to Completion of Vascularization

The time to the investigator's first assessment where the vascularization is complete will be analysed analougously to the approach described in Section 6.2.3.5. This will be analysed on subject level for all subjects, all bilateral cases as well as all unilateral cases.

6.2.3.11 Number of Visits required up to Week 24

The number of visits up to Week 24 including mandatory vists according to the protocol as well as additional optional visits will be described.

Accounting for eventual reduced follow-up times due to dropping-out of subjects, the number of visits up to Week 24 will be also displayed descriptively only for treatment completers (see Section 3 for definition).

6.2.4 Subgroup Analyses

For the primary and secondary efficacy endpoints, subgroup analyses by zones (Zone I/II/AP-ROP) will be performed. The subgroup will be defined twice, once based on investigator assessment and once for reading center assessment.

6.2.5 Other exploratory analyses

6.2.5.1 Recurrence following complete regression of ROP from Baseline to Week 24

Recurrence following complete regression of ROP until Week 24 is defined for this study as the presence of ROP at a visit after ROP has been documented as "absent" for at least one previous visit.

This binary endpoint will be analysed using the same Bayesian statistical model as described for the primary analysis of the primary endpoint.

Handling of missing data

Missing data regarding recurrence of ROP will be imputed analogously to the approach used for the primary endpoint and as described in the following as long as respective eyes did not have a recurrence before dropping out:

1. If at the last visit before dropping out the subject had no active ROP and Zone II was completely vascularized (aflibercept subjects) or laser treatment was completed (laser subjects) the respective eye will be considered as not having a recurrence.
2. Otherwise, if the subject dropped out at or after Week 16, the respective eye will be considered as not having a recurrence.

3. If the subject dropped out before Week 16, the missing information will be imputed as follows:
    a. If there was a clear documentation that the subject dropped out due to lack of efficacy, the respective eye will be considered as not having a recurrence;
    b. otherwise, a multiple imputation approach will be used giving the same probability of success as subjects having the same treatment group and initial staging (Zone I versus II versus AP-ROP).

6.3 Pharmacokinetics/Pharmacodynamics

Biomarkers for efficacy are not evaluated in this study.

Pharmacokinetic (PK) evaluations will describe plasma concentrations of free, adjusted bound and total aflibercept. Adjusted bound will be calculated from 'bound aflibercept' by multiplication with a factor of 0.717. Total aflibercept will be calculated as the sum of free and adjusted bound aflibercept.

The results of the bioanalytical analyses will be reported separately and attached to the final clinical study report as an appendix.

The objectives of the PK evaluations include:

- Describe the individual concentrations of free,adjusted bound and total aflibercept at several time points after the first IVT injection up to 24 weeks thereafter
- Explore the potential influence of demographic factors such as age and weight on systemic aflibercept concentrations
- Explore the relationship of systemic exposure and blood pressure (as a pharmacodynamic safety marker) and urine protein

PK evaluation

Concentrations of free,adjusted bound and total aflibercept will be listed by sample time point throughout the treatment period, and summarized by descriptive statistics overall and by baseline weight (<1000 gr, 1000 gr - < 1500 gr, 1500 gr - < 2000 gr, 2000 gr - < 2500 gr, ≥ 2500 gr), gender and race subgroups including the subgroup of Japanese patients.

Mean concentration time curves (plus/minus one SD) will be presented.

PK concentrations will only be considered for the analysis, if the actual measurement took place at the scheduled time point within the allowed time window according to the protocol.

I.e., PK samples will only be analysed if measured:

- D1 post treatment: For a maximum of 3 days after aflibercept injection
- Week 2: If measured 14 ± 3 days after aflibercept injection
- Week 4: If measured 28 ± 7 days after aflibercept injection
- Week 8: If measured 56 ± 7 days after aflibercept injection
- Week 12: If measured 84 ± 7 days after aflibercept injection
- Week 24: If measured 168 ± 30 days after aflibercept injection.

To describe the deviations from the planned PK assessment time point, descriptive summary statistics will be presented by time point for:

- Time deviation = (Actual PK sample date – aflibercept injection date) – planned timepoint (PTP) (with the PTP = 1 for Day 1, PTP = 14 for Week 2 and PTP = 28 for Week 4 as well as PTP = 56 for Week 8, PTP = 84 for Week 12 and PTP = 168 for Week 24).

PK concentrations from samples taken outside the defined time window will be listed together with the actual sampling date.

Relationship of systemic exposure and clinical outcome

The relationship of exposure to specific safety parameters (blood pressure, urine protein, treatment emergent serious AEs) will be explored graphically. For each sampling time point scatterplots for individual aflibercept concentrations (adjusted bound and free) vs. each pharmacodynamics safety marker (systolic, diastolic blood pressure, change from baseline in systolic blood pressure (SBP) and diastolic blood pressure (DBP)) will be presented. Aflibercept concentrations measured as below lower limit of quantification (<LLOQ) will be presented with a value ½ LLOQ. Subjects with unilateral and bilateral treatment will be presented differently, as well as subjects with treatment emergent serious AEs after the initial aflibercept administration (within 30 days after first injection) and subjects without.

Similar scatter plots will be created for the absolute values and change in SBP/DBP at Week 4 vs gestational age presenting subjects for the aflibercept group differently to subjects from the laser group. Similarly, additional scatter plots for absolute values and change in SBP/DBP at Week 4 vs baseline weight will be presented.

Optional Population Pharmacokinetic Analysis

Optionally, a population PK analysis may be considered as part of a meta-analysis of PK data obtained with aflibercept. In this case, aflibercept concentrations collected during the study will be analyzed using nonlinear mixed effects models. Mixed effects models, or population-type PK models, describe the relationship between dose and time and variables such as drug plasma concentrations. Both structural and random effects are involved in this relationship. A population PK compartmental model may be developed using the concentration of aflibercept as the dependent variable.

Results obtained by the optional population PK modeling may then be further specified and presented in a separate report that may also include PK data from other studies with Eylea and aflibercept.

CCI [redacted]

CCI [redacted]

6.4 Safety

All safety variables will be analyzed for the safety analysis set.

6.4.1 Adverse Events, Serious Adverse Events, and Device Events

All AEs will be coded using Medical Dictionary for Regulatory Activities (MedDRA). The latest available version at the time of coding will be used.

Adverse events will be summarized, at a minimum, on the level of system organ class (SOC) and preferred term (PT) and summaries will be provided displaying AEs within each system organ class (SOC) in decreasing order of total frequency according to the numbers of subjects reporting the AE (not number of events). Data will also be summarized according to intensity and investigator's causality assessment.

Ocular and systemic TEAEs and SAEs from baseline to Week 24 will be analyzed descriptively using summary statistics or frequency tables as appropriate. Thereby, TEAE is defined as AE that is observed or reported after the first and not later than 30 days after the last administration of study treatment. This definition includes that the length of the time window for potential treatement emergence varies from subject to subject.

Overview tables will be presented for both AEs and TE(S)AEs for ocular events and systemic events as well as for all adverse events (ocular and systemic). Ocular events will be presented overall, and for bilaterally and unilaterally treated subjects separately.

Ocular events on treated eyes will be presented by SOC and PT for:

- AE
- TEAE
- Serious adverse event (SAE)
- TESAE
- Aflibercept related TEAE
- Injection procedure-related TEAE
- Photocoagulation related TEAE
- Aflibercept related TESAE
- Injection procedure-related TESAE
- Photocoagulation related TESAE
- Non serious AE
- TEAE resulting in discontinuation from aflibercept
- TEAE resulting in discontinuation from photocoagulation
- AE by intensity

- TEAE by intensity
- TESAE by intensity
- Aflibercept related TEAE by intensity
- Injection procedure-related TEAE by intensity
- Photocoagulation related TEAE by intensity

Ocular events for non treated eyes will be reported with a reduced number of tables:

- AE
- SAE
- Non serious AE
- AE by intensity

The same analyses as for ocular events for treated eyes will be repeated for systemic adverse events.

Device events will be listed only.

A listing including all COVID-19 adverse events will be presented.

6.4.2 Surgical Procedures after Study Start

Any surgical procedures after study start and the type of any anesthesia used, will be displayed using number and percentage of subjects.

6.4.3 Ophthalmic Examinations

Any clinically significant abnormal findings:

- before treatment and changes from baseline over time in the anterior segment
- post treatment in the posterior segment (other than ROP)

will be displayed using descriptive statistics.

For subjects in the aflibercept arm, the intraocular pressure (IOP) will be measured prior to the injection and at least once post-injection. Summary statistics and changes from pre- to post-injection will be displayed. In case of multiple post injection IOP assessments per eye, the first post injection assessment will be analyzed. Analyze first and possible second and third injection per subject separately.

6.4.4 Physical Examinations and Vital Signs

Measurements from physical examinations (weight, body length, head circumference) and vital signs (body temperature, heart rate, respiratory rate, and blood pressure) will be displayed using descriptive statistics. Mean (plus/minus one SD) blood pressure time curves will be presented and repeated for the aflibercept arm for subjects with free or adjusted bound aflibcercept concentrations >2000 ng/mL at any time point.

6.4.5 Laboratory Safety Variables

Results of laboratory analyses (hematology, chemistry, urinalysis, proteinuria) will be listed. Proteinuria results over time will be displayed with frequency counts.

6.4.6 Central Nervous System Imaging

Procedure findings will be listed.

6.5 Development of Anti-drug Antibodies

Antibodies to aflibercept will be evaluated in serum samples collected from all participants at baseline and at Week 12.

Immunogenicity will be characterized by ADA responses and titers observed in patients in the ADA analysis set. ADA response categories and titer categories are defined as follows:.

- **ADA Negative:** all time points negative and those that exhibit a pre-existing response, regardless of any missing samples
  - CCI [redacted]
- **ADA positive:** defined as those that exhibit a treatment emergent or treatment boosted ADA response, regardless of any missing sample
  - CCI [redacted]
  - CCI [redacted]

ADA titer will additionally be summarized with number and percentage of sujects for categories:

- CCI [redacted]
- CCI [redacted]
- CCI [redacted]

NAb status will be evaluated for the samples that are positive in the ADA assay and have sufficient volume to analyse. The number and proportion of participants positive in the NAb assay will be summarized for the NAS.

If applicable, further analyses for ADA positive or NAb positive subjects will include:

- Plots of drug concentrations will be examined and the influence of ADAs and NAbs on individual PK profiles evaluated.
- Primary efficacy endpoint
- Ocular TEAE for study eye and fellow eye
- non-ocular TEAE

Potential association between immunogenicity variables and safety may be explored with a primary focus on the following safety events during the TEAE period:

- Injection site reaction (serious or severe and lasting 24 hours or longer)
- Infusion reactions

- Hypersensitivity (SMQ: Hypersensitivity [Narrow])
- Anaphylaxis (SMQ: Anaphylaxis [Narrow])

7. Document History and Changes in the Planned Statistical Analysis

- Version 1.0 of the SAP dated 10 JUL 2019

Summary of changes:

- Addition of COVID-19 pandemic related analyses
- Addition of frequentist analysis approach for primary efficacy variable
- Addition of further exploratory endpoint: "Recurrence following complete regression of ROP from Baseline to Week 24"
- Further details on analyses of PK and immunogenicity data
- Version 2.0 of SAP dated 09 FEB 2021

Summary of changes:

- Definition of screen failure added
- Version 3.0 of SAP data 24 FEB 2021

Summary of changes:

- Addition of data rules to correct the AP-ROP assessment of the RC to follow the RC charter
- Editorial changes to the PK figures

8. References

Smith LEH, Hellström A, Stahl A, Fielder A, Chambers W, Moseley J, Toth C, Wallace D, Darlow B, Aranda J, Hallberg B, Davis JM, for the Retinopathy of Prematurity Workgroup of the International Neonatal Consortium. Development of a retinopathy of prematurity activity scale and clinical outcome measures for use in clinical trials. JAMA Ophthalmol. 2018 Dec 13 (corrected 10 Jan 2019). Available from: doi:10.1001/jamaophthalmol.2018.5984.

Robertson T. and Cryer J.D. (1974). An iterative procedure for estimating the mode. J. Amer. Statist. Assoc., 69(348):1012-1016.

Koch GG, Carr GJ, Amara IA, Stokes ME, Uryniak TJ (1990). Categorical Data Analysis in Statistical methodology in the pharmaceutical sciences

International Committee for the Classification of Retinopathy of Prematurity. The International Classification of Retinopathy of Prematurity revisited. Arch Ophthalmol. 2005 Jul;123(7):991-9.

9. Appendix

9.1 Multiple imputation approach for missing data in primary efficacy analysis

For the primary efficacy analysis, a multiple imputation approach is planned in case that the efficacy data for Week 24 is not available and the otherwise provided data is not sufficient to assess the endpoint otherwise, i.e. if the subject was followed up only shortly and there is no clear information that the subject is a non responder (rescue treatment provided or documentation of ocular unfavorouble structural outcome).

For the multiple imputation approach only a single eye will be imputed (regardless whether the subject was uni- or bilaterally treated) and the same probability of success as subject's having the same treatment group and initial staging (Zone I versus II versus AP-ROP) will be assumed. For subjects where one eye is assessable and the second eye would need imputation, no further imputation will be considered.

For each subject with missing data, the imputation step will be repeated 20 times and the resulting posterior distributions for the probability of response will be combined from all imputation steps. Summaries of the resulting posterior distribution (mean, median and mode) will be derived from the combined posterior distribution.