NN103

Statistical Analysis Plan

B Cell Targeted Treatment in Myasthenia Gravis: A Phase II Trial of Rituximab in Myasthenia Gravis

Richard J. Nowak, MD, MS
Yale University School of Medicine
Protocol Principal Investigator

Christopher S. Coffey, PhD
University of Iowa Data Coordinating Center
Principal Investigator

Merit E. Cudkowicz, MD
Massachusetts General Hospital Clinical Coordinating Center
Principal Investigator

VERSION 1.0
November 17, 2017

STATISTICAL ANALYSIS PLAN SIGNATURE PAGE

NN103

[BEAT MG: B Cell Targeted Treatment in Myasthenia Gravis: A Phase II Trial of Rituximab in Myasthenia Gravis]

Principal Investigator Approval

Signature: _____ **Date:** _____
Name: Richard J. Nowak, MD, MS

NeuroNEXT Clinical Coordinating Center Approval

Signature: _____ **Date:** _____
Name: Merit Cudkowicz, MD

NeuroNEXT Data Coordinating Center Approval

Signature: _____ **Date:** _____
Name: Christopher S. Coffey, PhD

PREFACE

This Statistical Analysis Plan (SAP) describes the planned analyses for the NeuroNEXT NN103 (BEAT MG) study [National Institute of Neurological Disorders and Stroke (NINDS) grant # U01NS084495]. The planned analyses identified in this SAP are intended to support the completion of the Final Study Report (FSR) and will be included in regulatory submissions and/or future manuscripts. All final, planned analyses identified in this SAP will be performed only after the last randomized subject has completed the full 52 week study period. Once all week 52 data have been cleaned and verified, a "locked" version of the data will be used for reporting the final study results. Key statistics and study results will be made available to the Protocol Principal Investigator (PPI) following database lock and prior to completion of the final FSR. It is important to recognize that this SAP only applies to the primary 52 week study. Additional exploratory analyses added to the protocol as part of the extended follow-up study will be reported separately.

1. STUDY DESIGN

A previous study conducted at Yale demonstrated that 82% of subjects who received rituximab achieved at least a 75% reduction in their prednisone dose at 52 weeks (95% CI: 48%-98%). This study follows up on that finding with a multicenter randomized, double-blind, placebo controlled phase II clinical trial evaluating the safety and steroid-sparing effect of rituximab in MG utilizing a futility design. The specific primary objective of this study is to determine whether rituximab is a safe and beneficial therapeutic for MG that warrants further study in a phase III efficacy trial. The primary clinical endpoint will be the steroid sparing effect of rituximab, and the primary objective will be accomplished using a futility design which tests the hypothesis that subjects treated with rituximab will achieve at least an absolute 30% increase in the frequency of favorable responses (Levin, 2012). If "futility" is declared, then the results would imply that it is not cost effective to conduct a future phase III trial with this agent. If "futility" is not declared, then the study would suggest that there could be a potentially clinically meaningful effect of rituximab which should be explored in a larger, phase III follow-up study.

The study will enroll 50 AChR antibody positive generalized MG subjects, with subjects randomized in a 1:1 manner to receive either rituximab or placebo (25 per group). Each previously diagnosed generalized MG subject will be expected to be on a stable dose of prednisone (minimum dose of 15 mg/day) for at least 4 weeks (28 days) with stable symptoms at the time of enrollment. There will be two groups of standard of care treatment regimens allowed into the study:

- Prednisone Only
- Prednisone + Another IST: The subject must be on a stable dose for at least 6 months prior to baseline on one of the following IST: azathioprine, mycophenolate mofetil, cyclosporine, tacrolimus, or methotrexate

Subjects on pyridostigmine must be on a stable dose of ≤ 480 mg/day for a minimum of 2 weeks prior to the Screening Visit. Subjects must remain on a stable fixed dose for the duration of the study. The dose cannot be changed after study entry.

For the main study, subjects will be followed for 52 weeks. The study period of 52 weeks was chosen based on the delayed benefits observed following rituximab treatment, and in the setting of utilizing a two-cycle protocol. In order to assess safety in the B cell recovery period, as well as assess the long-term durability of response, there are two additional optional observational off study-intervention time points (weeks 72 & 96 – extended study follow-up).

The treatment group will receive a total of two cycles of rituximab (375mg/m^2 IV) separated by 6 months. Each cycle is defined as one infusion per week for four consecutive weeks. As such, cycle 1

will be administered weeks 0-3 and cycle 2 will be given at weeks 24-27. The placebo group will receive an infusion that contains only the vehicle components of the rituximab solution.

A predetermined, forced steroid taper schedule for both treatment (rituximab) and placebo groups will begin at the week 8 visit. At every 4-week assessment thereafter, the MGC score will be calculated, confirmed, and available during the study visit in order to make the steroid dose adjustment. Prednisone dose will be lowered following confirmation of clinical improvement or stable symptoms based on the MGS score (current MGC score is ≤ 2 points above the baseline visit or MGC score at the prior study visit). If the MGC score change is ≥ 3 points above the baseline visit score, the taper will be stopped the prednisone dose increased until symptoms resolve or are at least are stabilized to baseline status (baseline visit MGC score). If the MGC score is ≤ 2 points above the baseline visit score, but has increased ≥ 3 points from the MGC score at the previous study visit, the taper will be stopped and prednisone dose will either be held or increased (at the discretion of the Site Investigator). Once symptoms stabilize (MGC score is the same or less than the baseline visit score and ≤ 2 points above the prior study visit), the prednisone taper can again be resumed at the next scheduled assessment. If the Site Investigator does not taper per protocol, this will be recorded as a protocol deviation and will be corrected immediately. As this is linked with primary outcome, we wanted to make the decision on lowering the dose as objective as possible. A mechanism will be put in place to double check the MGC score calculation made at the visit, and whether or not the prednisone adjustment was made correctly. The dose of prednisone taken will be record by each subject daily and collected at each evaluation.

Subjects will have clinical evaluations at baseline and then every 4 weeks thereafter (week 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52). Clinical evaluations will be completed by a blinded evaluator. The dose of prednisone will be recorded by each subject daily and collected at each scheduled assessment. Post-intervention status will be assessed by measuring MGFA class, Myasthenia Gravis Composite (MGC), Quantitative Myasthenia Gravis (QMG), MG-Activities of Daily Living (MG-ADL), and MG-Quality of Life (MG-QOL) scores. Blood will be collected for safety, specialized and other studies at scheduled time points. Adverse effects will be monitored at each visit to assess safety and tolerability in this subject population per the NINDS Guidelines for Data and Safety Monitoring in Clinical Trials and Genentech guidelines.

If subject symptoms significantly worsen during the course of the trial and are not controlled by increased steroid doses (high dose prednisone), the subject can receive PLEX or IVIg as a rescue therapy. Subjects that could not be managed with steroids, IVIg, or PLEX and required additional immunotherapy (e.g. pulse IV steroids, azathioprine, etc.) would be considered treatment failures and likely withdrawn from the study.

1.1. Primary Outcomes

Primary Objective 1 – Steroid Sparing Effect: *Percent of subjects that achieve a ≥75% reduction in mean daily prednisone dose in the 4 weeks prior to week 52 and with clinical improvement or no significant worsening of symptoms (≤ 2 point increase in MG composite score) as compared to the 4-week period prior to randomization and initiation of treatment.*

Primary Objective 2 - Safety: *Percentage of subjects with treatment-related adverse events.*

1.2. Major Secondary Objective

The main secondary objective is to evaluate whether there is a trend towards clinical benefit at the end of the 52 week treatment period, as measured by MG-specific clinical outcome scales used as endpoints in prior clinical trials:

(1) Myasthenia Gravis Composite (MGC) Score

(2) Quantitative Myasthenia Gravis (QMG) Score

The clinical evaluators who determine the MGC & QMG will be blinded to treatment assignment. If successful, measures studied would lay the groundwork toward optimizing the design of a subsequent phase III efficacy trial of rituximab in MG.

2. PRIMARY ENDPOINTS

2.1. Prednisone Reduction

The first primary outcome measure for this study is the percent of subjects achieving a ≥ 75% mean daily prednisone reduction in the four weeks prior to week 52 (week 49-52) along with clinical improvement or no significant worsening of symptoms (≤2 point increase in MG composite score), as compared to the four week period prior to randomization. The primary endpoint will be a binary indicator of whether the subject achieved the definition above. The prednisone-sparing aspect of the endpoint will be computed by comparing the mean daily prednisone dose (per the protocol defined taper) during the four week period prior to randomization versus the four week period at the end of the study (weeks 49-52). For subjects that had their prednisone dose changed after the week 48 visit, or who missed their week 48 visit, the primary endpoint will be determined by comparing the prednisone dose reported at baseline to the last prednisone dose recorded prior to the week 52 visit. The MGC aspect of the endpoint will be computed by comparing the MGC obtained at baseline to the MGC obtained at the week 52 visit. For the primary analysis, we will take a conservative approach and impute an outcome of “failure” for any subject that either terminates the study early, for whom the prednisone dose in the last 4 weeks is unknown, or for whom the week 52 MGC score is missing.

2.2. Safety

The second primary outcome will assess the safety profiles of rituximab vs. placebo. Primary interest involves an examination:

- Treatment-related adverse events (AEs)
- Treatment-related serious adverse events (SAEs)

An AE is any untoward medical occurrence associated with the use of a drug in humans, whether or not considered drug related. Examples of AEs include new conditions, worsening of pre-existing conditions, clinically significant abnormal physical examination signs (i.e. skin rash, peripheral edema, etc), or clinically significant abnormal test results (i.e. lab values or vital signs). Stable chronic conditions (i.e. diabetes, arthritis) that are present prior to the start of the study and do not worsen during the trial are NOT considered AEs. Chronic conditions that occur more frequently (for intermittent conditions) or with greater severity, would be considered as worsened and therefore would be recorded as AEs.

AEs are generally detected in two ways:

- Clinical → Symptoms reported by the subject or signs detected on examination
- Ancillary Tests → Abnormalities of vital signs, laboratory tests, and other diagnostic procedures

All AEs should be reported within 5 working days / 7 calendar days of the site learning of a new AE. Similar timelines apply for reporting upon receipt of any updates for previously reported AEs. If discernible at the time of reporting, a specific disease or syndrome rather than individual associated signs and symptoms should be identified by the Site Investigator and recorded as an AE. However, if an observed or reported sign, symptom, or clinically specific disease or syndrome, then it should be recorded as a separate AE. Clinically significant laboratory abnormalities, such as those that require intervention, are those that are identified by the Site Investigator.

The study will utilize the CTCAE version 4.0 coding system for AE recording. AEs reported using CTCAE will be recoded into MedDRA terms by the DCC.

For the purposes of this study, a treatment-related AE (also referred to as an Adverse Drug Reaction) is defined as any noxious or unintended response to a medicinal product related to any dose. The phrase "responses to a medicinal product" means that a causal relationship between a medicinal product and an AE is at least a reasonable possibility, i.e., the relationship cannot be ruled out. Therefore, a subset of AEs can be classified as treatment related if there is thought to be a causal relationship to Rituximab. At the time of reporting, the relationship of the AE to the investigational product should be specified by the Site Investigator using the following definitions:

- Definitely Related: The reaction follows a reasonable temporal sequence from administration of investigational product; that follows a known or expected response pattern to the investigational product; and that is confirmed by improvement on stopping or reducing the dosage of the investigational product, and reappearance of the reaction on repeated exposure (suspected treatment related AE or ADR)
- Probably Related: The reaction follows a reasonably temporal sequence from administration of investigational product; is confirmed by discontinuation of the investigational product or by re-challenge; and cannot be reasonably explained by the known characteristics of the subject's clinical state (suspected treatment related AE or ADR)
- Possibly Related: The reaction follows a reasonably temporal sequence from administration of the investigational product and follows a known response pattern to the suspected investigational product; the reaction could have been produced by the investigational product or could have been produced by the subject's clinical state or by other modes of therapy administered to the subject (suspected treatment related AE or ADR)
- Unlikely to be Related: The reaction has little or no temporal sequence from administration of the investigational product, and/or a more likely alternative etiology exists
- Not Related: Concomitant illness, accident, or event with no reasonable association with treatment

As this is a double-blind study, the causality assessment should be made under the assumption that the subject is receiving active study medication. If considering unblinding, this assessment should be made prior to unblinding to avoid bias.

An AE is considered serious if it meets one or more of the following criteria:

- Results in death
- Is life-threatening (i.e., poses an immediate risk of death as the event occurred)
- Requires inpatient hospitalization or prolongation of existing hospitalization
- Results in persistent or significant disability or incapacity
- Results in a congenital anomaly or birth defect in the offspring of the subject (whether the subject is male or female)
- Necessitates medical or surgical intervention to preclude permanent impairment of a body function or permanent damage to the body structure
- Important medical events that may not result in death, are not life-threatening, or do not require hospitalization may also be considered SAEs when, based on appropriate medical judgment, they may jeopardize the subject and may require medical or surgical intervention to prevent one of the outcomes listed above. Examples of such medical events include blood dyscrasias or convulsions that do not result in inpatient hospitalization, or the development of drug dependency or drug abuse.

An inpatient hospital admission in the absence of a precipitating, treatment-emergent, clinical AE may meet criteria for “seriousness” but is not an adverse experience, and will therefore not be considered an SAE. An example of this would include a social admission (subject admitted for reasons other than medical, e.g. lives far from the hospital, has no place to sleep). The Site Investigator is responsible for initially classifying AEs as serious or non-serious. SAEs must be reported using the OEARS within 24 hours of the site learning of the SAE.

Dr. Michael Shy will serve as the Medical Safety Monitor (MSM) for this trial. Dr. Shy will work closely with the DCC, and will use the online AE reporting system to review all SAEs in near real time and evaluate them to identify the need for timely intervention. For any reported SAEs, an automatic email will be sent to Dr. Shy to prompt a review of the event for determination of whether the event meets the criteria for an SAE and, if so, whether the SAE is unanticipated and/or related to study drug. An unexpected SAE is any SAE for which the specificity or severity is not consistent with the current Investigators Brochure or package insert described in the protocol. An unexpected and treatment-related SAE is an unexpected SAE that, in the opinion of the MSM, has a reasonable possibility that the investigational product caused the event. With the assistance of the coordinators at the DCC, Dr. Shy has the option of requesting additional information about any SAE. He will complete a form for each review, and this information will be entered into the online data entry system.

Thus, in summary, the determination of whether an AE or SAE is treatment-related (at least possibly related to treatment) differs. Because the MSM only reviews SAEs in real-time, the determination of whether or not a non-serious AE is considered treatment-related will be made at the site level. However, for SAEs, the MSM determination of whether or not an SAE is treatment-related will take precedent over the classification at the site level.

3. MAJOR SECONDARY ENDPOINTS

3.1. Myasthenia Gravis Composite Score (MGC)

The Myasthenia Gravis Composite (MGC) score is a validated, patient- and physician-reported 10-item assessment tool for evaluating the symptoms and signs of MG. Physician assessment includes evaluation for ptosis (upward gaze), double vision on lateral gaze, eye closure, neck flexion or extension, should abduction, and hip flexion. Patient assessment includes self-report of impact (normal, mild, moderate, or severe) and is additionally weighted for clinical significance. Total score ranges from 0 to 50, with higher scores indicating a greater impact of MG on functional activities. A three point change is considered clinical meaningful. This brief assessment takes approximately 10 minutes to complete (Burns et al, 2008; Burns et al, 2010; Burns, 2012; Sadjadi et al, 2012).

3.2. Quantitative Myasthenia Gravis Score (QMG)

The QMG score is a validated, physician-reported 13-item disease-severity assessment tool. It evaluates muscle strength based on quantitative testing of sentinel muscle groups: ocular (two items), facial (one item), bulbar (two items), gross motor (six items), axial (one item), and respiratory (one item). Each item is graded on a scale of 0 to 3, with 3 being the most severe. Total score ranges from 0 to 39, with higher scores representing greater disease burden. A 3-point improvement in total score considered a clinically meaningful improvement. This assessment takes 30-40 minutes to complete, and is the most widely used tool in MG trials (Barohn et al, 1998).

4. ENROLLMENT & RANDOMIZATION

Subjects who meet the eligibility criteria and have given their consent will be randomized to one of the 2 treatment arms. Randomization will be performed through an interactive website, and will be stratified based on the steroid dose at baseline [moderate dose prednisone (15-35 mg/day) vs. high dose prednisone (>35 mg/day)] and treatment regimen at the baseline visit [prednisone only vs.

prednisone plus another immunosuppressive therapy (IST)]. Subjects will be assigned a study ID at the time of enrollment. The study ID includes the identification of the clinical study site and a unique subject ID. The DCC will generate a randomization table for each of the strata using a permuted block design with random block sizes.

5. PRELIMINARY TABLULATIONS

All subjects who provide informed consent will be accounted for in this study. Regularly generated enrollment reports will describe:

- Number of subjects consented, eligible, and randomized by site
- Ongoing study status of all randomized subjects
- Reasons for ineligibility
- Protocol deviations
- Early study terminations

The data set will also be summarized by treatment group with respect to important confounders. The distributions of categorical variables will be tabulated by treatment group and overall. Continuous variables will be summarized as mean, median, standard deviation, minimum, and maximum by treatment group and overall. Variables to be collected will include:

- Gender
- Race
- Ethnicity
- Age
- Baseline Prednisone Dose (mg/day)
- Baseline Myasthenia Gravis Composite Score (MGC)
- Baseline Quantitative Myasthenia Gravis Score (QMG)
- Baseline MG-Activities of Daily Living Score
- Baseline MG-Quality of Life Score (MG-QOL)
- Baseline MGFA Clinical Classification Grade
- Hand Preference
- Thymectomy Results

6. ANALYSIS POPULATIONS

Due to the exploratory nature of this study, all analyses to address the primary and major secondary objectives will be conducted at the 0.10 significance level. All analyses will be implemented using an intent-to-treat approach. Any subject who received a random treatment assignment will be included in all analyses.

7. PRIMARY EFFICACY ANALYSES

7.1 Primary Futility Hypothesis: *Subjects treated with rituximab will have at least a 30% absolute increase in the frequency of achieving at least a 75% reduction in mean daily prednisone dose with maintenance of minimal or no symptoms.*

The primary futility hypothesis being tested in this trial is that subjects treated with rituximab will achieve at least an absolute 30% increase in the frequency of favorable responses. Assuming a placebo response rate of 40% (as in the original sample size calculations – see section 9), this corresponds to

an odds ratio of 3.5. Therefore, the primary futility hypothesis will be assessed using a logistic regression model, adjusted for the two stratification variables, to estimate the log-odds of primary endpoint success in each group. The logistic regression model used for these purposes is stated here:

$$\text{logit}(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i}$$

where

- Y_i represents the binary variable indicating whether or not the i^{th} subject met the primary outcome requirement of a 75% or greater reduction in prednisone dose, with no significant worsening of symptoms (≤2 point increase in MG composite score at baseline)
- X_{1i} is an indicator variable for prednisone dose at baseline (=0 if moderate, =1 if high)
- X_{2i} is an indicator variable for treatment status at baseline (=0 if prednisone alone, =1 if prednisone plus IST)
- X_{3i} is an indicator variable for whether the i^{th} subject was randomized to the Rituximab group (=0 if placebo, =1 if rituximab)

Correspondingly, the primary futility hypothesis of interest can be assessed by performing the following hypothesis test:

$$H_0: \exp(\beta_3) \geq 3.5 \textit{ vs. } H_A: \exp(\beta_3) < 3.5$$

Results will be summarized in the following tables (number and percent of primary endpoint success are displayed in Table 7.1; Odds ratios and confidence intervals are displayed in Table 6.2).

Table 7.1: Number and Percent of Primary Endpoint Success

Treatment Group	Number in each group	Number and Percent of Success	Number and Percent of Failure	Number missing
Rituximab	xx	xx (xx%)	xx (xx%)	xx
Placebo	xx	xx (xx%)	xx (xx%)	xx

Table 7.2: Odds-ratios of Primary Endpoint Success

Comparison	Odds-ratio (1-sided 90% CI)	p-value
Rituximab vs. Placebo	xx.x (xx.x, xx.x)	x.xx

Rejecting the null hypothesis suggests 'futility' in the sense that it appears unlikely that conducting a future phase III clinical trial would lead to a significant effect with a magnitude at least as large as the specified clinically meaningful effect of interest. If we don't reject the null hypothesis, this would provide justification for proceeding to examine the superiority hypothesis, with the magnitude of the estimate and confidence intervals surrounding β_3 providing information helpful for planning the future phase III trial.

Due to randomization, it is unlikely that important covariates will be imbalanced in this study. However, given the small sample size, this cannot be dismissed. We will assess for important baseline imbalances, and if any imbalances exist, we will conduct sensitivity analyses to examine the impact on the results when the relevant covariate(s) are added to the logistic regression model.

7.2 Primary Safety Hypothesis: *There will be no increase in adverse experiences for the rituximab-treated vs. placebo subjects.*

As described elsewhere, general assessments of safety will occur throughout the trial in conjunction with the medical safety monitor. The primary assessment of safety will compare the percentage of subjects in each group with:

- Treatment-related adverse events (AEs)
- Treatment-related serious adverse events (SAEs)

This hypothesis will be assessed in two ways. First, the percentage of subjects who experience any treatment-related AE or SAE (overall and by MedDRA system organ class) in each group will be compared using a Fisher's exact test. If the null hypothesis is rejected, with a greater frequency observed in the rituximab group, we will conclude that rituximab was associated with a significantly greater frequency of treatment-related AEs. If the hypothesis is not rejected, we will conclude that the study does not provide sufficient evidence to conclude that rituximab was associated with a significantly greater frequency of treatment-related AEs. If there are significant differences between groups within any specific SOC, then additional tests will compare differences across groups for specific MedDRA preferred terms in order to further explore the cause of observed differences.

In addition to the comparison of percentages in the manner described above, the rates of treatment-related AEs in each group will be compared using the following Poisson regression model:

$$\log\left(\frac{Y_i}{T_i}\right) = \beta_0 + \beta_1 x_{1i} + \epsilon_i$$

where

- Y_i represents the number of treatment related AEs experienced by the i^{th} subject.
- T_i represents the number of days between the date of randomization and the date of last follow-up for the i^{th} subject.
- x_{1i} = 1 if i^{th} subject was randomized to rituximab, and 0 if the subject was randomized to placebo group
- ϵ_i is random error for the i^{th} subject

To determine if the rate of treatment related AEs differ across treatment group we will test the following hypothesis:

$$H_o: \beta_1 = 0 \text{ vs. } H_A: \beta_1 \neq 0$$

If the null hypothesis is rejected, the direction of β_1 will indicate the direction of the observed effect. Values of $\beta_1 > 0$ indicate an increased rate of treatment-related AEs associated with the rituximab group, while values of $\beta_1 < 0$ indicate a decreased rate of treatment-related AEs associated with the rituximab group.

Treatment-related SAEs will be analyzed in the same manner described above. Additional safety analyses will assess all treatment-emergent AEs, treatment-emergent SAEs, unanticipated SAEs, and treatment-related & unanticipated SAEs in a similar manner.

8. IMPACT OF MISSING DATA

The primary analysis will follow the intent-to-treat (ITT) paradigm. All enrolled subjects must be included in the primary ITT analysis, and will be analyzed in the treatment group to which they were initially randomized. As such, it will be critically important to minimize the occurrence of missing data. Obviously, the optimal strategy for dealing with missing data is to make every effort to obtain complete data during the conduct of the study. Our team of data managers and protocol coordinators will work diligently and use a variety of methods in order to minimize the percentage of missing data in this trial. Nevertheless, there is likely to be a small percentage of missing data. As specified above, we will take a conservative approach for the primary analysis and impute an outcome of "failure" for any subject that either terminates the study early, for whom the prednisone dose in the last 4 weeks is unknown, or for whom the week 52 MGC score is missing. We then propose a series of sensitivity

analyses to further assess the potential dependence of the results of the primary analysis on these missing values. This sensitivity analysis will employ multiple methods:

- **Using Only Observed Data:** Use only subjects who completed the study, for whom the prednisone dose in the last 4 weeks was known, and for whom the week 52 MGC score is known.
- **Tolerability/Imputation:** For all additional sensitivity analyses, all subjects who terminated from the study due to an AE or had clinical worsening (an MGC score >2 above baseline) at the time of termination will be considered "failures". For remaining subjects with missing data due to other reasons (lost to follow-up, discontinuation for reasons other than AE, etc.), outcomes will be imputed using a variety of methods
    - **Last Observation Carried Forward:** Last known prednisone dose status at the time of termination will be carried forward and used for the endpoint determination for these subjects
    - **Multiple Imputation**: For simplicity, we assume subjects requiring imputation will not have had clinical worsening had they stayed in the study. Thus, the imputation is focused solely on the missing prednisone dose information. To implement this multiple imputation model, we will impute week 52 prednisone dose data using a multiple imputation model based on the prednisone dose strata and treatment status at baseline, as well as prednisone dose data computed at each intermediate time point for all subjects with observed data. Imputed values will be derived using the MCMC method with multiple chains, adequate burn-in iterations, and a non-informative prior distribution. Once the imputed prednisone dose data have been obtained, the binary outcome variable will be generated for all subjects and fit using the same model described in section 6.1. We will conduct five separate iterations, and the mean of the parameter estimates from the five imputed data sets will be used as the estimate for the final analysis. Variances for the primary parameter estimate will be estimated using standard formulas as a function of within imputation and between imputation variable (Little & Rubin, 2002).
    - **Best-case scenario**: Assume all subjects missing prednisone dose information during the last 4 weeks, or missing a week 52 MGC score, in the rituximab group are "successes" (did achieve ≥ 75% dose reduction); Assume all subjects missing prednisone dose information during the least 4 weeks, or missing a week 52 MGC score, in the placebo group are "failures" (did not achieve ≥ 75% dose reduction).
    - **Worst-case scenario**: Assume all subjects missing prednisone dose information during the last 4 weeks, or missing a week 52 MGC score, in the rituximab group are "failures" (did not achieve ≥ 75% dose reduction); Assume all subjects missing prednisone dose information during the last 4 weeks, or missing a week 52 MGC score, in the placebo group are "successes" (did achieve ≥ 75% dose reduction)

Results will be reported from both the primary analysis and all sensitivity analyses in order to inform how robust the overall trend observed in the study is to the missing data. For example, if the final analysis suggests a non-futile study, future researchers might be more comfortable proceeding to a phase III study if that finding of non-futility is also supported by the majority of the sensitivity analyses. On other hand, if a primary finding of non-futility is not supported by the sensitivity analyses, further exploration of the data might be needed prior to embarking on a future phase III trial. These results will be displayed in the following table:

Table 8.1: Number and Percent of Subjects Meeting Primary Endpoint w/ Odds Ratios & One-Sided 90% CIs

Imputation Method	Reduction in mean daily prednisone dose ≥ 75%		Odds Ratios (1-Sided 90% CI) Rituximab vs. Placebo
	Rituximab N (%)	Placebo N (%)	
Primary	xx (xx.x%)	xx (xx.x%)	x.xx (x.xx, x.xx)
Observed	xx (xx.x%)	xx (xx.x%)	x.xx (x.xx, x.xx)
Last Observation Carried Forward	xx (xx.x%)	xx (xx.x%)	x.xx (x.xx, x.xx)
Multiple Imputation	xx (xx.x%)	xx (xx.x%)	x.xx (x.xx, x.xx)
Best Case	xx (xx.x%)	xx (xx.x%)	x.xx (x.xx, x.xx)
Worst Case	xx (xx.x%)	xx (xx.x%)	x.xx (x.xx, x.xx)

9. MAJOR SECONDARY ANALYSES

The secondary objective of the study is to evaluate whether there is a trend towards clinical benefit as measured by MG-specific clinical outcome scales used as endpoints in prior clinical trials. Specifically, we will determine if rituximab can significantly improve the scores of the following MG-specific clinical outcome measures: 1) Myasthenia Gravis Composite (MGC), and 2) Quantitative Myasthenia Gravis (QMG). These studies measures would lay the groundwork towards optimizing the design of a subsequent phase III efficacy trial of rituximab in MG.

9.1. Major Secondary Objective #1 – Myasthenia Gravis Composite: *Rituximab-treated subjects will have clinically significant improvement in their Myasthenia Gravis Composite (MGC) scores at the end of the 52 week treatment period.*

The first secondary hypothesis assesses the change in MGC scores at the end of the 52 week study period. This objective will be assessed longitudinally comparing the final score at the end of the study to the score obtained at baseline. The outcome will be defined as the change from baseline to week 52 in the MGC. This hypothesis will be assessed using a linear regression model, adjusted for baseline MGC score. For example, the following model will be fit to these data:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i}$$

where

- Y_i represents the change from baseline in the MGC score for the i^{th} subject
- X_{1i} is an indicator variable for prednisone dose at baseline (=0 if moderate, =1 if high)
- X_{2i} is an indicator variable for baseline treatment (=0 if prednisone alone, =1 if prednisone plus IST)
- X_{3i} is the baseline MGC score for the i^{th} subject
- X_{4i} is an indicator variable for whether the i^{th} subject was randomized to the Rituximab group (=0 if placebo, =1 if rituximab)

Correspondingly, the secondary hypothesis of interest can be assessed by performing the following test:

$$H0: \beta_4 = 0 \text{ vs. } HA: \beta_4 \neq 0.$$

The results will be displayed in the following table

Table 9.1: Change in MGC Scores from Baseline to 52 Weeks

MGC Score	Rituximab	Placebo	Model Adjusted Difference (90% CI)
(52 week – baseline) **Mean (SD)** **Min. – Max** **Missing**	 xx (xx) xx – xx xx	 xx (xx) xx – xx xx	 xx.x (xx.x, xx.x)

9.2. Major Secondary Objective #2 – Quantitative Myasthenia Gravis: *Rituximab-treated subjects will have clinically significant improvement in their Quantitative Myasthenia Gravis (QMG) scores at the end of the 52 week treatment period.*

The second secondary hypothesis assesses the change in QMG scores over the course of the 52 week study period. Because the only difference between this and the first secondary hypothesis is the choice of outcome, the analysis will proceed in the same manner described above for the first secondary hypothesis.

9.3. Exploratory Analyses

A number of additional exploratory analyses are also planned to monitor effectiveness as well as evaluate other endpoints that may be useful in optimizing future MG trial designs, but will not be included as part of the FSR. These additional analyses may include, but are not limited to:

- Other Previously Validated MG-specific outcome measures
    - MG-Activities of Daily Living (MG-ADL) – This 8 point scale assesses the subject's ability to perform daily activities (Wolfe et al, 1999).
    - MG-Quality of Life (MG-QOL) – The subject completes a 15-question questionnaire and reports the effect of MG on their quality of life (Burns et al, 2008).
- Other Previously Used Measures of Steroid-Sparing Effect:
    - Mean daily prednisone dose at each scheduled assessment (every 4 weeks)
    - A delayed start of the area under the dose-time curve (AUDTC), starting at week 8
    - Percentage of subjects achieving a ≥ 50% mean daily prednisone reduction with maintenance of minimal or no symptoms in 4 weeks prior to week 52
    - Body mass index (screening visit and weeks 24 & 52)
    - HbA1C (screening visit and week 52)
- MG Flare Rate (Failure of Therapy)
    - Percentage of subjects requiring rescue treatments (PLEX or IVIg)
    - Percentage of subjects requiring prednisone dose increase
    - Rate of subjects requiring prednisone dose increase
    - Percentage of subjects with a ≥ 3 point increase in the MGC score

Additional exploratory analyses will be conducted as part of the extended follow-up study. The primary focus of the two observational off study-intervention time points (weeks 72 & 96) will be to assess B cell recovery/repopulation as a safety measure. Specifically, we will examine: (1) Percentage of subjects achieving normal B cell counts at weeks 72 & 96; and (2) Percentage of subjects returning to at least baseline (pre-treatment) B cell counts at weeks 72 & 96.

10. SAMPLE SIZE JUSTIFICATION

Below, we introduce some key notation that we use to describe the analysis plan for the proposed trial:

- Let p_P represent the true (unknown) percentage of subjects treated with placebo who will achieve success on the primary endpoint
- Let p_R represent the true (unknown) percentage of subjects treated with rituximab who will achieve success on the primary endpoint

Based on a prior study completed by Sanders et al (2008) on MMF in AChR+ MG, 38.6% of placebo treated case achieved a treatment response. The placebo start point was 34.1 mg prednisone; hence, 38.6% of placebo recipients had a reduction of prednisone dose by at least 78%. Also the MMF start point was 30.7 mg prednisone; therefore, 44.3% of MMF recipients had a reduction of prednisone dose by at least 76%. Based on this information, we assume that 40% of placebo recipients will achieve a 75% or greater prednisone dose reduction ($p_P = 0.40$).

The design of this trial was somewhat restricted due to the fact that the company that produces rituximab only agreed to provide 25 doses to the investigators. As a consequence, this study required more of a sample size justification (for the sample size fixed by external factors) as opposed to a standard sample size calculation that determines the required sample size for a fixed target power. Using the notation above, the one-sided futility hypothesis that the treatment achieved the desired clinically meaningful level of interest may be stated as:

$$H_0: p_R - p_P \geq 0.30 \textit{ vs. } H_A: p_R - p_P < 0.30$$

Therefore, rejecting the null hypothesis suggests 'futility' in the sense that it appears unlikely that conducting a future phase III clinical trial would lead to a significant effect with a magnitude at least as large as the specified clinically meaningful effect of interest. If we don't reject the null hypothesis, this would provide one of the "go" conditions for conducting a future phase III study.

The table below shows the power computed across a range of assumed values for the true response rate in the rituximab subjects. The calculations assume a type I error rate of 10%, $p_P = 0.40$, and a conservative assumption of up to 20% missing data. The table below demonstrates the benefits of using the futility design. When the true success rate for rituximab is near or below the true success rate for placebo, the study will declare "futility" with high probability. Likewise, when the true success rate for rituximab is well above the true success rate for placebo, the study has a very low chance of incorrectly declaring "futility". Given the sample size limitations mentioned above, we feel that this provides a reasonable chance of having a successful study – where "success" is defined as answering the main question of interest regarding whether there is clear evidence to rule out an effect of rituximab in this population, or to provide enough evidence to justify a larger trial in the future.

Table 10.1: Power of Futility Test as a Function of True Rituximab Rate

Rituximab Rate (*p*R)	**30%**	**35%**	**40%**	**45%**	**50%**	**55%**	**60%**	**65%**	**70%**	**75%**	**80%**	**85%**	**90%**
Pr (Futility)	92%	84%	74%	63%	50%	37%	25%	16%	10%	4%	2%	1%	<1%

11. SAFETY MONITORING

The monitoring of subject safety and data quality will follow the NINDS Guidelines for Data and Safety Monitoring in Clinical Trials. Subjects will be monitored through regular physical examinations, vital signs, laboratory tests, and incidence and severity of adverse events. Infections will be treated symptomatically. Cardiovascular risk factors will be assessed prospectively by recording risk factors (e.g. family history, smoking history, and status). Additional safety evaluations will be conducted on conventional safety variables, such as adverse events, laboratory tests, and vital sign changes. In

addition, B cell counts, immunoglobulin levels, infusion-related reactions, and thromboses infections will be carefully examined. Tolerability will be determined by the ability to complete the study on the assigned experimental medication.

11.1. Adverse Experience Reporting

The adverse event (AE) definitions and reporting procedures for this study comply with all applicable United States FDA regulations and International Conference on Harmonization (ICH) guidelines. Adverse events will be reviewed and recorded at each study visit and infusion visit. Information on AEs of medication and on inter-current events will be determined at each visit by direct questioning of the subjects, clinical examination, and laboratory tests. The Site Investigator will carefully monitor each subject throughout the study for possible AEs. All AEs will be documented on CRFs designed specifically for this purpose. It is important to report all AEs, especially those that result in permanent discontinuation of the investigational product being studied, whether serious or non-serious.

Subjects will be monitored for AEs from the time they sign consent until the Week 52 visit. All SAEs must be followed until resolution, or until the Week 52 Visit, whichever comes first. If the SAE is still ongoing at Week 52, it will be Resolved with Sequelae.

- SAEs that are discovered less the 90 days prior to the Week 52 Visit will be followed until resolution or for a minimum of 90 days, whichever comes first, even if it is past the Week 52 Visit. A repeat Termination Visit does not need to occur.
- If a new SAE is discovered at the Week 52 Visit, it must be followed until resolution, or for a minimum of 90 days, whichever comes first.
- Subjects that are withdrawn from the study or have intervention discontinued due to an SAE will have reduced follow-up. This will include being followed monthly via telephone or in person for a minimum of 90 days or until SAE resolution, whichever comes first, after which a Termination Visit will be conducted. The Termination visit will mirror the Week 52 Visit.
    - If the SAE is resolved within 90 days, the Termination Visit can occur earlier.
    - If the SAE is not resolved within 90 days, the SAE will be Resolved with Sequelae, and the Termination Visit will occur

  During the 90 day (or less) SAE follow-up period, no new AEs will be recorded. Existing AEs will be followed until resolution, or the Termination Visit, whichever comes first.

During the optional week 72 & 96 visits, only those AEs that in the opinion of the investigator are deemed related to study procedures will be reported.

Each Clinical Study Site Investigator and research team (co-investigators, research nurse, clinical trial coordinator) are responsible for identifying AEs, reporting them through the DCC Online Adverse Event Reporting System (OEARS), and determining the relationship of the AE to the study drug/study procedures (as described in section 2.2). Investigators are also responsible for complying with the NeuroNEXT Central IRB (CIRB) reporting requirements for all safety reports. Copies of each report and documentation of IRB notification and receipt will be kept in the investigator's study file.

For the purposes of this study, before randomization only those AEs (serious and non-serious) that, in the opinion of the investigator, are deemed related to study procedures will be reported. Non-serious adverse events that are reported to or observed by the investigator or a member of their research team will be submitted to the DCC in a timely fashion (within 5 working days / 7 calendar days). Investigators must report any SAE within 24 hours of learning of the event. Upon entry of an SAE by a site investigator, the DCC Online Adverse Event Reporting System (OEARS) will immediately notify the Medical Safety Monitor (MSM). The MSM will review the SAE report, and may request further information if necessary. The OEARS maintains audit trails and stores data and communication related to the review of any AE reported in the study. The MSM may determine that the SAE requires

expedited reporting to the FDA. For example, the Sponsor-Investigator are required to notify the FDA of any fatal or life-threatening AE that is unexpected and assessed by the MSM to be at least possibly related to the use of Rituximab. If expedited reporting is required, the DCC will prepare a MedWatch safety report for submission to the FDA and Genentech. If warranted, the MSM will notify the DSMB chair. The DSMB may suggest changes to the protocol or consent form to the PPI as a consequence of SAEs.

11.2 Medical Safety Monitor

As previously indicated, Dr. Michael Shy will serve as the MSM for this trial. In addition to performing real-time reviews of all SAEs (as described in section 2.2), Dr. Shy will also receive quarterly tabulations, by blinded treatment group, of all AEs/SAEs for the purpose of determining if any safety trends exist that may raise concerns. Aggregate reports, blinded by treatment group, will be provided by severity, attribution (anticipated or unanticipated), and relationship to study treatment. The percentage of subjects who experience any AE will be compared by body system across the two groups. The additional questions related to whether the AE/SAE is related to treatment and/or unanticipated will be used to subset these into a series of additional tables. The quarterly review will identify any disconcerting discrepancy in the frequency of any AE/SAE between the two groups.

11.3 Data and Safety Monitoring Board

A Data and Safety Monitoring Board (DSMB), appointed by the NIH/NINDS, will meet at approximately six-month intervals (or as determined by the NINDS) to review partially unblinded study data provided by the study statistician. The DSMB will periodically review and evaluate the accumulated data for participant safety, adverse events, study conduct, and study progress. The DSMB may suggest changes to the protocol or consent form to the Study Chair as a consequence of AEs. The DSMB may also make recommendations to NINDS concerning continuation, modification, or termination of the study. The frequency and format of DSMB meetings, reports, and guidelines for interim analysis will be agreed upon prior to study subject enrollment.

11.4 Study Hold Rules (Safety)

Individual study subjects may be withdrawn from the study medication, but continue to be followed for safety, if subjects develop a grade 3 or more suspected toxicity as graded by the National Cancer Institute's (NCI) Common Terminology Criteria for Adverse Events (CTCAE) Version 4.0 or an SAE related to study medication as determined by the MSM. Grade 3 adverse events are severe or medically significant, but not immediately life-threatening and may cause hospitalization or prolongation of hospitalization indicated. Descriptions of CTCAE grading criteria are included in the Manual of Operations and SAEs are specifically defined in section 2.2. Subjects will be allowed to resume participation in the study if their suspected toxicity or AE resolves completely, and in the judgement of the investigator and MSM it is safe for the subject to continue.

12. INTERIM STOPPING RULES

The study will be permanently stopped, and no further administration of rituximab will be given, if the investigator, CIRB, DSMB, and/or any other institutional or regulatory body deems it inappropriate for the study to resume due to a significant number of randomized subjects developing safety concerns that cannot otherwise be attributed to MG, infections, disease relapse, or pre-existent comorbidities as deemed by the MSM.

REFERENCES

Barohn RJ, McIntire D, Herbelin L et al (1998). “Reliability Testing of the Quantitative Myasthenia Gravis Score”, *Annals of the New York Academy of Science*, 841:769-772.

Burns TM, Conaway MR, Cutter GR et al (2008). “Construction of an Efficient Evaluative Instrument for Myasthenia Gravis: The MG Composite”, *Muscle Nerve*, 38:1553-1562.

Burns TM, Conaway MR, Cutter GR et al (2008). “Less Is More, Or Almost As Much: A 15-Item Quality-of-Life Instrument for Myasthenia Gravis”, *Muscle Nerve*, 38:957-963.

Burns TM, Conaway M, Sanders DB et al (2010). “The MG Composite: A Valid and Reliable Outcome Measure for Myasthenia Gravis”, *Neurology*, 74:1434-1440.

Burns TM (2012). “The MG Composite: An Outcome Measure for Myasthenia Gravis for Use in Clinical Trials and Everyday Practice”, *Annals of the New York Academy of Science*, 1274:99-106.

Levin B (2012). “Selection and Futility Designs”. In: Ravina B, Cummings J, McDermott MP, and Poole M ed. *Clinical Trials in Neurology*. Cambridge: Cambridge University Press, pp. 78-90.

Little RJ and Rubin DB (2002). *Statistical Analysis with Missing Data*. Hoboken, New Jersey: John Wiley & Sons.

Sadjadi R, Conaway M, Cutter G et al (2012). “Psychometric Evaluation of the Myasthenia Gravis Composite Using Rasch Analysis”, *Muscle Nerve*, 45(6):820-825.

Sanders DB, Hart IK, Mantegazza R et al (2008). “An International, Phase III, Randomized Trial of Mycophenolate Mofetil in Myasthenia Gravis”, *Neurology*, 71:400-406.

Wolfe GI, Herbelin L, Nations SP et al (1999). “Myasthenia Gravis Activities of Daily Living Profile”, *Neurology*, 52: 1487-1499.