# Diabetic Retinopathy Clinical Research Network

# Intravitreous Anti-VEGF Treatment for Prevention of Vision Threating Diabetic Retinopathy in Eyes at High Risk

# Statistical Analysis Plan

| VERSION NUMBER | AUTHOR | APPROVER | EFFECTIVE DATE | REVISION DESCRIPTION |
|---|---|---|---|---|
| 1.0 | Danni Liu | Michele Melia | 17 October 2017 | Initial version for Protocol version 4.0 |
| 2.0 | Wesley T. Beaulieu | Michele Melia | 10 January 2019 | Revisions for Protocol version 5.0 and consistency with other DRCR.net SAPs following DSMC review. Key changes include the following. (1) No interim analysis is planned. (2) Four key outcomes defined with a hierarchical approach to controlling Type I error. (3) Changed the minimum sample size for subgroup analyses to 20 per treatment per level of subgroup covariate. (4) Will impute missing data for secondary analyses of visual acuity and OCT data. |
| 3.0 | Kristin Josic | Michele Melia | 04 Mar 2021 | Revised to extend the 48-month visit analysis close window from 3 to 6 months due to concerns about delayed visits during COVID. Note: revision implemented after the primary 2-year manuscript was submitted for publication and did not affect the 2-year publication. |

## 1.0 Introduction

This document specifies the statistical analyses to be performed for the Diabetic Retinopathy Clinical Research Network (DRCR.net) study evaluating anti-vascular endothelial growth factor (anti-VEGF) treatment for prevention of vision-threatening diabetic retinopathy in high-risk eyes (Protocol W). Technical details of the analyses reported in the primary manuscript will be documented separately in a technical analysis plan.

This study has two primary objectives. First, to determine the efficacy and safety of intravitreous aflibercept injections versus sham injections (observation) for prevention of proliferative diabetic retinopathy (PDR) and central-involved diabetic macular edema (CI-DME) with vision loss in high-risk eyes. Second, to compare long-term vision outcomes in eyes that receive anti-VEGF therapy early in the course of disease with those that are observed initially and treated only if PDR or CI-DME with vision loss develop.

Study eyes are randomly assigned to one of two treatment groups: sham injections or intravitreous 2-mg aflibercept injections. Study participants may have one or two study eyes. Participants with two study eyes receive sham injections in one eye and intravitreous aflibercept injections in the other eye.

Randomization is stratified as follows:

- Study participants with one study eye are randomly assigned with equal probability to one of two treatment groups: sham injections or intravitreous aflibercept injections.
    - Randomization for participants with one study eye is stratified by reading center grading of diabetic retinopathy (DR) severity level (43, 47A, 47B-D, 53 with no neovascularization in the periphery, or 53 with neovascularization in the periphery).
- Study participants with two study eyes (both eyes must be eligible at the time of randomization) are randomized with equal probability to one of the following:
    - Sham injections in the eye with greater DR severity and intravitreous aflibercept injections in the eye with lower DR severity.
    - Intravitreous aflibercept injections in the eye with greater DR severity and sham injections in the eye with lower DR severity.
        - If both eyes have the same DR severity, then the right eye is considered the eye with the greater DR severity.

For the purpose of analysis, the randomization stratification variables will be modeled as two categorical variables, defined as laterality (one or two eyes randomized) and DR severity level based on reading center assessment of digital fundus photographs (43, 47A, 47B-D, 53 with no neovascularization in the periphery, or 53 with neovascularization in the periphery). If there are not at least 20 eyes per treatment group in each of the DR severity levels specified above, then

adjacent categories may be combined (e.g., 47A and 47B-D; 53 without peripheral neovascularization and 53 with peripheral neovascularization).

## 2.0 Efficacy Analysis Plan

*Primary outcome*: development of PDR or DME defined as the first occurrence of any of the following (composite time-to-event outcome):

- PDR Outcomes:
    - Neovascularization within the 7-modified Early Treatment Diabetic Retinopathy Study (ETDRS) fields on fundus photography or fluorescein angiography, confirmed by a masked grader at the central reading center
        - At non-annual visits, fundus photography and fluorescein angiography will only be submitted to the reading center to assess for this component of the primary outcome if the investigator thinks treatment is necessary.
    - Neovascularization of the iris (at least 2 cumulative clock hours), definitive neovascularization of the angle, or neovascular glaucoma on clinical exam (photographic documentation not required)
    - Other outcomes presumed to be from PDR and documented: traction retinal detachment, vitreous hemorrhage, or pre-retinal hemorrhage greater than ½ disc area
    - Procedures undertaken for the treatment of PDR (when present or presumed to be present): PRP, anti-VEGF, or vitrectomy
- DME Outcomes:
    - CI-DME on clinical exam with at least 10% increase in central subfield thickness from baseline and either (1) at least a 10-letter decrease in visual acuity from baseline at a single visit or (2) a 5-to-9-letter decrease in visual acuity from baseline at 2 consecutive study (i.e., not unspecified) visits at least 21 days apart, with vision loss presumed to be from DME
    - Non-topical treatment for DME performed without meeting the above criteria, including focal/grid laser or intravitreous injections for DME

The primary outcome analysis will be performed when the last enrolled participant reaches 2 years of follow up and will include all available follow-up data. The treatment groups will be compared using the hazard ratio.

*Other Key Outcomes:*

- Development of PDR or DME outcome through 4 years
- Mean visual acuity change from baseline at 2 years
- Mean visual acuity change from baseline at 4 years

*Type I Error Rate Control*

The overall Type I error rate for the primary outcome and all key outcomes will be controlled at 5%. To control the Type I error rate for each time point, 2.5% Type I error will be allocated to the 2-year analysis, and 2.5% will be allocated to the 4-year analysis. To control the Type I error rate for the multiple key outcomes, a hierarchical approach will be used. The visual acuity outcome will be formally compared (i.e., with a *P* value) only if there is a significant treatment group difference in the anatomic outcome at the same time point. If the visual acuity outcome is not compared because the PDR/DME outcome is not significant, then only point estimates and 97.5% confidence intervals for within and between group changes in visual acuity from baseline will be computed at the time point.

## 2.1 Primary Outcome Analyses

*PDR/DME Outcome*

Comparison of the PDR/DME composite time-to-event outcome will be based on the hazard ratio from a marginal Cox regression model. The analysis will adjust for laterality and retinopathy severity. The correlation between eyes of participants having two study eyes will be modeled with a robust sandwich estimate of the covariance matrix (Lee, Wei, and Amato 1992). The primary analysis is an intention-to-treat analysis. Data from participants not observed to meet outcome criteria who are lost to follow up will be censored at the time of the last completed visit. The hazard ratio and 97.5% confidence interval for the treatment effect will be presented.

*Visual Acuity Outcome*

If there is a significant difference in the PDR/DME composite outcome at 2 or 4 years ($P \leq .025$), a treatment group comparison of the difference in the mean change in visual acuity from baseline to the outcome visit will be conducted at the same time point with alpha of .025. A linear mixed model will be used to estimate the mean treatment group difference and 97.5% confidence interval. The analysis will adjust for baseline visual acuity, laterality, and retinopathy severity. The correlation between eyes of participants having two study eyes will be modeled using random intercepts. This will be an intention-to-treat analysis that includes all randomized eyes. Multiple imputation will be used to impute missing data. The imputation model will include laterality, retinopathy severity, baseline visual acuity, and change in visual acuity from baseline at each protocol assessment visit up to and including the analysis time point. For the 2-year analysis, visual acuity measured beyond 2 years, if available, will not be included in the analysis or imputation. If the PDR/DME outcome is not significant at the same time point, then the *P* value will not be reported, but the 97.5% confidence interval will be reported.

### 2.1.1 Sensitivity Analyses

A sensitivity analysis of the key visual acuity outcomes including only observed data from participants completing the visit (2 or 4 years) will be conducted (i.e., complete-case analysis). If the analyses of imputed and observed data differ substantially, then exploratory analyses will be

performed to evaluate factors that may have contributed to the difference. The sensitivity analysis of completers will only be performed if more than 10% of randomized participants would be excluded by these criteria.

Multiple imputation assumes that data are missing at random (MAR). In the present study, this means that whether change in visual acuity is missing may be a function of observed characteristics included in the imputation model, but not a function of the unobserved data being imputed. This assumption cannot be tested directly since these data are unknown. However, a tipping point analysis for the key visual acuity outcomes will be conducted to adjust the imputed values using a shift parameter and thereby determine how severe the departure from MAR must be to change the outcome of the analysis with respect to rejecting or failing to reject the null hypothesis. The tipping point analysis will only be conducted if more than 10% of randomized participants would be excluded by these criteria.

A shift parameter will be applied to the imputed values in the aflibercept group to determine the tipping point at which the results of the primary analysis are nullified. That is, if one group is found to be superior, the tipping point will identify the shift parameter necessary to negate the superiority. Conversely, if the null hypothesis is not rejected, two tipping points will be identified – one that would make aflibercept superior and one that would make sham superior. In either case, this tipping point will be evaluated to determine if it is plausible. If not, then the MAR assumption is likely reasonable. For example, if the tipping point were 100 letters, then this would be evidence that the MAR assumption is reasonable.

### 2.1.2 Per-protocol Analysis

Per-protocol analyses for the PDR/DME composite and visual acuity outcomes will be performed including only eyes that received at least 80% of injections (sham or intravitreous) according to protocol and no other treatment for DR or DME. The limited cohort for the per-protocol analysis will be described in the technical plan. Missing data will not be imputed. The per-protocol analysis will be conducted only if at least 10% of randomized participants would be excluded by these criteria.

The intention-to-treat analyses are considered the primary analyses. If the intention-to-treat and per-protocol analyses yield similar conclusions, then the per-protocol analyses will be used to provide supportive evidence of the magnitude of the treatment effect among participants who had good adherence to the treatment. If the results of the two methods differ, then exploratory analyses will be performed to evaluate factors that could have contributed to the differences.

### 2.1.3 Confounding

Imbalances between groups in important covariates are not expected to be of sufficient magnitude to produce confounding in the primary analysis and other key analyses. However, the presence of confounding in the primary and other key analyses will be evaluated in additional regression models by adding baseline covariates that are potentially associated with the outcome. These include but are not limited to the following:

- Age
- Duration of diabetes
- HbA1c
- Mean arterial blood pressure
- Visual acuity
- Prior treatment for DME
    - Note that eyes with prior DME treatment within 12 months of randomization or more than 4 prior intraocular injections were ineligible
- OCT central subfield thickness
- Each of the following within 500 µm of the center of the macula on OCT as graded by the reading center (minimum 20 eyes in the cohort with the characteristic):
    - Epiretinal membrane
    - Vitreomacular traction
    - Cystoid abnormalities
    - Subretinal fluid
- Hard exudates within 1800 µm of the center of the macula on fundus photography as graded by the reading center (minimum 20 eyes in the cohort with the characteristic)

Additional variables associated with the outcome will be included if there is an imbalance in the variables between groups. Imbalance by treatment group will not be judged using statistical testing, but will be based on judgment regarding whether the size of the imbalance is clinically important, i.e., whether it is large enough that it could have a clinically important effect on visual acuity or worsening to PDR or DME.

### 2.1.4 Subgroup Analyses

The treatment effect will be assessed at the 2- and 4-year visits in subgroups determined by baseline factors in pre-planned subgroup analyses. These analyses will repeat the primary and other key analyses, with the exception that multiple imputation for missing outcome data will not be performed for the visual acuity outcome. Unless the imputation process is done separately for each treatment group and the subgroup factor is included in the imputation model, the analysis is biased towards the null hypothesis of no interaction when an interaction is present (Sullivan et al., 2016). It is recognized that analyzing only observed data also might be biased if data are not missing at random; however, the imputed analysis is unavoidably biased in the presence of interaction.

A term for the main effect of the baseline subgroup factor and an interaction term for baseline subgroup factor by treatment will be added to the models used for the primary outcome and all other key outcomes. If the interaction *P* value is less than .025, the estimated treatment effect and 97.5% confidence interval will be obtained from the interaction model for each subgroup. Summary statistics will be presented for each outcome by subgroup, regardless of significance.

Baseline factors to be evaluated for possible subgroup effects include the following:

- Prior anti-VEGF treatment: yes vs. no
- Prior DME treatment: yes vs. no
- Diabetic retinopathy severity (as graded by the photograph reading center): less than 53 vs. 53 with no neovascularization in the periphery vs. 53 with neovascularization in the periphery
- Non-central DME: yes vs. no

To increase statistical power and reduce the risk of spurious results, which are more likely in a small sample size, subgroups will only be analyzed if there are at least 20 eyes in each treatment group for each subgroup. Interaction p-values will be calculated using continuous and ordinal variables, where possible, in addition to the categorizations described above.

Subgroup analyses will be conducted to determine whether the overall treatment effect is similar to the treatment effects seen in these subgroups. For each subgroup, the hypothesized direction of effect is for a larger treatment difference among eyes with more severe characteristics (i.e., prior anti-VEGF treatment, prior DME treatment, more severe DR level, or non-central DME).

It is recognized that the study is not powered to detect subgroup effects and that lack of significance is not necessarily an indication that subgroup effects do not exist. In the absence of a significant treatment effect in the primary or other key analyses, assessment of subgroups will be interpreted with caution.

There are no data to suggest that the treatment effect will vary by gender or race/ethnicity. However, both factors will be evaluated in exploratory subgroup analyses.

### 2.1.5 Center Effects

The number of study participants per center is expected to be small for many centers. Therefore, center effects will not be included in the statistical model. However, for centers with a large number of study participants ($N \geq 20$ in either treatment group), heterogeneity across centers will be explored using random center effects by estimating empirical best linear unbiased predictors with 95% confidence intervals.

## 2.2 Secondary Outcome Analyses

The treatment groups will be compared on secondary outcomes of interest at 2 and 4 years. In general, analyses will be adjusted for the baseline measure of the outcome, where appropriate, and laterality and retinopathy severity. For each secondary outcome, the hypothesis test of no difference between treatment groups will be conducted and the estimated treatment effect with a 97.5% confidence interval will be calculated. Descriptive statistics will be presented at 1 and 3 years without formal statistical testing.

Binary outcomes will be analyzed using logistic regression and robust variance estimation. Potential correlations between two study eyes of the same participant will be modeled using

generalized estimating equations (GEE) with an exchangeable correlation structure. The number and percentage of eyes meeting the outcome at the visit (observed data only) will be reported. The treatment effect will be estimated as an odds ratio.

Continuous outcomes will be analyzed using a linear mixed model with robust variance estimation. Potential correlations between two study eyes of the same participant will be modeled using a random intercepts. Means and standard deviations will be reported (observed data only). The treatment effect will be estimated as a mean difference.

Time-to-event outcomes will be analyzed using a marginal Cox proportional hazards regression model. Potential correlations between two study eyes of the same participant will be modeled using a robust sandwich estimate of the covariance matrix. The number and percentage of eyes meeting the outcome at or before the visit (observed data only) will be reported. The treatment effect will be estimated as a hazard ratio.

For each outcome, a plot showing the proportion, mean, or cumulative survival by treatment group over time will be constructed without imputation of missing data.

### 2.2.1 Visual Acuity

Additional analyses will be conducted on the visual acuity data (Table 1). The primary purpose will be to aid in interpretation of the key visual acuity outcome. If the statistical comparison of the mean change in visual acuity is not performed at a given time point because the anatomic outcome comparison is not statistically significant, then all analyses of visual acuity will be considered exploratory at that time point. Analyses will use the imputed data sets created for the mean change analysis.

**Table 1. Secondary Visual Acuity Outcomes**

| Outcome | Analysis Method |
|---|---|
| Failure proportion: visual acuity loss ≥ 10 letters | Logistic regression with GEE |
| Failure proportion: visual acuity loss ≥ 15 letters | Tabulation only |
| Success proportion: visual acuity gain ≥ 5 letters at both the time point and the previous study visit | Tabulation only |
| Failure proportion: visual acuity loss ≥ 5 letters at both the time point and the previous study visit | Tabulation only |
| Success proportion: visual acuity letter score ≥ 84 (approximate Snellen equivalent 20/20 or better) | Tabulation only |
| Success proportion: visual acuity letter score ≥ 69 (approximate Snellen equivalent 20/40 or better) | Tabulation only |
| Failure proportion: visual acuity letter score ≤ 38 (approximate Snellen equivalent 20/200 or worse) | Tabulation only |
| Change in VA from baseline AUC | Linear mixed model |

Abbreviations: AUC, area under the curve.

Change in visual acuity from baseline area under the curve will be calculated using the trapezoidal method:

$$AUC = \sum_{i=1}^{n} \left( \frac{V_i + V_{i+1}}{2} \times d \right)$$

Where $V_i$ is the truncated (see Section 7.4) change in visual acuity from baseline measured at the $i^{th}$ visit, $d$ is the number of days between visits $i$ and $i$+1, and $n$ is the number of common visits included in the analysis. For example, the 2-year outcome has $n = 9$ as the analysis will include visits at baseline, 1, 2, 4, 8, 12, 16, 20, and 24 months. For presentation, AUC will be divided by the number of days between baseline and the 2-year (or 4-year) visit so that the value shown will have units of letters rather than letter·days (e.g., 730 days at 2 years). This statistic can then be interpreted as the average change in visual acuity over the time between baseline and the 2-year (or 4-year) visit.

### 2.2.2 Development of PDR or DME

Additional analyses will be conducted on the components of the composite PDR/DME outcome to aid in interpretation.

**Table 2. Secondary PDR and DME Outcomes.**

| Outcome | Analysis Method |
|---|---|
| Development of PDR or PDR-related outcome (as defined within the composite time-to-event outcome) | Marginal Cox proportional hazards regression |
| Development of DME or DME-related outcome (as defined within the composite time-to-event outcome) | Marginal Cox proportional hazards regression |
| Mean change in OCT central subfield thickness from baseline* | Linear mixed model |
| Mean change in OCT volume from baseline* | Linear mixed model |
| Proportion of eyes with at least 2-step worsening of DR severity level (scale for individual eyes) by central reading center from baseline | Logistic regression with GEE |
| Proportion of eyes with at least 2-step improvement of DR severity level (scale for individual eyes) by central reading center from baseline | Logistic regression with GEE |

*Analyses will use multiply imputed data sets created similarly as for visual acuity, but substituting central subfield thickness or retinal volume for visual acuity.

The following outcomes will include descriptive statistics without statistical comparison of treatment groups:

- Development of PDR or DME based only on the objective components defined in the composite outcome, including OCT, visual acuity, and reading center assessment of photos and FA (i.e. not including investigator-only assessments)
- Development of each component of the composite outcome assessed individually
- Development of CI-DME on clinical exam with at least 10% increase in central subfield thickness and at least a 25 μm increase from baseline, regardless of visual acuity change
- Proportion of eyes with at least 3-step worsening of DR severity level (scale for individual eyes) by central reading center from baseline
- Proportion of eyes with at least 3-step improvement of DR severity level (scale for individual eyes) by central reading center from baseline
- Level of retinopathy on color photos

The definitions for at least 2-step and 3-step improvement and worsening of diabetic retinopathy for individual eyes on photographs, graded by central reading center, are shown in Tables 3 and 4, respectively. Note that only levels 43 to 53 are enrolled in Protocol W following reading center grading. If an eye outside of this range were to be enrolled, it will be excluded from the analyses of improvement and worsening.

**Table 3. Definitions for 2-Step Improvement and Worsening of Diabetic Retinopathy on Photographs for Individual Eyes.**

| Baseline | | Worsening (if FU ≥) | Improvement (if FU ≤) |
|---|---|---|---|
| **NPDR** | 43 | 53 | 14/15/20 |
| | 47 | 60 | 35 |
| | 53 | 60 | 43 |

Abbreviations: FU, follow up.

**Table 4. Definitions for 3-Step Improvement and Worsening of Diabetic Retinopathy on Photographs for Individual Eyes.**

| Baseline | | Worsening (if FU ≥) | Improvement (if FU ≤) |
|---|---|---|---|
| **NPDR** | 43 | 60 | 10/12 |
| | 47 | 65 | 14/15/20 |
| | 53 | 71 | 35 |

Abbreviations: FU, follow up.

### 2.2.3 Workplace Productivity and Activity Impairment Questionnaire

Outcomes from the Workplace Productivity and Activity Impairment Questionnaire will be compared between treatment groups at 2 and 4 years. For functional outcomes measured at the participant level, bilateral participants are non-informative with respect to the treatment comparison and will not be included in the analyses. Analyses will be conducted with a general linear model and robust variance estimation. Baseline level of the score being analyzed, laterality, and retinopathy severity will be included as adjustments. Only participants completing the corresponding visit will be included in the analysis, and there will be no imputation of missing data. The adjusted treatment effect will be presented along with a 95% confidence interval and *P* value. The following outcomes will be evaluated:

- Mean change from baseline in the percentage of work time missed due to vision problems over the past week (Absenteeism score)
    - Tabulated without statistical comparison
- Mean change from baseline in the percentage of impairment while working due to vision problems over the past week (Presenteeism score)
    - Tabulated without statistical comparison
- Mean change from baseline in the percentage of overall work impairment (Absenteeism and Presenteeism scores combined) due to vision problems over the past week (Work Productivity Loss score)
- Mean change from baseline in the percentage of activity impairment due to vision problems over the past week (Activity Impairment score)

## 3.0 Outcomes within Treatment Groups

Within each treatment group, the following outcomes will be tabulated without formal statistical comparison.

- Distribution and median (inter-quartile range) number of intravitreous injections performed up to 12, 24, 36, and 48 months as well as the intervening periods for eyes completing any visit at or beyond the upper limit (e.g., for injections through 36 months, eyes must have completed the 36-, 40-, 44-, or 48-month visit).
    - Intervals will be closed on the left and open on the right (e.g., for injections through 12 months, an injection given at 12 months will not be counted towards the total; however, an injection given at 12 months will count for the interval of 12 to 24 months).

## 4.0 Economic Analysis

The purpose of the economic analysis is to compare the treatment groups with respect to cost and workplace productivity loss. Data from the clinical trial on number of clinic visits completed, number of procedures performed (e.g., OCT, fundus photography), and number of aflibercept

injections will be used to estimate an average cost per patient for each treatment arm, using the Medicare Fee Schedule to estimate medical costs. The cost estimates, in combination with the percentage of productivity loss for each treatment arm (estimated from the WPAIQ), will be incorporated into the analysis.

## 5.0 Safety Analysis Plan

Adverse events will be categorized as study eye, non-study eye, and systemic. All randomized eyes will be included in the safety analysis and analyzed according to treatment group assignment at randomization. A tabulation of all study eye ocular, non-study eye ocular, and systemic adverse events by treatment group as defined above will be created. An additional tabulation will be made for adverse events possibly related to study treatment. For all analyses, the null hypothesis of no difference between treatment groups will be evaluated.

### 5.1 Ocular adverse events

The ocular adverse events below will be tabulated by treatment group for study eyes. A separate tabulation will be made for non-study eyes receiving study aflibercept. The frequency of the event occurring at least once per eye will be calculated. Ocular adverse events will be compared between treatment groups using logistic regression with GEE to account for the potential correlation between study eyes of bilateral participants. If there are convergence issues with the GEE model due to low event rates for one or more outcomes, then Barnard's Unconditional Exact Test may be used for analysis of all ocular adverse events.

The following ocular adverse events are of primary interest:

- Endophthalmitis
- Any retinal detachment (rhegmatogenous, traction, combined rhegmatogenous and traction, or not otherwise specified)
    - Rhegmatogenous retinal detachment (tabulated without formal analysis)
    - Traction retinal detachment (tabulated without formal analysis)
- Traumatic cataract due to intravitreal injection (limit to eyes that are phakic at baseline)
- Vitreous hemorrhage
- Ocular inflammation
- Intraocular pressure (IOP) elevation (any of the following)
    - Increase of IOP $\geq$ 10 mmHg from baseline (at a follow-up visit)
    - IOP $\geq$ 30 mmHg (at a follow-up visit)
    - Initiation of glaucoma medications
    - Glaucoma procedure
- Neovascularization of the iris

- Neovascular glaucoma

## 5.2 Systemic adverse events

Systemic adverse events will be reported in three groups: (1) unilateral participants randomized to sham, (2) unilateral participants randomized to aflibercept, and (3) bilateral participants. The frequency of the event occurring at least once per participant will be calculated. Statistical comparisons for systemic adverse events will include only unilateral participants. Analysis of systemic adverse events will be conducted with Barnard's Unconditional Exact Test.

Primary:

- Death
- Serious adverse event (at least one)
- Hospitalization (at least one)
- Cardiovascular and cerebrovascular events according to the Antiplatelet Trialists' Collaboration (excerpted from BMJ Jan 8, 1994):
    - Nonfatal myocardial infarction
    - Nonfatal stroke (counted only if symptoms lasted at least 24 hours)
    - Death attributed to cardiac, cerebral, hemorrhagic, embolic, other vascular (does not need to be ischemic in origin), or unknown cause
    - At least one event (nonfatal myocardial infarction, nonfatal stroke, or death attributed to potential vascular or unknown cause)

Note that transient ischemic attack, angina, possible myocardial infarction, and possible stroke are not counted. Nonfatal myocardial infarction and nonfatal stroke require that the patient is alive at the end of the study. If not, then only the death is counted.

Secondary (for tabulation without formal statistical comparison):

- Hypertension
- Frequency of at least one event per participant in each Medical Dictionary for Regulatory Activities (MedDRA) system organ class

Sensitivity analyses will replicate the systemic analyses above by whether a participant was randomized to receive aflibercept (in either eye) or not. The formal comparison groups will be unilateral participants randomized to aflibercept and bilateral study participants versus unilateral participants randomized to sham.

# 6.0 Additional Tabulations

The following will be tabulated according to treatment group:

- Baseline demographic and clinical characteristics

- Annual visit completion rate (excluding deaths)
- Treatment adherence

## 7.0 General Principles for Analysis

### 7.1 Analysis Cohort

Unless otherwise specified, treatment group comparisons will follow the intention-to-treat principle with all randomized eyes included and each eye analyzed according to the treatment assignment at randomization, regardless of the actual treatment received.

### 7.2 Visit Windows for Analysis

All participants are required by protocol to have assessment visits at baseline, 1, 2, and 4 months. After the 4-month visit, protocol assessment visits will be every 4 months at 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, and 48 months (Table 5).

**Table 5. Analysis Windows**

| Visit (Protocol Window) | Target | Analysis Window | |
|---|---|---|---|
| 1 month ± 1 week | 30 days | 16 – 44 days | (1 month ± 2 week) |
| 2 months ± 1 week | 61 days | 54 – 68 days | (2 months ± 1 week) |
| 4 months ± 8 weeks | 122 days | 66 – 178 days | (4 months ± 8 weeks) |
| 8 months ± 8 weeks | 244 days | 188 – 300 days | (8 months ± 8 weeks) |
| 12 months ± 8 weeks | 365 days | 281 – 449 days | (12 months ± 12 weeks) |
| 16 months ± 8 weeks | 487 days | 431 – 543 days | (16 months ± 8 weeks) |
| 20 months ± 8 weeks | 609 days | 553 – 665 days | (20 months ± 8 weeks) |
| 24 months ± 8 weeks | 731 days | 647 – 815 days | (24 months ± 12 weeks) |
| 28 months ± 8 weeks | 852 days | 796 – 908 days | (28 months ± 8 weeks) |
| 32 months ± 8 weeks | 974 days | 918 – 1030 days | (32 months ± 8 weeks) |
| 36 months ± 8 weeks | 1096 days | 1012 – 1180 days | (36 months ± 12 weeks) |
| 40 months ± 8 weeks | 1218 days | 1162 – 1274 days | (40 months ± 8 weeks) |
| 44 months ± 8 weeks | 1339 days | 1283 – 1395 days | (44 months ± 8 weeks) |
| 48 months ± 8 weeks | 1461 days | 1377 – 1629 days (48 months - 12 weeks, + 24 weeks) | |

If multiple visits fall within the same analysis window, then protocol assessment visits will be prioritized. If there is no protocol assessment visit in the analysis window, then priority will be given as follows: (1) a DME outcome assessment visit, (2) a treatment assessment visit, and (3) an unspecified visit. If there are multiple visits of the same type in the analysis window, then

whichever is closest to the target will be used. The DME outcome assessment visits take priority after treatment assessment visits because they occur prior to the primary outcome. Unspecified visits have the lowest priority because not all study procedures are required at these visits. To account for overlapping analysis windows, annual visits will be assigned before non-annual visits.

### 7.3 Missing Data

The strategy for handling missing data generally is included with the description of each analysis. If not otherwise specified, only participants with non-missing data will be included in the analysis (i.e., complete-case analysis).

### 7.4 Outliers

To help ensure that statistical outliers do not have undue impact on analyses of continuous visual acuity and optical coherence tomography (OCT) data, outcomes will be truncated to ± 3 standard deviations based on the mean and standard deviation at 2 years for 2-year completers, irrespective of treatment group. Change in visual acuity from baseline, change in OCT central subfield thickness from baseline, and change in OCT retinal volume baseline will be truncated. Truncation will be performed after imputation of missing data, where applicable (i.e., raw data will be used for imputation).

### 7.5 Model Assumptions

All model assumptions will be verified including linearity, normality of residuals, and homoscedasticity. The proportional hazards assumption for the marginal Cox regression model will be verified by testing the interaction between treatment and time. If model assumptions are not reasonably satisfied, then covariates may be categorized or excluded, and a nonparametric approach, transformation, or robust method may be considered.

### 7.6 Type I Error

For the primary outcome and other key outcomes, strict control of the Type I error rate is described in Section 2.0. There is no formal adjustment for multiplicity among secondary or safety outcomes to compensate for the number of outcomes being compared. All comparisons are conducted at alpha level .05 unless otherwise noted.

### References

Lee EW, Wei LJ, Amato DA. *Cox-type regression analysis for large numbers of small groups of correlated failure time observations*. Dordrecht: Kluwer Academic; 1992.

Sullivan TR, White IR, Salter AB, Ryan P, Lee KJ. Should multiple imputation be the method of choice for handling missing data in randomized trials? *Stat Methods Med Res*. 2016. DOI: https://doi.org/10.1177/0962280216683570.