STATISTICAL ANALYSIS PLAN

TITLE: NAD+ Precursor Supplementation With Exercise Training to Increase Aerobic Capacity in Friedreich's Ataxia

NCT: NCT04192136

DOCUMENT DATE: 04/02/2025

STATISTICAL CONSIDERATIONS (SHANA MCCORMACK, FRDA TRIAL)

Putt, Mary finalized 4/2/2025 (prior to unblinding)

/Users/maryputt/Library/CloudStorage/Box-Box/iddrc_new/iddrc/Mccormack Freidrich's Ataxia/FA_trial/Statistical_Analysis_Plan/STATISTICAL CONSIDERATIONS.docx

1.1 Primary Endpoint

The primary outcome for **Aim 1** will be:

- within-participant change in peak VO_2, the highest oxygen uptake achieved on adaptive cardiopulmonary exercise testing, with peak VO_2 expressed in L/min (unadjusted for individual lean mass)

Peak VO_2 refers to the highest oxygen uptake achieved during specific exercise testing conditions (versus VO_2max, which is the theoretical maximal oxygen uptake achievable by the individual, typically confirmed by demonstrating a plateau in VO_2 during while work continues to increase, along with a combination of other factors, PMID: 39873542). Not all individuals who perform maximal test according to our criteria will demonstrate this plateau. In this ramp, symptom limited test, participants continued until maximal volition or until limiting symptoms occurred (PMID: 38515223). Ratio methods for scaling VO2 to size will produce lower values in those with higher body weight or body fat, as reviewed in PMID: 39873542.

Key Secondary endpoints for Aim 1 include within-participant changes in:

- within-participant change in peak VO_2, the highest oxygen uptake achieved on adaptive cardiopulmonary exercise testing, with peak VO_2 expressed in mL/min, accounting for individual lean mass (appendicular, lower extremity only), the latter of which is a key driver of peak VO_2
- max workload in Watts, with and without accounting for individual lean mass (appendicular, lower extremity only),
- max O_2 pulse in ml/beat,
- AT in ml/min, with and without accounting for individual lean mass (appendicular, lower extremity only),
- muscle strength (dynamometry in N), with and without accounting for individual lean mass (appendicular, lower extremity only)
- fatigue (Fatigue Impact Scale[57]),
- physical activity (PADS-R [58]),
- resting heart rate,
- FA-HI (PMID: 37646046),
- heart rate variability via MCOT

1.2 Secondary Endpoints

The primary outcome for **Aim 2** will be

- within-participant change in whole body insulin sensitivity (WBISI), reflecting mostly the contribution from skeletal muscle.

Other key secondary outcomes include within participant changes in:

- whole body insulin sensitivity S_i[76],
- fasting glucose,
- fasting insulin,
- fasting lactate,
- glucose area under the curve (AUC),
- insulin AUC,
- first- and second-phase insulin secretion estimates from OGTT[78],
- disposition index,
- insulin clearance,
- lactate AUC,
- whole-body lactate clearance (K_{01})[77]

1.3 Statistical Methods

1.3.1 Baseline Data

Baseline and demographic characteristics will be summarized by standard descriptive statistics (i.e., means and standard deviations or medians and interquartile ranges for continuous variables such as age, percentages for categorical variables such as gender) and graphically.

1.3.2 Outcome Analysis

The primary outcome for **Aim 1** is the change in peak VO_2 from baseline to 12 weeks.

Rationale for selection of the primary outcome:

- CPET measures are widely used, objective, and highly associated with health.[79]
- Peak VO_2 demonstrates excellent reproducibility and reflects disease progression in FA.[80]
- As our preliminary data show, peak VO_2 contributes to both exercise capacity and daily function.[81]

Key secondary outcomes (change from baseline to 12 wks):

Primary analysis: We will fit a linear model with the outcome of change in peak VO_2 and covariates of treatment arm, baseline peak VO_2, and randomization stratum (child vs. adult). In a modification of previous strategies for the analysis of factorial designs[82], the family-wise Type I error rate is maintained at <5% by using a two-stage approach: a Holm-Bonferroni correction is applied in Stage 1, and a single contrast is pre-specified in Stage 2 (**Table 3**). In Stage 1, we test if mean peak VO_2 in any of the active treatment arms differs from control (no-exercise + placebo). In Stage 2, if the exercise+NR arm differs from the control arm, we then compare

exercise+NR to exercise alone (exercise + placebo), to determine whether NR increases mean peak VO_2 beyond the effect of exercise alone (exercise + placebo).

Secondary analyses: The primary analysis will be repeated with each secondary outcome. Additional models will explore whether change from baseline can be explained by either baseline demographic or clinical covariates (age at randomization, sex, FRDA disease activity – mFARS upright stability score), baseline characteristics of muscle quality (appendicular lean mass – lower extremity, muscle OXPHOS capacity – CrCEST recovery time constant, NAD^+ content of blood or muscle), or features of the intervention in the relevant intervention groups only (final dose of NR, final training intensity in Mets per week). ***Sex as a biological variable:*** We will include sex as a covariate. We will report the results of sex-stratified analyses for the entire cohort and describe age/puberty effects for children.

Omaveloxolone (OMAV) was licensed in 2022 in the U.S. for the treatment of FRDA in individuals at least 16 years of age. "OMAV" is expected to be used eventually by most patients with FRDA for whom it is approved. In this study, a small fraction of participants are expected to be prescribed OMAV because of when the study was started (first participant enrolled in 2020, when OMAV was not yet approved) and its age range (including individuals under age 16 years, for whom OMAV is not yet approved). We will adjust for OMAV prescription as a pre-specified secondary analysis.

The main outcome for **Aim 2** is change in S_i[76] (reflecting mostly muscle) from baseline to 12 weeks. *Rationale for selection:*

Key secondary outcomes (change from baseline to 12 wks): fasting glucose, insulin, lactate; glucose, insulin, GLP1, lactate AUC; insulin secretion estimates from OGTT[78]; insulin clearance; FFA flux; CGM time spent with blood glucose above 140 mg/dL, CGM glucose coefficient of variation

Analysis of primary outcome and key secondary: The same two-stage approach as outlined in **Aim 1** will be used to analyze **Aim 2** outcomes (See **Table 4** from the protocol below). To assess how aerobic capacity is associated with glycemic outcomes, we will also describe the incremental effect of adding peak VO_2 as a covariate to statistical models. ***Sex as a biological variable*** will be addressed as in Secondary analyses above and as a subgroup (see below).

Missing Data

We will explore document rates of missing data and explore differences between the baseline covariates of individuals with and without missing data, both overall and by treatment arm. Using multiple imputation, we will carry out a sensitivity analysis for the primary and secondary analyses, exploring the impact of the missingness under the missing at random assumption.

Safety Analysis

All participants enrolled in the study who receive the supplement/placebo will be included in the safety analysis. The frequencies of AEs by type and severity will be summarized by arm and the relationship to the supplement/placebo and to exercise will be summarized by grouping across arms. 'Falls' are pre-specified as a particular AE of interest. SAEs (if any) will be described in detail.

Subgroup Analyses

The four pre-specified subgroups are: Age (child vs. adult, stratified at age 18 years), omaveloxolone (OMAV) status (prescribed or not), biological sex (male vs. female) and Respiratory Exchange Ratio (RER) achieved during CPET (greater than/equal to vs. less than 1). For the primary outcome, the analysis will include an hypothesis test of the interaction between the subgroups and study arm. These analyses will be unadjusted for multiple comparisons. A forest plot including two-sided 95% CI will be created.

Using a linear model, the association of duration of OMAV prescription with outcome will be explored among individuals who are prescribed the drug.

1.4 Sample Size and Power

This section is included for completeness.

Sample size calculations for **Aim 1** appear in **Table 3**. For these calculations (RStudio version 1.1.456), we used our preliminary data that the mean VO_2 max in individuals with FRDA (n=19) undergoing cycle ergometry was 20.2 mL/kg/min (SD 7.1); after one year, the mean change in VO_2 max was -0.5 mL/kg/min (SD of within-participant change, 2.4). The post-hoc comparisons to be tested are shown in **Table 3**. For the main comparison (NR, plus exercise vs. placebo for NR, plus exercise), a sample size of n=32 individuals (16 in each group) has 80% power to detect a clinically relevant difference of 2.6 mL/kg/min (this represents a ~16% of the cohort mean of 20.2) in the change in VO_2 max, with a conservative inflated SD of 3 and a two-sided =0.05 (using a two-sample t-test) Bonferroni adjusted for multiple comparisons. A previous case report of exercise training alone in an individual FRDA demonstrated an improvement of 27% in VO_2 max, thus this effect size is plausible.[81] A previous investigation in FRDA estimated that an increase of 30% represents the upper limit of the adaptation possible with aerobic exercise training.[87] Thus, we will randomize 72 individuals to achieve 64 with complete data (allowing 12% attrition). In addition, over an estimated 4 years of conducting study procedures, this sample size requires that our team conduct 3 in-person study visits per month, which is eminently feasible with the existing research infrastructure.

Sample size calculations for **Aim 2** appear in **Table 4**. We calculated the distribution of S_i in individuals with FRDA.[16] Mean S_i was 359 ml/min/m^2 ± 88 SD, similar to previous.[76] The 12-wk within-subject correlation of S_i was 0.81[76], thus the SD of the 12-wk change in S_i will be ~56 ml/min/m^2. With our design, we have >80% power to detect administration-related changes in S_i of 18-34%, accounting for multiple testing with a family-wise Type I error rate <5%. These changes are plausible based on previous studies. For example, S_i increased 22% after 16 weeks

of training in previously sedentary healthy adults.[88] In another study, S_i doubled with insulin-sensitizing therapy in individuals with pre-DM without FRDA.[86]

1.5 Interim Analysis

There are no pre-specified outcome analyses. Adverse events will be summarized at each DSMB meeting and the DSMB will make recommendations regarding any decisions to modify or stop the study related to safety considerations. The data will be explored for completeness and for an initial cleaning at a point roughly half-way through accrual.

TABLES

TABLE 3. Analytic approach & sample size considerations, Aim 1

Scenario	**Mean D in VO_2max (ml/kg/min)[1] over 12 weeks; D is presented in 2 ways: absolute D, and in parentheses, as a % of mean baseline VO_2max, 20.2 ml/kg/min**				**Stage 1: Power to detect differences between administration arms vs. control arm (%)[2]**			**Stage 2: Power to detect NR + exercise vs. Placebo for NR plus exercise [4]**
	Cntrl (Placebo for NR, No exercise)	Placebo for NR, Plus Exercise	NR, No exercise	NR, Plus Exercise	Three effects[2]	At least 2 effects[2]	At least 1 effect[2]	
1	0	3.1 (15)	3.1 (15)	3.1 (15)	81	90	96	NA[3]
2	0	3.2 (16)	0	3.2 (16)	NA[3]	81	94	NA[3]
3	0	2.9 (14)	2.9 (14)	5.5 (27)	80	92	100	81
4	0	0	0	3.0 (15)	NA[3]	NA[3]	80	84

[1]Calculations assume a 12wk SD of 2.5 ml/kg/min for change in VO_2max from baseline to 12 wks. This SD value is based on our pilot data (1y SD, 2.4 ml/kg/min).

[2]Power to detect at least 1, 2 or 3 administration arms to be different from control (Cntrl), as relevant to the scenario. For example, in Scenario 2 where exercise alone (exercise + placebo) is effective and NR alone is not effective, the power to detect both the differences between exercise alone (exercise + placebo) vs. control and exercise + NR vs. control is 81%. Power to detect a difference between ≥1 administration arm vs. control is 94%.

[3]"NA" means "not applicable", for example in Scenario 2 where only 2 administration arms differ from control, we do not compute the statistical power to detect 3 effects. Throughout, the family-wise Type I error rate (the probability of falsely rejecting at least one null hypothesis) is controlled at <5%.

[4]Testing central hypothesis: exercise + NR vs. exercise alone (exercise + placebo)

TABLE 4. Analytic approach & sample size considerations, Aim 2

	Mean D in S_i (ml/min/m^2)[1] over 12 weeks; D is presented in 2 ways: absolute D, and in parentheses, as a % of mean baseline S_i, 359 ml/min/m^2				**Stage 1: Power to detect differences between intervention arms vs. control arm (%)[2]**			**Stage 2: Power to detect NR + exercise vs. placebo for NR + exercise[4]**
Scen-ario	Cntrl (Placebo for NR, No exercise)	Placebo for NR, Plus Exercise	NR, No exercise	NR, Plus Exercise	Three effects[2]	At least 2 effects[2]	At least 1 effect[2]	
1	0	72 (20)	72 (20)	72 (20)	81	90	96	NA[3]
2	0	67 (19)	0	67 (19)	NA[3]	81	94	NA[3]
3	0	65 (18)	65 (18)	123 (34)	80	92	100	81
4	0	0	0	67 (19)	NA[3]	NA[3]	80	84

[1]All calculations assume a 12wk SD of 56 ml/min/m^2 for **D** in S_i from baseline to 12 wks.

[2]Power to detect at least 1, 2 or 3 intervention arms to be different from control (Cntrl), as relevant to the scenario. For example, in Scenario 2 where placebo for NR, plus exercise is effective and NR, no exercise is not effective, the power to detect both the differences between placebo for NR, plus exercise vs. control (placebo for NR, no exercise) and NR, plus exercise vs. control (placebo for NR, no exercise) is 81%. Power to detect a difference between ≥1 intervention arm vs. control (placebo for NR, no exercise) is 94%.

[3] "NA" means "not applicable", for example in Scenario 2 where only 2 intervention arms differ from control (placebo for NR, no exercise, we do not compute the statistical power to detect 3 effects. Throughout, the family-wise Type I error rate (the probability of falsely rejecting at least one null hypothesis) is controlled at <5%.

[4]Testing central hypothesis: NR, plus exercise vs. placebo, plus exercise