Statistical Analysis Plan for the Primary Endpoint

Sponsor: University of Maryland through a grant from the NIH/NIA

Protocol Title: Non-Invasive Treatment of Abdominal Aortic Aneurysm Clinical Trial

Document Version / Date: 1.0: 26-Mar-2017

TABLE OF CONTENTS

1 Signature:

William C. Blackwelder, Ph.D. Senior Biostatistician	Date
Diana Lam, Ph.D. Biostatistician	Date
David Prince, Ph.D. Biostatistician	Date
Michael L.Terrin, MD, MPH, Director Data Coordinating Center	Date
B. Timothy Baxter, MD, Study Principal Investigator	Date
Jon Matsumura, MD, Director Imaging Core Laboratory	Date
John Curci, MD, Director Biomarkers Core Laboratory	Date
Ruth McBride, Director Data Management Unit	Date

1 Signature:

William C. Blackwelder, Ph.D. Senior Biostatistician — Date: April 4, 2017

Diana Lam, Ph.D. Biostatistician — Date: April 4, 2017

David Prince, Ph.D. Biostatistician — Date

Michael L.Terrin, MD, MPH, Director Data Coordinating Center — Date: April 4, 2017

B. Timothy Baxter, MD, Study Principal Investigator — Date

Jon Matsumura, MD, Director Imaging Core Laboratory — Date

John Curci, MD, Director Biomarkers Core Laboratory — Date

Ruth McBride, Director Data Management Unit — Date

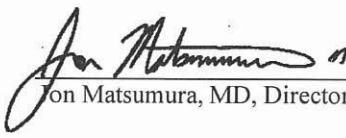
1 Signature:

William C. Blackwelder, Ph.D. Senior Biostatistician	Date
Diana Lam, Ph.D. Biostatistician	Date
David Prince, Ph.D. Biostatistician	04/05/2017 Date
Michael L.Terrin, MD, MPH, Director Data Coordinating Center	Date
B. Timothy Baxter, MD, Study Principal Investigator	Date
Jon Matsumura, MD, Director Imaging Core Laboratory	Date
John Curci, MD, Director Biomarkers Core Laboratory	Date
Ruth McBride, Director Data Management Unit	04/05/2017 Date

1 Signature:

William C. Blackwelder, Ph.D. Senior Biostatistician	Date
Diana Lam, Ph.D. Biostatistician	Date
David Prince, Ph.D. Biostatistician	Date
Michael L.Terrin, MD, MPH, Director Data Coordinating Center	Date
B. Timothy Baxter B. Timothy Baxter, MD, Study Principal Investigator	4/5/17 Date
Jon Matsumura, MD, Director Imaging Core Laboratory	Date
John Curci, MD, Director Biomarkers Core Laboratory	Date
Ruth McBride, Director Data Management Unit	Date

1 Signature:

William C. Blackwelder, Ph.D. Senior Biostatistician	Date
Diana Lam, Ph.D. Biostatistician	Date
David Prince, Ph.D. Biostatistician	Date
Michael L.Terrin, MD, MPH, Director Data Coordinating Center	Date
B. Timothy Baxter, MD, Study Principal Investigator	Date
 Jon Matsumura, MD, Director Imaging Core Laboratory	4/10/17 Date
John Curci, MD, Director Biomarkers Core Laboratory	Date
Ruth McBride, Director Data Management Unit	Date

1 Signature:

William C. Blackwelder, Ph.D. Senior Biostatistician	Date
Diana Lam, Ph.D. Biostatistician	Date
David Prince, Ph.D. Biostatistician	Date
Michael L.Terrin, MD, MPH, Director Data Coordinating Center	Date
B. Timothy Baxter, MD, Study Principal Investigator	Date
Jon Matsumura, MD, Director Imaging Core Laboratory	Date
John Curci, MD, Director Biomarkers Core Laboratory	Date 4/16/17
Ruth McBride, Director Data Management Unit	Date

2 List of Abbreviations and Definitions of Terms

AAA: Abdominal Aortic Aneurysm

ANCOVA: Analysis of covariance

BMI: Body Mass Index

CRP: C-reactive protein

CT scan: Computerized tomography scan

DCC: Data Coordinating Center

DSMB: Data Safety and Monitoring Board

GEE: Generalized Estimating Equations

ICL: Imaging Core Laboratory

IFN-γ: Interferon gamma

MMP-9: Matrix metallopeptidase 9

MTD: Maximum Transverse Diameter

SAP: Statistical Analysis Plan

SF-36: The Short Form (36) Health Survey

3 Introduction

This Statistical Analysis Plan (SAP) describes the analyses to be conducted for the primary endpoint of the Non-Invasive Treatment of Abdominal Aortic Aneurysm Clinical Trial (N-TA^{3}CT).

3.1 Primary Aim:

Determine if doxycycline (100 mg bid) will inhibit the increase in greatest transverse diameter of small abdominal aortic aneurysms (3.5-5.0 cm in men, 3.5-4.5 cm in women) over a 24-month period of observation in comparison to a placebo-treated control group.

3.2 Endpoints

The primary outcome is the change in abdominal aortic aneurysm maximum transverse diameter (MTD) on CT scan from baseline to the follow-up assessment CT scan two years after randomization, as measured in the Imaging Core Laboratory (ICL).

Secondary outcomes will derive from central ICL analyses of the CT scans performed every six months on patients and from the clinical follow-up of randomized patients, from clinical observation, local laboratory findings, study visit quality of life assessments, and from biomarker analyses to be performed in the Biomarkers Core Laboratory (e.g., changes from initial MMP-9 levels, and MMP-9 levels at 24 months of follow-up).

3.3 Study Design

3.3.1 Study population

The study population will include patients 55 years of age or older who are found to have small (3.5-5.0 cm among men and 3.5 - 4.5 cm among women in the largest transverse diameter) abdominal aortic aneurysms on quantitative CT scans.

Patients will be excluded from the study if they are unable to give their own informed consent to participate; have symptoms related to abdominal aortic aneurysm; have other intra-abdominal vascular pathology that may require repair within 24 months (e.g., renal artery stenosis, large iliac artery aneurysms, iliac occlusive disease, aneurysmal involvement of the renal artery); have had previous abdominal aortic aneurysm repair by

open surgical or endovascular technique; have an active malignancy with life expectancy less than two years; have an allergy to tetracycline; are currently or have been recently treated (previous six months) with tetracycline derivatives; they are currently taking anti-seizure medicines metabolized by pathways influenced by doxycyclines (e.g., carbamazepine, phenytoin, and barbiturates); stage II hypertension (patients whose blood pressure is persistently in the range of systolic > 160 mm Hg or diastolic > 100 mm Hg despite primary physician's best effort to achieve adequate therapy); have dialysis dependent renal failure or impending dialysis treatment for renal insufficiency; have a chronic infection requiring long-term (> 2 weeks) antibiotics, have known genetic syndromes responsible for the abdominal aortic aneurysm (e.g., Marfan's Syndrome), are under treatment with systemic immunosuppressive agents, could become pregnant, are not good candidates for clinical trial participation or are enrolled in another clinical trial.

3.3.2 Treatment groups and dosing

Patients will be assigned to doxycycline 100 mg p.o., b.i.d., or matching placebo. Patients randomly assigned to doxycycline will receive bottles containing a sufficient supply of 100 mg doxycycline capsules to take one capsule twice a day until the next appointment (about 100 days). Patients randomly assigned to placebo will receive a similar appearing supply of placebo capsules.

3.3.3 Study visits and assessments, study baseline

Under the original protocol, participants were to be followed in-clinic every 3 months until a common termination when the last participant randomized had been followed for 24 months. Under Protocol Revision 1.6.0, participants are expected to have 10 visits for this study: one baseline, eight at quarterly intervals while on treatment, and one post-treatment follow-up. Follow-ups will be conducted as in-clinic visits except that at 9, 15 and 21 months they *may* be conducted by telephone. The appendix of the protocol contains a summary of the timing of the visits and the assessments and data that are collected at each visit.

3.3.4 Randomization and Blinding or Masking

The DCC staff prepared randomization schedules for each Clinical Site participating in N-TA^{3}CT. The program for generating randomization schedules has the following characteristics: 1. There will be two randomization strata – one for men and one for women – at each clinical site; 2. Treatments are assigned in random order within blocks sizes two, four, six or eight with equal numbers of patients assigned to the doxycycline or placebo within each block. 3. Block sizes (two to eight patients per block) are randomly selected with the probability of each block size specified by DCC staff.

The DCC staff maintained a web-based randomization system for Clinical Site staff to request treatment allocations as eligible patients are identified. Only study-certified clinic staff with individually assigned userid/password can access the randomization and drug dispensing

application. At randomization, after selecting the participant ID to randomize, clinic site staff are asked to enter the participant's "letter code", gender and date of birth. These are verified against the data entered into the electronic data capture application for this participant. Only participants meeting the inclusion/exclusion criteria according to data entered in the eDC system, with a CT scan read by the Imaging Core Laboratory confirming eligibility, and with complete baseline data, including record of collection of the baseline biomarker sample are listed among the participant IDs eligible for randomization. The date and time of the completion of the randomization treatment assignment is the time of study entry for each patient.

In the Clinical Sites, Clinical Coordinating Center, Imaging Core Laboratory and Biomarkers Core Laboratory, the patients, directors (i.e., Clinical Site lead investigator), clinical site coordinators, and other study staff involved with direct patient care will be blinded to treatment assignments. Staff of the Data Coordinating Center will have access to individual patient treatment assignment on a "need-to-know" basis. The DSMB, the Director of the Data Coordinating Center will remain blinded to individual participant treatment assignment and to by-treatment-arm interim results. Only Data Coordinating Center staff involved in the production and distribution of the Closed Report to the DSMB or development and maintenance of the randomization system will have access to treatment code assignments and by-treatment-arm results.

The Data Coordinating Center will maintain records of each patient's treatment assignment.

3.3.5 Sample Size and Power

The study is designed to have greater than 90% power with a total sample size of 248 patients for a one-sided 0.025 test for an assumed reduction of 40% in the mean increase in aneurysm diameter from 2.5 mm/year to 1.5 mm/year. The calculations allow for a 10% non-adherence rate in the doxycycline treatment group and for 15% of participants to have a missing two-year CT scans. Additional details regarding the sample size and power calculations are contained in the protocol.

4 General Consideration for Data Analyses

4.1 Analysis Populations

4.1.1 Intention to Treat

The intention-to-treat population includes all patients randomized, analyzed by treatment assigned at randomization.

4.1.2 Full Analysis Population

The full analysis population includes all patients randomized except those with no follow-up contact and no follow-up CT scan, analyzed by treatment assigned at randomization. Because

treatment was never started and because the primary endpoint is not known and cannot be reasonably imputed in participants with no follow-up and no follow-up CT scan, analysis of the primary endpoint will be conducted on the *full analysis population*.

4.1.3 Per-Protocol Population

A per-protocol population will be defined as all participants randomized who

1. Complete 24 months of follow-up with a 2-year CT scan available for analysis;
2. Or undergo repair within 24 months (+ 90 days) of randomization;
3. Or die or suffer rupture of their AAA within 24 months (+90 days) of randomization;
4. And reported to have taken at least 80% of their expected study drug doses over the course of their follow-up period.

4.2 Statistical Analysis Issues

4.2.1 Strata and Covariates

Randomization is stratified by gender and site. Gender will be included in the primary efficacy analysis model.

4.2.2 Multiple Comparisons

For the primary analysis, we consider a one-sided p-value of 0.0247 to be significant. It is less than 0.025 to account for the interim analyses. To account for the multiplicity of hypotheses being tested in secondary and exploratory analyses, a p-value < 0.01 will be required to consider there to be evidence of differences present, i.e. to reject the null hypothesis. We consider $p<0.001$ as strong evidence.

4.2.3 Multi-center Studies

There are a total of 22 sites participating in the study. Analyses will not adjust for site.

4.2.4 Study Baseline and Visits

Study baseline corresponds to measurements collected at the baseline visit, prior to randomization and initiation of treatment. Follow-up visits are scheduled based on the date of randomization and are expected to occur 45 days before or after the target time-point; 90 days are allowed for CT scans or primary outcome clinical events (i.e., death or surgical repair).

5 Interim Analysis and Data Safety and Monitoring Board

Two formal interim analyses will be performed: one when approximately the first third of outcome data is available and the next when approximately the second third of outcome data is available. Formal rules for stopping for efficacy, futility, or harm are described in the protocol.

In addition, at each DSMB meeting an analysis of six-month CT scan results will be presented. The DSMB may recommend early termination if the aneurysm growth is higher in patients receiving doxycycline than in placebo recipients, with a one-sided p-value less than or equal to 0.05 ($Z = 1.645$).

6 Subject Disposition

6.1 Disposition of Participants

A CONSORT diagram (https://www.consort-statement.org/consort-statement/flow-diagram) will depict the disposition of all participants screened and randomized.

6.2 Extent of Exposure

Compliance with assigned treatment arm will be assessed by capsule counts at each follow-up visit. For the full analysis set, for each treatment arm and visit, the percentages of patients who have taken greater than or equal to 80% of the expected capsules, based on the capsule count, will be reported.

Plasma samples will be collected approximately every six months. For patients assigned to doxycycline, these samples will be analyzed for measurement of doxycycline levels. At each of the four six-month time points, the percentages of patients on the doxycycline arm with detected doxycycline will be reported, along with mean, standard deviation, median and interquartile range of the doxycycline values.

Time to permanent discontinuation of study drug will be depicted by Kaplan-Meier plot comparing the two treatment arms.

6.3 Protocol Violations

Protocol violations will be reported for the full analysis population using counts and percentages of violations. The difference in count between the intention to treat population and the full analysis population will be presented.

7 Baseline Data

7.1 Demographics and Baseline Characteristics

Baseline characteristics by treatment group for the intention-to-treat population will be compared using counts, percentages and for continuous variables, descriptive statistics. The following characteristics will be reported: gender, age (mean, standard deviation, median and interquartile range), ethnicity, race, maximum diameter and volume of AAA separately by gender (most recent prior to baseline with mean, standard deviation, median and interquartile range), vital signs including systolic and diastolic blood pressure, height, weight and BMI by gender, medical histories, concomitant medications, baseline quality of life as measured by the SF-36 Quality of Life instrument, and baseline biomarker levels. Characteristics will be reviewed for the full analysis population as well.

8 Efficacy Analyses

8.1 Primary Efficacy Analysis

The primary analysis will be performed to test the null hypothesis of no difference in growth of abdominal aortic aneurysms between the two treatment groups (doxycycline and placebo) after 24 months of follow-up as measured by CT scans analyzed in the Imaging Core Laboratory. This analysis will use the full analysis population. In order to include patients whose 24-month abdominal CT scans are missing because of death, endovascular repair or for other reasons in the primary analysis, we will base our analysis on normal scores for percentile of rank status. We will assign worst ranks to deaths, in order of time from study entry to death, next worst ranks to rupture, next worst ranks to endovascular repair for clinical presentation, next worst ranks to endovascular repair for reasons of aneurysm growth beyond 5.5 cm in men or 5.0 cm in women or other presentation free of clinical indication, all in order of time from study entry within category. The order of ranks are summarized in Table 1. Clinical presentation will be assessed by the three vascular surgeon Principal Investigators, blinded to treatment assignment, independently; agreement by two of the three or all three will be taken as final. Outcomes for patients missing the 24-month CT scan will be imputed, as described in the next section.

Table 1. Summary of ranks.

Order of rank assignment (worst to best)	Description
1	Death (from day of study entry to day of death)
2	Aneurysm rupture (from day of study entry to day of repair)
3	Endovascular repair for reasons related to clinical presentation (from day of study entry to day of repair)

4	Endovascular repair for growth beyond 5.5 cm in men or 5.0 cm in women or other presentation free of clinical indication (from day of study entry to day of repair)
5	Change from baseline in largest, transverse diameter of the aneurysm, using normal scores as described below.

The ranks for patients assigned to doxycycline will be compared to the ranks of patients assigned to placebo using normal scores. If N is the total number of randomized patients, a rank of k will be assigned the value corresponding to k/(N+1) in a standard normal distribution; for example, a rank of 100 out of 248 patients will receive the rank score corresponding to 0.4016, or a normal score z = 0.2492.

The change in normal score from baseline -- i.e., the two-year normal score for patients randomized to doxycycline and placebo minus the baseline normal score -- will be analyzed using a linear regression model (analysis of covariance -- ANCOVA), with variables for the randomized treatment assignment (0 for placebo, 1 for doxycycline), the normal score for baseline diameter, and gender in the model.

Prior to performing the primary outcome analysis, an interaction term for the randomization stratum (gender) with treatment assignment will be assessed in the ANCOVA model. If there is significant (at two-sided α=0.05) evidence of differences in the effect of treatment according to gender, other than a small quantitative difference (less than half a normal score difference in treatment effect) between effects that are in the same direction for each gender, the analysis will be performed separately for men and for women. In the event of a treatment by gender interaction being observed, we will review study results by gender to determine the nature of the differences in treatment effects. If there is no interaction, the ANCOVA will be used to estimate the overall treatment effects across all clinical sites and both strata (genders). The primary analysis will be performed on the normal scores corresponding to percentile rank of each patient at baseline and follow-up.

8.1.1 Imputation for the Primary Efficacy Analysis

In the event a patient does not complete two years of CT scan follow-up for cause(s) other than death, rupture, open repair or endovascular repair (e.g., withdrawal of consent or loss to follow-up, expected to be infrequent events), the patient's previous CT scans and other characteristics will be used for the imputation of a measurement of the abdominal aortic aneurysm diameter to be included in the primary analysis. The model used for imputation will be developed as described below.

For the multiple imputation of missing 24-month CT scans, we first assume that event status is known for all participants. Therefore, we assume we are imputing the 24-month CT scan score only for patients who did not have an event. First, we will assess the 24-month CT scans of patients who did not have an event for normality and transform them if necessary. If the data

show egregious departures from normality, possible transformations are natural log transformations, square root transformations, or Box-Cox transformations. Second, we will model the MTD from the 24-month scan using fully conditional specification, a type of multiple imputation using chained equations (MICE), implemented with SAS 9.4's PROC MI under the FCS option. At a minimum, we will include MTD from the baseline CT scan, gender, treatment, and gender by treatment interaction, and MTD from CT scans obtained at 6-month, 12-month and 18-month follow-ups. The interaction of treatment by gender is included because the test of interaction is planned in the final analysis.

Additional variables that will be considered for the multiple imputation model are baseline history of coronary artery disease, peripheral vascular disease, cancer, chronic obstructive pulmonary disease, diabetes mellitus, hypercholesterolemia, baseline use of beta blockers, ACE inhibitors or angiotensin receptor blockers, calcium channel blockers, diuretics, non-steroidal anti-inflammatory drugs, use of aspirin or anti-platelet, adherence score, and current smoking status. Adherence scores over 100 will be censored to 100. A continuous variable will be added to the multiple imputation model if its Type III Wald Chi-Square test is significant at the 0.05 level in linear regression of the MTD from the 24-month CT scan after adjusting for baseline MTD, gender, treatment, and gender by treatment for complete cases. Alternatively, if a continuous variable's partial correlation with 24-month CT scan is above 0.4 or above $0.75*R^2$ after adjusting for baseline MTD, gender, treatment, and gender by treatment interaction, then it will be added to the multiple imputation model. The R^2 is the correlation of 24-month CT scan result with baseline MTD, gender, treatment and gender by treatment interaction.

To be considered for the imputation model, a categorical variable will be required to have at least 5 of each group in the data set. Any of the above listed categorical variables that is missing in 20% or more in the imputed cases will be excluded. If the final list of variables includes more than 3 different types of medications, the list will be sent to the Director for the Data Coordinating Center for a check of collinearity between variables on a biological basis. This information and consultation will in no way unblind the Director of the Data Coordinating Center to the results of the analyses. If there are collinear variables, only the variable with the higher partial correlation will be included in the modeling process.

Finally, using complete cases (i.e., cases with no missing values for the 24-month MTD or any of the candidate independent variables), we will start with a model for 24-month MTD that includes all the candidate independent variables and perform backward elimination to test whether or not there are variables that can be deleted. The initial model will include MTD from the baseline CT scan, 6-month scan, 12-month scan and 18-month scan, treatment, gender, treatment by gender variable, and any other variable that meets the above statistical requirements. Variables other than baseline MTD, gender, treatment and gender by treatment interaction will be included in the final imputation model if they have a p-value less than 0.05 after the backward elimination.

The process of choosing candidate variables and developing a final model for imputing 24-month MTDs will be implemented before each interim analysis of the primary endpoint, as well

as at the final analysis. Thus, variables that are used for the multiple imputation at one analysis may or may not be used at a different stage, as data on more participants are available.

Data from study participants who have a clinical endpoint event (death, rupture, or repair) before the 6-month follow-up CT scan will not be included in the fitting of the model for imputing the 24-month MTD for other participants. Data from participants who have an event after the earliest significant follow-up MTD will be included in the fitting of those models, but their rank for the primary analysis will be determined by the event they experienced – i.e., their data will be used in imputing the 24-month MTD for other participants, but their clinical event will be used in determining their rank in the primary analysis. Data from participants excluded from the multiple imputation modeling will be added back into the imputed data set, so that they will be included in the primary analysis.

In the imputation model, the MTDs from CT scans at 6, 12, and 18 months will be constrained to have a minimum value of 3.5 and a maximum value of 7, to reflect the plausible range of MTD's in participants who do not experience a clinical event. The maximum value for MTD's was purposefully set high in order to not constrain the within and between imputation variance. All imputed MTD values will be rounded to the nearest hundredth (0.01) to reflect the same accuracy as the observed values.

For the multiple imputation regression modeling, we will use the default SAS settings (linear regression for continuous variables, logistic regression for dichotomous variables, and discriminant function for nominal variables with more than 2 categories). Since we are using the default, there is no need to specify a method on the FCS statement of PROC MI. The default number of burn-in iterations of 10 will be used (NBITER=10). Variables other than final and intermediate follow-up MTDs will be listed on the VAR statement from most complete to least complete. MTDs from follow-up CT scans will be listed in chronological order, regardless of missingness, with MTD at 24 months listed last. A seed of 384830 will be used, and 100 imputed values will be created for each missing observation in the multiple imputation model.

The multiple imputation data set will be used to test the significance of both the treatment variable and the treatment by gender interaction. For both the test of the treatment by gender interaction and the subsequent test of treatment effect or treatment and treatment by gender interaction, the degrees of freedom will be adjusted. The adjusted degrees of freedom is a function of the complete-data degrees of freedom and the between-imputation and within-imputation variance estimates. In the test of treatment by gender interaction, the complete-data degrees of freedom is the number of subjects in the analysis minus one. If the test of interaction is not significant at the 0.05 level, then we test only the treatment effect and the complete data degrees of freedom is the same (number of subjects minus one). If the test of interaction is significant at the 0.05 level, then we test the treatment and treatment by gender interaction, and the complete-data degrees of freedom is the number of participants minus two, as we are estimating two parameters in this case.

8.2 Analysis *Without* Imputation

Using the same normal scores for percentiles of rank methodology described above (section 7.1), we will compare the treatment groups *without imputation of any missing 24-month CT measurements of the MTD of the aneurysm.*

8.3 Analysis of Per-Protocol Population

An analysis will be performed using the methodology described above for the primary endpoint, based on the per-protocol population.

8.4 Changes from Protocol-Specified Efficacy Analyses

The primary efficacy analysis population has been renamed the "full analysis population" rather than the "intention-to-treat population" to avoid dispute over whether exclusion of patients for whom there are no follow-up visits or follow-up CT scans represents an imperfection in an "intention-to-treat" analysis plan or a different study population.

9 Interpretation of the Primary Outcome

Interpretation of the primary outcome analysis must be cognizant of N-TA^{3}CT as a Phase IIb clinical trial. The primary outcome is based primarily on anatomical measurements instead of clinical events. Benefit in the primary outcome could influence physician behavior for off label use, but is not going to result in a labeled indication. In particular it cannot be expected that we will know whether or not doxycycline effects on abdominal aortic aneurysm diameter are associated with the occurrence of rupture or the diameter at which risk of rupture outweighs the risk associated with repair. Effects on rupture or mortality could be observed fortuitously, and would have to be interpreted as secondary outcomes. If the investigators reject the null hypothesis, they have a considerable responsibility to assure that they have not introduced bias and explain to the medical community their basis for believing that bias is not the reason for rejecting the null hypothesis. If the investigators fail to reject the null hypothesis, they have a considerable responsibility to assure that they have performed a sensitive test, i.e., not introduced bias through defects in study design (especially power calculation assumptions) and performance (especially adherence or missing data), and explain to the medical community their basis for believing that bias is not the reason for failing to reject the null hypothesis.

Interactions:

With gender: If there is a statistically significant interaction ($p \leq 0.05$) between the effect of doxycycline in men and the effect in women, there will be a search for explanations for the interaction. For example, adherence will be examined. The gender strata will be analyzed separately in the event of an important interaction. If there are important, non-significant differences between the gender strata, e.g., no effect in one stratum and a 40% effect size in the

other, there will be a search for explanations for the differences between the strata, and the separate strata will be analyzed according to the standards of secondary analyses.
With Clinical Site: Influence of clinical site on treatment effects will be inspected. If there are outlier clinical sites, explanations for the inhomogeneity will be sought (e.g., differences in adherence) and any differences or inhomogeneities found will be presented with the primary analysis.

The findings of the primary outcome analysis in the full analysis population may reject or fail to reject the null hypothesis. Rejection of the null hypothesis with a median growth $\geq$40% less in the doxycycline group than in the placebo group would be evidence of both a statistically significant and clinically significant effect.

Results of analysis of the Full Analysis Population Conflict with Results from the Per-Protocol Population
If the primary outcome analyses based on the full analysis population and per protocol population are in agreement rejecting the null hypothesis or failing to reject the null hypothesis, we will reject or fail to reject the null hypothesis accordingly. If the analysis based on the full analysis population fails to reject the null hypothesis but the analysis based on the per protocol population rejects the null hypothesis, we will interpret the findings to indicate that doxycycline may be but has not been demonstrated to be efficacious for the reduction of growth of small abdominal aortic aneurysms. N-TA^{3}CT would be inconclusive with regard to treatment effect but may be informative as a Phase IIb clinical trial with regard to defining the defects in performance, e.g., insufficient treatment adherence, that must be addressed in designing a Phase III clinical trial. If the full analysis population analysis leads us to reject the null hypothesis and the per protocol population analysis leads to an effect size of similar magnitude to the full analysis population analysis but does not reject the null hypothesis, we will conclude that the per protocol analysis population analysis was insensitive and reject the null hypothesis. If the full analysis population analysis and the per protocol population analysis are contradictory – i.e., no similarity of effect size and no agreement on rejection of or failure to reject the null hypothesis – we will examine the results closely for an explanation of these pathological findings.

If we reject the null hypothesis with ≥ 40% less growth in doxycycline group:

- An analysis without imputation will be performed. If the results of the analysis without imputation do not return an effect size similar to the primary outcome analysis, the rejection of the null hypothesis will be called into question as too highly dependent on imputation.
- Also, a per protocol analysis with an effect size similar to or larger than 40% would be supportive of the interpretation that a clinically significant effect was observed in the full analysis population, but if the per protocol analysis effect size were less than 40%, the clinical significance of the finding should be viewed as questionable.

If we reject the null hypothesis with a median growth less than 40% reduced in the doxycycline group as compared to the placebo group:

- Should the per protocol analysis be associated with an effect size greater than 40%, there could be reason for belief that a clinically significant benefit exists that would justify further investigation.
- Also, rejection of the null hypothesis with an effect size less than 40% but a large difference (e.g., doubling) of the proportion of patients whose aneurysms grew less than 1.0 mm or a large reduction (e.g., halving) of the proportion of patients whose aneurysms grew more than 6 mm would suggest the possibility of clinically significant benefit that would justify further investigation.

An alternative interpretation to be kept in mind whenever the null hypothesis is rejected is the occurrence of a difference by chance.

Failure to reject the null hypothesis:
Failure to reject the null hypothesis in the primary outcome analysis would be interpreted to mean that doxycycline did not have a beneficial effect size $\geq$40% or that the design assumptions and performance were not of the quality necessary for a definitive test of the null hypothesis. If the effect size is $\geq$40% but we do not reject the null hypothesis, we will examine the design assumptions and study performance carefully to report any deficiencies that may influence future investigations.

10 SAP Revision

Identify major changes. Only changes after sign-off need to be recorded.

Revision Date	Section	Summary of Revision	Reason for Revision