Clinical Protocol Statistical Analysis Plan (SAP)

Study Title:	The WISE Trial – Walking Improvement for SCI Exoskeleton
Protocol Number:	105333
NCT:	NCT02943915
Sponsor:	Ekso Bionics, Inc. 1414 Harbour Way S Ste 1201 Richmond, CA 94804
Author:	John J. Carlow Discovery Statistics 31434 West Nine Dr. Laguna Niguel, CA 92677
Date:	7/15/15

1 STATISTICAL ANALYSIS PLAN

1.1 Data Analysis Methods

The following guidance documents and standards have been consulted for the development of this study design and statistical analysis plan:

- Title 21 – Food and Drugs Chapter 1 – Food and Drug Administration Department of Health and Human Services, Subchapter H – Medical Devices, Part 890 – Physical Medicine Devices, Subpart D – Physical Medicine Prosthetic Devices, Sec. 890.3480 Powered Lower Extremity exoskeleton.
- Guidance for Industry and/or FDA Reviewers/Staff – Guidance Document for the Preparation of IDEs for Spinal Systems, January 13, 2000.
- Guidance for Industry and FDA Staff – Spinal System 510(k)s, May 3, 2004.
- Investigational Device Exemptions (IDEs) for Early Feasibility Medical Device Clinical Studies, Including Certain First in Human (FIH) Studies: Guidance for Industry and Food and Drug Administration Staff, October 1, 2013.
- Design Considerations for Pivotal Clinical Investigations for Medical Devices: Guidance for Industry, Clinical Investigators, Institutional Review Boards and Food and Drug Administration Staff, November 7, 2013.
- Guidance for Industry, E9 Statistical Principles for Clinical Trials, FDA, 1998.
- Guidance for Industry, Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims, FDA, December 2009
- Statistical Guidance for Clinical Trials of Non-Diagnostic Medical Devices, FDA, 1996.
- Clinical Investigation of Medical Devices for Human Subjects –Good Clinical Practice ISO 14155:2011).

This study design involves outcome variables measured on binomial, ordinal, and continuous scales. For the continuously scaled outcome variables (i.e., 10MWT, 6MWT, WISC-II Score, etc.), the general situation applies analysis of variance and covariance when there are *g* distinct groups with a sample of observations for each group. The general null hypothesis is that the outcome variable distribution is the same for the treatment and the active control group. The assumptions associated with the analysis of variance are: 1) the *g* samples are independent random samples, 2) the observations in group *i* (for each *i* = 1,2,...,g) are a random sample from a normal probability distribution with mean μ_i and variance σ_i^2 , and 3) the *g* population variances, σ_1^2,....,σ_g^2, are equal to a common variance σ^2. Inherent to these assumptions is the assumption of a linear additive model with equal sample sizes. There are a variety of statistical procedures for the comparison of means in an analysis of variance setting. In the case of unequal sample sizes and multiple endpoints, the Scheffe's or Bonferroni procedures are appropriate to control for experiment-wise multi-comparison error (Woolson, 1983). If the assumption of normal distributions is not met, rank analysis of variance is used. More common rank analysis of variance tests are Kruskal-Wallis and Friedman tests. In addition, paired t-Test will be used for paired data since this is a single group matched-pairs design with the assumption that the pairwise differences are normally distributed.

Multiple comparison analysis (repeated measures) of ordinal or nominal-scale data (i.e., 5-point Likert type scales, numeric rating scales, etc.) generally assumes a logit model. This type of model uses cumulative logits, which automatically take the ordinality of the data into account. If the assumptions associated with this model are met, tests of uniform association and proportional odds can be computed. If, however, maximum likelihood estimation is desirable, generalized logits are computed from a log-linear model (Agresti, 1990 and Stokes et al., 1995).

The standard formulation for the analysis of binomial variables (i.e., Occurrence of an Adverse Event, Yes or No, etc.) typically entails testing a null hypothesis of no association between the treatment groups and outcomes. Chi-square and likelihood ratio statistics, for which the appropriateness depends on the linearity of the data, are computed for this analysis. In addition, two-sample chi-square tests of equivalence will be computed. Moreover, since single-group pre- and post-test comparisons are anticipated, McNemar's test for matched pairs is appropriate since the assumption of independent samples cannot be met (McNemar, 1947). Agresti (1990, pp. 348) has also proposed several alternative methods for comparing dependent proportions. Exact probabilities will be reported when appropriate.

The study design includes a run-in period where the first 1 to 4 participants at each site will be assigned. Those participants who meet study entrance criteria during that run-in period will be randomized on a 2:2:1 ratio to one of three treatment groups; the Ekso G Intervention Group, the Active Control Group, or the Passive Control group. The participants who are assigned to the Ekso intervention will receive GT robotic gait training 3 times per week for 12 weeks (36 sessions). The goal is a minimum of 300 steps of gait training in the Ekso GT per 45 minute session. Overground walking starts when Participants require only minimal assistance of one therapist and one aide to assist with assistive device, without any BWS, for at least 10 meters. This will be assessed every 3rd session during the 10MWT. At this point, sessions will consist of 30 minutes of gait training in the Ekso, followed by 15 minutes of standard overground gait training without BWS for a total of 45 minutes of walking.

The participants randomized to the Active Control Group will receive a matched number of sessions of standard gait training. Sessions will consist of 45 minutes of walking three times per week for 12 weeks (36 sessions). Standard gait training will be a combination of body-weight supported treadmill training and overground training without BWS, with a goal of a minimum 300 steps during BWSTT per session. The Passive Control participants will continue with daily activities as normal over 12 weeks. No new gait training, mobility therapy, nor new medications (including Botox) are commenced during the study period. Participants in this group will come to the study sites for evaluations at baseline, 6 and 12 weeks. After the 12-week evaluation, the participants in this group will have an opportunity to choose to receive either Ekso or standard gait training therapy for 12 weeks.

Four sets of data will be analyzed:

1. Clinical data as derived from the observation examinations.
2. Data captured on validated quality of life questionnaires.
3. Patient Reported Outcomes and observational data.
4. Physical therapist ratings of effort and injuries.

Other appropriate analytical methods and data presentations will include:

- Descriptive statistics and graphical plots (i.e., box and whisker plots) to assess the central tendency and variability of the raw data
- Distribution diagnostics and outlier test to determine degree of nonlinearity and other systematic error
- Robust methods are used to assure adherence to assumptions associated with parametric tests used
- Poolability testing across investigators and sites with appropriate statistical and clinical justification, such as homogeneity of treatment outcomes across investigators and sites.
- One-way Nested Analysis of Variance (ANOVA) as well as generalized linear model testing adjusted for sample distributions and the unknown correlation matrix associated with placing multiple implants in one mouth
- Subgroup Analysis to include:
    - left and right lower extremity single support time, initial double support time
    - Step length
    - Stride width performed in centers using GAITRite equipment and software
- Presentation of time course distributions for:
    - Patient accounting
    - Primary and secondary endpoints
    - All complications
    - Individual patient success rates

All of the above analyses will be performed for the following study populations:

Intent-to-treat population:	Subjects who met the inclusion criteria and were randomized to an intervention arm. Thus, enrolled in the study.
All enrolled population:	Subjects who were enrolled in the study (randomized to an intervention arm) and completed at least one day of the intervention. This group of subjects will be used for safety analysis.
Per protocol population:	All subjects who were enrolled, completed one of the three intervention arms, and were followed out to last follow-up visit after following intervention.

1.2 Working Study Hypotheses

Participants undergoing exoskeleton training for 12 weeks /36 sessions will demonstrate equal progress in walking speed as those participants undergoing standard gait training for 12 weeks/36 sessions. Participants in both the exoskeleton group and the standard gait training group will show greater progress after 12 weeks/36 sessions than the participants in the passive control group.

1.3 Primary Endpoint

Efficacy: The primary endpoint is the mean increase in gait speed demonstrated during the 10MWT after 12 weeks/36 sessions of training compared to baseline and compared between groups. Both self-selected and fast speeds will be performed on the 10MWT, with the fast speed taking precedent.

Hypothesis #1: The mean change in the Ekso G intervention group will be statistically superior to that in the passive control group, as follows:

$$H_0: \mu_E \leq \mu_p \; vs \; H_a: \mu_E > \mu_p$$

where μ_E and μ_p are the mean change in the Ekso G intervention group and the passive control group, respectively.

Hypothesis #2: Using the two one-sided (TOST) procedure, the null hypothesis of non-equivalence or non-inferiority of means will be tested for the Ekso GT (μ_E) versus the Active Control (μ_A) using a two-group design with equal sample sizes in the two groups. We want to demonstrate that the mean gait speeds will have similar distributions on the average for the two intervention groups . We specify that the practical difference in means is therefore defined to be 0.055 m/s, which is one-half of the hypothesized treatment effect of Ekso G over passive control as stated below, and one-eighth of the threshold of 0.44 m/s (1.0 mph) commonly used to represent gait speed associated with community ambulation. The two null hypotheses are specified below:

$H_{01}: \mu_E - \mu_A \leq \delta_L$ or $H_{01}: \mu_E - \mu_A + \delta_U \leq 0$ vs $H_{02}: \mu_E - \mu_A \geq \delta_L$ or $H_{02}: \mu_E - \mu_A - \delta_U \geq 0$

The composite null hypothesis is that the two groups have unequal mean gait speeds. The alternative hypothesis is that the mean gait speeds are equivalent for both the Ekso G Intervention group and the active control group.

Reject H_{01} if the observed difference in the two means is too large. That is, reject if the lower bound for two-sided 100(1-2α)% confidence interval for $\mu_E - \mu_A$ is greater than the group mean multiplied by the assumed practical difference of 0.055 m/s.

Reject H_{02} if the observed difference in means is too small. That is, reject if the upper bound for the two-sided 100(1-2α)% confidence interval for $\mu_E - \mu_A$ is less than the group mean multiplied by the assumed practical difference of 0.055 m/s.

The one-sided alpha levels are fixed at 0.05 for each of the two one-sided tests and desire a power of 80% when the two gait speeds are the same true mean gait speed.

Safety: The occurrence of adverse events, if any, will be recorded, thoroughly investigated, documented and included in the study safety analysis. A determination will be made regarding the relationship of the adverse event to the intervention.

Null Hypothesis: The difference in the proportion of active control group subjects who experience a device-related adverse event and the proportion of subjects in the Ekso G intervention group to have experienced a device-related adverse event will be zero.

$$H_0: p = p_0 \ \ vs \ \ H_a: p \neq p_0$$

1.4 Individual Subject Success and Overall Study Success Criteria

Study success takes into consideration the purpose of the treatment and comparison to a control group as well as the study goals (e.g., superiority or equivalency). All primary evaluation parameters, at minimum, as well as safety information, should be accounted for in the definition of study success.

"Clinically meaningful improvement" (Intervention Responder) is defined for this proposed study as those subjects who 1) achieve no worse than one-eighth of the threshold of 0.44 m/s (1.0 mph) commonly used to represent gait speed associated with community ambulation, and 2) are free of a intervention-related complication in the active control group and the Ekso G intervention group.

Moreover, the Ekso G intervention and the active control group intervention will be considered an overall success if at least 85% of the per protocol population are defined as "Responders" and achieve the criteria specified for "Clinically Meaningful Improvement".

Individual and overall success rates will be provided at each post-intervention evaluation interval (e.g., 6, 12, and 24 weeks).

1.5 Overall Study Hypothesis

The study hypothesis is directly related to the success criteria.

Accordingly, the null hypothesis of the study posits that there is no difference between the success probabilities for the Ekso G intervention and the active control intervention. Therefore, this trial of equivalence compares case success probabilities where the investigational intervention will be considered acceptable only if it can be demonstrated with 95% confidence (α = 0.05) that it is, at worst, 15% inferior to the active control intervention (δ = 0.15).

H_0: $\pi_{\text{Ekso}} - \pi_{\text{Active Control}} \leq -0.15$ vs. H_a: $\pi_{\text{Ekso}} - \pi_{\text{Active Control}} > -0.15$

where π_i = the probability of subjects who achieve the success criteria that the difference in success probabilities is no more than 15% using the Blackwelder approach to testing equivalence of paired proportions.

1.6 Multiplicity

Although there is only one efficacy study endpoint and one safety study endpoint, the assessment of primary and secondary endpoints as well as the repeated measurement of the primary efficacy endpoint over time warrants the implementation of a conservative alpha allocation strategy. Common methods for handling endpoint multiplicity as suggested by Bonferroni and Tukey are useful adjustments for establishing significance levels and hypothesis testing (Hochberg et al). Moreover, a hierarchical test procedure (i.e., fixed sequence procedure) will be followed in order to reflect the relative clinical importance of the null hypothesis for the primary and secondary endpoints. The fixed sequence procedure tests hierarchically ordered hypothesis in sequence until at a 0.05 significance level until first non-rejection. Once a hypothesis is not rejected, no further testing will be performed. It is assumed that

the overall study hypothesis is more important than the primary and secondary endpoint hypotheses and the primary endpoint hypotheses are more important than the secondary endpoint hypotheses.

All analysis will be performed using SAS Analytics Pro, Version 9.3 and JMP Statistical Software, Version 13.2 (SAS Institute Inc., Cary, NC or MedCalc Statistics for Biomedical Research, Version 18.1, MedCalc Software, Belgium, EU.

1.7 Lost to Follow-up

Data on subjects who are lost to follow-up or who withdraw from the study will be maintained and analyzed up to the point at which they discontinued. The reason for withdrawal will be recorded if known. Subjects who discontinue participation, for whatever reason, will remain in the study and be subject to follow-up in the same manner as those who complete the study except as noted above. The only confirmed lost to follow-up subject will be the subject who dies or refuses to continue to participate in the study and thereby withdraws.

1.8 Sample Size Estimate

Computations for sample size and power are based on effect sizes derived from internal study data and the relevant clinical literature for the primary endpoint. For the Ekso G intervention group, the postulated effect (derived from internal study data) is a mean change from baseline in gait speed of 0.11 m/s, with a corresponding standard deviation of 0.18 m/s. For the active control group, the postulated effect is a mean change from baseline in gait speed of 0.078 m/s (derived from the relevant clinical literature), with a corresponding standard deviation of 0.108 m/s.

The passive control group is postulated to have a mean change of zero with the same standard deviation as active control, or 0.108 m/s. Sample size and power are then computed for both statistical tests cited above: superiority of Ekso G to passive control and non-inferiority of Ekso G to active control.

Under 2:2:1 randomization with a desired power of 80%, the required sample size for superiority of Ekso G intervention to passive control is 38 subjects with evaluable data in the Ekso G intervention group and 19 in passive control (incorporating the 2:1 Ekso G: passive control randomization). For non-inferiority of Ekso G to active control, the groups are of equal size and the required sample size is 37 per group.

Taking the greater of these numbers for each randomized group, total sample size under 2:2:1 randomization is therefore 38 for Ekso G intervention, 38 for active control and 19 for passive control, a total of 95. To account for possible attrition as well as potential variance from the postulated effects, up to 127 subjects in the randomized group will be enrolled.

With the sample size fixed at 38 in both the Ekso G intervention Group and the Active Control Intervention Group, the power estimate for the detection of a significant difference in the **Overall Study Success Rate** is as follows

- The statistical hypotheses are

  H_0 (Null hypothesis): P_E= 0.850

H_a (alternative hypothesis): $P_A = 0.550$

- A two group chi-square test with a 0.05 two-sided significance level will have 82% power to detect the difference between a Ekso G Intervention group proportion, P_E, of 0.850 and the Active Control Group intervention proportion, p_A, of 0.550 (odds ratio of 0.216) when the sample size in each group is 38.

When testing the **Overall Study Hypothesis of Equivalence,** a paired test with a 0.025 one-sided significance level will have 80% power to reject the null hypothesis that the proportions are not equivalent (the difference in proportions, $P_E - P_A$, is -0.150 or farther from zero in the same direction) when the expected difference in proportions is 0.00, assuming that the proportion of discordant pairs is 0.160 and the sample size is 56 in each intervention group.

1.9 Statistical Analysis References

1. Agresti A. (1990). Categorical data analysis. John Wiley & Sons, New York, New York.
2. Glass GV, McGraw B, and Smith ML (1981). Meta-analysis in social research. Beverly Hills, Sage Publications.
3. Dixon WJ and Massey FJ (1983). Introduction to Statistical Analysis. 4th Edition. McGraw-Hill. Pages 286-288.
4. Eliasziw M and Donner A (1991). Application of the McNemar test to non-independent matched pair data. Stats. in Medicine, **10**, 12, 1981-1991.
5. Hochberg Y and Sharper A. Bonferroni procedure for multiple tests of significance. Biometrika 1988; 75: 800-802.
6. McNemar Q (1947). Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika, **12**, 153-157.
7. Mosteller F and Chalmers TC (1992). Some progress and problems with meta-analysis of clinical trials. Stat Sci, **7**, 2, 227-236
8. Stokes ME, Davis CS, and Koch GG (1995). Categorical data analysis using the SAS system. SAS Institute, Inc., North Carolina.