Official Title: Cochlear - Implanted Recipient Observational Study Cochlear-IROS

NCT Number: NCT02004353

Document Date: 05 November 2019

Statistical Analysis Plan

Clinical Investigation Title: Publications from IROS database

Clinical Investigation Short Title: IROS Papers

Clinical Investigation Number: NA

Related CIP: CEL5277, V7

Clinical Project Manager (CPM):

Statistical Analysis Plan author/s:

SAP Version: 1.0

Contents

INTRODUCTION

This document is a companion document to the Clinical Investigation Plan CEL5277 but does not replace any previous statistical analysis plan: It is to describe the intended method of analysis of data in the adult IROS database with the aim to produce four publications.

It includes a comprehensive description of the recorded sample sizes, the intended statistical analyses with reference to the primary and secondary hypotheses, and additional statistical considerations such as the treatment of missing data.

There is no need to report deviations from this statistical analysis plan in any reports relating from the IROS data source.

The planned analysis for each publication is described in separate sections.

1 REAL LIFE OUTCOMES VERSUS AGE FOR CI

1.1 Overall objectives

- Clinical Evidence Strategy: Healthy Aging
- Pivotal Hypothesis: Are there differences in outcomes with CI for different age ranges, and impact on subject's life
- Primary measures: HUI3, SSQ (complete scales)
- Secondary measures: Changes in HUI3 >0.03, changes in SSQ >2 (i.e. dichotomisation of outcomes), changes in employment status pre and post (1 year, and sustained at 2 years), Medical leave after surgery, employment and other specific questions and listening effort (e.g. from SSQ).
- Age groupings for comparisons:
  - Use ABS-type age groupings 18-34, 35-44, 45-54, 55-64, 65+
  - Note that the younger age groups were combined due to sample size.
- Descriptive statistics of sub populations/ predictors (to be compared: pre and post, pre vs post, by age group, and as potential predictors in regression analyses)
  - Onset of deafness
  - Etiology
  - Hearing loss degree
  - HA use/presence (in the implanted ear at baseline and contralateral ear at baseline and visit 1)
  - Telephone use
  - Presence of tinnitus
  - Presence of dizziness
  - Employment status
  - Whether hearing ability negatively affected work
  - Has ability to do job changed post implant
  - Medical leave (taken and amount)
- Considerations:
  - Bilateral implant summaries were also of interest but were not possible due to small numbers.
  - Some of the above summaries may not be possible due to (subgroup) sample sizes.
  - The questions around employment and leave may not have been captured in a usable form, as such these variables will be explored more carefully, and analyses should be interpreted cautiously.

1.2 Study population

All available data will be used for longitudinal and pairwise comparisons.

1.3 Statistics

1.3.1 Sample Size

The minimum clinically important change/difference in HUI3 is 0.03 units and in SSQ the minimum clinically important change/difference is 2 units. For HUI3 this change is very small and larger differences are typically observed. As such for these analyses we are interested in differences/changes of at least 0.1 units.

Calculating sample sizes for longitudinal designs is complicated and requires more assumptions that are difficult to ascertain than simpler designs. For these sample size calculations, we use a simple approach understanding that longitudinal designs are often more efficient and hence requiring smaller sample sizes than those we might estimate below.

Estimating the SD

For the sample size calculations, we require estimated standard deviations (SD) which are taken as follows:

HUI3: From Figure 9E pp326 of the UK Cochlear Implant Study Group (UKCISG) paper in Ear and Hearing (2004, Vol 25, Issue 4, pp310-315) we can calculate the SD for each age group from $SD = \sqrt{N} * \frac{(\text{upper CI} - \text{lower CI})}{3.92}$ and then pool all of the SDs:

Table 1: Estimates from Figure 9E of UKCISG paper

Age group	20	30	40	50	60	70
N	39	35	59	75	70	33
Mean	0.24	0.22	0.21	0.16	0.20	0.18
SE=(upper-lower)/3.92	0.031	0.036	0.026	0.026	0.026	0.046
SD=SE*SQRT(N)	0.191	0.211	0.196	0.221	0.213	0.264
P-value for change in each age group	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Then the pooled standard deviation for the difference between groups is equal to **0.22**. The mean change varies between 0.16 and 0.24 with smaller mean changes observed in older groups. There is a significant change (paired t-test) over time in each age group.

From Cochlear data HUI3 and SSQ category data are available pre and post-surgery for a sample of subjects though not by age group. Summaries are as follows:

Table 2: Estimates of the mean and standard deviation from a sample of participants

Outcome measure	HUI3 pre	HUI3 post	Diff (Paired)	Spatial pre	Spatial post	Spatial diff
N	29	38	27	47	49	38
Mean	0.53	0.65	0.09	3.55	4.60	1.18
SD	0.188	0.168	0.182	1.907	2.104	1.356
P-value			0.016			<0.001
Outcome measure	Speech pre	Speech post	Speech diff	Qual pre	Qual post	Qual diff
N	49	51	40	52	50	40
Mean	4.53	5.62	1.18	2.85	4.65	1.80
SD	1.707	1.685	1.919	1.355	1.726	1.289
			<0.001			<0.001

For SSQ the SD for change varies between 1.3 and 1.9. A conservative choice for the SD might then be 1.9, as per the speech subscale. Note that there was a significant change from ore to post in all outcomes.

HUI sample size calculations:

Interest is in the difference in change between the 5 age groups of interest. We can look for differences in change between age groups using a one-way analysis of variance where the omnibus test will show if there is a significant difference between any pair of groups.

We use a 5% significance level, power of 80%, and assume a SD=0.22 (from UKCISG) within each group and use the data from UKCISG change in means for each age group (where 0.23 is used for the youngest age group). So, the biggest difference between any pair of groups here is 0.07 units. With these specifications, a sample size of 995 would be required to find a significant difference between at least one pair of age groups.

If the largest difference between the oldest and youngest group was 0.1 with 0.025 difference between all adjacent groups, then a sample size of 470 would be needed to find a significant difference between at least one pair of age groups.

SSQ sample size calculations:

We have no data on the mean and SD for each age group but, using a 5% significance level and power of 80%, and assuming a SD=1.9 if the difference between the oldest and youngest age group is 2 units and there is a 0.5 difference between adjacent groups (so that the youngest group has the MCID in change but the oldest group makes no gain [note that this is not realistic since we know all ages gain]) then a sample size of 95 would be needed to find a significant difference between at least one pair of age groups.

The true difference between age groups is likely to be much smaller. Suppose the youngest age group gains 3 SSQ units for the speech subscale but the oldest age group gains only 2

units. Assume adjacent groups differ by 0.25. Then a sample of 350 participants would be needed to find a significant difference between at least one pair of age groups.

Repeated measures methods, which will be used to analyse the data are more efficient and so it is likely that smaller sample sizes will be needed than stated in the above calculations.

1.3.2 Analyses

1.3.2.1 Pass/Fail Criteria

Not applicable.

1.3.2.2 Primary Hypothesis and Analysis

The effectiveness of cochlear implantation may vary by adult age group. Effectiveness is defined by the gain in total SSQ or HUI3 scores. As the outcome is likely neutral with respect to implications the hypotheses are presented to look for significant differences rather than to prove equivalence: We are neither worried that older subjects receive less benefit than younger patients, or vice versa since overall benefit is already know for these. However, differences in the benefits may imply revision of rehabilitation and/or counselling for prospective candidates. Thus:

H_0: There is no interaction effect between age group and visit on SSQ/HUI3 scores.

H1: There is an interaction effect between age group and visit on SSQ/HUI3 scores.

The proposed method of analysis is to use linear mixed-effects models (LMM). This allows determination of main (fixed) effects (visit, age group) and interaction effects, such as to answer the main objective while utilising all available data and controlling for repeated measures on the same individuals via a random effect for individual.

Tukey pairwise differences will be used to examine differences between age groups (if there is no evidence of an interaction effect) after adjusting for time if there is a significant main effect of age group. If an interaction effect is evident then differences between age groups at each time point can be estimated while controlling appropriately for pairwise testing.

The LMM approach allows the possibility of observing the influence of any continuous or categorical covariates of interest.

1.3.2.3 Secondary Hypotheses

Secondary objectives involve the dichotomisation of HUI and SSQ changes from pre to post where the threshold used is an increase at or above the minimum clinically important difference (>0.03 for HUI and >2 for SSQ). This can be done for baseline to first visit and baseline to second visit (probably not beyond that because of lack of follow-up data).

Bivariate responses will be modelled using mixed-effects logistic regression. Specifically:

H_0: There is no interaction effect between age group and visit on the odds of achieving a clinically significant gain in SSQ/HUI3 score.

H_1: There is an interaction effect between age group and visit on the odds of achieving a clinically significant gain in SSQ/HUI3 score.

Additional covariates can be included in these models to determine their effects (such as etiology, tinnitus etc).

Alternatively, standard logistic regression models can be used to determine the presence of MCID changes between a pair of time points only (for example baseline to visit 1 or baseline to visit 2).

1.3.3 Analysis Datasets

1.3.3.1 Intent-to-Treat

Linear mixed-effects models provide unbiased estimates of effects even in the presence of missing data provided that the data may be assumed to be missing at random. All available data (to the last time point of interest – possibly 2 or 3 years due to limited follow-up beyond that time). The data will be examined for evidence of informed missingness.

1.3.3.2 Per Protocol dataset

These are observational, retrospective data and therefore per protocol is not well defined. As above, the data will be examined to determine whether characteristics are different between those who are present or absent at the various follow-up time points.

1.3.4 Additional Statistical Considerations

1.3.4.1 Missing, Unused or Spurious Data

Differences in the characteristics of populations with follow-up versus those without should be described. The effect of missing descriptive data on the population characteristics should be estimated.

1.3.4.2 Planned Interim Analysis

Not applicable.

1.3.4.3 Criteria for Termination of the Clinical Investigation

Not applicable.

1.3.4.4 Additional Statistical Analyses

Additional analyses (using some or all variables described in section 1.1) will include:

- McNemar's test for a change in proportions over time in paired categorical variables.
- Fisher's exact tests will be used to look for associations between age group and other categorical variables.

1.3.5 Conduct of Statistical Analysis

[redacted] will conduct the analysis. [redacted] will advise and discuss the approaches and provide domain knowledge.

R software will be used to analyse the data (versions of base and extension packages will be noted).

1.3.6 Quality control on statistical analysis

[redacted] will conduct quality control on the R code used, and the summarised data.

1.3.7 Presentation of data

The report will conform to a conventional format with a number of data completeness and population summaries (such as demographics).

For the age groupings, ABS age ranges will be utilised unless sample sizes are too small. In this situation, one or more groups will be combined. While quintiles were of interest the age breakpoints were not useful and not considered further.

For all linear and linear mixed-effects models, estimates of the marginal means and plots of change over time will be presented. P-values for interaction effects and differences between groups (where relevant because of significance or interaction or main effects) will be presented. For logistic models estimated odds ratios and p-values will be presented.

Tables of summaries at each time point will be presented for the overall cohort as well as for those present or absent at a subsequent time point. Tables of cross-tabulations will also be presented.

2 DEVICE CHOICE BY AUDIOLOGICAL PROFILE

2.1 Overall objectives

The aim is a description of the characteristics of patients using different implantable hearing solutions provided by Cochlear. Primarily we wish to compare audiological configurations across device types, and within types if feasible numbers. Measures are:

- Audiometry air-conduction thresholds
- Audiometric bone-conduction thresholds
- Degree of hearing loss
- Etiology
- Use of hearing aids
- Duration of HL
- Age

Air (AC) and bone conduction (BC) thresholds will be used to indicate “sensorineural” (SN) and “conductive” (CO) components of HL. BC thresholds indicate the sensorineural HL component. The air-bone gap or difference AC-BC indicate the conductive component.

Etiology should correspond to particular SN and CO mappings. Similarly, the technical specifications of each device type can be mapped to SN and CO values based on transducer power and other factors (see Hypothesis).

The secondary objective is to characterise the baseline SSQ and HUI3 total scores across device types. In addition, telephone use, presence of tinnitus, and dizziness may be different across device types.

2.2 Study population

All available baseline data will be used. Firstly, the devices will to be combined into device types as below. The larger **overall** groups are defined as below.

BAHA

Baha Connect: BI210, BI300, BI<u>A</u>400

Baha Attract : BI<u>M</u>400

AMEI

Codacs : Codacs (DI110)

MEI : Carina (MET7000), Met (MET1000)

CI

Contour : CI512, CI513, CI612, Freedom CI24RE(CA) (CI24RECA), 24 Contour (CI24RC), 24 Contour Advance (CI24RCA)

Straight : CI422, CI522, CI622

Modiolar : CI532, CI632

Other : CI24RE(ST) (CI24REST), CI24RST (CI24RST), CI24REH (Hybrid L), CI other

ABI

ABI : ABI541

Bilaterally implanted patients can be simply divided into CI-CI and Baha-Baha, and any other combination for numbers. If sufficient (i.e. N>30), CI-CI can be treated as a separate group, etc.

2.3 Statistics

2.3.1 Sample Size

There are a total of 1361 baseline records available in the IROS database, with 1192 with some exploitable threshold data (BC value or HL degree category missing for 50 Baha and 115 CI). There are approximately 247 of the total set with BC data, of which 84 are CI. Thus, most of the CI group have HL degree but no specific AC or BC data. The exploitable AC and BC data will be used to provide individual data points. The HL degree data for CI can be summarised in terms of a range of values.

It is known that BC thresholds above ~80 dB HL cannot be measured due to transducer saturation. When this occurs often an AC threshold alone is given, but these are generally of high values.

2.3.2 Analyses

The primary aim is a description of the characteristics. No formal hypothesis testing will be performed.

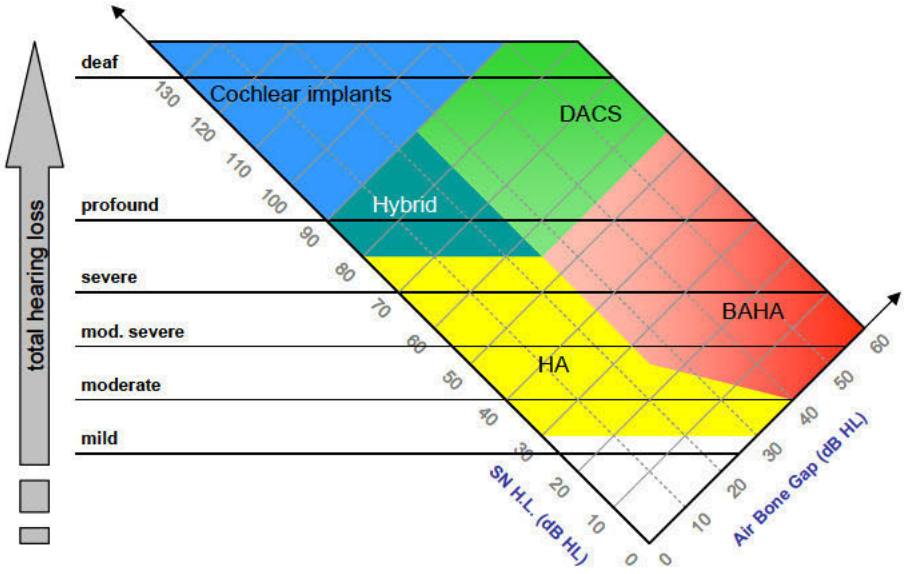
2.3.2.1 Pass/Fail Criteria

NA.

2.3.2.2 Primary Hypothesis

A schema originally produced by [redacted] serves as a reference for the design specifications of different Cochlear hearing solutions in terms of what level of SN and CO HL components may be addressed (Figure below).

We hypothesise that the SN and CO values in the IROS database should match those for the device labelled in the figure. We do not have a good handle on Hybrid HL in IROS database at present. We see that a mapping area for conventional acoustic hearing aids (HA) is indicated (yellow). Generally speaking, HAs can provide up to 60 dB corrective gain. Approximately half of the SN HL needs to be corrected and all of the CO HL needs to be corrected. Acoustic implants such as Carina and Codacs (DACS) bypass much of the the CO HL, as do Baha devices. However, Baha devices are more limited in corrective gain for SN HL (~30 dB).



To answer the hypothesis a new device type mapping will be created from current fitting ranges. Then the ipsilateral SN and CO values from IROS will be plotted onto the schema with specific point and colour types each device type (or overall device type).

2.3.2.3 Secondary Hypotheses

Summarise of the characteristics of patients using each device type will be made. Such as the etiology and demographics, and also baseline SSQ and HUI3 scores.

3 Implantable Hearing Solution Utility Gain

Not currently applicable.

4 Factors Influencing QoL Outcomes with CI

To differentiate from previous publication the approach will be to establish odds of improvement in HUI Mk III scores. The minimum clinically important change/difference in HUI3 is 0.03 units, but an alternative higher cut-off may be used (or in addition, see below). The aim is to give recommendations on the characteristic/s which are most likely to provide significant health utility from a CI.

4.1 Statistics

4.1.1 Sample Size

The decision of success/fail in significant impact of CI on HUI3 scores requires paired pre/post data in order to test the difference scores. Currently one centre has good follow-up, ZA-01, and overall a high number of patients with 161 at baseline and 117 at 1 year as of December 2019. Thus approximately 110 datapoints are available.

4.1.2 Analyses

A logistic regression approach will be used to determine which factors may significantly influence gains in HUI3 scores.

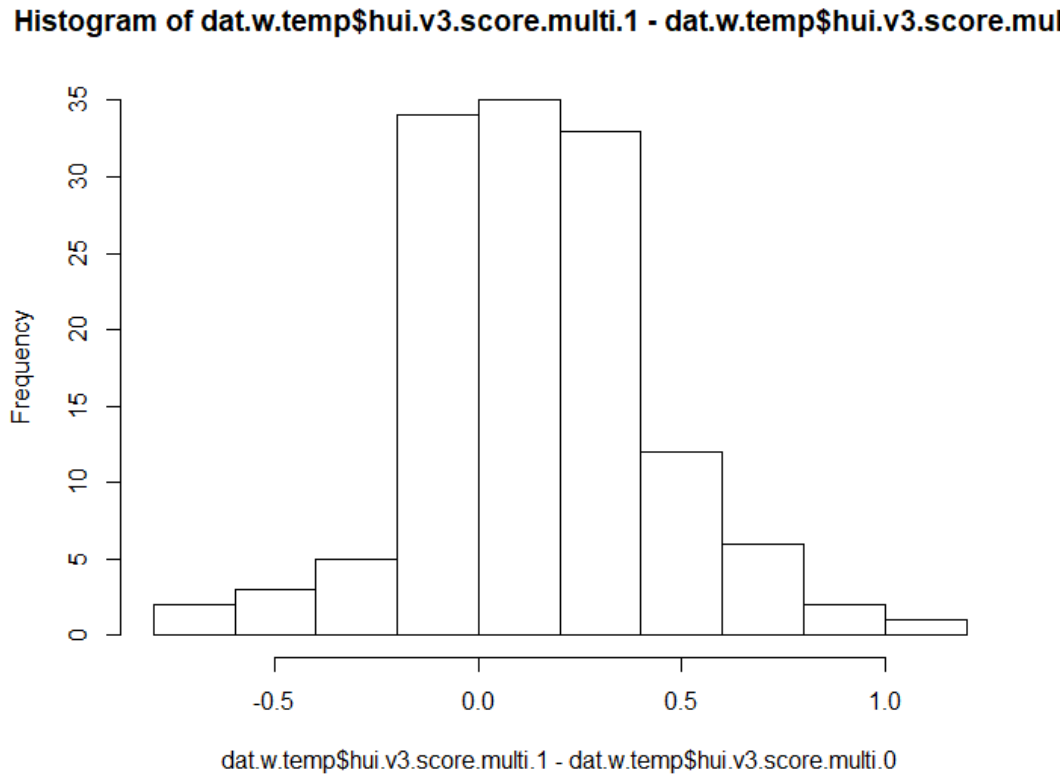
4.1.2.1 Pass/Fail Criteria

The minimum clinically important change/difference in HUI3 is 0.03 units. For HUI3 this change is very small and larger mean differences are typically observed. As such for these analyses we are interested in differences/changes of at least 0.1 units. Below is summarised the simple paired comparison of baseline versus 1-year HUI3 scores:

```
> t.test(dat.w.temp$hui.v3.score.multi.1,dat.w.temp$hui.v3.score.multi.0, p
aired=TRUE)

        Paired t-test

data:  dat.w.temp$hui.v3.score.multi.1 and dat.w.temp$hui.v3.score.multi.0
t = 5.9659, df = 132, p-value = 2.105e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.09963182 0.19847490
sample estimates:
mean of the differences
              0.1490534
```



The mean difference score is 0.149, with a confidence interval of ~0.1 to ~0.2. As stated above, 0.1 is generally considered to indicate a highly significant individual gain in HRQoL. Based on the histogram above, somewhat more than 50% of cases achieved a HUI3 gain of >0.1. An approximate cut-off representing 50% of cases is a good basis for a success/fail criterion for a logistic analysis (i.e. median gain in HUI3 for example). However, it may be useful to tune the criterion to improve the predictive performance of the model once initial results are obtained for gain = 0.1.

4.1.2.2 Primary Hypothesis and Analysis

Once the criterion for “significant gain” in HUI3 is defined, the baseline characteristics of those meeting and not meeting the criterion can be compared. Then it can be hypothesised that those characteristics which are divergent for the group may influence significant gain in HUI3 scores.

Typical factors to consider are etiology, age, duration of hearing loss/deafness, onset of hearing loss, use of hearing aids, tinnitus status, and dizziness status. There are too many etiology categories compared to the total N of subjects so they will be grouped as in Table 1.

For the effects of age, continuous variable or range groups may be evaluated (i.e. approximate ABS age as in Section 1 above will be used unless sample sizes are too small).

Onset of deafness in childhood, or “congenital” deafness may be calculated from age and duration of hearing loss. In consultation with the centre, “pre-linguistic” onset of deafness is

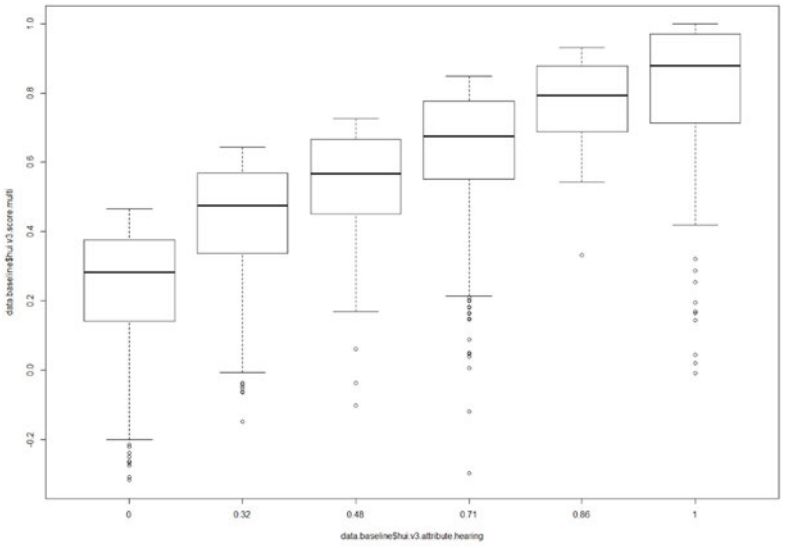
defined as <10 years of age. Arguably “congenital” deafness should be limited to ages <2 (for the purposes of identification). The other defined onsets are progressive and sudden HL.

Table 1. Etiologies (implanted ear) at baseline grouped for ZA-01.

Familial	7
Vestibular (Large vestibular aqueduct Menieres syndrome)	8
Meningitis	8
Viral, (measles and mumps rubella)	9
Ototoxic drugs	11
Middle ear: Cholestheatoma, Congenital atresia, Otosclerosis	15
Other	27
Unknown	75

4.1.2.3 Secondary Hypotheses

One factor which might influence HUI3 gain is the baseline HUI3 score. However, much of the HUI3 multi attribute score variation in this population is produced by the 6-point Hearing attribute (figure below – all CI subjects pre and post scores). This allows us to avoid comparing baseline and gain directly in HUI3 scores (partially correlated).



The hearing attribute can be divided into two broad classifications: ≤3 (0.71-1 in the figure), where all three indicate that the subject can understand what is said in a group conversation with at least three other people, with or without using a hearing aid, and >3 (0-0.48) , the converse, including not hearing at all either one-on-one or with three other people.

Table 2. Numbers of patients in HUI3 Hearing attribute categories 1-6; baseline rows, at visit 1 columns.

Value	***6***	***5***	***4***	***3***	***2***	***1***	***Status classification***
6	4	7	0	22	2	2	No hearing
5	4	9	1	31	1	1	Group No/with HA
4	0	0	0	2	0	0	Group No/with HA
3	2	3	0	32	1	2	Group OK/with HA
2	0	1	0	1	0	0	Group OK/with HA
1	0	0	1	3	0	1	Group OK/No HA

The baseline hearing attribute classification may be correlated with demographic characteristics and thus it is worthwhile to test this, as well as the influence of baseline hearing attribute on gain in multi-attribute score.

We see that 32 subjects are in category 3 both at baseline and at visit 1. The SSQ scores can be used to evaluate with more precision whether this group benefitted from cochlear implantation. Also the HUI3 multi-attribute score may indicate some change.

Thirdly, to look at the outcome questions in IROS such as whether an improvement in hearing was noted by patients.

4.1.3 Conduct of Statistical Analysis

[redacted] will conduct the analysis. [redacted] will advise and discuss the approaches and provide domain knowledge.

R software will be used to analyse the data (versions of base and extension packages will be noted).

4.1.4 Quality control on statistical analysis

[redacted] will conduct quality control on the R code used, and the summarised data.

4.1.5 Presentation of data

The report will conform to a conventional format with data completeness and population summaries (such as demographics).

Summary statistics will be provided for baseline characteristics for those patients reaching the 0.1 HUI3 gain at visit 1, and those who did not. Similarly, for those with baseline HUI hearing attribute categories ≤3 and those >3.

Univariate logistic models will be produced with estimated odds ratios and p-values will be presented. If feasible, multivariate logistic models will be produced. In addition, the predictive

capability of the model can be tested, and/or the threshold criterion varied(e.g. via RoC) to obtain higher predictive power.

5 HEARING PRESERVATION PER DEVICE TYPE

Not currently applicable.

6 REFERENCES

6.1 Internal References

ID	Document Title	Number
	CEL5277 Cochlear IROS Registry Plan Version 7	

6.2 External References

ID	Document Title	Number

7 CHANGE HISTORY

Version	Change	Author	Date
1	Original version	[redacted]	5 Nov 2019

8 DEFINITIONS

Term	Description
LMM	Linear mixed model, for repeated measure analyses
GLM	General linear model, for example employing a non-linear link function
GLMM	General linear mixed model, for repeated measure analyses
HUI	Health Utilities Index (in this case mark III, or HUI3 for short)
SSQ	Speech, spatial and qualities (score from questionnaire)