

Document Type:	Statistical Analysis Plan
Official Title:	An open-label, randomized, active-controlled, parallel-group, Phase-3b study of the efficacy, safety, and tolerability of three different treatment regimens of 2 mg aflibercept administered by intravitreal injections to subjects with diabetic macular edema (DME)
NCT Number:	NCT02818998
Document Date:	08 November 2018



An open-label, randomized, active-controlled, parallel-group, Phase-3b study of the efficacy, safety, and tolerability of three different treatment regimens of 2 mg aflibercept administered by intravitreal injections to subjects with diabetic macular edema (DME)

Efficacy and safety of three different aflibercept regimens in subjects with DME

Bayer study drug BAY86-5321/aflibercept / VEGF Trap-Eye (Eylea)

Study purpose: Posology comparison

Clinical study phase: IIIb **Date:** 08 Nov 2018

Study No.: BAY86-5321/17613 **Version:** 4.0

Author: PPD
Syneos Health
81929 Munich, Germany

INC Project Code: 04.6000.1005566A

Confidential

The information provided in this document is strictly confidential and is intended solely for the guidance of the clinical investigation. Reproduction or disclosure of this document, whether in part or in full, to parties not associated with the clinical investigation or its use for any other purpose without the prior written consent of the sponsor is not permitted.

Throughout this document, symbols indicating proprietary names (®, TM) are not displayed. Hence, the appearance of product names without these symbols does not imply that these names are not protected.

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

1	Introduction.....	6
1.1	Background.....	6
1.2	Rationale of the study	6
1.3	Benefit-risk assessment	6
1.4	List of documents used	7
2	Study Objectives.....	7
3	Study Design	7
3.1	Design overview	7
3.2	Administration schedule	7
3.3	Treatment cessation	8
3.4	Treatment re-start	8
3.5	Randomization.....	9
3.6	Visit overview.....	9
3.7	Assessments overview	10
4	General Statistical Considerations	11
4.1	General Principles.....	11
4.2	Handling of Dropouts	11
4.3	Handling of Missing Data.....	12
4.3.1	Adverse events	12
4.3.2	Prior and concomitant medication.....	12
4.3.3	Efficacy analysis	13
4.4	Interim Analyses and Data Monitoring	15
4.5	Data Rules.....	15
4.6	Validity Review	18
5	Analysis Sets and Subgroups	19
5.1	Analysis Sets.....	19
5.2	Definition of subgroups	19
5.2.1	Subgroups for efficacy and safety analyses	19
5.2.2	Subgroups for efficacy analyses only.....	20
5.2.3	Subgroups for safety analyses only	20
6	Statistical Methodology	21
6.1	Population characteristics	21
6.1.1	Screening failures	21
6.1.2	Disposition	21
6.1.3	Demography	22
6.1.4	Baseline characteristics	23
6.1.5	Medical/ophthalmic history.....	23
6.1.6	Medical history of diabetes and DME.....	24
6.1.7	Prior and concomitant medications	24



6.1.8 Exposure.....25

6.2 Efficacy.....26

6.2.1 Primary efficacy variable26

6.2.2 Secondary efficacy variables.....28

6.2.3 Exploratory efficacy variables.....28

6.2.4 Further exploratory efficacy analyses29

6.2.5 Subgroup analyses.....31

6.3 Pharmacokinetics/pharmacodynamics31

6.4 Safety32

6.4.1 Adverse events (AEs).....32

6.4.2 Laboratory evaluations and pregnancy tests34

6.4.3 Further Safety35

7 Document history and changes in the planned statistical analysis.....37

8 References38

9 Appendix.....39

9.1 Calculation of the NEI VFQ-25 scores.....39

**9.2 Process to derive cut-off data for the Week 52 analysis from the global study database
.....41**



Abbreviations

2PRN	2 mg aflibercept pro re nata (as needed)
2Q8	2 mg aflibercept administered every 8 weeks
2Q8ext	2 mg aflibercept at injection intervals \geq 8 weeks
2Q8fix	2 mg aflibercept at injection intervals of exactly 8 weeks
AE	Adverse Event
ANCOVA	Analysis of Covariance
APTC	Antiplatelet Trialists' Collaboration
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Chemical Classification System
ATE	Arterial Thromboembolic Event
AUC	Area under the Curve
BCVA	Best-Corrected Visual Acuity
BUN	Blood Urea Nitrogen
CI	Confidence Interval
CNV	Choroidal neovascularization
CRF	Case Report Form
CRT	Central Retinal Thickness
DME	Diabetic Macular Edema
DMP	Data-Management-Plan
DR	Diabetic Retinopathy
DRSS	Diabetic Retinopathy Severity Scale
eCRF	Electronic Case Report Form
EMA	European Medicines Agency
ETDRS	Early Treatment Diabetic Retinopathy Study
EU	European Union
FA	Fluorescein Angiography
FAS	Full Analysis Set
FP	Fundus Photography
HbA1c	Glycohemoglobin A1c
HDL	High Density Lipoprotein
HL	Normal Range Higher Limit
IOP	Intraocular Pressure
IVT	Intravitreal(ly)
IxRS	Interactive Voice / Web (either/both) Response System
LDM	Lead Data Manager
LOCF	Last Observation Carried Forward
LL	Normal Range Lower Limit
LS	Least Squares
MAH	Marketing Authorization Holder
MCH	Mean Corpuscular Hemoglobin
MCHC	Mean Corpuscular Hemoglobin Concentration
MCV	Mean Corpuscular Volume
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputations
MMRM	Mixed-Effect Model Repeated Measure
MSSO	Maintenance and Support Services Organisation
NEI VFQ-25	National Eye Institute Visual Functioning Questionnaire 25
NPDR	Non-Proliferative Diabetic Retinopathy
OAD	Operational Acquisition Data
OCT	Optical Coherence Tomography



Statistical Analysis Plan



Protocol No.: **BAY86-5321 / 17613**

Page: 5 of 42

PDR	Proliferative Diabetic Retinopathy
PPS	Per-Protocol Set
PRN	As needed (pro re nata)
QoL	Quality of Life
SMQ	Standardised MedDRA Query
SAE	Serious Adverse Event
SAF	Safety Analysis Set
SAP	Statistical Analysis Plan
SAS	Statistical Analysis System
SD-OCT	Spectral Domain Optical Coherence Tomography
SF	Screening Failure
SOC	System Organ Class
TEAE	Treatment-Emergent Adverse Event
US	United States
V	Visit
VA	Visual Acuity
VEGF	Vascular Endothelial Growth Factor
VRM	Validity Review Meeting
WHO-DD	World Health Organization Drug Dictionary
Wk	Week



1 Introduction

This analysis plan describes the planned analyses for the study reports that will be prepared

- after all subjects have finalized their first year of treatment (Week 52) and
- after the end of the study (Week 100).

The analyses will be mainly described for the Week 52 report. Details for cut-off of the data for the Week 52 analysis are given in the Appendix 9.2.

All analyses will then be repeated for the Week 100 CSR using Week 100 as the respective analysis time point, if not specified otherwise.

1.1 Background

Diabetic retinopathy (DR) is a major cause of visual impairment. Diabetic macular edema (DME) is a manifestation of DR and is the most frequent cause of blindness in young and mid-aged adults.

Vascular endothelial growth factor (VEGF) induces vascular leakage and neovascularization. While neovascularization is the most severe manifestation of DR, vascular leakage leading to macular edema is an important cause of reduced visual acuity (VA).

VEGF Trap-Eye has a high binding affinity for VEGF and can neutralize VEGF mediated biological activity. Therefore, VEGF Trap-Eye may effectively block a key pathway in DME pathophysiology.

The approved European Union (EU) labeling for Eylea (aflibercept) states:

The recommended dose for Eylea is 2 mg aflibercept, equivalent to 50 microliters.

Eylea treatment is initiated with one injection per month for five consecutive doses, followed by one injection every two months. There is no requirement for monitoring between injections.

After the first 12 months of treatment with Eylea, the treatment interval may be extended based on visual and anatomic outcomes. The schedule for monitoring should be determined by the treating physician.

If visual and anatomic outcomes indicate that the subject is not benefiting from continued treatment, Eylea should be discontinued.

More background information is available in the study protocol section 3.1.

1.2 Rationale of the study

The approved dosing schedule for the first year of DME treatment has been thoroughly studied during the clinical development program and is well supported by the available data as originally submitted.

For additional support of the approved dosing schedule for the second year of treatment and beyond, the marketing authorization holder (MAH) seeks to collect further data including information on treatment cessation and restart. To this end, the present study is initiated. It is designed to meet requests made by European Medicines Agency (EMA) as part of a post-approval commitment.

1.3 Benefit-risk assessment

Throughout the entire study, all subjects enrolled will receive active treatment approved for DME with close medical supervision according to established local standard of care.



Taken together, participation in this study is not expected to bear an undue risk for the enrolled subjects.

1.4 List of documents used

- Clinical Study Protocol: No. BAY86-5321/17613, version 1.0, dated 07 Dec 2015
- Case Report Form (CRF) : 17613 VIOLET aCRF final version 1.0, dated 05FEB2016
- Study Protocol Deviation Document (PDD): PDD 17613 Final Version 1.0, dated 26Apr16
- Database Cutoff Specification v1.0, 15Sept2017

2 Study Objectives

Primary objective

To evaluate the efficacy of long-term treatment with 2 mg aflibercept via different intravitreal (IVT) treatment regimens to subjects with DME pre-treated with 2 mg aflibercept every 8 weeks after 5 initial monthly injections for approximately 1 year or more (according to the EU label for the first year of treatment)

Secondary objectives

To assess the safety and tolerability of different treatment regimens of aflibercept in this population

3 Study Design

3.1 Design overview

This is a randomized, 3-arm, active-controlled, parallel-group, open-label, multicenter, Phase-3b study to be conducted in subjects pre-treated with aflibercept for 1 year or more.

Aflibercept 2 mg administered at a fixed schedule every 8 weeks (2Q8fix) is regarded as the reference arm.

Two alternative regimens will be compared to the reference arm: a regimen with a gradually extended dosing interval according to the current EU label (2Q8ext) and a pro re nata dosing scheme (2PRN) with monthly monitoring (see Figure 1).

3.2 Administration schedule

2Q8fix: Fixed injection intervals: 8 weeks throughout the entire treatment period

After the baseline visit, the subjects return every 8 weeks to the investigational site. Starting at the baseline visit, they receive an injection at each visit. The last injection will be at Week 96. The final safety and efficacy assessments are done at Week 100 (final visit).



2Q8ext: Flexible injection intervals \geq 8 weeks (no upper limit)

At the baseline visit and at each subsequent visit, the investigator will determine the duration until the subject's next injection, based on visual and anatomic outcomes at the current visit.

Injection intervals can be extended if visual and anatomic measures are stable. When/if edema recurs, the treatment interval should be reverted to the last treatment interval when the disease was inactive. The increments and decrements for the injection intervals are at the investigator's discretion; typically, increments of 2 weeks are recommended.

2PRN: 2 mg aflibercept pro re nata (as needed)

After randomization, the subjects return every 4 weeks to the investigational site. At each visit, the investigator will determine, based on pre-specified re-treatment criteria, whether an injection is to be given or not. Injection decision will be based on the investigator's judgment.

Further details regarding the scheduled administrations are given in the study protocol Section 5, Table 1.

3.3 Treatment cessation

2Q8fix: Fixed injection intervals - 8 weeks throughout the entire treatment period

Usually, treatment should not be ceased in this reference arm. However, temporary treatment cessation can be considered upon investigator's discretion.

2Q8ext: Flexible injection intervals \geq 8 weeks (no upper limit)

Cessation of treatment can be considered if the interval between two injections reaches or exceeds 16 weeks and if this duration can be maintained for at least two consecutive intervals without need for shortening of the interval. If treatment is halted, the subject should return for monitoring visits at least every 16 weeks.

2PRN: 2 mg aflibercept pro re nata (as needed)

If there is no recurrence of active disease, treatment is terminated until such time, if ever, that the disease recurs. The monthly monitoring is continued throughout.

Further details regarding the cessation of the treatments are given in the study protocol Section 5, Table 1.

3.4 Treatment re-start

2Q8fix: Fixed injection intervals - 8 weeks throughout the entire treatment period

Treatment should be restarted once the condition leading to cessation of treatment is resolved. In case of recurrence of DME it is proposed to begin with the same regimen as a de novo treatment, i.e. five monthly injections followed by a 2Q8 regimen.

2Q8ext: Flexible injection intervals \geq 8 weeks (no upper limit)



Treatment should be restarted if a deterioration of visual and anatomic outcome parameters occurs. If treatment has to be restarted due to recurrence of DME, it is proposed to begin with the same regimen as a de novo treatment, i.e. five monthly injections followed by a 2Q8 regimen.

2PRN: 2 mg aflibercept pro re nata (as needed)

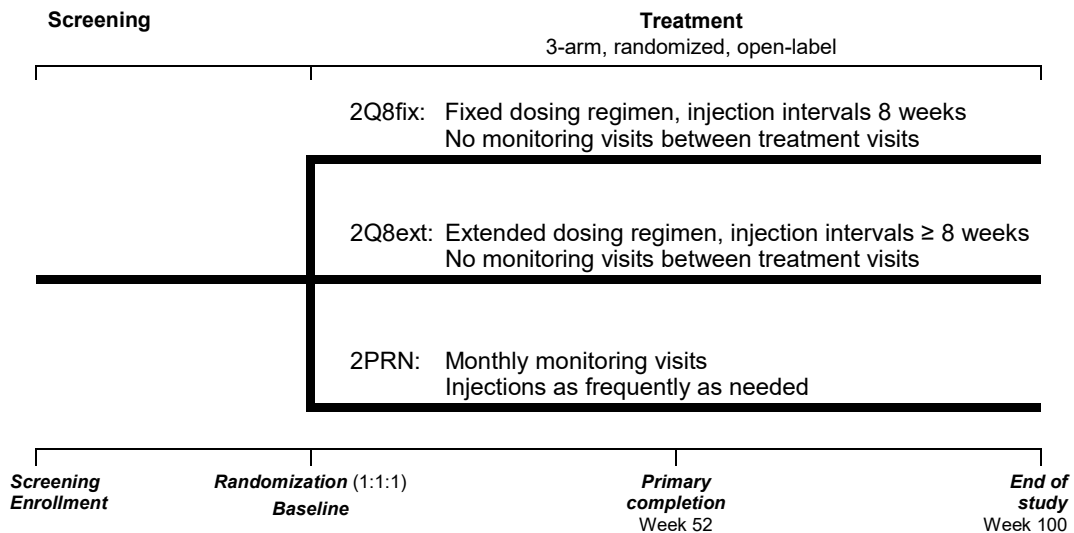
Treatment will be resumed with single injections each time the re-treatment criteria are met. Further details regarding the re-start of the treatments are given in the study protocol Section 5, Table 1.

3.5 Randomization

Randomization will be stratified by 10-letter gain from start of aflibercept treatment (“yes”/”no” as obtained from available medical documentation). For each group, the first injection of study medication will be given at the baseline visit.

Owing to the nature of the different dosing regimens, the three arms will not have the same visit and treatment schedule. Therefore, masking of the treatments is not feasible.

Figure 1: Overall study design



3.6 Visit overview

In general, the time window for all visits is ± 3 days relative to baseline. An overview of visits and assessments is given in Table 5 of the study protocol. In Section 9.2 of the study protocol scheduling and conduct of the following visits are described in detail:

- Screening



- Baseline
- Visits after baseline and before Week 52
- Week 52
- Visits after Week 52 and before Week 100
- Week 100 (final visit or early termination)

If all data needed for enrollment are available, the screening visit and the baseline visit may take place on the same day. In such a case, procedures scheduled for both visits will be conducted only once.

The scheduling of visits „after baseline and before Week 52“ and „after Week 52 and before Week 100“ differs among treatment groups:

- 2Q8fix: Fixed schedule for each subject in this treatment group: Visits take place in strict accordance with the Q8 schedule, i.e. at Weeks 8, 16, 24 etc.
- 2Q8ext: Individualized visit schedule: Visit interval ≥ 8 weeks
- 2PRN: Fixed schedule for each subject in this treatment group: Visits every 4 weeks

The treatment schedule in the 2Q8fix and 2Q8ext groups may deviate from the schedule if cessation and re-start become necessary. If a subject returns for a monitoring visit during temporary treatment cessation, the procedures described for a visit in the 2PRN treatment group should be performed. If treatment is re-started, the paradigms described in Section 3.4 should be followed.

For each treatment group, the assessment of the primary endpoint will be performed at the mandatory visit at Week 52 (± 3 days) after baseline.

The Week 100 visit (± 3 days) after baseline is a mandatory visit, too. After the final visit or after early termination, all subjects return to standard-of-care treatment outside of this study.

The end of the study as a whole will be reached as soon as the last visit of the last subject has been reached in all centers in all participating countries (EU and non-EU).

3.7 Assessments overview

Overall safety of the subjects will be assessed by monitoring

- ocular and non-ocular adverse events (AEs)¹ and
- prior and concomitant medications throughout the study
- vital signs assessed at all study visits and
- laboratory (haematology, chemistry, urinalysis) at screening, Week 52 and Week 100.

¹ All potential arterial thromboembolic events (ATEs) will be adjudicated according to the Antiplatelet Trialists' Collaboration (APTC) endpoints of non-fatal myocardial infarction, non-fatal stroke, and fatal vascular events.



The following ophthalmic examinations will be assessed at all study visits:

- Best-Corrected Visual Acuity (BCVA)
- Optical Coherence Tomography (OCT)
- intraocular pressure (IOP)
- indirect ophthalmoscopy
- slit lamp biomicroscopy
- Fluorescein angiography (FA)
- Fundus photography (FP)

4 General Statistical Considerations

The Statistical Analysis Plan (SAP) of this open-label study was finalized before the first subject received any study drug, to avoid additional reporting bias.

The statistical evaluation will be performed by using the software package Statistical Analysis System (SAS) release 9.3 or higher (SAS Institute Inc., Cary, NC, USA).

4.1 General Principles

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. Mean and standard deviation will be reported to one decimal place greater than the data were collected. Quartiles, median, minimum and maximum values will be reported with the same precision as they were collected.

Frequency tables will be generated for categorical data. These include the counts and percentages of each category including the category 'missing' as a separate category, if applicable. Percentages will be calculated using a denominator of all subjects in the specified population and treatment arm, and the percentage values will be reported to one decimal place.

If not stated otherwise, ANCOVA tables will include the least squares (LS) means and the LS mean changes relative to Baseline for all treatment arms, and the difference in the LS means between 2Q8fix and 2Q8ext or between 2Q8fix and 2PRN group as point estimate and as 95% confidence-interval (CI). If comparisons between 2Q8ext and 2PRN are performed, they are considered as exploratory.

4.2 Handling of Dropouts

Criteria for withdrawal of subjects from the study are specified in Section 6.3.1 of the protocol. Subjects who withdraw from the study will not be replaced. Premature permanent discontinuation from study medication implies premature discontinuation from study participation.



Depending on the time point of withdrawal, a withdrawn subject is referred to as either “Screening failure” or “Drop-out” as specified below:

Screening failure (SF)

A subject who, for any reason (e.g. failure to satisfy the selection criteria), terminates the study before the time point used for the definition of “drop-out” (see below) is considered a “screening failure.” Re-screening is only allowed under special conditions, specified in Section 6.3.1 of the study protocol.

Drop-out

A subject who discontinues study participation prematurely for any reason is defined as a “drop-out” if the subject has already been randomized.

The number of SF will be summarized and the reasons will be listed. The number of subjects who withdrew, as well as the reasons for drop-out of study treatment will be summarized.

4.3 Handling of Missing Data

Data from subjects who drop out of the study will be included in all summaries where possible.

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

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

4.3.1 Adverse events

In general, data will not be imputed for safety analysis. If dates of adverse experiences (clinical or laboratory untoward events) are 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. If it is clearly stated that the adverse event occurred prior to first dose of study drug under this protocol or more than 30 days after the last dose of study drug, it will be treated as not treatment emergent.

4.3.2 Prior and concomitant medication

Completely missing start and stop dates of medication are considered missing and no replacement is generated. A medication with a complete missing start date will be assumed to start before first application of the study drug. A complete missing stop date will be handled as “ongoing”.

In addition, to flag the medications correctly, the worst-case scenario will be used:

- for incomplete start dates: if only day is missing - first of month, if day and month missing - first day (1st January) of year.
- for incomplete stop dates: if only day is missing - last of month, if day and month missing - last day (31st December) of year or “ongoing”, when last day of year is after last study-day of the subject.



4.3.3 Efficacy analysis

The primary method for replacing missing values for all efficacy analyses will be “last observation carried forward” (LOCF), see Section 4.3.3.1.

The following other methods will be applied to account for missing data for the primary endpoint (assuming that the missing values are missing at random):

- Multiple imputations (MI), see Section 4.3.3.2.
- Repeated measurements model, see Section 4.3.3.3.

Moreover, an observed case analysis (OC), see Section 4.3.3.4, will be provided.

4.3.3.1 Last observation carried forward

Missing ETDRS BCVA letter scores will be replaced by LOCF using the last available scheduled post-baseline values.

After the imputation the respective dataset will be filled and the data will be analyzed as if the missing values have been observed (as the last available value). For example, the primary efficacy variable will be analyzed as described in Section 6.2.1.

4.3.3.2 Multiple Imputation

Multiple imputation methods involve three steps:

I. Imputation

i.e., the generation of multiple copies of the original dataset by replacing missing values using an appropriate stochastic model.

- a. First missing data will be imputed in order to achieve a monotone missing pattern using the Markov Chain Monte Carlo (MCMC) method, using SAS-procedure proc MI similarly as below.

```
PROC MI DATA=<indata> SEED=24680 OUT=out1 NIMPUTE=20;  
    MCMC impute=monotone;  
    VAR base t1 t2 tx ... ;  
RUN;
```

- b. Subsequently missing data will be imputed by a regression model.

```
PROC MI DATA=out1 SEED=13579 OUT=full nimpute=1;  
    BY _Imputation_;  
    CLASS treatment;  
    MONOTONE method=reg;  
    VAR treatment base t1 t2 tx ... ;
```



RUN;

II. Analysis

i.e., the analysis of the multiple imputed datasets as complete sets. The analysis step is performed for each of the multiply imputed datasets. Since all imputed datasets are complete there are no issues with missing data. On each imputed dataset the efficacy analysis will be performed as described.

III. Pooling

i.e., the combination of the different parameter estimates across the multiple datasets based on Rubin's rules to produce a unique point estimate and standard error taking into account the uncertainty of the imputation process using SAS procedure PROC MIANALYZE.

Please note, that this MI analysis assumes that the missing data is missing at random.

After the imputation the respective efficacy dataset will be filled and analyzed as described in Section 6.2.1, with the exception of the use of Hochberg procedure to account for multiplicity, since MI is only supportive. For descriptive purposes the un-adjusted p-values will also be displayed.

4.3.3.3 Repeated measurements model

The repeated measurements model (RMM) does not employ formal imputation. RMM has been extensively used in the analysis of longitudinal data especially when missing data is a concern and missing at random (MAR) is assumed (6)

For RMM, if SAS mixed model is used, the sample SAS codes will be like the following:

```
PROC MIXED DATA=<indata>;
  CLASS subject treatment(ref=2Q8) time-window stratum;
  MODEL chg = base treatment stratum time-window treatment*time-window
  /ddfm=kr;
  repeated time-window / sub = subject type = un;
  lsmeans treatment / cl diff at time-window = Week52;
RUN;
```

Where the treatment difference is obtained with the LSMEANS statement for the treatment differences at time Week52. Actual classification² of "10-letter gain from start of aflibercept treatment to baseline (yes/no)" will be used as stratum. Generally, the time-windows from Table 2 will be taken. To get robust results, the values from the mandatory visit "Week 52" will be used instead of the respective time-windows and the values from time-window "Week 49-51" will be dropped.

Because the RMM is only supportive it is not necessary to account for multiplicity.

4.3.3.4 Observed cases

² The wording "actual classification" refers to classification according to the BCVA-values given in the eCRF. It is used in opposite to the supportive analysis "stratified via IxRS".



The primary analysis will be repeated on observed cases, without any imputation.

4.4 Interim Analyses and Data Monitoring

No formal interim analysis in the sense of adaptive or group-sequential design will be conducted.

The data will be analyzed at Week 52 and will serve as basis for the clinical study report to be written at this point, while the final analysis of the Week-100 data will be conducted at that point. The data collected up to Week 52 of this study (after at least two years of treatment with aflibercept) will be reported in a stand-alone clinical study report to make this important information available to the community as soon as possible. The data collected between Week 52 and Week 100 will be reported in a separate clinical study report.

4.5 Data Rules

Generally, pre-treatment values recorded at Visit 2 (Week 0/Day 1) will be used as baseline values. If no baseline values are available, then the last available screening values are used.

Change from baseline is calculated as the value at the post-baseline time point minus the baseline value, i.e. value at time point – value at baseline.

Handling of repeated measurements at the same visit/time-window

If measurements were repeated at the same scheduled visit or in the same time window (see Table 2) the value actually flagged as scheduled will be the

- Last non-missing repeated measurement, if measurement is before start of treatment for this protocol, and
- First non-missing repeated measurement, if measurement is after start of treatment for this protocol.

Generally, only scheduled measurements will be used for statistical summaries and analyses. Unscheduled measurements will not be used for analyses; however, they will be listed.

Handling of time-windows

Visit schedules may deviate by ± 3 days. Scheduled visits should not be altered due to the deviation of the previous visit.

Posology

Treatment posology is detailed in Table 1.



Table 1: Treatment Posology

	2Q8fix		2Q8ext		2PRN		
	Visit	Treatment	Visit	Treatment	Visit	Treatment	
Screening	mandatory Visit 1	no treatment	mandatory Visit 1	no treatment	mandatory Visit 1	no treatment	
Baseline	mandatory Visit 2	mandatory	mandatory Visit 2	mandatory	mandatory Visit 2	mandatory	
Week 4	no visit		no visit		mandatory Visit 3		
Week 8	mandatory visit with treatment		<i>between Week 8 and Week 52, visits (incl. treatment) can be scheduled at any time point (provided injection intervals are ≥ 8 weeks)</i>		mandatory Visit 4		
Week 12	no visit				mandatory Visit 5		
Week 16	mandatory visit with treatment				mandatory Visit 6		
Week 20	no visit				mandatory Visit 7		
Week 24	mandatory visit with treatment				mandatory Visit 8		
Week 28	no visit				mandatory Visit 9		
Week 32	mandatory visit with treatment				mandatory Visit 10		
Week 36	no visit				mandatory Visit 11		<i>throughout all visits between Week 4 and Week 96, treatment is optional according to the re-treatment criteria given in Protocol Section 5 Table 1,</i>
Week 40	mandatory visit with treatment				mandatory Visit 12		
Week 44	no visit				mandatory Visit 13		
Week 48	mandatory visit with treatment				mandatory Visit 14		
Week 52	mandatory	no treatment			mandatory	optional	
Week 56	mandatory visit with treatment		<i>between Week 52 and Week 100, visits (incl. treatment) can be scheduled at any time point (provided injection intervals are ≥ 8 weeks)</i>		mandatory Visit 16		
Week 60	no visit				mandatory Visit 17		
Week 64	mandatory visit with treatment				mandatory Visit 18		
Week 68	no visit				mandatory Visit 19		
Week 72	mandatory visit with treatment				mandatory Visit 20		
Week 76	no visit				mandatory Visit 21		
Week 80	mandatory visit with treatment				mandatory Visit 22		
Week 84	no visit				mandatory Visit 23		
Week 88	mandatory visit with treatment				mandatory Visit 24		
Week 92	no visit				mandatory Visit 25		
Week 96	mandatory visit with treatment		mandatory Visit 26				
Week 100	mandatory	no treatment	mandatory	optional	mandatory Visit 27	no treatment	

The reasons for treatment cessation – and if applicable re-start – will be documented. The treatment schedule in the 2Q8fix and 2Q8ext groups may deviate from the schedule proposed if cessation and re-start become necessary. However, the visits at Week 52 and 100 are mandatory for all subjects regardless of any temporary cessation of treatment.

Defining time-windows for the summaries

Study days will be calculated relative to baseline. Study days prior to first injection date are calculated as (actual date – first injection date), after first injection date as (actual date – first injection date)+1.



For the 2Q8fix- and the 2PRN-groups, the mandatory visits (see Table 1) will be used for the summaries of the efficacy (Section 6.2) and the safety variables (Section 0), with the exception of Adverse Events.

The following time-windows (mostly 8-weeks- intervals) will be used for summarizing the respective results in the 2Q8ext-group:

Table 2: Time-windows for the 2Qext-group.

Screening:	before first Study drug administration
Baseline:	study day 1
Week 1-8:	study days 2 to 59
Week 9-16:	study days 60 to 116
Week 17-24:	study days 117 to 172
Week 25-32:	study days 173 to 228
Week 33-40:	study days 229 to 284
Week 41-48:	study days 285 to 340
Week 49-51:	study days 341 to 361
Week 52:	study days 362 to 368
Week 53-60:	study days 369 to 424
Week 61-68:	study days 425 to 480
Week 69-76:	study days 481 to 536
Week 77-84:	study days 537 to 592
Week 85-92:	study days 593 to 648
Week 93-99:	study days 649 to 697
Week 100:	study days 698 to 704
> Week 100:	study days > 704

In case of multiple scheduled visits in a specific time-window, the following procedures will be performed to get only one value per subject in a window:

- Efficacy variables: the measurement of the last scheduled visit before or at the upper bound of the time interval will be displayed in each time interval.
- Safety variables:
 - Intra-ocular pressure (IOP): The highest values will be used.
 - Indirect ophthalmoscopy: “Abnormal” values and their descriptions will be used, if available, and the highest stage/numbers of vitreous cells.
 - Slit lamp biomicroscopy: “Abnormal” values and their descriptions will be used, if available, and the highest stage/numbers of anterior chamber flare or cells.

For repeated measurement analysis (Section 4.3.3.3), time-windows are also applied to the 2Q8fix and 2PRN treatment groups.



Early termination (ET)

If a subject is terminated early, data from the early termination visit will be mapped to the next scheduled visit where collection of this data was mandatory (LOCF-approach), if he/she was randomized to the 2Q8fix- or the 2PRN-group. For the 2Q8ext-group the time-windows from Table 2 apply. Only variables that were mandatory at the visit will be shown in descriptive tables. All variables mandatory at the ET will be used for LOCF analyses.

Pooling centers

All centers will be combined for the purposes of the analyses.

Calculation of durations

Durations are calculated relative to baseline, if not specified otherwise. Durations will be presented and used (e.g. in summary tables) as integer values.

The integer value of the durations will be listed and summarized, if not specified otherwise.

Coding

The following panels will be coded verbatim by the latest version of Medical Dictionary for Regulatory Activities (MedDRA) available before database lock:

- Medical history
- Adverse events
- Surgeries after start of study

Prior and concomitant medications will be coded by the latest version of World Health Organization Drug classification Dictionary (WHO-DD) available before database lock.

Presentation

Listings will be sorted by treatment group, unique subject identifier, and date time if applicable.

Dates will be formatted as DDMMYYYY. Partial dates will be presented on data listings as recorded on CRFs.

Rounding for all variables will occur only as the last step, immediately prior to presentation in listings and tables. No intermediate rounding will be performed on derived variables. The standard rounding practice of rounding down numbers ending in 0-4 and rounding up numbers ending 5-9 will be employed.

Every table, listing and figure will be produced with an electronic date stamp to document when it was produced.

4.6 Validity Review

Validity Review Meetings (VRMs) are performed according to Bayer Standard Operating Procedures (SOP) and will be led by Lead Data Manager (LDM) of Syneos Health. Details are available in the Data Management Plan (DMP).



The results of the VRM will be documented in the Validity Review Report and may comprise decisions and details relevant for statistical evaluation. Any changes to the statistical analysis prompted by the results of the VRM will be documented in an amendment or, if applicable, in a supplement to this SAP.

5 Analysis Sets and Subgroups

5.1 Analysis Sets

Populations for analysis are defined as follows:

Full Analysis Set (FAS)

The FAS will include all randomized subjects who received any study drug and have a baseline BCVA assessment and at least one post-baseline BCVA assessment. The FAS will be analyzed as randomized.

Per-Protocol Set (PPS)

The PPS will include all FAS subjects who have at least one BCVA assessment at Week 36 or later and do not have a major protocol deviation until Week 52. The PPS will be analyzed as randomized.

Major protocol deviations are described in the Study protocol deviation document, version 1.0, dated 26 April 2016.

Safety Analysis Set (SAF)

The SAF will include all subjects who receive any study drug under this protocol. The safety analysis set will be analyzed as randomized.

5.2 Definition of subgroups

5.2.1 Subgroups for efficacy and safety analyses

- Sex
- Baseline age: <55years; ≥55 to <65years; ≥65 to <75years; ≥75years
- Race: White vs. Other³
- Ethnicity: Hispanic or Latino (no/yes)
- Geographic region: North America vs. Europe
- Baseline HbA1c: above/below 8% (> 8% and ≤ 8%)
- Prior participation in AQUA study: yes / no
- 10-letter gain from start of aflibercept treatment: yes / no (according Interactive Voice / Web (either/both) Response System (IxRS))
- 10-letter gain from start of aflibercept treatment: yes / no (actual values)

³ Race = „Not reported“ is not included here



5.2.2 Subgroups for efficacy analyses only

- Baseline visual acuity (VA) category:
 - < 40 letters
 - ≥ 40 letters to < 55 letters
 - ≥ 55 letters to < 65 letters
 - ≥ 65 letter

5.2.3 Subgroups for safety analyses only

These subgroups will be identified using standardised MedDRA queries (SMQs) of the Maintenance and Support Services Organisation (MSSO):

- Medical history of hypertension (HT): yes / no
 - MSSO-SMQ (20000147) ‘Hypertension (SMQ)’
- Medical history of cerebrovascular disease e.g. cerebrovascular accident (CVA)/stroke: yes / no
 - MSSO-SMQ 20000060 ‘Cerebrovascular disorders’ including sub-SMQs 20000061, 20000063 and 20000064, i.e ‘Central nervous system haemorrhages (SMQ)’, ‘Cerebrovascular conditions (SMQ)’ and ‘Haemorrhagic cerebrovascular conditions (SMQ)’
- Medical history of ischemic heart disease e.g. myocardial infarction: yes / no
 - MSSO-SMQ (20000043 and 20000047) ‘Ischaemic heart disease (SMQ)’ or ‘Myocardial Infarction (SMQ)’
- Medical history of renal impairment (normal/mild/moderate/severe) will be classified with Creatinine Clearance (CLCR) and dialysis.
 - Normal: CLCR >80 ml/min
 - Mild: CLCR >50-80 ml/min
 - Moderate: CLCR >30-50 ml/min
 - Severe: CLCR ≤30 ml/min or requiring dialysis

‘Requiring dialysis’: Preferred term from product-specific Bayer MedDRA query (SMQ_1274) ‘Medical history of renal impairment requiring dialysis’.

CLCR will be calculated using baseline values (creatinine, age, weight, sex) for the Cockcroft-Gault equation:

$$\text{Males: CLCR} = (140 - \text{age}) * \text{body weight} / (72 * \text{creatinine})$$
$$\text{Females: CLCR} = (140 - \text{age}) * \text{body weight} * 0.85 / (72 * \text{creatinine}).$$
- Medical history of hepatic impairment: yes / no
 - MSSO-SMQ 20000005 ‘Hepatic disorders’ including all sub-SMQs except sub-SMQ 20000018 ‘Pregnancy-related hepatic disorders’



6 Statistical Methodology

If not specified otherwise, all tables will be summarized overall and by treatment arms.

6.1 Population characteristics

If not specified otherwise, these variables will be summarized for all 3 analysis populations (overall and stratified by 10-letter gain from start of aflibercept treatment), depending on the type of data as described in Section 4.1.

6.1.1 Screening failures

All available data from Screening Failures (SF) will be listed. At minimum, the following data from SF should be displayed:

- demographic information
- date of informed consent
- reason for premature discontinuation
- date of last visit

For SF with an SAE, the following data will be listed in addition to the data specified above:

- All information related to the SAE such as:
 - Concomitant medication
 - Medical history
- Other information needed for SAE complementary page

6.1.2 Disposition

Disposition will be summarized for all screened subjects.

An overview table for all randomized subjects will be given, displaying the number and percentages of subjects in each treatment arm and overall:

- Completed
- Not completed
 - withdrawn after randomization / including to Week 52
 - withdrawn after Week 52
 - withdrawn during study
 - Primary reason:
 - Adverse Event



- Lost to follow-up
- Withdrawal by subject
- Physician decision
- Pregnancy
- Study terminated by sponsor
- Death
- Protocol violation
- Other

Another overview table for all screened subjects will be given, displaying the number and percentages of subjects in each treatment arm and overall:

- Enrolled⁴
- Screening failures
- Safety Analysis Set (SAF)
- Randomized
- Full Analysis Set (FAS)
- Randomized, but not FAS
 - Reason for exclusion from FAS
- Per-Protocol Set (PPS)
- FAS, but not PPS
 - Reason for exclusion from PPS

The screened but not randomized subjects (screening failures, SF) will be summarized together with the reason(s) for non-treatment. Possible reasons are:

- Adverse event
- Death
- Withdrawal by subject
- Lost to follow-up
- Not met inclusion criteria / met exclusion criteria
- Other

6.1.3 Demography

⁴ defined as subjects with approved informed consent.



The following demographic variables are recorded at screening. In case of repeated measures the last available value before baseline will be used for summary tables.

- Age (years)
- Sex
- Childbearing potential
- Ethnicity
- Race
- Height
- Weight
- Body Mass Index (BMI)
- Study eye - left or right
- Smoking status (yes/no)

6.1.4 Baseline characteristics

The following baseline characteristics are observed at screening and/or baseline. In case of repeated measures the last available value before the first application of the study drug will be used for summary tables.

- BCVA letter scores (study eye)
- Central retinal thickness (CRT) (study eye)
- Diabetic retinopathy severity scale (DRSS) (study eye)

Additionally, the categorized BCVA (5-letter intervals) for the study eye will be presented for baseline values for each of the analysis populations (FAS, PPS, SAF).

6.1.5 Medical/ophthalmic history

Medical history findings (i.e. previous diagnoses, diseases, or surgeries) meeting all criteria listed below will be collected for SAF and FAS:

- Start date before signing of the informed consent
- Considered relevant for the subject's study eligibility

The number and percentages of subjects affected as well as the number of events will be displayed. The following variables are of interest:

- Any medical history finding
- System Organ Classes (SOCs)



- Preferred Terms (PT)

The SOC's will be sorted by descending frequency of subjects, within each SOC the PTs will be sorted by descending frequency of subjects affected.

The tables will be repeated for ocular medical history (i.e. ocular medical history ticked as Yes) by eye (study eye, fellow eye and both eyes). Ocular medical history findings in both eyes will be assigned to the study eye as well as to the fellow eye.

6.1.6 Medical history of diabetes and DME

Medical history of diabetes and DME will be summarized for SAF and FAS, and listed for all subjects, with screening failures on a separate page.

The following variables will be summarized by number of subjects and percentages.

- Diabetes mellitus type 1
- Diabetes mellitus type 2
- DR – study eye
- DR – fellow eye
- DME – study eye
- DME – fellow eye
- Baseline HbA1c classified ($> 8\%$, $\leq 8\%$)

The following variables will be summarized as described in Section 4.1.

- Duration of diabetes until baseline
- Duration of DR until baseline
- Duration of DME until baseline
- Baseline HbA1c

6.1.7 Prior and concomitant medications

Summaries of all prior and concomitant medications recorded will be presented in tabular form for SAF using 3-digit Anatomical Therapeutic Chemical (ATC) classification codes and preferred drug name via the WHO-DD (latest version available before database lock).

The medications will be classified as

- Concomitant: Medications that are ongoing at, began after the start of study drug under this protocol, or medications that were started after end of study drug.
- New Concomitant: Medications that began after the start of study drug under this protocol, and those that were started after end of study drug.



- Prior: Medications that started and stopped before the start of study drug under this protocol.

For each of these categories, a table consisting of medication class and preferred name will be created, sorted by descending frequencies and stratified by

- ocular medication (yes/no) and
- eye (study eye vs. fellow eye)

6.1.8 Exposure

The following parameters will be summarized for all 3 populations (SAF, FAS, PPS):

- The number of study drug injections into the study eye under this protocol.
- The total amount of aflibercept injected into the study eye under this protocol.
- The number of injections into the study eye under this protocol per subject-year: defined as $52 * (\text{Number of iv injections until Week 52}) / (\text{Number of weeks participating})$.
- Study drug exposure into the study eye under this protocol (weeks): calculated as $(\text{date of last injection} - \text{date of first injection at/after randomization} + 28) / 7$.
- Study drug exposure into the study eye in total (weeks): defined as $(\text{date of last injection} - \text{date of first injection of aflibercept} + 28) / 7$ and will be tabulated the same way. The date of first injection of aflibercept shall be recorded as concomitant medication.

For subjects with a treatment cessation, the reasons for cessation as well as the reason for a restart (if applicable) will be summarized in a listing.



6.2 Efficacy

The efficacy variables and the ranking of their statistical analyses at the different time points (primary, secondary, exploratory) are specified in Table 3.

Table 3: Efficacy variables

Variable	Ranking	
	Week 52	Week 100
Change from baseline in ETDRS BCVA letter score for the study eye	Primary	Secondary
Change from baseline in CRT in the study eye	Secondary	Secondary
Proportion of subjects who gained ≥ 10 , ≥ 15 letters	Secondary	Secondary
Proportion of subjects who lost ≥ 30 letters	Secondary	Secondary
Total number of intravitreal injections required in the study eye	Exploratory	Exploratory
Total number of visits	Exploratory	Exploratory
Proportion of subjects who lost ≥ 0 , ≥ 5 , ≥ 10 , ≥ 15 letters	Exploratory	Exploratory
Proportion of subjects who gained ≥ 0 , ≥ 5 letters	Exploratory	Exploratory
Total number of OCT / FA / FP procedures	Exploratory	Exploratory
Mean change from baseline in NEI VFQ-25 total score	Exploratory	Exploratory

BCVA: best corrected visual acuity; CNV: choroidal neovascularization; CRT: central retinal thickness; ETDRS: Early Treatment Diabetic Retinopathy Study

Efficacy analyses will use LOCF-imputed datasets as the datasets of first choice.

6.2.1 Primary efficacy variable

The primary efficacy variable is the change from baseline in ETDRS BCVA letter score for the study eye at Week 52.

It will be assessed at the end of the first year of treatment under this protocol. As the protocol is designed, subjects completing one year in the study will have also completed the end of their second year of treatment given that subjects entering this study will have received a first year of treatment outside of this study.

The primary analysis of the primary efficacy variable will be conducted on the full analysis set (FAS). The analyses will be repeated on the PPS to provide supportive evidence (sensitivity analysis). Additionally, MI, RMM and OC will be performed on the primary efficacy variable for the FAS.

Statistical testing will be conducted to prove the non-inferiority of each of the two extended-dosing regimens (2Q8ext, 2PRN) to the 2Q8 fixed-dosing regimen. The corresponding hypotheses are



- 1. P_{2Q8ext} : Null hypothesis $H_{01}: \mu_{2Q8ext} \leq \mu_{2Q8fix} - D$ versus
Alternative hypothesis $H_{11}: \mu_{2Q8ext} > \mu_{2Q8fix} - D,$
- 2. P_{2PRN} : Null hypothesis $H_{02}: \mu_{2PRN} \leq \mu_{2Q8fix} - D$ versus
Alternative hypothesis $H_{12}: \mu_{2PRN} > \mu_{2Q8fix} - D,$

where

- D = non-inferiority margin chosen as 4 letters (see below for a justification)
- μ_{2Q8ext} = true mean change in BCVA letter score for the study eye from baseline to Week 52 in the 2Q8ext treatment regimen
- μ_{2Q8fix} = true mean change in BCVA letter score for the study eye from baseline to Week 52 in the 2Q8fix dosing treatment regimen
- μ_{2PRN} = true mean change in BCVA letter score for the study eye from baseline to Week 52 in the 2PRN treatment regimen

The two hypotheses stated above will be tested based on an analysis of covariance (ANCOVA) model with the baseline measure as a covariate and treatment group and stratum = “10-letter gain from start of aflibercept treatment to baseline (yes/no)” as a fixed factor:

The observation (BCVA-change from baseline to Week 52) Y_{it} of subject i receiving treatment t can be written as follows:

$$Y_{it} = \mu_t + \gamma_j + x_i\beta + \varepsilon_{it}$$

with

- μ_t denoting the treatment effect for treatment t (2Q8ext, 2Q8fix, 2PRN),
- γ_j denoting the effect of the stratification variable “10-letter gain from start of aflibercept treatment (yes/no)” (actual classification)
- x_i denoting the baseline BCVA of subject i ,
- β denoting the coefficient associated with baseline BCVA and
- $\varepsilon_{it} \sim N(0, \sigma_i^2)$ denoting the residual error.

In the data analysis, ANCOVA can be performed using SAS PROC MIXED procedure with the following key SAS statements:

```
PROC MIXED data=<input dataset>;
  CLASS treatment stratum;
  MODEL chg = base treatment stratum / DDFM=KenwardRoger;
  REPEATED / group = treatment
  LSMEANS treatment / PDIFF=CONTROL('2Q8fix') CL ALPHA=0.05;
RUN;
```

To control the overall type-I error rate for multiple comparisons, Hochberg procedure (4) as implemented in PROC MULTTEST in SAS will be used to adjust for multiplicity.

```
data a;
  input Test$ Raw_P @@;
  datalines;
  test01 <P2Q8ext>
  test02 <P2PRN>;
```



```
run;  
PROC MULTTEST inpvalues=a hoc;run;
```

with p_{2Q8ex} and p_{2PRN} as the p-values of the respective treatment-effects. Non-inferiority of the respective treatment-group versus 2Q8fix is indicated if the lower bound of the 95% CI is above the non-inferiority margin of -4 letters.

Because of possible unequal variances, the Kenward-Roger approximation will be used to the degrees of freedom for the reference distribution.

6.2.2 Secondary efficacy variables

The secondary variables will be analyzed descriptively on the FAS using LOCF, if applicable. They will be summarized overall and by treatment group, including 95% CI.

Though no confirmatory hypothesis testing will be performed, the same model as outlined for the primary efficacy variable will be applied for the following secondary efficacy variables to calculate 95% CI using normal approximation for the differences between the 2Q8ext or 2PRN, respectively and the 2Q8fix group.

- Change from baseline in ETDRS BCVA letter score for the study eye to Week 100
- Change from baseline in CRT in the study eye to Week 52
- Change from baseline in CRT in the study eye to Week 100

The proportions derived from ETDRS BCVA letter scores are summarized per treatment group and overall, including Cochran-Mantel-Haenszel 95% CI for the differences between the flexible regimen and the 2Q8fix group, adjusting for the stratification variable “10-letter gain from start of aflibercept treatment (yes/no)” (actual stratification).

- Proportion of subjects who gained ≥ 10 at Week 52
- Proportion of subjects who gained ≥ 10 at Week 100
- Proportion of subjects who gained ≥ 15 at Week 52
- Proportion of subjects who gained ≥ 15 at Week 100
- Proportion of subjects who lost ≥ 30 letters at Week 52
- Proportion of subjects who lost ≥ 30 letters at Week 100

6.2.3 Exploratory efficacy variables

The exploratory variables will be analyzed descriptively on the FAS using LOCF, if applicable. They will be summarized overall and by treatment group, including descriptive 95% CI. No analysis for the differences between the 2Q8ext or 2PRN, respectively and the 2Q8fix will be performed.



6.2.3.1 Total number of IVT injections required in the study eye

The total number of IVT injections is defined as the number of administered injections per subject at/after randomization until Week 52 and Week 100, respectively, without using LOCF.

In this context, compliance will be summarized similarly (no LOCF, no treatment-differences). Compliance (%) is calculated as the number of intravitreal injections in the study eye at the scheduled injection visits (+/- 3 days) at or after randomization divided by number of scheduled injection visits (multiplied by 100). Scheduled injection visits in the 2Q8ext group are those visits for which the subject was scheduled by the investigator.

6.2.3.2 Total number of visits

The total number of visits as the number of scheduled and un-scheduled visits including Week 52 and Week 100, without using LOCF.

6.2.3.3 Categorized changes of ETDRS BCVA letter score for the study eye

The following proportions will be summarized descriptively using LOCF.

- Proportion of subjects who lost ≥ 0 , ≥ 5 , ≥ 10 , ≥ 15 letters
- Proportion of subjects who gained ≥ 0 , ≥ 5 letters

6.2.3.4 Total number of OCT / FA / FP procedures

OCT, FA and FP procedures are mandatory at screening, Week 52 and Week 100. Additionally, OCT is mandatory in the 2PRN group only throughout the study. Otherwise OCT, FA and FP procedures will be done if deemed necessary by the investigator or required by local medical practice.

- The number of these procedures will be summarized on the FAS without using LOCF, overall and by treatment group, including descriptive 95% CI.

6.2.3.5 Mean change from baseline in NEI VFQ-25 total score

NEI VFQ-25 total scores at Week 52 and Week 100 will be calculated without LOCF. For details see Appendix 9.1. The summary will be done on FAS and PPS without any CI.

In addition, the near and distance activities scale will be presented in the same way as the total score.

6.2.4 Further exploratory efficacy analyses

6.2.4.1 Diabetic retinopathy severity scale as assessed by FP



The 7-Field-FP for both eyes will be transmitted to the central reading center for ETDRS diabetic retinopathy severity scale (DRSS) grading according to the following scale for both eyes.

- 10 – DR absent
- 14 – DR questionable
- 15 – DR questionable
- 20 – Micro-aneurysms only
- 35 – Mild Non-Proliferative Diabetic Retinopathy (NPDR)
- 43 – Moderate NPDR
- 47 – Moderately severe NPDR
- 53 – Severe NPDR
- 61 – Mild Proliferative Diabetic Retinopathy (PDR)
- 65 – Moderate PDR
- 71 – High-risk PDR
- 75 – High-risk PDR
- 81 – Advanced PDR: fundus partially obscured, center of macula attached
- 85 – Advanced PDR: posterior fundus obscured, or center of macula detached
- 90 – Cannot grade, even sufficiently for level 81 or 85

In general, a value of “90” will be treated as missing in the summary tables and statistical analysis, though it will be listed as “90”.

The following summaries will be performed by treatment arm for the study eye:

- Shift tables will be created, displaying the number and percentages of subjects in each class at baseline and at Week 52.
- A “ ≥ 2 step improver” is defined as a subject whose DRSS-category will decrease 2 classes, e.g. from “65 – Moderate PDR” to “53 – Severe NPDR”. The definition of “ ≥ 3 step improvers” is equivalent. Both variables will be displayed.

The proportion of subjects progressing to ≥ 61 in ETDRS DRSS will be displayed together with exact binomial 95% CI, restricted to subjects < 61 in ETDRS DRSS at baseline (LOCF, FAS).

6.2.4.2 Change in BCVA from start of aflibercept treatment

Descriptive statistics for the change in BCVA from start of aflibercept treatment will be provided by visit and treatment group.

6.2.4.3 Average change (AUC) from baseline to Week 52 (Week 100) in BCVA

For each subject the average change from baseline to Week 52 (Week 100) will be calculated using the following formula.

Let x_{i1}, \dots, x_{in_i} be the changes from Baseline in BCVA for subject i measured at the (ordered) planned time points (weeks) t_{i1}, \dots, t_{in_i} , and let $x_{i0} = 0$ and $t_{i0} = 0$. Then the average change is calculated as



$$avg_i = \left(\sum_{j=0}^{n_i-1} (t_{ij} - t_{ij-1}) \frac{x_{ij} - x_{ij-1}}{2} \right) / t_{ini}$$

This corresponds to the area under the individual ‘change from Baseline in BCVA’ curve, if the points are connected with straight lines.

Only observed values (no LOCF) are used for this calculation.

The individual averages will be summarized by descriptive statistics. Furthermore, the averages will be analyzed by an ANCOVA model using treatment and stratification factor (actual classification) as fixed factor and baseline BCVA as covariate.

6.2.5 Subgroup analyses

Descriptive statistics will be displayed for subgroups defined in Section 5.2 for the primary and all secondary efficacy endpoints in Year 1, Year 2 and split by treatment and visit, based on the FAS, and on PPS for the primary endpoint.

For the primary endpoint ANCOVAs will be conducted on the FAS following key SAS statements:

```
PROC MIXED data=<input dataset>;
  CLASS treatment stratum;
  By subgroup;
  MODEL chg = base treatment stratum / DDFM=KenwardRoger;
  REPEATED / group = treatment
  LSMEANS treatment / PDIFF=CONTROL('2Q8fix') CL ALPHA=0.05;
RUN;
```

6.3 Pharmacokinetics/pharmacodynamics

Not applicable.



6.4 Safety

The safety analysis will be conducted on the SAF (as defined in Section 5). The following safety variables will be assessed:

- Adverse events
- Clinical laboratory parameters
- Further safety analyses (vital signs, Surgeries, indirect ophthalmoscopy, slit lamp biomicroscopy)

6.4.1 Adverse events (AEs)

The definitions of AEs and serious AEs (SAEs) are provided in Section 9.6.1.1 of the study protocol,. The classifications according to seriousness, intensity, causality, action taken, other specific treatment and outcome are provided in Section 9.6.1.2 of the study protocol,.

Treatment-emergent adverse events (TEAEs) are AEs that start after the first application of aflibercept under this protocol.

Potential arterial thrombotic events (ATEs) will be evaluated by a masked adjudication committee according to criteria formerly applied and published by the Anti-Platelet Trialists' Collaboration (ATPC). Further details are described in the adjudication committee charter.

A summary table by treatment and overall will be produced for the following categories for all TEAEs:

- Subjects with at least one TEAE
- Subjects with at least one ocular TEAE
- Subjects with at least one ocular TEAE in the study eye
- Subjects with at least one ocular TEAE in the fellow eye
- Subjects with at least one non-ocular TEAE
- Subjects with at least one serious TEAE
- Maximum intensity for TEAEs Maximum intensity for study drug-related AEs
- Subjects with at least one (treatment emergent) ATPC event
- Subjects with at least one TEAE causally related to study drug
- Subjects with at least one TEAE causally related to injection procedure
- Subjects with at least one TEAE causally related to other procedures required by the protocol
- Deaths
- Study-drug related SAEs
- SAEs related to procedures required by the protocol



- Discontinuation of study-drug due to AEs
- Discontinuation of study-drug due to SAEs

The following tables will present the respective TEAEs by MedDRA preferred term (PT) within primary system organ class (SOC) and summarized by treatment arms. Subjects may be counted under multiple system organ classes and preferred terms, but for each system organ class and preferred term, subjects are only counted once.

- TEAEs
- Ocular TEAEs
- Ocular TEAEs in the study eye
- Ocular TEAEs in the fellow eye
- Serious TEAEs
- TEAEs by maximum severity
- Serious ocular TEAEs in the study eye
- Serious non-ocular TEAEs in the study eye
- TEAEs by maximum severity
- Treatment-emergent APTC events
- TEAEs in the study eye causally related to study drug
- TEAEs in the study eye causally related to injection procedure
- TEAEs in the study eye causally related to other procedures required by the protocol
- Serious TEAEs in the study eye causally related to study drug
- Serious TEAEs by maximum severity
- TEAEs causally related to study drug by maximum severity
- Serious TEAEs causally related to study drug by maximum severity
- TEAEs by worst outcome
- Serious TEAEs by worst outcome
- TEAEs resulting in discontinuation of aflibercept
- Non-Serious TEAEs

The number and percentages of subjects affected as well as the number of events will be displayed. SOCs will be sorted by descending frequency of subjects, within each SOC the PT will be sorted by descending frequency of subjects affected.

Subgroup analyses

Reference Number: BHC-RD-OI-119

Supplement Version: 7



Subgroup analyses for TEAEs will be performed for the subgroups described in Section 5.2 **Error! Reference source not found.**, for each of the following types of TEAE:

Number of subjects with

- ocular TEAEs study eye
- non-ocular TEAEs
- Serious ocular TEAEs study eye
- Serious non-ocular TEAEs

Tables will be displayed by treatment, MedDRA SOC and PT.

6.4.2 Laboratory evaluations and pregnancy tests

Hematology, chemistry and urinalysis will be conducted by local laboratories at screening, Week 52 and Week 100. Safety laboratory parameters to be evaluated are summarized in Table 4.

Table 4: Laboratory safety parameters

Chemistry	Urinalysis	Hematology
Sodium	Glucose	Hemoglobin
Potassium	Protein	Hematocrit
Chloride	Specific Gravity	Red blood cell count
Calcium	Blood	Mean corpuscular volume (MCV)
Glucose	Ketones	Mean corpuscular hemoglobin concentration (MCHC)
HbA1c	Protein:Creatinine Ratio (UPCR)	Mean corpuscular hemoglobin (MCH)
Albumin		Leukocyte count
Total Protein, Serum	Pregnancy test for women of childbearing potential (either urine or serum test)	Differential count
Creatinine		Neutrophils
Blood urea nitrogen (BUN)		Lymphocytes
Aspartate aminotransferase (AST)		Monocytes
Alanine aminotransferase (ALT)		Basophils
Alkaline phosphatase		Eosinophils
Total bilirubin		Platelet count
Amylase		
Total cholesterol		
HDL cholesterol		

HDL: High density lipoprotein

According to current ICH guidelines, deviations from the reference range should be evaluated for clinical significance in each individual case. The reference ranges and the units and methods for all variables will be provided by the laboratory.

Deviations of laboratory values from the laboratory reference ranges will be flagged on the laboratory listings.

Treatment-emergent laboratory abnormalities, defined as laboratory abnormalities after randomization under this study protocol, will be summarized by laboratory category and treatment group.



Summary statistics for the laboratory values will only be provided, when they are converted in comparable SI values. Values below “Normal Range Lower Limit” (LL) will be set to $0.5 \times LL$ for the calculation of summary statistics. Values above “Normal Range Higher Limit” (HL) will be set to HL.

Pregnancy tests will be performed for women of childbearing potential; a serum pregnancy test will be performed at screening and urine tests will be performed at baseline, Week 52 and Week 100. Beyond that schedule, pregnancy tests are to be done in women of childbearing potential as frequently as requested.

Results of the pregnancy tests for women of childbearing potential will be listed only.

6.4.3 Further Safety

6.4.3.1 Vital signs

Vital signs (body temperature, systolic and diastolic blood pressure, heart rate) will be analyzed descriptively including changes from baseline.

6.4.3.2 Surgeries after start of study

Separate tables for treatment-emergent surgeries will be presented, overall and split by study eye and fellow eye.

6.4.3.3 Intra-ocular pressure (IOP)

At visits with study drug administration, pre-injection and 30 to 60-min post-injection IOP will be assessed for the study eye. All IOP measurements will be classified as follows:

- 25 mmHg
- ≥ 30 mmHg
- ≥ 35 mmHg
- ≥ 40 mmHg

Any increase from pre-injection/baseline of ≥ 10 mmHg will be flagged.

The IOP analysis will include

- original measurements
- changes from pre-injection
- changes from baseline
- classified measurements
- flagged measurements

This analysis will be performed descriptively at all visits depending on the type of data as described in Section 4.1.

Additionally, the changes of the measurements between pre- and post-injection will be analyzed descriptively depending on the type of data as described in Section 4.1.



6.4.3.4 Slit lamp biomicroscopy

The slit lamp examination will be performed according to local medical practice and applicable medical standards at the site.

Frequency tables for normal/abnormal slit lamp biomicroscopy findings as well as for the grading of 'anterior chamber flare', 'anterior chamber cells' and 'anterior vitreous cells' will be provided.

6.4.3.5 Indirect ophthalmoscopy

At visits with study drug administration, pre-injection indirect ophthalmoscopy will be assessed for the study eye. Post-injection assessments will be carried out 30-60 minutes following the IVT injection.

Frequency tables for normal/abnormal findings in the pre-injection assessment will be provided. The cup-to-disc ratio will be presented by summary statistics. All other data will be listed.



7 Document history and changes in the planned statistical analysis

Statistical Analysis Plan, version 1.0, 31st May 2016 was not approved.

The main changes to Statistical Analysis Plan, version 2.0, 1st June 2016 are as follows:

- a) Section 4.3.3.3: (SAS-Code of) Repeated Measurement Model has been clarified
- b) Section 4.5: Table 2: Time-windows for the 2Qext-group. Eight-week (instead of ten-week) windows have been implemented.
- c) Section 5.2.2: Subgroups for “10-letter gain from start of aflibercept treatment” added
- d) Section 6.17: Prior and concomitant medications: Analysis for FAS has been dropped.
- e) Section 6.1.8: Exposure: Clarifications
- f) Section 6.2.4: Other ophthalmic examinations has been shifted:
 - “Diabetic retinopathy severity scale as assessed by FP” now in “Further exploratory efficacy analyses”
 - “Indirect ophthalmoscopy”, “Slit lamp biomicroscopy” and “Intra-ocular pressure (IOP)” in “Other Safety Analysis”.
- g) The parameters from “Indirect ophthalmoscopy” and “Slit lamp biomicroscopy to be summarized in tables have been clarified.
- h) Section 6.2.5: Subgroup-Analysis: ANCOVAs implemented.
- i) Section 9.2: Process to derive cut-off data for the Week 52 analysis from the global study database has been updated. Now a reference document from Bayer was available.

The main changes to Statistical Analysis Plan, version 3.0, 26 Jan 2018 are as follows:

- j) Section 4.3.2: The worst-case scenario for incomplete dates has been explained in detail.
- k) Section 4.3.3.3: The RMM has been clarified, derivations are needed to get converging results.
- l) Section 5.2.3: The derivations of the subgroups have been described.
- m) Section 6.4.2: The analysis of laboratory values below LL has been described.
- n) Appendix 9.2: The process to derive cut-off data for the Week 52 analysis has been clarified.



8 References

- 1) Clinical Study Protocol No. BAY86-5321/17613, version 1.0, 07 December 2015
- 2) Study Protocol Deviation Document (PDD) version 1.0, 26 April 2016
- 3) COMMITTEE FOR MEDICINAL PRODUCTS FOR HUMAN USE (CHMP), EMA: GUIDELINE ON MISSING DATA IN CONFIRMATORY CLINICAL TRIALS, 2009.
- 4) Hochberg Y: A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 1988; 75: 800-802.
- 5) Mangione CM: NEI VFQ-25 Scoring Algorithm – August 2000, Version 2000: The National Eye Institute 25-Item, Visual Function Questionnaire (VFQ-25) Drug Information Journal dij.sagepub.com, August 2000.
- 6) Mallinckrodt CH, Lane PW, Schnell D, Peng Y, Mancuso JP: Recommendations for the Primary Analysis of Continuous Endpoints in Longitudinal Clinical Trials. *Therapeutic Innovation & Regulatory Science*, July 2008, vol. 42 no. 4 303-319.



9 Appendix

9.1 Calculation of the NEI VFQ-25 scores

The calculation for NEI VFQ-25 sub-scale scores and total score will be performed according to the “NEI VFQ-25 Scoring Algorithm - August 2000“ (5). The most important instructions are displayed below:

The NEI VFQ-25 consists of a base set of 25 vision-targeted questions representing 11 vision-related constructs, plus an additional single-item general health rating question.

Additionally, the NEI VFQ-25 contains the single general health rating question which has been shown to be a robust predictor of future health and mortality in population-based studies.

Scoring NEI VFQ-25 is a two-step process:

1. First, original numeric values from the survey are re-coded following the scoring rules outlined in Table 5. All items are scored so that a high score represents better functioning. Each item is then converted to a 0 to 100 scale so that the lowest and highest possible scores are set at 0 and 100 points, respectively. In this format scores represent the achieved percentage of the total possible score, e.g. a score of 50 represents 50% of the highest possible score.
2. In step 2, items within each sub-scale are averaged together to create the 12 sub-scale scores. Table 6 indicates which items contribute to each specific sub-scale. Items that are left blank (missing data) are not taken into account when calculating the scale scores. Sub-scales with at least one item answered can be used to generate a sub-scale score. Hence, scores represent the average for all items in the subscale that the respondent answered.

Table 5: Scoring Key: Recoding of Items

Item Numbers	Change original response category ^(a)	To recoded value of:
1,3,4,15c ^(b)	1	100
	2	75
	3	50
	4	25
	5	0
2	1	100
	2	80
	3	60
	4	40
	5	20
	6	0
5,6,7,8,9,10,11, 12,13,14,16,16a	1	100
	2	75
	3	50



	4	25
	5	0
	6	*
17,18,19,20,21, 22,23,24,25	1	0
	2	25
	3	50
	4	75
	5	100

(a) Pre-coded response choices as printed in the questionnaire.

(b) Item 15c has four-response levels, but is expanded to a five-levels using item 15b.

Note: If 15b=1, then 15c should be recoded to "0"

If 15b=2, then 15c should be recoded to missing.

If 15b=3, then 15c should be recoded to missing

* Response choice "6" indicates that the person does not perform the activity because of non-vision related problems. If this choice is selected, the item is coded as "missing."

To calculate an overall composite score for the NEI VFQ-25, simply average the vision-targeted subscale scores, excluding the general health rating question. By averaging the sub-scale scores rather than the individual items we have given equal weight to each sub-scale, whereas averaging the items would give more weight to scales with more items.

Table 6: Averaging of Items to Generate NEI VFQ-25 Sub-Scales

Scale	Number of items	Items to be averaged (after recoding per Table 5)
General Health	1	1
General Vision	1	2
Ocular Pain	2	4, 19
Near Activities	3	5, 6, 7
Distance Activities	3	8, 9, 14
Vision Specific:		
Social Functioning	2	11, 13
Mental Health	4	3, 21, 22, 25
Role Difficulties	2	17, 18
Dependency	3	20, 23, 24
Driving	3	15c, 16, 16a
Color Vision	1	12
Peripheral Vision	1	10

**Scoring example**

Items 5, 6, and 7 are used to generate the near activities sub-scale score (Table 6). Each of the items has 6 response choices.

- Response choice 6 indicates that the respondent does not perform the activity because of reasons that are unrelated to vision. If a respondent selects this choice, the answer is treated as missing and an average of the remaining items is calculated.
- Response choice 5 indicates that an activity is so difficult that the participant no longer performs the activity. This extremely poor near vision response choice is to “0” points before taking an average of all three items.
- To score all items in the same direction,
- In step 2, items within each sub-scale are averaged together to create the 12 sub-scale scores. Table 6 indicates which items contribute to each specific sub-scale. Items that are left blank (missing data) are not taken into account when calculating the scale scores. Sub-scales with at least one item answered can be used to generate a sub-scale score. Hence, scores represent the average for all items in the subscale that the respondent answered.
- Table 5 shows that responses 1 through 5 for items 5, 6, and 7 should be recoded to values of 100, 75, 50, 25, and 0 respectively.
- If the respondent is missing one of the items, the person's score will be equal to the average of the two non-missing items.

Formula:

$$\text{Mean} = \frac{\text{Score for each item with a non-missing answer}}{\text{Total number of items with non-missing answers}}$$

Example:

$$\text{With responses converted:} = (25 + 100 + 25) / 3 = 50$$

Note: 100 = Best, 0 = Worst possible score.

9.2 Process to derive cut-off data for the Week 52 analysis from the global study database

General rules from the guidelines “Database Cutoff Specification v1.0, 15Sept2017” will be followed:

- All variables which are set by the timing concept, e. g., relative days, as well as the treatment emergent flag will be re-calculated after the cutoff is done.
- If there is a domain, which depends on another, the parent domain is always cut first, the dependent domain will then be cut accordingly⁵.

⁵ For example, after removal of an AE after cut-off, the respective concomitant medication for this AE shall also be removed



- If an observation has a partial date, the observation will be cut only if the partial date is without any doubt after the cutoff date.

There are some exceptions with regard to the guideline mentioned above.:

Section 5.1: *Cutoff Visit: The visit which indicates any data collected by this visit will be used for analysis*

As described in Section 6, the interim data release will be based on a planned observation time for each subject, i.e., a visit-based cut-off will be used. In addition to the definitions given above, study drug injections given at the cut-off visit as well as data of any assessments scheduled post-injection at this visit will not be included.

Section 7.1: *Date based cutoff: Specification for different Operational Acquisition Data (OAD) Domains*

Usually authorities request a special listing of all subjects who died after cutoff date. For this purpose the information related to “death” (primary cause of death, death date, etc.) should not be removed even if after the cutoff date.

This rule is applicable in oncological studies and will not be used here. Therefore, in this study fatal AEs are not treated differently to other serious AEs.

In addition, some algorithms for specific domains have been created:

- a. In case of an injection at Week 52, this exposure and the IOP after injection will not be included in the cut data. Last injection will be the respective previous injection of this subject.
- b. In case of missing Week 52 date, study-day $365+7=372$ will be used as the cutoff date in general
- c. AE-dataset (to account for the AEs starting in the time-window 30 days after last injection)
 - if Week 52 is available, the maximum of last injection-date + 30 days and Week 52-date will be used as the cutoff date
 - if Week 52 is not available, the maximum of last injection-date + 30 days and study-day $365+7=372$ will be used as the cutoff date.