Document Type:	Statistical Analysis Plan
Official Title:	Open-label Phase-4 study to examine the change of vision-related quality of life in subjects with diabetic macular edema (DME) during treatment with intravitreal injections of 2 mg aflibercept according to EU label for the first year of treatment
NCT Number:	NCT02581995
Document Date:	04 May 2017

Open-label Phase-4 study to examine the change of vision-related quality of life in subjects with diabetic macular edema (DME) during treatment with intravitreal injections of 2 mg aflibercept according to EU label for the first year of treatment

Investigation of the change of vision-related quality of life in subjects treated with aflibercept according to EU label for DME (AQUA)

Bayer study drug	BAY86-5321/ aflibercept / VEGF Trap-Eye (Eylea)
Study purpose:	Assessment of quality of life

Clinical study phase:	4	**Date:**	04 May 2017
Study No.:	BAY86-5321/ 17850	**Version:**	3.0
Author:	PPD INC Research PPD Germany		
INC Project Code:	04.6000.1005565	**EUDRAC:**	2014-005119-17

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.

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

Table of Tables

Abbreviations

AE	adverse event
ANCOVA	analysis of covariance
APTC	Antiplatelet Trialists' Collaboration
ATC	Anatomical Therapeutic Chemical
ATE	Arterial thrombotic event
BCVA	Best corrected visual acuity
BMI	Body mass index
BSE	Better-seeing eye
CRF	case record form
CRT	central retinal thickness
DME	diabetic macular edema
DR	diabetic retinopathy
DRSS	diabetic retinopathy severity score
EMA	European Medicines Agency
EoS	end of study/early termination
ET	early termination
ETDRS	Early Treatment Diabetic Retinopathy Study
EU	European Union
EV	equal vision
FA	fluorescein angiography
FAS	full analysis set
FP	fundus photography
IOP	intraocular pressure
IVT	intravitreal(ly)
LOCF	last observation carried forward
MCID	minimal clinically important difference
MedDRA	Medical Dictionary for Regulatory Activities

MI	multiple imputation
NEI VFQ-25	The National Eye Institute 25-Item Visual Function Questionnaire
OCT	optical coherence tomography
PAES	postapproval efficacy study
PDR	proliferative diabetic retinopathy
PT	Preferred term
SAE	serious adverse event
SAF	safety population
SAP	statistical analysis plan
SmPC	summary of Product Characteristics
SOC	system organ class
SOP	standard operating procedure
UPCR	urine protein/creatinine ratio
VA	visual acuity
VEGF	vascular endothelial growth factor
VIVID (DME)	A randomized, double masked, active controlled, phase III study of the efficacy and safety of repeated doses of intravitreal VEGF Trap-Eye in subjects with diabetic macular edema
VISTA (DME)	A double- masked, randomized, active-controlled, Phase 3 study of the efficacy and safety of intravitreal administration of VEGF Trap-Eye in patients with diabetic macular edema
VRM	Validity Review Meeting
WHO-DD	World Health Organization Drug Dictionary
WSE	worse-seeing eye

1. Introduction

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. It is estimated that 4.8% of the global population has diabetic retinopathy, while 3% to 4.1% of Europeans are affected.

Details of background information are available in the study protocol.

1.2 Rationale of the study

Little is known about QoL outcomes in DME patients treated with aflibercept per the European Summary of Product Characteristics (SmPC), or the relative influence of the visual acuity of better- and worse-seeing eyes on QoL in such patients.

Details of the study rationale are available in the study protocol.

1.3 List of documents used

- Clinical Study Protocol No. BAY 86-5321 /17850, version 1.0, 10 February 2015
- Clinical Study Protocol No. BAY 86-5321 /17850, version 2.0, 07 September 2015

2. Study Objectives

Primary objective

To evaluate the change in quality of life (NEI VFQ-25) in subjects with DME during the first year of treatment with aflibercept according to the EU label for DME.

Secondary objectives

- To assess further the safety and tolerability of aflibercept in this population
- To assess the change in the diabetic retinopathy severity score (DRSS) from baseline to Week 52
- To support subject recruitment for the EMA-requested post-approval efficacy study in DME

3. Study Design

Design overview

Single-arm, multicenter study administering aflibercept according to EU label for the first year of treatment, i.e. 2 mg every 4 weeks for 5 consecutive doses, and dosing every 8 weeks thereafter.

Justification of the design

Open-label setting: This is a single arm study without masking requirements. Treatment in this study is according to EU label for DME.

The length of the observation period (1 year) is based on the primary-endpoint results of the pivotal DME studies (VISTA DME, VIVID DME), where most of the benefits after start of treatment with aflibercept occurred during the initial year of treatment.

End of 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).

Visit overview

The scheduled assessments and procedures are displayed in the following Table 1.

Table 1: Schedule of assessments and procedures

	Screening Visit 1	Baseline Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7 to Visit 10	End of study / early termin. Visit 11
Acceptable deviations relative to BL	4 weeks before BL	Day 1	Week 4 ± 5 days	Week 8 ± 5 days	Week 12 ± 5 days	Week 16 ± 5 days	Week 24 to Week 48 ±10 days[a]	Week 52c ± 10 days
Initiation procedures								
Informed consent	●							
Demographic data	●							
Medical / ophthalmic history	●							
Check of enrollment criteria	●	●						
Study medication								
Administration of study drug		●	●	●	●	●	●	no treatment
Ophthalmologic assessments (●● = bilaterally; ●○ = study eye only [additional assessments may occur outside of this protocol])								
BCVA (ETDRS chart starting at 4 m) [b]	●●	●○	●○	●○	●○	●○	●○	●●
Optical coherence tomography	●○	●○						●○
Fluorescein angiography	●○							●○
Fundus photography	●●							●●
Indirect ophthalmoscopy [c]	●○	●○	●○	●○	●○	●○	●○	●○
Slit lamp biomicroscopy	●●	●●	●○	●○	●○	●○	●○	●○
Intraocular pressure (IOP) [c]	●○	●○	●○	●○	●○	●○	●○	●○
Patient-reported outcomes								
NEI VFQ-25	●	●	●	●	●	●	●	●
Standard safety								
Prior / concomitant medications	●	●	●	●	●	●	●	●
Adverse events [d]	●	●	●	●	●	●	●	●
Hematology / chemistry	●							●
Urinalysis / UPCR	●							●
Pregnancy test – serum [e] (women of childbearing potential only)	●							
Pregnancy test – urine dipstick (women of childbearing potential only) [f]		●						●
Vital signs (body temperature, blood pressure, pulse)	●	●						●

BCVA = best corrected visual acuity; BL = baseline; ETDRS = Early Treatment Diabetic Retinopathy Study; UPCR = urine protein / creatinine ratio

a   The intervals between Visits 6 to 9 are 8 weeks ±10 days.
     The interval between Visit 9 and Visit 10 must be ≥ 56 days (8 weeks).

b: Refraction to be done at each visit

c   Also post injection

d: Any AE occurring up to 4 weeks after the last injection of aflibercept has to be documented, regardless of the causal relationship to the study drug or the seriousness of the event and reported in accordance with this protocol (i.e. not as a spontaneous report). For any drug-related AE occurring after 4 weeks after the last application of aflibercept, the standard procedures that are in place for spontaneous reporting will be followed. All potential arterial thrombotic events (ATEs) will be adjudicated according to the Antiplatelet Trialists' Collaboration (APTC).

e   The test is to take place within 7 days before the first injection of study medication

f   The test is to be repeated as frequently as required

4. General Statistical Considerations

The Statistical Analysis Plan (SAP) for this open-label study was finalized before First-Patient-First-Visit to avoid additional reporting bias. Later updates afterwards are of cosmetic type and/or rely on updates of other core documents.

The statistical evaluation will be performed using the software package 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, minimum, quartiles, median, and maximum will be calculated for metric data.

Frequency tables will be generated for categorical data.

4.2 Handling of Dropouts

Depending on the time point of withdrawal, a withdrawn subject is referred to as either "screening failure" or "dropout"

A subject who discontinues study participation and/or study drug prematurely for any reason is defined as a dropout, if the subject has already received at least one dose of study medication.

Subjects must or might be withdrawn from study for different reasons, which are specified in the protocol in Section 6.3.1. Subjects who withdraw from the study will not be replaced. The number of subjects who withdrew, as well as the reasons for drop-out of study treatment will be summarized.

A subject who, for any reason (e.g. failure to satisfy the selection criteria), terminates the study before first study drug application is regarded as a screening failure. Re-screening of screening failures may be acceptable under the specific conditions, which are specified in the protocol in Section 6.3.1. A subject may be re-screened once only. To be eligible, re-screened subjects must meet all selection criteria at the re-screening visit.

The number of screening failures will be summarized, the reasons will be listed.

4.3 Handling of Missing Data

All missing or incomplete data will be presented in the subject data listings as they are recorded on the Case Report Form (CRF). In general they will not be substituted or replaced, except for the parameter described in this Section, Sections 4.5 and 6.2.

General rules

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

In case of missing data from safety variables all replacements will be performed according to a worst case scenario. Missing drug relationship will be replaced by drug related and missing intensity will be replaced by severe intensity.

Adverse events

In case the AE onset date is missing or incomplete, the date will be imputed for the calculation according to the worst case approach. If the AE onset date is missing it will be imputed with the start of treatment date. For incomplete AE onset dates the missing day will be imputed as the first day of the month and the missing month will be imputed as January. If the replaced starting date is prior to the start of treatment date, the starting date will be set to the start of treatment date. Exceptions are given where the partial dates would allow to state otherwise.

In case the AE end date is completely missing it will not be imputed. If the AE end date is partially missing, the missing day will be imputed with the last day of the month, and the missing month will be imputed with December.

Prior and concomitant medication / medical history

Completely missing start and stop dates of medication / medical history event are considered missing and no replacement is generated. A medication / medical history event 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".

If only the day is missing, it will be replaced with the first day of the month for start dates, and with the last of the month for stop dates. If day and month are missing, and the year is non-missing, the date will be completed with '01 January' for start dates and with '31 December' for stop dates.

4.4   Interim Analyses and Data Monitoring

No formal interim analysis will be conducted.

4.5   Data Rules

Determination of baseline values

Generally, pre-treatment values recorded at Visit 2 (Day 1) will be used as baseline values. This visit should take place within 4 weeks of the screening visit. If no baseline value is available then last available scheduled 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

If measurements were repeated at the same scheduled visit, the value actually flagged as scheduled will be the

- Last non-missing repeated measurement, if visit is before start of treatment, and

- First non-missing repeated measurement, if visit is after start of treatment.

Generally, only scheduled measurements will be used for statistical summaries and analysis. Unscheduled measurements will not used for analysis, but only listed.

Handling time-windows

The screening visit must occur within 4 weeks of the baseline visit (Day 1). Screening values collected more than 4 weeks before baseline will be flagged in the patient listings, but used for summary tables and analysis, if no unscheduled screening measurement are available.

For the scheduling of Visit 3 to Visit 10 (Week 4 to Week 48), the following deviations relative to baseline are foreseen:

- Monthly post-baseline visits through week 16 (Visits 3 to 6): ± 5 days

- Regular Q8 schedule (Visits 7 to 10) ± 10 days

- The interval between Visit 9 and Visit 10 must be at least 56 days (8 weeks).

Visit 11 (Week 52) will be the last scheduled visit under this protocol. It is to be scheduled at Week 52 (± 10 days) after baseline.

Values outside these windows will not be dropped from summary tables and analysis, but flagged as "outside time-window" in the patient listings.

Early termination (ET)

Visit based information of the early termination visit will be mapped to the closest missing visit according to the visit schedule. This can result in data for visits at which this variable was not scheduled to be collected. This data will nevertheless be included into the (LOCF-) analyses and be reported as observed value for the respective visit.

Pooling centers

No pooling of centers will be performed.

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.

Classification of the study eye

It is up to the investigator to decide which eye will be treated with the study drug (study eye) and which not (fellow eye). At baseline and at end of study/early termination (EoS) the BCVA assessment will be performed in both eyes. In the other scheduled visits only the study will be examined.

The following classification of the study eye (Table 2), first published from Bressler et al (2) and recommended from Hirneiss (3) will be performed:

Table 2: Definition of Better-seeing eye, Worse-seeing eye and equal vision

	Better-seeing eye (BSE)	Worse-seeing eye (WSE)	Equal vision (EV)
Baseline VA letter score in both eyes $\geq$ (20/100)*	Baseline VA of the study eye is better than that of the fellow eye by $\geq$ 5 letters	Baseline VA of the study eye is worse than that of the fellow eye by $\geq$ 5 letters	Baseline VA of the study eye is within +/- 4 letters of that of the fellow eye
Baseline VA letter score in one or both eyes < (20/100)*	Baseline VA of the study eye is better than that of the fellow eye by $\geq$ 10 letters	Baseline VA of the study eye is worse than that of the fellow eye by $\geq$ 10 letters	Baseline VA of the study eye is within +/- 9 letters of that of the fellow eye

* A Snellen score of 20/100 is equivalent to an ETDRS letter of 50.

This classification, which takes the variability of VA measurements through implementing the EV-category into account, will equivalently used for EoS.

Coding

The verbatim of the following panels will be coded 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 unique subject identifier and date if applicable.

Dates will be formatted as DDMMMYYYY. 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 numbers ending in 0-4 down and numbers ending 5-9 up 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's Standard Operating Procedures (SOP) and will be led by the INC Research Lead Data Manager. Details are available in the Data-Management-Plan.

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 validity review meeting will be documented in an amendment and, if applicable, in a supplement to this SAP.

5.   Analysis Sets

5.1   Assignment of analysis sets

Populations for analysis will be defined as follows:

Full Analysis Set (FAS)

The Full Analysis Set (FAS) will include all subjects who received at least one injection of study drug and have completed the baseline and at least one post-baseline NEI VFQ-25 questionnaire. The FAS will be the primary dataset for all efficacy parameter.

Safety analysis set (SAF)

The Safety Analysis Set will include all subjects who have received at least one injection of study drug. The SAF will be the primary dataset for all safety variables.

6.   Statistical Methodology

No hypothesis testing will be performed. 95% confidence intervals will be provided as appropriate. No stratification is planned.

6.1   Population characteristics

If not mentioned otherwise, the population characteristics will be summarized in FAS and SAF depending on the type of data as described in Section 4.1.

6.1.1   Screening Failures and protocol deviations

Screening Failures will be listed together with the reason for their exclusion from study and all available data (Date of informed consent, Reason for screening failure, Date of last visit). For screening failures with an SAE all information related to the SAE will be listed.

Protocol deviations will be summarized for all enrolled subjects.

6.1.2   Subject validity status

An overview-table will summarize the subject validity and primary reasons for exclusion from analysis for all subjects assigned to treatment, displaying the number and percentages of:

- subjects valid for safety analysis
- subjects valid for FAS analysis
- excluded from SAF (by reason)
- excluded from FAS (by reason)

6.1.3   Subject disposition

The number and reason for withdrawal will be obtained and summarized in a frequency table referring to

- end of screening (all enrolled subjects),

- end of treatment (all randomized subjects), and

- end of study (all randomized subjects)

In addition, subjects disposition (all enrolled subjects) will be summarized to get the number of subjects enrolled, with screening failures, randomized, treated, and completed the study.

6.1.4   Demography and baseline characteristics

Demographic variables will be summarized for FAS and SAF and listed for all subjects, with screening failures on a separate page, and patients excluded from FAS flagged.

The following demographic parameters will be summarized:

- Gender
- Age
- Race
- Ethnicity
- Weight
- Height
- Body mass index (BMI)
- Smoking history – never, former, current
- Study eye – left or right
- Study eye – BSE, WSE or EV

The following baseline characteristics are observed at screening and/or baseline and will be tabulated for FAS and SAF. In case of repeated measures the last available value before 1st application of the study drug will be used for summary tables.

- baseline NEI VFQ-25 total score
- baseline BCVA letter scores (study eye)
- baseline DRSS (study eye)
- baseline CRT (study eye)

Baseline HbA1c will be classified (> 8%, <=8%), and summarized as categorized and as metric variable.

6.1.5   Medical history

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

Overall medical history will be summarized based on the MedDRA version available before database lock. 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
- System Organ Classes (SOCs)
- Preferred Terms (PTs)

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

The table will be repeated for ocular medical history by eye (study eye, fellow eye). Bilateral findings are counted as findings of the study eye as well as the fellow eye.

6.1.6   Medical history of diabetes, diabetic retinopathy and DME

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

For the following variables number of patients (percentage) and duration of the disease will be summarized.

- Diabetes Mellitus Type 1
- Diabetes Mellitus Type 2
- Diabetic Retinopathy – Study Eye
- Diabetic Retinopathy – Fellow Eye
- Diabetic Macular Edema – Study Eye
- Diabetic Macular Edema – Fellow Eye

6.1.7   Prior and concomitant medication

Summaries will be presented in tabular form using 3-digit Anatomical Therapeutic Chemical Classification (ATC) classification codes and subclass via the World Health Organization Drug classification Dictionary (WHO-DD), latest version available before database lock.

The medications will be classified as follows

- Concomitant: Medications that are ongoing at, began after the start of study drug, or medications that were started after end of study drug
- New concomitant: Medications that began after the start of study drug, and those that were started after end of study drug
- Prior: Medications that started and stopped before the start of study drug

For each of these categories a table for the SAF-population will be created overall and stratified by ocular medication (yes/no) and study eye (study eye vs. fellow eye) consisting of ATC class and subclass used by the subjects, sorted by descending frequencies. Bilateral medications are counted as medications for the study as well as the fellow eye.

6.1.8   Exposure

The number of injections, total amount of aflibercept and the duration of exposure will be tabulated for the FAS and the SAF.

6.1.9   Smoking History

History of cigarette smoking (never, former, current) will be tabled as baseline characteristics.

The following parameters will be listed only.
- Number of cigarettes per day
- Age at start of habitual cigarette use
- Duration of habitual cigarette use
- Any other tobacco smoking type (never, former, current)

6.2   Efficacy

Efficacy parameter will be listed for all subjects summarized for the FAS and listed for all subjects, with patients excluded from FAS flagged.

6.2.1   Primary Efficacy Variable – NEI VFQ-25 total score

The primary efficacy variable is

- The change from baseline to Week 52 in the NEI VFQ-25 total score.

The NEI VFQ-25 total score will be evaluated at each visit (Visit 1 to Visit 11). 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" (4). The most important instructions are displayed in appendix 9.1.

The scores and their changes from baseline to all post-baseline visits in the NEI VFQ-25 total score will be summarized descriptively, including 95% confidence intervals for the changes.

Primary analysis

The primary efficacy variable will be summarized descriptively. Ninety-five percent confidence intervals will be provided based on the t-distribution assuming that the changes from baseline are normally distributed.

To be comparable with publications of other studies in this indication, missing values will be imputed with the last observed post-baseline value collected before the missing value (LOCF).

Sensitivity analysis

Several sensitivity analyses will be performed to support results of the primary analysis.

Non-parametric analysis

The Central Limit Theorem implies that the limiting distribution of the sample mean is normal regardless of the distribution of the original data (under some conditions), so that the normal approximation is quite accurate in large sample sizes.

Nevertheless, the assumption of normality of the primary efficacy variable will be described and checked with the Shapiro-Wilk test. If the assumption of normality can be rejected a non-parametric 95%-CI, consisting of 2.5%- and 97.5%-quartiles as confidence limits, will be calculated and Hodges-Lehmann estimations will be performed.

The primary method (LOCF) is not necessarily conservative because subjects dropping out due to lack of efficacy are considered as stable. Therefore, several further sensitivity analyses will be provided, which are described below

Observed Cases (OC)

The primary analysis will be repeated on the observed values, i.e., only including patients that have a NEI VFQ-25 total score result at Week 52 (+/- 10 days), i.e. without LOCF-imputation.

Categorized Analysis

Frequencies of subjects worsening or improving, additionally for not achieving, achieving or exceeding the minimally clinically important difference (MCID) will be computed for each time point. MCID is defined as 5.

The change in QoL will be classified according to the two categories, change $\geq$ or $<$ MCID. The category $<$ MCID includes no change in QoL or deteriorations.

Multiple Imputation (MI)

MI Instead of LOCF-imputation will be performed to estimate the missing values. The imputations will be based on the baseline value as well as all observed changes from baseline. Subsequently, the primary efficacy analysis will be repeated.

It is assumed that any systematic difference between the missing values and the observed values can be explained by differences in observed data, i.e. the missing efficacy data are "Missing at random".

Multiple imputation aims to allow for the uncertainty about the missing data by creating several different plausible imputed data sets and appropriately combining results obtained from each of them. The validity of results from multiple imputation depends on such modeling being done carefully and appropriately.

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. The missing data will be imputed by a two step procedure.

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

```
PROC MI DATA=<indata> SEED=3456 OUT=full NIMPUTE=20;
                                *(=number of imputations m)*;
BY <...>;
mcmc impute=monotone;
VAR y1 y2 y3 y4 y5;              *(results depend on order);
RUN;
```

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

```
PROC MI DATA=out1 SEED=5678 OUT=full nimpute=1;
BY <imputation_ ...>;
MONOTONE reg;
VAR y1 y2 y3 y4 y5;
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 is no need to bother with any missing data.

```
proc means data=full alpha=0.05 clm mean std;
VAR y1 y2 y3 y4 y5;
run;
```

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, if applicable.

Repeated Measurement Analysis

The effect of the type of eye treated (better-seeing or worse-seeing at baseline) during the
study shall be described within a repeated-measurements-model including

- visit
- baseline value
- baseline*visit interaction and
- eye

as fixed factors. "Eye" denotes the study eye class (BSE, EV or WSE) at baseline. The model
is:

$$Y_i = X_i\beta + \varepsilon; \text{ with}$$

i = 1, …, n (subjects)

$\varepsilon \sim N(0, \sigma^2)$

Y = changes from baseline

X = vector of fixed effects (visit, baseline, interaction, eye)

The analysis will be performed for OC and LOCF. SAS code similar to the following will be used.

```
PROC MIXED DATA = <indata>;
CLASS avisitn subjid eye;
MODEL chg = base avisitn avisitn*base eye/ DDFM=kr;
REPEATED avisitn / SUBJECT = subjid TYPE=un;
RUN;
```

Exploratory analysis

Subgroups

The effect of the choice of the study eye (BSE, WSE or EV), gender and prior use of anti-VEGF on the quality of life will be investigated at each post-baseline visit as follows.

The change to baseline in the NEI VFQ-25 total score will be summarized descriptively in each subgroup at each visit.

Ninety-five percent confidence intervals will be provided based on the t-distribution assuming that the changes from baseline are normally distributed. LOCF and OC will be analyzed.

Intra-individual Variability

Further exploratory analyses will be conducted to investigate the intra-individual variability in the score over time. This will be done graphically displaying the distribution of the standard deviation of the individual changes from baseline over time with a box-plot or scatter-plot.

Correlation BCVA and QoL

The correlations between the BCVA-changes to baseline for the following eyes and the NEI VFQ-25 total score-changes to baseline will be performed using Pearson's Correlation Coefficient and displayed graphically using scatter-plots on the LOCF-data.

- Study eye
- BSE at baseline
- WSE eye at baseline
- BSE at EoS
- WSE at EoS

Furthermore the correlation between the change in NEI VFQ-25 total score and
- 'Minimum of BCVA in left eye and BCVA in right eye at EOS' minus 'Minimum of BCVA in left eye and BCVA in right eye at Baseline'

- 'Maximum of BCVA in left eye and BCVA in right eye at EOS' minus 'Maximum of BCVA in left eye and BCVA in right eye at Baseline'

will be investigated.

In addition, the mean change from baseline in BCVA will be evaluated, and a 95% confidence interval for the difference between both MCID-categories, change $\geq$ or $<$ MCID, will be calculated.

Possible confounder

The effect of the following parameters on the NEI VFQ-25 total score-changes to baseline the will be analyzed, using Pearson's Correlation Coefficient and displayed graphically using scatter-plots

- duration of diabetes mellitus (years),
- duration of diabetic retinopathy (years),
- duration of diabetic macular edema (years),
- age (years).

To describe the influence of gender the distribution (mean, SD) of the NEI VFQ-25 total score-changes to baseline will be displayed graphically over time, overall and stratified by gender. The same will be done regarding the prior use of anti-VEGF (Y/N).

6.2.2   Secondary Efficacy Variables

In general, the secondary efficacy variables will be analyzed descriptively in analogy to the primary efficacy variables, if applicable.

The secondary efficacy variables are

- The change from baseline to Week 52 in the NEI VFQ-25 near activities subscale
- The change from baseline to Week 52 in the NEI VFQ-25 distant activities subscale
- The change from baseline to Week 52 in BCVA (ETDRS letter score)
- The change from baseline to Week 52 in CRT measured by optical coherence tomography (OCT)
- Proportion of subjects progressing to $\geq$ 61 ETDRS diabetic retinopathy severity scale (DRSS) as assessed by fundus photography (FP)

6.2.2.1   NEI VFQ-25 near activities subscale

The NEI VFQ-25 near activities subscale will be calculated according to Table 5 (see Appendix 9.1). The statistical analysis strategy will be performed equivalently to those of the primary endpoint (Section 6.2.1), except the exploratory analysis.

6.2.2.2   NEI VFQ-25 distant activities subscale

The NEI VFQ-25 distant activities subscale will be calculated according to Table 5 (see Appendix 9.1). The statistical analysis strategy will be performed equivalently to those of the primary endpoint (Section 6.2.1), except the exploratory analysis.

6.2.2.3   Best corrected visual acuity (BCVA)

The values might range from 0 to 100. Primary and sensitivity analysis will be performed according to the analysis of the primary endpoint, if applicable.

The categorized analysis will be performed with regard to the following classes during the course of time:

- ≥ 15 letter gain (Yes/No)
- ≥ 10 letter gain (Yes/No)
- ≥ 5 letter gain (Yes/No)
- ≥ 0 letter gain (Yes/No)
- ≥ 0 letter loss (Yes/No)
- ≥ 5 letter loss (Yes/No)
- ≥ 10 letter loss (Yes/No)
- ≥ 15 letter loss (Yes/No)

The exploratory analysis described in 6.2.1 is fitted to the QoL-scores and might not be applicable.

6.2.2.4   Central retinal thickness (CRT) measured by OCT

The retinal thickness will be recorded at screening, baseline and end of study/early termination (EoS) at the study eye only. Absolute changes to baseline will be calculated. Missing post-baseline values will not be imputed. No additional analysis except the descriptive summaries will be performed.

6.2.2.5   ETDRS diabetic retinopathy severity scale (DRSS)

The ETDRS DRSS will be assessed by FP at screening and at EoS according to the following scale for both eyes.

- 10 – DR absent
- 14 – DR questionable
- 15 – DR questionable
- 20 – Micro-aneurysms only
- 35 – Mild NPDR
- 43 – Moderate NPDR
- 47 – Moderately severe NPDR
- 53 – Severe NPDR
- 61 – Mild 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".

Shift tables will be created, displaying the number and percentages of subjects in each class at baseline and at Week 52, stratified by study eye vs. fellow eye.

A "≥ 2 step improvers" 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 as described in Section 4.1, stratified by study eye vs. fellow eye.

According to exclusion criterion 5 subjects suffering on an active proliferative diabetic retinopathy (PDR) in the study eye will be screening failures and excluded from analysis. Therefore, regarding the study eye, subjects progressing to ≥ 61 in ETDRS DRSS is equivalent to subjects getting a PDR during the study. Some subjects might be dropped from this analysis of progression in the fellow eye, because they already start at baseline with active PDR (exclusion criterion only valid for the study eye). .

The proportion of subjects progressing to ≥ 61 in ETDRS DRSS will be displayed together with exact binomial 95% confidence intervals in both eyes separately. The comparison of both proportions and their CI will be provided.

Because of the type of this parameter (classified, only one post-baseline measurement) the analyses regarding the primary endpoint (Section 6.2.1) are not applicable.

6.2.3 Exploratory Efficacy Variables

6.2.3.1 Average changes of NEI VFQ-25 during the study

The average of the changes with regard to baseline will be calculated for OC and LOCF in the total score and both sub-scores (NEI VFQ-25 near activities subscale and NEI VFQ-25 distant activities subscale) for each subject in the FAS. These average changes will be displayed as described in Section 4.1.

6.2.3.2 Indirect ophthalmoscopy

Pre- and post-injection results of the indirect ophthalmoscopy in the both eyes (vitreous body, optic nerve head, macula, peripheral retina and retinal tear, break, detachment or hemorrhage) will only be listed.

6.2.3.3 Fluorescein angiography (FA)

The FA parameters (leakage, central ischaemia, neovascularization elsewhere and neovasularization disc) will only be listed.

6.2.3.4 Fundus photography (FP)

Results of the FP will be listed for the following parameters

- microaneurysms,
- hemorrhages and microaneurysms,

- hard exudates entire field,
- hard exudates center of macula,
- soft exudates,
- intraretinal micro-vascular abnormalities ,
- neovascularization disc,
- fibrous proliferation on disc,
- neovascularization elsewhere,
- fibrous proliferation elsewhere,
- pre-retinal hemorrhage,
- venous beading,
- venous loops and/or reduplication,
- prior pan-retinal photocoagulation and/or
- vitreous hemorrhage

6.2.3.5   Refraction

Results of the Refraction will be listed only.

6.2.3.6   Optical coherence tomography (OCT)

With the exception of CRT (see Section 6.2.2.4) results of the OCT for the study eye (macular edema, cysts visible, subretinal fluid, vitro-macular traction, macular hole/pseudohole/lamellar hole fluid) will only be listed.

6.2.3.7   Classification of the study eye

At Baseline and EoS the study eye can be categorized in BSE, WSE or EV (see Table 2). A shift table will be provided for Week 52 (using LOCF data) to summarize the change in the classification from Baseline to Week 52.Pharmacokinetics / pharmacodynamics

 Not applicable

6.3   Safety

Safety parameters will be analyzed and listed for the Safety analysis set (SAF).

6.3.1   Adverse Events (AEs)

Treatment emergent AEs are defined as AEs that started after the first injection of study-treatment but not more than 30 days after the last injection of study-treatment under this protocol. A summary table of number of subjects with AEs and another one with treatment emergent AEs (TEAE) will be produced displaying the following categories:

- Subjects reporting at least one (TE)AE
- Subjects reporting at least one ocular (TE)AE in the study eye
- Subjects reporting at least one ocular (TE)AE in the fellow eye
- Subjects reporting at least one non-ocular (TE)AE
- Subjects reporting at least (treatment emergent) APTC event

- Subjects reporting at least one causal related to aflibercept (TE)AE
- Subjects reporting at least one causal related to iv injection (TE)AE
- Subjects reporting at least one causal related to other procedures (TE)AE
- Maximum intensity for any (TE)AE
- Maximum intensity for any study drug-related (TE)AE
- (Treatment emergent) deaths
- Subjects reporting at least one serious (TE)AE
- Subjects reporting at least one study-drug related serious (TE)AE
- SAEs related to procedures required by the protocol
- Discontinuation of study drug due to (TE)AEs
- Discontinuation of study drug due to SAEs

Treatment emergent AEs will be presented by MedDRA preferred term within primary system organ class (SOC) and summarized. Intensity and causal relationship to the investigational product will be analyzed descriptively.

The following tables are foreseen:

- TEAEs
- Ocular TEAEs
- Ocular TEAEs in the study eye
- Ocular TEAEs in the fellow eye
- Non-ocular TEAEs
- Serious TEAEs
- Serious ocular TEAEs
- Serious ocular TEAEs in the study eye
- Serious ocular TEAEs in the fellow eye
- Serious non-ocular TEAEs
- TEAEs by maximum severity
- Treatment emergent APTC events
- Causal related to aflibercept TEAE
- Causal related to iv injection TEAE
- Causal related to other procedures TEAE
- Serious study-drug related TEAEs
- Serious TEAEs by maximum severity
- Study-drug related TEAEs by maximum severity
- TEAEs by worst outcome
- Serious TEAEs by worst outcome
- Treatment emergent deaths
- TEAEs resulting in discontinuation of aflibercept
- Non-serious AEs

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.

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 PTs will be sorted by descending frequency of subjects affected.

Potential arterial thrombotic events (ATEs) will be evaluated by an adjudication committee according to criteria formerly applied and published by the Anti-Platelet Trialists' Collaboration (APTC) (1). The definition of ATEs as well as further details are described in the adjudication committee charter. They will be presented separately.

In the tables, 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.

In addition, subject listings will be performed for all deaths, serious AEs and AEs resulting in discontinuation of aflibercept. These specific and mandatory listings will be included in the tables section of the statistical output.

6.3.2   Pregnancy test

Results of the pregnancy tests (serum and urine dipstick) will be listed only.

6.3.3   Laboratory Tests

The following hematology, chemistry and urinalysis parameter will be summarized by visit. They will be converted to the Bayer Standard Units given in brackets.

Table 3: Laboratory safety parameters

Chemistry	Urinalysis	Hematology

Sodium (MMOL/DL)
Potassium (MMOL/DL)
Chloride (MMOL/DL)
Calcium (MG/DL)
Glucose (MG/DL)
HbA1c (%)
Albumin (G/DL)
Total Protein, Serum (G/L)
Creatinine (MG/DL)
Blood urea nitrogen (MG/DL)
Aspartate aminotransferase (U/L)
Alanine aminotransferase (U/L)
Alkaline phosphatase (U/L)
Total bilirubin (MG/DL)
Amylase (U/L)
Total cholesterol (MG/DL)
High density lipoprotein cholesterol (MG/DL)
	Protein (MG/DL)
	Specific Gravity (G/ML)
	Protein : Creatinine Ratio (MG/G CRE)
		Hemoglobin (G/DL)
		Hematocrit (%)
		Red blood cell count (T/L)
		Mean corpuscular volume (FL)
		Mean corpuscular hemoglobin concentration (G/DL)
		Mean corpuscular hemoglobin (PG)
		Leucocytes count (GIGA/L)
		Differential count
		   Neutrophils (GIGA/L)
		   Neutrophils/Leukocytes (%)
		   Lymphocytes (GIGA/L)
		   Lymphophils/Leukocytes (%)
		   Monocytes (GIGA/L)
		   Monophils/Leukocytes (%)
		   Basophils (GIGA/L)
		   Basophils/Leukocytes (%)
		   Eosinophils (GIGA/L)
		   Eosinophils/Leukocytes (%)
		Platelet count (GIGA/L)

Shift tables will be performed for the categorical parameters, glucose, blood and ketones from the urinalysis.

Additionally, treatment emergent high or low laboratory values will be tabled.

6.3.4 Vital Signs

Vital signs (body temperature, blood pressure [diastolic and systolic], pulse) will be summarized by visit.

6.3.5 Intraocular Pressure (IOP)

Pre- and post- injection IOP and change from baseline in IOP will be summarized by visit.

Classified pre- and post- injection IOP (> 25 mmHg, ≥ 30 mmHg, ≥ 35 mmHg, ≥ 40 mmHg) in the study eye will be summarized by visit.

6.3.6 Slit lamp biomicroscopy

Slit lamp biomicroscopy data for the study eye (Frequency of abnormal findings regarding anterior chamber and lens) will be summarized as shift tables (baseline vs. Week 52).

Specifications of all slit lamp biomicroscopy –related findings will be listed.

7. Document history and changes in the planned statistical analysis

- SAP final draft version 1.0, dated 19 Mar 2015 (not finalized)
- SAP final version 2.0, dated 20 May 2015

8. References

1. Antithrombotic Trialists' Collaboration. Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. Br Med J. 2002; 324: 71-86.

2. Bressler NM, Chang TS, Suner IJ, et al. Vision-related function after ranibizumab treatment by better- or worse-seeing eye: clinical trial results from MARINA and ANCHOR. Ophthalmology. 2010;117:747–756.

3. Hirneiss C. The impact of a better-seeing eye and a worse-seeing eye on vision-related quality of life. Clinical Ophthalmology 2014; 8: 1703–1709

4. Mangione, CM. NEI VFQ-25 Scoring Algorithm – August 2000, Version 2000, The National Eye Institute 25-Item, Visual Function Questionnaire (VFQ-25)

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" (4). 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, which are also presented in the protocol in Section 16.1.

The NEI VFQ-25 generates the following vision-targeted subscales (number of questions):

- global vision rating (1),
- difficulty with near vision activities (3),
- difficulty with distance vision activities (3),
- limitations in social functioning due to vision (2),
- role limitations due to vision (2),
- dependency on others due to vision (3),
- mental health symptoms due to vision (4),
- driving difficulties (3),
- limitations with peripheral (1),
- color vision (1), and
- ocular pain (2).

Additionally, the 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 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 4. 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 5 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 4: Scoring Key: Recoding of Items

Item Numbers	Change original response category[a]	To recoded value of:
1,3,4,15c[b]	1 2 3 4 5	100 75 50 25 0
2	1 2 3 4 5 6	100 80 60 40 20 0
5,6,7,8,9,10,11, 12,13,14,16,16a	1 2 3 4 5 6	100 75 50 25 0 *
17,18,19,20,21, 22,23,24,25	1 2 3 4 5	0 25 50 75 100

(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 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 5: Averaging of Items to Generate VFQ-25 Sub-Scales

Scale	Number of items	Items to be averaged (after recoding per Table 4)
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
   Mental Health
   Role Difficulties
   Dependency
	
	2
	4
	2
	3
		
		11, 13
		3, 21, 22, 25
		17, 18
		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 5). 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 recoded to "0" points before taking an average of all three items.

- To score all items in the same direction, Table 4 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.