Official Title: **Retrospective post-market hearing performance outcome in a cohort of CI532 recipients**

NCT Number: **NCT03039283**

Document Date: **30 June 2016**

Statistical Analysis Plan

Retrospective post-market hearing performance outcome in a cohort of CI532 recipients

Version: 1.0
Date: 30 June 2016

Study title	Retrospective CI532 Hearing Performance
Study number, version:	CEL5661
CPM	[redacted]

Table of Contents

Table of Tables

1 INTRODUCTION

This is the statistical analysis plan for a Multi-center retrospective data collection of already available routine clinical data to assess hearing performance outcome in a cohort of CI532 recipients.

The primary objective is to collect available speech perception data obtained in quiet and in noise for the pre-operative daily listening condition and at 3 and 6 months post CI532 implantation best aided conditions.

1.1 Objectives

1.1.1 Primary Objective

To assess available speech understanding in quiet and in noise pre-operative (daily listening condition) and at 6 months post CI532 implantation (best aided conditions).

1.1.2 Secondary Objective

1. To compare available aided (daily listening condition) and unaided pure tone audiometric thresholds obtained before surgery and at 6 months after CI532 surgery
2. To get insight for available CI532 device characteristics (electrode impedances, NRT, MAP T and C levels) and patient characteristics (datalogging) (cdx files)
3. To assess subjective surgeon's opinion and clinical experience with CI532 via a questionnaire

1.2 Endpoints

1.2.1 Primary Endpoint

1. Change from pre-op (daily listening condition) baseline speech understanding in quiet and in noise at 6 months post-op (best aided conditions).
2. The proportion of the recipient cohort in percent examined showing post-operative improvement per test and listening condition.

1.2.2 Secondary Endpoint

1. Change from pre-op baseline aided (daily listening condition) and unaided thresholds at 6 months post-operative
2. Defined normative CI532 device characteristics (electrode impedances, NRT, MAP T and C levels) and patient characteristics (datalogging)
3. Combined subjective clinical experience and opinion about CI532 from surgeons

1.3 Hypotheses

1.3.1 Primary hypotheses of retrospective study

1. H_0: Speech understanding in quiet and in noise at the 6 months post-operative visit are equal with speech understanding in quiet and noise at the pre-operative evaluation visit
   H_1: Speech understanding in quiet and in noise at the 6 months post-operative visit is better than speech understanding in quiet and noise at the pre-operative evaluation visit
2. H_0: The proportion of the recipient's cohort examined showing post-operative improvement per test and listening condition is below 80% for the 6 months post-operative visit.
   H_1: The proportion of the recipient's cohort examined showing post-operative improvement per test and listening condition is ≥ 80% (Dowell, Hollow & Winton, 2004) for the 6 months post-operative visit.

2 STATISTICS

A total of n=150 patients will be assessed. It is expected to enrol approximately 30 subjects per site to compensate the variability of speech understanding tests in noise across sites with a decent power per site.

At different points during follow-up data exports and interim analyses will be conducted. The first data export with available datasets is expected to take place in August 2016. Over the study period it is expected to gradually increase the number of available datasets and therefore also the number of datasets for following interim analyses. The final analysis is expected to cover 150 datasets by the end of July

The interim analysis will be of descriptive nature only to maintain the nominal type I error of 5%. Formal hypothesis testing will be confined to the final analysis.

The level of significance for all statistical tests is α=5%, 2-sided.

2.1 Sample size calculation

For the 1st primary hypothesis about speech understanding a power calculation was carried out to estimate the ability to detect a difference amounting to 20% for a clinically relevant benefit with the Freiburger monosyllabics test between the pre-op evaluations and 6 months evaluation covering test-/retest-variability and the respective deviation. For speech understanding in noise the clinically relevant difference between pre-op evaluation and 6 months visit is expected to be 1,4 dB SNR-SRT for OLSA, 30% for HSM.

A sample of n=150 patients is deemed sufficient to detect meaningful differences based on dB threshold values and % correct values. The standard deviation of the difference is based upon an earlier OLSA study (CAG5149 MP3000 Optimization Trial) and assumed as 3.0 to be on the conservative side.

Table 1: Numeric Results for Paired T-Test

Null Hypothesis: Mean of Paired Differences = 0, Alternative Hypothesis: Mean of Paired

Differences ≠ 0

Power	N	Alpha	Beta	Mean of Paired Differences	S	Effect Size
0.91002	100	0.05000	0.08998	1.0	3.0	0.333
0.95888	125	0.05000	0.04112	1.0	3.0	0.333
0.98196	150	0.05000	0.01804	1.0	3.0	0.333

Source for the study parameters:
ACE OLSA Data taken from MP3000 trial:
n=58, stdev.=3.3, rho=0.76, => stddev_diff = 2.29

$$\sigma_d = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\,\sigma_2}$$

To test for an improvement of 20% (0.2) in speech understanding, as compared to baseline, a reference proportion of 50% (0.5) and a correlation within patient of rho=0.5 is assumed. Based upon a sample of n=150 patients, the power to detect such a difference amounts to 99.9%, c.f Table 2.

Table 2: Power estimation using McNemar exact test for paired proportions

The POWER Procedure, McNemar Exact Conditional Test

Fixed Scenario Elements	
Distribution	Exact conditional
Method	Exact
Proportion 1	0.5
Proportion 2	0.7
Correlation	0.5
Number of Sides	2
Null Ratio of Discordant Proportions	1
Nominal Alpha	0.05

Computed Power			
Index	**N Pairs**	**Actual Alpha**	**Power**
1	100	0.0300	0.981
2	125	0.0322	0.995
3	150	0.0348	0.999

For the 2nd primary hypothesis claiming that the proportion of patients showing post-operative improvement is ≥80%, this translates into a one-sample test with a null hypothesis reference level of 80% (i.e. 0.8).

Table 3: Numeric Results for testing H_0: P = P0 versus H_1: P ≠ P0, 2-sided: Exact Test

Power	N	Proportion Given H0 (P0)	Proportion Given H1 (P1)	Difference (P1 - P0)	TargetActual Alpha	Alpha	Beta	Reject H0 If <=R|>=R
0.3947	100	0.8000	0.8750	0.0750	0.0500	0.0326	0.6053	71|89
0.6058	125	0.8000	0.8750	0.0750	0.0500	0.0440	0.3942	90|109
0.6761	150	0.8000	0.8750	0.0750	0.0500	0.0411	0.3239	109|130
0.7030	100	0.8000	0.9000	0.1000	0.0500	0.0326	0.2970	71|89
0.8811	125	0.8000	0.9000	0.1000	0.0500	0.0440	0.1189	90|109
0.9279	150	0.8000	0.9000	0.1000	0.0500	0.0411	0.0721	109|130
0.9287	100	0.8000	0.9250	0.1250	0.0500	0.0326	0.0713	71|89
0.9876	125	0.8000	0.9250	0.1250	0.0500	0.0440	0.0124	90|109
0.9958	150	0.8000	0.9250	0.1250	0.0500	0.0411	0.0042	109|130

Assuming a difference of 10% (i.e. 0.1) with respect to H_0 stating an improvement rate of 0.8, a sample of n=150 patients yields a power of 92.8% (Other scenarios are shown for completeness, c.f Table 3).

Power calculations were carried out with software PASS 11: Hintze, J. (2011). PASS 11. NCSS, LLC. Kaysville, Utah, USA and with SAS software proc power, Vs. 9.4.

3 CONDUCT OF STATISTICAL ANALYSIS

The statistical analyses including reformatting datasets and derivation of outcomes data-management will be carried out by numerics data gmbh, [redacted] (i.e. subcontractor).

[redacted]

The speech understanding data will be collected as % correct for either in-quiet or in-noise environments, or speech reception threshold in dB SNR for an in-noise environment.

Electrode location due to proximity to the modiolus, along with duration of deafness and age at onset of hearing loss, will be used as potential variables for the speech understanding. Therefore different categories will be implemented separating patients with a duration of ≤ 30 and > 30 years of age; onset of hearing loss at ≥ 5 and < 5 years of age, close proximity to the modiolus and distanced from the modiolus at the basal turn.

The primary hypothesis will be tested using a one sample (paired) t-test with the difference of the 6 month values versus the baseline values as outcome. This will be done for % speech understanding and dB threshold values. Results will be reported with mean and a 95% CI. The analysis will also be complemented by an analysis of covariance approach, that is, the pre-treatment values are treated as covariate while the values at month 6 are treated as outcome variable. This allows for a sensitive approach to account for baseline covariates such as age, gender and clinic, enabling meaningful average

estimates for specific subgroups with the aim to explore whether the treatment effects hold for the whole study population or are restricted to specific subgroups only. Least squares mean differences are calculated for this purpose.

The secondary hypothesis will be examined using a binomial exact test with a p=0.8 as H_0 reference level. This rather crude test will be complemented by regression analyses as described in section 8.2.

The analyses will be carried out with the SAS software (SAS Version 9.4, SAS Institute, Cary, NC, U.S.A.).

4 QUALITY CONTROL ON STATISTICAL ANALYSIS

The quality control of the statistical analyses will be conducted according to the SOPs of the subcontractor numerics data gmbh.
The method of validation is based upon the criticality level which is a combination of the complexity of the underlying SAS program and the impact of errors (importance of delivery):

1. Method 1: Reprogramming on the basis of the study specification, or on the basis of the requirements according to the study SAP.

2. Method 2: Review of the programming code based on the SAS Programming Standards and study specifications.

3. Method 3: Validation of output listings on the basis of the corresponding raw datasets (i.e. spot checks of selected data).

4. Method 4: Validation of results against the listings or related tables.

Code review and output review will be carried out by an independent person (i.e. not identical to the primary programmer).

5 DERIVED DATA

Derived data will include the binary response as a difference between baseline (pre-op) and 6 months post-operative visits, respectively. A response (y=1) will be defined as better speech understanding with respect to pre-op, a non-response (y=0) will be defined as a worse or equal speech understanding. In addition, paired differences of pre-op values vs. month 6 will be carried out.

6 PRESENTATION OF DATA

Tables will include data for continuous variables and discrete variables.

Continuous variables will be reported presenting: n, mean, stdev, median, a 95 % confidence interval (CI), a 25% and 75% percentile along with a minimum and maximum value.

Discrete variables will be presented showing the counts along with %-ages and a 95 % CI.

Specific tables will be produced for the regression model results containing information about model fitness, degrees of freedom etc.

Figures include box-plots, scatter-plots and mean +/- standard error plots of differences for visits pre-op, month 3 and month 6. Moreover, profile plots of averages of electrode impedances over the electrode contact location range will be produced. Likewise, profile plots will be produced to illustrate the hearing thresholds over the respective frequency ranges.

All figures and tables/listings will be produced as rtf-files allowing for implementing in MS WORD.

7 ANALYSIS POPULATIONS

ITT: the intention to treat population will comprise all patients with a baseline assessment.

PP: the per-protocol will include only patients with a non-missing assessment at pre-op and at visit 3 months or 6 months.

8 EFFICACY ANALYSIS

8.1 Tables, figures and lists

Demographic section

Age and gender distribution by center and overall.

Efficacy outcomes

Descriptive tables of the various outcome variables in the format described under chapter 6 Presentation of data.

Specific tables summarizing the results of the regression analysis (see section below) with estimates +/- standard error of the time effects (visits) and covariates.
Note that time effects will be considered as treatment effects (visit 3 and visit 6 against visit pre-op).

8.2 Statistical methods

Hypothesis testing

Formal testing of the 1st primary hypothesis claiming a difference of 20% or more for speech understanding will be carried out by using a paired t-test on differences of speech understanding (i.e. values at visit 6 months - values at pre-op) with a reference proportion of 20%. Number of patients with values greater or less than 20% will be reported.

Formal testing of the 2nd primary hypothesis claiming an improvement of 80% or more of patients showing post-operative improvement will be carried out by using a binomial test (one-sample) of improvement rates with a reference proportion of 80%.

These tests will be carried out at the final analysis only. For interim analyses, descriptive rates along with 95% confidence levels will be produced reporting the frequencies of responses along with the 95% CI based on exact methods (Clopper-Pearson) for visits pre-op, month 3 and month 6.
At the final analysis, in addition refined analyses will be carried out based on a modeling approach. This enables a more sensitive analysis allowing for covariates (such as baseline characteristics). Regression analyses will be carried out for this purpose incorporating the paired structure of observations. To this end the SAS proc genmod will be used (procedure glimmix will be used as a sensitivity analysis as it estimates random effects using a different approach). Furthermore, this modelling approach allows for:

- estimating consolidated treatment effects comparing post-baseline vs. pre-op performance.
- considering center specific deviations from the thresholds used by keeping these threshold as covariate for the binomial response (success rate).

Analyses will be carried out both based upon the ITT population and the PP set. Due to the nature of the data, i.e., paired data analyses based upon ITT, this will be considered as secondary and carried out in the perspective of a sensitivity analysis. For this purpose a random effects model is required to be used treating the patient as random effect.*.

* Brown, H. & Prescott, R. (1999). *Applied Mixed Models in Medicine*, New York: John Wiley & Sons.

Analyses based on continuous outcomes such as audiometric thresholds or electrode impedances will be carried out in a similar manner using the SAS proc mixed to account for random effects of normally distributed data.

Special care will be directed to check for model adequacy by examining the residual diagnostics (specific figures to this end will be presented in the appendix section of the report).

9 SAFETY ANALYSIS

Not applicable as no specific safety analyses are planned.