[redacted]

Statistical Analysis Plan for the Cochlear Implant and Healthy Ageing Observational Study- CEL5671

Commercial in Confidence

Version 1.0

Prepared for Cochlear Limited

21st January, 2018

[redacted]

Table of Contents

CI & HEALTHY AGEING OBSERVATIONAL STUDY - STATISTICAL ANALYSIS PLAN

Glossary

CI	Cochlear Implant
CAP-II	Capabilities of Auditory Performance test
DSST	Digit Symbol Substitution Test
GDS-15	Geriatric Depression Scale-15
HHIE-S	Hearing Handicap Inventory for the Elderly Scale
HUI 3	Health Utility Index (HUI) Mark (version) III
L-iADL	Lawton-instrumental Activities of Daily Living
MMSE	Mini Mental State Examination
PTA	Pure Tone Audiometry
SFT	Sound Field Thresholds
SSQ	Speech Spatial and Qualities questionnaire
Trail B	Trail Making Test Version B
TUG	Timed Up and Go test

1. Source documents

This Statistical Analysis Plan (SAP) is based on

i. The Cochlear® document observational study protocol document entitled: **Cochlear Implant and Healthy Aging: A Multinational, Multicentre Observational Study** written by [redacted]

ii. The sample case report forms

2. Overview of the protocol

With an ageing population comes an increasing burden of disease and a high associated economic cost. Among the leading contributors to burden of disease in people over 60 are sensory, mental and neurological disorders as well as unintentional injury with hearing loss considered one of the 15 most burdensome disorders (Prince et al, 2015).

Cochlear Implant (CI) treatment restores hearing function which enables communication. The ability to communicate reduces the potential for social isolation. This study aims to show that CI treatment improves overall health related quality of life and general well-being in CI candidates aged 60 or older with bilateral post-lingual deafness.

Using a series of standardised forms and a variety of questionnaires, participants will have data gathered pre-implant and then at 12 and 18 months post-implant.

The aim, as described in the protocol, is to provide transparent and comparable medical evidence that can support health care policy makers to take informed decisions on the provision of health services for the treatment of hearing loss. A brief summary of the protocol follows.

3. Objectives of the Study

The primary objective is to evaluate the change in health-related quality of life following CI treatment by using the generic Health Utilities Index Mark III (HUI 3) multi-attribute index score. All pairwise comparisons are of interest i.e. preimplant to 12 months postimplant, preimplant to 18 months postimplant, and the change from 12 to 18 months.

A secondary objective involves evaluating the impact of CI treatment in the elderly on domains that may be associated with healthy aging and overall well-being such as

- hearing ability (using the Capabilities of Auditory Performance (CAP-II) test, the Speech Spatial and Qualities (SSQ) questionnaire, the Hearing Handicap Inventory for the Elderly Scale (HHIE-S), speech recognition tests in the quiet and in noise, pure tone audiometry (PTA) and sound field thresholds (SFT))
- independence (using the Lawton-instrumental Activities of Daily Living (L-iADL) tool)
- cognition (using the Mini Mental State Examination (MMSE) score, the Digit Symbol Substitution (DSST) test and the Trail Making Test version B (Trail B))
- falls (using the Timed Up and Go test (TUG))
- depression (using the Geriatric Depression Scale-15 (GDS-15), and the De Jong Loneliness Scale)
- CI usage measures.

All pairwise comparisons between time points, as defined above, are of interest here also.

A tertiary objective is to identify healthcare resource utilisation that is impacted by comparing costs post CI treatment with those prior to treatment.

For all of these objectives, interest is also in whether various demographic variables are associated with the change in outcome scores observed.

In summary:

Primary hypothesis

CI treatment significantly improves the overall health related quality of life of elderly individuals compared to their pre-implant condition (and again from 12 to 18 months following surgery) as measured by the HUI3 multi-attribute score.

Secondary hypothesis

CI treatment significantly impacts healthy aging domains and wellbeing of elderly individuals compared to their pre-implant condition (and again from 12 to 18 months following surgery) as measured by clinically standard evaluation tools in the geriatric and audiological fields (as described above)

Tertiary hypothesis

CI treatment significantly reduces healthcare resource utilisation in elderly individuals over time.

Other hypotheses

There are covariates associated with the change in health-related quality of life and other healthy aging domains and related well-being.

4. Study Design

The study is a prospective observational longitudinal study performed to evaluate the benefits for overall health following CI treatment in the elderly population treated in routine clinical practice. Multiple centres are included in the study

Implantation with commercially available cochlear implant devices (i.e. products approved for market release) available for routine clinical treatment of deafness in all ages in collaborating investigator clinic sites is made as part of the study and no additional interventions are applied.

Change in outcome scores for various standard assessment tools and health domains post CI treatment compared to the pre-implant status are made using participants as their own controls. All questionnaires of interest are administered at each visit (up to 2 months pre-implant, 12 months post-implant [+/- 1 month] and 18 months post-implant [+/- 1 month]).

Whilst a recommended evaluation schedule is provided, participating clinics may select the which optional evaluation measures they wish to use for their patients from the tools made available on the study specific electronic platform at baseline registration of each patient.

Participating in the study is voluntary and the participant is free to electively cease participation in the registry at any time. To be included in the study, participants must meet all local and study specific inclusion criteria and have none of the exclusion criteria. Participants remain in the study for a period of up to 2 years.

5. Sample size

Assuming that a paired t-test of the change in HUI3 multi attribute score from pre-implant to 12 or pre-implant to 18 months following implantation is of interest, then with 90% power and a 5% level of significance a change of 0.1 with standard deviation of change equal to 0.25, would require 68 participants.

Because all pairwise tests are of interest one method for undertaking this is to perform multiple tests and then adjust the significance level for the number of comparisons (in this case 3). This is equivalent to having a significance level of 0.05/3 = 0.0167 and is a form of Bonferroni correction. If we recompute the sample size required with this significance level then at least 88 participants would be needed to find a change of 0.1 units in the HUI3 significant.

6. Study outcomes

The outcomes measures described in section 3 are detailed in the Protocol document and are briefly described here. This section could be updated when a better understanding of the distribution of the outcome measures is obtained through the data collected.

HUI 3

The HUI3 is a 15-item questionnaire. Responses to items are converted to levels and a scoring function applied. The ultimate result is a series of domain scores and a multi-attribute score. Domain scores typically don't have continuous distributions and need to be analysed with care. The multi-attribute score may be treated as continuous and is the focus of this study. Higher scores indicate less impairment.

CAP-II

The CAP-II comprises 9 questions placing each participant into one of 10 ordinal categories that determines their auditory performance at the time of assessment. Scores range from 0 (no awareness of environmental sounds or voice) through to 9 (use of telephone with unknown speaker in unpredictable context) so that higher scores imply better performance. These scores are an ordinal categorical scale and should not be treated as continuous.

SSQ

The SSQ is a 49-item questionnaire utilising a 10-point numerical rating scale for each question. The SSQ comprises three main subscales; speech perception (using the first 14 questions), spatial hearing (using the next 17 questions) and other qualities of hearing (using the last 18 questions).

Each subscale is the average of the scores from each subscale item; so, for example, scores on the speech component are between 0 and 10. These scores may be assumed to be continuous, higher scores imply better performance.

HHIE-S

The HHIE-S is a 10-item short form that assesses the impact of hearing impairment on emotional and social adjustment. Higher scores indicate greater handicap. The distribution is described as continuous.

Speech recognition in quiet

This tool measures the percentage of items correctly identified at 65dB SPL or the speech perception threshold level (dB SPL) at which 50% of speech items can be correctly recognised. The distribution of the first outcome may be binomial while the second is likely continuous. These outcomes need to be explored in more detail when data are available.

Speech recognition in noise

This tool measures the percentage of items correctly identified at 65dB SPL with noise in the background or the speech perception threshold level (dB SPL) at which 50% of speech items can be correctly recognised with noise in the background. The distribution of the first outcome may be binomial while the second is likely continuous. These outcomes need to be explored in more detail when data are available.

PTA

Pure tone audiometry are the unaided hearing thresholds for pure-tone stimuli via air conduction. Various frequencies will be examined. For analysis purposes, these outcomes will be presented as a descriptive mean group threshold per interval.

SFT

Sound field thresholds are the aided hearing thresholds for warble-tone stimuli presented in the sound field. Various frequencies will be examined. For analysis purposes, these outcomes will be presented as a descriptive mean group threshold per interval.

TUG

The timed up and go (TUG) test measures the time a person takes to stand up from a chair, walk three metres, turn around, walk back to the chair and sit down again. This is a continuous measure.

GDS-15

The Geriatric Depression Scale -15 (GDS-15) is a 15-question self-report. A 1 is assigned to the "negative" (i.e. more depressive) responses to each item and a 0 to the "positive" responses. The sum of the items is taken. This is likely to be able to be treated either as a continuous or a binomial response though the responses will need to be explored further to determine the distribution.

MMSE

The mini-mental state examination is a 30 point cognitive screening test resulting in scores between 0 and 30. The distribution is typically skewed to the left with most scores being close to the boundary of 30 in a cognitively healthy population. The protocol notes that analysis will need to adjust for age, gender and education level as these variables impact the MMSE score.

DSST

The digit symbol substitution test is a working memory test that is used as an indicator of brain damage, dementia and depression. The number of correct responses in a given time span is collected. This is likely to be able to be treated as a continuous response though the distribution of responses should be examined carefully.

Trail B

The Trail B test is a cognitive test that assesses executive function. The time taken to correctly complete a "trail" of letters and numbers is recorded up to a maximum of 5 minutes. This is a continuous response that is likely to be normally distributed in a healthy population. The distribution should be examined before analysis in the sample explored in this study.

L-iADL

The Lawton Instrumental Activities of Daily Living scale gauges level of disease and a patients' ability to look after him or herself. Activities assessed require mental and physical capacity and determine what a person could rather than does do. Scores on 8 items (given either a 1 or 0) are summed to give a total score, higher scores indicate more independence. This may be able to be treated as a continuous response though the distribution of responses should be examined carefully prior to analysis.

De Jong Loneliness

The De Jong Loneliness scale is a 6-item questionnaire. Each question has 3 possible responses one of which is negatively worded and the others considered positive or neutrally worded. Neutral and positive responses are scored a 1 on the negatively worded items and neutral and negative responses are scored a 1 on the positively worded items. A sum of the responses is obtained giving a range of scores from 0 to 6 in which higher scores indicate greater loneliness. This may be able to be treated as a continuous response though the distribution of responses should be examined carefully prior to analysis.

CI Usage

Daily usage data are logged in the implanted device and will be recorded at post-implant assessments.

Patient Profile Data

A variety of patient demographics will be collected. The table below describes some of these variables and the classification (as categorical or otherwise) of these variables.

Table 1: Patient Profile Data

Question number	Variable	Classification
1.1	Native language	Categorical (nominal)
1.2	Other languages	Categorical (nominal)
1.3	Local town/ city population	Categorical (ordinal)
1.4	Household description	Categorical (nominal)
1.5	People in household Age Dependent?	 Continuous Categorical (binary)
1.6	Education level	Categorical (ordinal)
1.7	Work status	Categorical (nominal)
1.8	Work level	Categorical (ordinal)
1.9	Job title	NA
1.10	Job classification	Categorical (nominal)
2.1	Weight (kg)	Continuous
2.2	Height (cm)	Continuous
3.1	How often smoking?	Categorical (ordinal)
3.2	Smoking level	Discrete (counts)
3.3	Ever smoked?	Categorical (binary)
3.4	Years regular smoking	Continuous
4.1	Alcohol how often?	Categorical (nominal, because of refusal)
4.2	Frequency of more than 6 drinks	Categorical (nominal, because of refusal)
5.1	Tinnitus frequency	Categorical (nominal, because of DK)
5.2	Does tinnitus affect activities?	Categorical (nominal, because of DK)
5.3	Does tinnitus affect sleep	Categorical (nominal, because of DK)
6.1	Exercise level	Categorical (ordinal)
6.2	Exercise undertaken	Categorical (nominal)
6.3	Hours of gentle exercise	Categorical (ordinal)
6.4	Hours of moderate exercise	Categorical (ordinal)
6.5	Hours of vigorous exercise	Categorical (ordinal)

HRU Data

Healthcare resource utilisation data will be collected. The variables in this data set

Table 2: Health resource utilisation data

Question number	Variable	Classification
CSL1	Several yes/no questions about insurance modalities	Categorical (binary)
CSL2	Cost of voluntary private insurance per month	Continuous
CSL3	Cost of employer provided insurance per month	Continuous
CSL4	Percentage of various insurance options covered	Continuous
CSL5a	Employment status 6 months ago	Categorical (nominal)
CSL5b	Employment status 6 months ago	Categorical (nominal)
AP1	Number of visits to various primary care or emergency care centres Number of visits Number of visits due to ears	 Discrete Discrete
AP2	How to travel to centre? Primary care Emergency	 Categorical (nominal) Categorical (nominal)
AP3	How long to travel to centre by foot? Primary care Emergency	 Categorical (ordinal) Categorical (ordinal)
AP4	Cost of travel to primary care	Continuous
AP5	Primary care out of pocket cost	Continuous

AP6	Emergency care out of pocket cost	Continuous
AP7	Who accompanies you?	
	Primary care	Categorical (nominal)
	Emergency	Categorical (nominal)
ES1	Number of visits to specialists	
	Type of specialist	Categorical (nominal)
	Number of visits	Discrete
ES2	Number of outpatient or day procedures	
	Type of procedure	Categorical (nominal)
	Number of occurrences	Discrete
ES3	How to travel to specialist/outpatient centre?	Categorical (nominal)
ES4	How long to travel to centre by foot?	Categorical (ordinal)
ES5	Cost of travel to specialist/outpatient centre	Continuous
ES6	Specialist/outpatient centre out of pocket cost	Continuous
ES7	Who accompanies you to specialist/outpatient centre	Categorical (nominal)
PD1	What type of functional/special tests (by test)	Categorical (nominal)
	Number of tests	Discrete
	Number of tests due to ears	Discrete
	Out of pocket cost of tests	Continuous
	What type of radiology (separately by test)	Categorical (nominal)
	Number of tests	Discrete
	Number of tests due to ears	Discrete
	Out of pocket cost of tests	Continuous
	What type of laboratory tests (separately by test)	Categorical (nominal)
	Number of tests	Discrete
	Number of tests due to ears	Discrete
	Out of pocket cost of tests	Continuous
H1	In hospital tests in last 6 months	
	Admission and discharge dates	Dates[1]
	DRG code	Categorical (nominal)
	Hired person accompanying?	Categorical (nominal)
	Family accompanying?	Categorical (nominal)
H4	Out of pockets costs for hospitalization in last 6 months	
	Not related to hearing	Continuous
	Related to hearing	Continuous
MP1	Prescription medications taken (per type)	Categorical (Nominal)
	Start and end date	Dates[2]
	Daily dose	Continuous
	Proportion out of pocket	Continuous
	Cost per month	Continuous
	Related to hearing?	Categorical (binary)
MNP	Non-prescription medications taken (per type)	Categorical (Nominal)
	Start and end date	Dates[2]
	Daily dose	Continuous
	Proportion out of pocket	Continuous
	Cost per month	Continuous
	Related to hearing?	Categorical (binary)
PHT1	Received physiotherapy/rehab in last 6 months	Categorical (binary)
PHT2	How many weeks of physio/rehab?	Discrete
PHT3	Days per week of physio/rehab?	Discrete
PHT4	Hours per day of physio/rehab?	Continuous
PHT5	Percentage of sessions due to ear-related problems	Continuous
PHT6	Had out of pocket costs for physio/rehab?	Categorical (binary)
PHT7	How much out of pocket per month?	Continuous
LG1	Received speech therapy in last 6 months	Categorical (binary)

[1] The date variables will be used to calculate length of stay, a continuous variable

[2] The date variables for MP1 will be used to calculate duration of treatment etc; likely continuous variables

LG2	How many weeks of speech therapy?	Discrete
LG3	Days per week of speech therapy?	Discrete
LG4	Hours per day of speech therapy?	Continuous
LG5	Percentage of sessions due to ear-related problems	Continuous
LG6	Had out of pocket costs for speech therapy?	Categorical (binary)
LG7	How much out of pocket per month?	Continuous
PSI1	Received psychological therapy in last 6 months	Categorical (binary)
PSI2	How many weeks of psychological therapy?	Discrete
PSI3	Days per week of psychological therapy?	Discrete
PSI4	Hours per day of psychological therapy?	Continuous
PSI5	What % of these sessions were due to ear problems	Binomial (proportion)
PSI6	Had out of pocket costs for psychological therapy?	Categorical (binary)
PSI7	How much out of pocket per month?	Continuous
SC1	How many times attending adult day care in last 6 months?	Discrete
SC2	Is the stay due to hearing difficulties?	Categorical (nominal)
SC3	Out of pocket costs for adult day care?	Categorical (binary)
SC4	Out of pocket cost per month for adult day care	Continuous
RES1	Lived in a residential care home in the last 6 months	Categorical (binary)
RES2	How many months in residential care in last 6 months	Continuous
RES3	Is residential care stay related to hearing difficulties?	Categorical (nominal)
RES4	Out of pocket expenses in residential care home?	Categorical (binary)
RES5	How much out of pocket per month?	Continuous
AD1	Hope help/social services in the last 6 months	Categorical (binary)
AD2	How many months home help/social services in last 6 months	Discrete
AD3	How many days/week home help/social services in last 6 months	Discrete
AD4	How many hours per day home help/social services in last 6 months	Continuous
AD5	Is home help/social services related to hearing?	Categorical (nominal)
AD6	Out of pocket expenses for home help/social services?	Categorical (binary)
AD7	How much out of pocket per month?	Continuous
CI1	Limited in daily activities due to disability/health problem in last 6 months	Categorical (nominal)
CI2	What percentage of health or disability problems relates to ear problems	Continuous
CI3	Hours per week people help you with your needs	Continuous

Other data:

The forms also collect information on:

- Gender (categorical; binary)
- Date of birth
- Inclusion/exclusion criteria
- Hearing, device and medical histories (a combination of categorical and continuous variables)

7. Statistical Analysis Plan

7.1 Handling of Dropouts or Missing Data

All participants should be included in the analysis regardless of missing values and no imputation (for example, last value carried forward) should be utilised. All possible care should be taken to ensure that missing values are not encountered at baseline especially with respect to subgroups of interest. The analysis methods to be suggested in the following section take into account all available data and produce unbiased estimates of effect even in the presence of missing values.

7.2 Summary Statistics

In preliminary work, summary statistics and plots should be examined for each variable to be included in a given analysis at each time point. This helps to describe the population and identify potential issues with the data. Appropriate summaries include:

- For one categorical variable (such as CAPII, gender, work status, education level) the count and proportion of participants in each category of the given variable should be tallied and the mode could be identified. Bar charts should also be created.
- For one continuous or discrete (integer only) variable (for example HUI3 multi attribute score, out of pocket costs, smoking level, age) the mean, standard deviation, median and interquartile range (IQR) could be calculated. Boxplots and histograms can be used to summarise the continuous and discrete (including response) variables.
- Non-normality of a given response at any time point is not necessarily a problem since the analysis to be used does not make assumptions about the distribution of the response scores.

For the primary and secondary hypotheses

- The change in the response variable over time should be examined by using boxplots with the response as the y-axis and time as the grouping variable or using scatterplots with a longitudinal trend included. Each persons' observations could be joined to show the variability in individual response over time. The purpose of this is to determine whether a linear change over time is apparent. This will impact on the form of the analyses undertaken.

For the tertiary and other hypotheses

- For categorical predictors of interest, boxplots at each time point should be used to determine whether the response variable (on the y-axis) appears to indicate differences between the categories (on the x-axis) of the categorical predictor.
- For continuous predictors of interest, scatter plots at each time point should be used to determine whether a relationship between the response variable (on the y-axis) and the predictor variable (x-axis) is evident.

The purpose of the above steps is to investigate the relationship between the predictor variable and the given response variable.

7.3 Statistical Analyses

Primary hypothesis analysis

If comparison between the HUI3 multi-attribute score from pre-surgery to only one follow-up time point is of interest then paired t-tests can be used. A check of the assumption that the change from pre-surgery is normally distributed should be performed. The distribution of change scores should be approximately symmetric. The disadvantage of this approach is that it uses only complete data at any given pair of time points

If all pairwise differences are of interest (as per the hypothesis) and the scatter plot of change over time appears (close to) linear then a linear mixed-effects model with a fixed effect of time (treated as a continuous variable) and a random effect for participant should be used to explain the change in response variable over time. Tukey pairwise differences between time points can be specified (we recommend using the *multcomp* package in R as it does not require balanced designs i.e. the same sample size across time points). Appropriate diagnostic checks of the linear mixed-effects model (for example, normality of the residuals, normality of the random effects) should always be undertaken. An advantage of the linear mixed-effects model is that it uses all available data even for subjects who have dropped out at a given time point and then returned.

If the change over time was evidently not linear from the scatterplot of response against time then alternative approaches should be used. For example:

1. A quadratic or cubic term for time could be added to the linear mixed-effects model and checked to determine if that improves the fit of the model (eg via reduced AIC, better diagnostic checks).
2. The pre-surgery HUI3 multi attribute score value can be used as a predictor variable (with the response variable then only including the post-surgery visits). A significant coefficient would imply that the amount of change is dependent on the pre-surgery score. An interaction between baseline score and post-surgery time points could also be included – this would show whether the change in mean score over time is impacted by baseline score. This approach will remove any participant without baseline data from the analysis.
3. Use visit as categorical variable. This essentially fits a linear model between time points but makes no assumption that the change from baseline is linear. Tukey pairwise comparisons between all pairs of time points (if time is found to be a significant predictor) are straight forward with this approach and should be undertaken using the multcomp package in R, which does not require balanced designs.

Secondary hypotheses analysis

The approach described for the primary hypothesis can be used for all continuous responses described in the secondary hypotheses. Failure of the assumptions may indicate that that a linear model is not appropriate and that other response distributions need to be explored. At this stage it is likely best that a statistician is consulted.

For ordinal responses such as the CAP II score an ordinal mixed-effects model including time as a fixed effect and participant as a random effect. The results of this model will show the

odds of being in a higher category compared to the current category over time (this interpretation is package dependent).

Tertiary hypotheses analysis

The approach described for the primary hypothesis is appropriate here also for continuous outcomes. As described above, for ordinal responses an ordinal mixed-effects model including time as a fixed effect and participant as a random effect. If the response variable is nominal (categorical without natural ordering) a multinomial mixed-effects model can be used and for binary responses a mixed effects logistic regression can be undertaken.

If only one pair of time points is of interest then for binary responses a test for the change in proportions can be undertaken and for continuous responses a paired t-test can be undertaken.

Analysis for other hypotheses

The other hypotheses involve the effect of covariates on a response and on the change over time. To assess the effect of a given covariate (for example age, gender, etc) the covariate can be added to the model as a fixed main effect in addition to the time variable and as an interaction with the time variable.

While there is no widely accepted standard approach for creating a statistical model, in the situation where many covariates are of interest, one approach is to use a model that includes only one covariate at a time (in addition to time) and assess its significance. Combinations of covariates with p-values less than 0.1 could then be added to the model at the same time (eg age and gender). Finally, those without significant p-values at the 5% level could be removed from the final model. If an interaction is significant but the main effect for that covariate is not significant, the main effect should be retained in the model. If an interaction is not significant it can be removed from the model.

For each covariate included in the model a rule of thumb suggests having at least 20 participants. In other words, if the sample size is 100 participants then a maximum of 4 covariates plus time could be included in the model. This approach aims to retain sufficient power to find a significant result if an association between these variables indeed exists.

7.3 Multiple Comparisons

The statistical methods described above involve many comparisons using related measures on the same sets of participants. This can result in occasional spuriously significant results. Analyses specifically for the purpose of publication should

1. limit the number of analyses undertaken and presented or
2. use a more stringent level of significance that is defined a priori. For example, a significance level of 0.01 could be used rather than the usual 0.05 significance level. A very limited number of pre-specified analyses would generally not require multiple comparison corrections.
3. a Bonferroni-type correction could be applied in which the significance level is adjusted for the number of comparisons.

References

Prince, Martin J et al. (2015). The burden of disease in older people and implications for health policy and practice. **The Lancet**, 385(9967): 549 – 562.