IGF-1 and Bone Loss in Women with Anorexia Nervosa

NCT # 01406444

Version Date: January 16, 2019

VI. Biostatistical Analysis

Experimental Endpoint

The primary endpoint will be PA spine BMD, and the primary analysis strategy will be to use a linear random effects model to estimate and test the effects of the three therapies on the study endpoints. The model that we propose is one where each patient has a linear trajectory from baseline to 6 months and then a possibly different linear trajectory from 6 months to a year. The 3 parameters that describe this trajectory are the intercept at baseline, the first slope, and the second slope. We assume that these three parameters have a multivariate normal distribution, with means dependent on the treatment group, the parameters of which we will estimate using SAS PROC MIXED. Since there are two slope parameters that depend on treatment, the mean slope in the first period and the mean slope in the second period, it is necessary to define the contrast of these parameters that will be used as the primary test of treatment effect. The contrast we will use will be the one that estimates the difference in the change in the endpoint at 12 months between the treatment groups. The advantage of using this analysis method rather than an analysis of covariance using only the 12 month and baseline data is that the method we propose will be more powerful and will also accommodate missing data better. It will give unbiased estimates whenever data are missing at random, while last observation carried forward (LOCF) or a complete case analysis would be biased (89). We will examine goodness of fit to this model using tests based on adding quadratic terms and by residual plots. If necessary we will add higher order terms to the model. We would use LOCF or complete case analysis as a secondary analysis. We use a model which has a break point at 6 months for all patients even for those on continuous therapy (placebo, risedronate) because having models that differ for the 3 treatments requires strong assumptions that lead to biases (90). An additional advantage of this model is that it relatively easy to add baseline covariates. The primary analysis would be intent to treat. Patients who wish to stop therapy will be encouraged to return for all evaluations. Our primary comparison for BMD is PA spine BMD, and other comparisons are considered secondary. In this way, we avoid having to correct each comparison using a multiple comparison correction such as the Bonferroni inequality. We consider the comparison of sequential therapy to risedronate and the comparison of each of these to placebo to be separate scientific questions, so we do not plan to correct these 3 planned contrasts using a multiple comparison procedure (91).

Markers of bone metabolism

The analysis will focus on 2 periods – one from 0 to 6 months, where we will have 2 post-treatment measurements, and one from 6 to 12 months, where we will also have 2 post-treatment measurements. The primary analysis will be to compare the bone markers in patients post risedronate values to the post-rhIGF-1 values using a repeated measures analysis of covariance where the covariant is the baseline value for each period and the primary question is on average over the period whether there is a difference in markers. The significance of the analysis in the first period is whether these markers are affected differently by the 2 treatments, rhIGF-1 and risedronate. The second question, for the second period, is whether pre-treatment with rhIGF-1 changes the response to risedronate.

Muscle mass

Body composition parameters (thigh muscle mass, total, subcutaneous and visceral abdominal fat) by cross-sectional CT at 0, 6, and 12 months. The primary analysis of this measure will be by analysis of covariance with the 6-month measures as the dependent variable and the baseline measure as a covariate.

Bone microarchitecture and strength
There will be two analyses of these data, one comparing 6 months to baseline with an analysis of covariance, and one comparing 12 months to baseline. The first analysis isolates the effect of IGF-1 vs. risedronate and placebo. The second measures the effect of the sequential therapy.

Power Analyses: Determination of an n of 100 subjects for the primary aim.
The primary aim of the study is Specific Aim 1A, and the primary endpoint will be PA spine BMD. This aim will compare rhIGF-1 followed by risedronate (for anabolic consolidation and further BMD increases) to risedronate alone and to placebo. The study will determine 1) whether sequential therapy is superior to risedronate, 2) whether sequential therapy is superior to placebo and 3) whether risedronate administration is superior to placebo. A total of 100 patients will enter this three treatment parallel-design study. With 36 evaluable patients in each of the two active treatment groups (assuming a 10% drop-out rate), the probability is >80% that the study will detect a difference between sequential therapy and risedronate at a two-sided 5% percent significance level, if the true difference between the treatments is 2.5%. This is based on the assumption that the SD of the response variable (BMD in women with AN receiving oral contraceptives) is 2.69%, as we found in our previous study (24) and on the 2.5% difference we found in our interim analysis of our preliminary study of sequential rhIGF-1 followed by risedronate (4.8%) compared with risedronate alone (2.3%) at 12 months. In addition, it is consistent with the increase of 2.54% from baseline (3.05% compared with placebo) at the PA spine at 6 months in subjects who received rhIGF-1 and OCPs (see Background and Figure 1). Risedronate increased BMD at 6 months followed by a decrease of 0.3% between 6 and 12 months (see Progress Report). Thus, we expect that the principal difference between sequential therapy and risedronate alone will be the 6-month effect we found of rhIGF-1 in this estrogen-treated population of 2.54% plus 0.4%. With 36 subjects in each treatment group and 18 subjects in the placebo group (assuming a 10% drop-out rate), we expect the difference between the rhIGF-1 followed by risedronate and placebo groups to be approximately twice that between the two treatment groups and therefore we will have adequate power even though we have half the number of subjects in the placebo group. The power for this comparison will be 89%. We also have >80% power to detect a difference between the risedronate and placebo groups. These sample size estimates were based on using an analysis of covariance as the analysis method. The power calculated with this method will be somewhat less than the power achieved using the random effects model we propose.

Markers of bone metabolism
With 76 evaluable patients randomized between risedronate and rhIGF-I for the first six months, we will have an 80% chance of detecting a difference of 13 units in PINP. This represents a difference of 25% of the baseline value of PINP in patients with AN. For the difference between these groups (sequential therapy and risedronate) over the 12-month period, we will be able to detect a difference of 0.13 units in CTX at a 5% significance level. For CTX, the difference is 19% of the baseline level. The power calculation was based on a repeated measures analysis of covariance with the baseline value as a covariate and two post baseline measurements. We used the programs provided by [redacted] (92). The variance covariance matrix of the measurements were (350,360;360,535) for PINP and (0.05,0.012;0.012,0.049) for CTX and are from our pilot study with 36 patient completers in the risedronate only and double placebo groups.

Muscle mass
With 40 women in the rhIGF-1 group and 20 in the placebo group, the probability is 80% that the study will detect a treatment difference at a two-sided 5.0% significance level, if the true difference between the treatments is 4.3%. This is based on the assumption that the SD of the response variable is 5.2%, as we found when we examined the 6-month change from baseline in women of reproductive age receiving placebo (93). We conclude that this is feasible because it is smaller than the 6% change in fat-free mass that was produced by IGF-1 in our previously published study in patient with AN (24).

GH and IGF-1
With 40 women in the rhIGF-1 group and 20 in the placebo group, the probability is 80 percent that the study will detect a treatment difference at a two-sided 5.0% significance level, if the true difference between the treatments is 0.7 times the SD of change from baseline.

Bone microarchitecture: Spine
The literature on trabecular thickness uses percent change as the unit of effect. With 36 evaluable patients in each treatment group the 95% confidence interval on the change in trabecular thickness of the spine will be 2.7% for each of the two active treatments based on Graeff *et al.* who estimated the difference between two spine microarchitectural endpoint measurements 12 months apart to be 8.8%, depending on the measure, and the SD of the change of 8.2% (73). Based on this we will be able to estimate trabecular thickness at the spine to sufficient precision to be confident as to whether the treatments affect this variable.

Radius and Tibia
With 40 patients in each treatment group and 20 in the placebo group entering the study and a 10% drop-out rate, the probability is 80 percent that the study will detect a treatment difference at a two-sided 5.0% significance level, if the true difference between the treatments is 2%. This assumes that the within subject SD is 3%, which is the interscan variation described for the instrument (94). If the within patient SD is twice this, e.g. 6%, we would be able to detect a change in trabecular thickness of 4%, which is smaller than the difference between women with and without fractures of 6% (95). In addition, for both of the comparisons with placebo, we will be able to detect a 2.5% difference between groups with an 80% probability. If the within patient SD is twice this, we would be able to detect a change in trabecular thickness of 5%. Thus, we think that this measurement has the potential of detecting small clinically important differences in this parameter. The SD of 6% is greater than the SD found in the EUROFORS study using similar high-resolution CT scanners after 6 months of treatment with PTH of 5.6% (73). In the powering, we focus on the pairwise comparisons of the treatment arms because the scientific questions are distinct for each pairwise comparison.

Bone strength (FEA)
The literature on FEA standardizes measurements using the population SD standard deviation. We refer to one population SD as a unit. For strength, with 36 evaluable patients in each treatment group the 95% confidence interval of the change in FEA of the spine will be ± 0.33 units for each of the two active treatments. This is based on Graeff *et al.* who estimated the difference between two measurements 12 months apart to be 0.43 to 0.82 units, depending on the measure, and the SD of the change of one unit (73). Orwoll *et al.* demonstrated that 1 unit produces a hazard ratio of 3.196. Based on this, we will be able to estimate FEA to sufficient precision to be confident as to whether the treatments affect this variable.

Exploratory Aim

Sequential therapy with physiologic rhIGF-1 administration replacement followed by a bisphosphonate will increase spine bone strength in women with AN compared with bisphosphonate or placebo. We will develop a novel FEA model to determine spine strength using high resolution MDCT data. [redacted] has recently published the development and validation of such a model for wrist and leg strength in AN (78), and [redacted] is a world-renowned expert in this area (82,84,96) and will serve as a consultant on this project. For power and data analyses, please see sections on FEA of wrist above.