Official Title: **Exploration of the specificity of an algorithm to detect anomalous electrode position**

NCT Number: **NCT03294681**

Document Date: **29 March 2017**

Statistical Analysis Plan

Exploration of the specificity of an algorithm to detect anomalous electrode position

[redacted]
Date: 29 *March 2017*

Study title: ***Exploration of the specificity of an algorithm to detect anomalous electrode position***

Study number, version: *CLTD5676*

Author: [redacted]
Cochlear Limited
1 University Avenue
Macquarie University NSW 2109
AUSTRALIA

Clinical Project Manager (CPM): [redacted]
Cochlear Deutschland GmbH & Co. KG
Karl-Wiechert-Allee 76 A
30625 Hannover (Germany)
[redacted]
[redacted]

Table of Contents

1 Introduction

1.1 Overview

This is the statistical analysis plan for a multi-centre exploratory, prospective study collecting and analysing CT, NRT and trans-impedance and intra-cochlear voltage measurements.

1.2 Objectives

1.2.1 Primary objective

To explore the specificity (percentage of cases in which a true negative gives a negative result) of a candidate algorithm that has the ability to detect low incidence deviations from a normal electrode position using trans-impedance measurements.

1.2.2 Secondary objectives

1. To explore the association between the trans-impedance and intra-cochlear voltage measurements, NRT thresholds and the intra-cochlear electrode array position.
2. To investigate changes in the trans-impedance and intra-cochlear voltages along with NRT thresholds over time to identify the stability of the measurements.
3. To establish a database of trans-impedance and intra-cochlear voltages, NRT thresholds and CT images for educational, research and development purposes.

1.3 Endpoints

1.3.1 Primary endpoint

Trans-impedance and intra-cochlear voltage measurements for each electrode contact after electrode insertion with post-operative CT/DVT Scan.

1.3.2 Secondary endpoints

1. NRT thresholds for each electrode contact during surgery and pre-operative CT/DVT Scan co-registered with the post-operative CT/DVT Scan collected for the primary endpoint.
2. Trans-impedance and intra-cochlear voltage measurements and NRT thresholds for each electrode contact at first activation and 3 months post-operative
3. Validated database entry of trans-impedance and intra-cochlear voltage measurement and NRT thresholds during surgery, at first activation and three months post-operative as well as the pre-operative and post-operative CT/DVT Scans respectively.

1.4 Hypotheses

- H_0: The specificity of an algorithm that will detect low incidence deviations from a normal intra-cochlear electrode insertion will be < 98%.
- H_1: The specificity of an algorithm that will detect low incidence deviations from a normal intra-cochlear electrode insertion will be ≥ 98%.

1.5 CE marked devices

The CSEP software, CP900 series sound processor and CI532 and CI512 implants are all used routinely by cochlear implant and hearing care professionals.

2 Statistics

2.1 Overview

A total of 154 patients will be assessed in the statistical analysis of the data. This will serve to collect a representative sample of the variability of surgical techniques across clinics and surgeons. The confidence interval to be used in the analyses will be a 90% two-tailed. This confidence interval was deemed to be satisfactory due to the exploratory nature of the study. The statistical analyses conducted for the primary and secondary objectives will lead to the endpoints of the study.

The statistical analysis will be conducted by [redacted], Cochlear).

2.2 Sample size calculation

The goal of the candidate algorithm is to indicate to surgeons and medical staff whether an electrode tip fold-over or electrode outside cochlea appears to have occurred, which should be confirmed radiographically. A target specificity of 98% and confidence level of 90% have been selected.

The 98% specificity target initialises the calculation of the required sample size (N). Since the specificity of the algorithms can only be estimated from a sample population, there is required to be a reasonable expectation that the specificity estimate is equal to or greater than the specificity target. To ensure that the reasonable expectation is fulfilled, the 98% specificity target is set to be the lower bound of the 90% binomial confidence interval for an estimated or measured specificity of 100%. The selected sample size must fulfil this requirement. The binomial confidence interval will be calculated using the Clopper-Pearson exact method and is denoted by Equation 1 and Equation 2:

$$P_{LB} = p : \sum_{i=0}^{x} \binom{n}{i} p^i (1-p)^{n-i} \approx \frac{\alpha}{2}, \qquad for\ 0 \le p \le P \tag{1}$$

$$P_{UB} = p : \sum_{i=x}^{n} \binom{n}{i} p^i (1-p)^{n-i} \approx \frac{\alpha}{2}, \qquad for\ P \le p \le 1 \tag{2}$$

where P_{LB} is the lower bound of the binomial confidence interval, P_{UB} is the upper bound of the binomial confidence interval, P is the measured probability of success, p is the probability of success, n is the number of trials, i is the number of successes, x is the measured number of successes and $\alpha = 0.1$ is the level of significance for the 90% confidence interval.

In the Clopper-Pearson method, n is the unknown input in the equation. To find n requires taking into account that there may be occurrences of electrode anomalies in the sample. The range of the possible numbers of anomalies must be estimated such that there is an expected sufficient number of samples (n) to estimate specificity. The sufficient number of samples is calculated by Equation 3 and Equation 4:

$$n = N - \mathrm{m} \tag{3}$$

$$m = i : \sum_{i=0}^{N} \binom{N}{i} p^i (1-p)^{N-i} \approx 0.90, \qquad for\ 0 \leq i \leq N \tag{4}$$

where n is the number of sufficient samples required to estimate specificity, N is the total sample size or number of patients required, i is the number of anomalies and p is the probability ofan anomaly. We assume that the most prevalent anomaly condition will be a tip foldover. For *p* we use an estimate of 0.04139 or 4.14% which is based on the perimodiolar weighted average percentage between Grolman et al. (2009) and Zuniga et al. (2017). m is then approximately an upper bound on the number of expected tip fold-overs based on the 90th percentile of the cumulative binomial distribution. In other words, there is at least an approximate 90% chance that there will be m number of tip fold-overs or less.

Determining the sample size N is an iterative process. N is increased from 0 and for each value of N, n is calculated using Equations 3 and 4 which allows for P_{LB} to be calculated using Equation 1. Once P_{LB} is approximately 0.98 or 98%, the corresponding N is selected as the sample size.

Using the denoted method (see Appendix A) a sample size of $N = 154$ was selected which had a 89% chance of 9 or less tip fold-overs taking place, resulting in an $m = 9$ and $n = 145$. With a measured specificity (i.e. estimated specificity) of 100% and $n = 145$, the $P_{LB} = 0.9797$ which is approximately 98%.

With an algorithm target specificity of at least 98%, a target sensitivity of at least 90%, an estimated probability of a tip fold-over of 4.14%, the positive predictive value (PPV) is equal to 66.02%. This entails that after the algorithm has detected a tip fold-over, it is estimated that two thirds subset of patients will be confirmed by imaging to have a tip fold-over. The calculation of PPV is calculated by Equation 5:

$$PPV = \frac{(prevalence * sensitivity)}{(prevalence * sensitivity) + ((1 - prevalence) \times (1 - specificity))} \tag{5}$$

$$PPV = \frac{(0.0414 * 0.90)}{(0.0414 * 0.90) + ((1 - 0.0414) \times (1 - 0.98))} = 0.6602 \tag{6}$$

where in this application $prevalence$ is the estimated rate of tip fold-overs.

3 Conduction of statistical analysis

3.1 Primary objective

The primary objective of the study is to explore the specificity of an algorithm that detects low incidence anomalies such as tip fold-overs or an electrode migration. The target maximum specificity of the algorithm has been selected to be 98%. The following steps will be run to explore the specificity up to this upper bound:

1. Run the candidate algorithms on each of the patient's data
    - Patients with 'normal' insertions (89% chance of at least 145 patients)
2. Collect the output of the algorithm

- The outputs of the algorithms will be Boolean in nature; 0 represents no tip fold-over or electrode outside the cochlea and 1 represents an tip fold-over or electrode outside the cochlea

3. Count the number of patients that have been classified as not having an anomaly by the algorithm.
4. Conduct radiographic analysis of each insertion to classify as either normal or abnormal. Document in the latter case the reason for the classification in the anomaly category. .
5. Calculate the percentage of patients (P) that have been correctly classified as not having an anomaly
6. Calculate the lower bound of the 90% binomial confidence interval (P_{LB}) for the true success rate (specificity) of the algorithm using Equation 1

If the calculated lower bound of the 90% binomial confidence interval around p is equal to or greater than 98%, the candidate algorithms can be stated to have an expected specificity equivalent to or greater than 98% with 90% confidence. As explained in Section 4, the candidate will have reached the upper limit of specificity of 98% with 90% confidence if 100% of at least 145 normal insertions are correctly classified as normal.

3.2 Secondary objective 1

The first secondary objective of this study is to explore the association or relationships between the trans-impedance, intra-cochlear voltage measurements and NRT thresholds and the intra-cochlear electrode position. This will be conducted by:

1. Measuring the intra-cochlear electrode array position for 22 electrodes for each patient from CT images
2. Univariate and multivariate analysis
    a. Visualisation (descriptive statistics)
    b. Feature extraction
    c. Correlation and ANOVA analysis
    d. Statistical modelling

Univariate analysis will first involve visualising the data and comparing the visualisation between electrode positing. The visualisation can be conducted by using simple scatterplots with the independent variables, or subcomponents of the independent variables, on the x-axis and the dependent variable on the y-axis. With multidimensional measurements, plots such as mesh grids or surface plots will yield more value. This will aid the statistician to visually identify any patterns, trends or key features between the variables and note them for further investigation.

More "*complex*" data or measurements can be processed, transformed or reduced to a more meaningful single value or a set of values that represents the original data and may have explanatory power. This process is feature extraction. Feature extraction can be governed by the visualisation stage or engaging in a set of routines as calculating the determinant or Eigen values of multidimensional data, means, medians, variances, converting continuous

variables into interval variables etc. Key values extraction may also include filtering or smoothing the data.

Electrode position and associated values of independent variable (e.g. measurements and/or key features) and be composed into arrays. This allows for the Pearson's correlation statistics to be calculated. The correlation statistic indicates whether or not linear relationship between two variables and the sign of the relationship. The statistic ranges from -1 to 1. A value of -1 entails that the two variables are negatively linearly related. A value of 1 entails that the two variables are positively linearly related. A value of 0 entails that there is no linear relationship. The resultant statistics will be tested for statistical significance. If the relationship is not statistically significant with a 90% confidence interval or has a low correlation, it is very unlikely it will be a candidate to be a primary explanatory variable. The correlation statistic may be indicative of a higher order relationship.

Interval, nominal or ordinal variables will be analysed using parametric or non-parametric methods depending on the observed distribution of the experimental data. Electrode positioning data will be grouped according to their associated value of the interval, nominal or ordinal variables. The values of the nominal or ordinal variables will be considered as treatments. The treatment effect will be analysed for magnitude and statistical significance. Distributions will be tested for normality using the Shapiro-Wilk test. Parametric ANOVA may include the use of Welch's t-test. Non-parametric tests may include Mann-Whitney U test or the Kruskal-Wallis test.

The previous steps will have identified key variables to be included in statistical models that describe relationships between measurements and electrode positioning. These key variables may be related to one another. Visualisation and correlation analysis will aid to identify variables that are related to one another. Variables that are identified to have possible relationships will be controlled for when statistically modelling.

Statistical modelling will involve determining a relationship the independent variables that result in a best fit or least error. The visualisation phase, key feature extraction and correlation analysis will define the variables that should be included in the model and what is the relationship between the variables. Candidate models will include regression, logistics regression or machine learning techniques such as feed-forward back-propagation neural networks. Models will be developed in a stepwise approach. Variable inclusion will proceed with the variables identified to be the most significant to the least significant. For regression analysis, for each step statistical tests will be performed such as f-test and t-tests on regression coefficients. If the output of the test does not indicate that the coefficients are statistically significant, the variables will be discounted. Accuracy statistics such as R^2, correlation and root mean square error will be calculated. If the accuracy does not increase, the variables will be discounted.

The end results will be statistical models that explain the relationship between intra-cochlear electrode positions and trans-impedance, intra-cochlear voltage and NRT thresholds. This information will substantially influence the development of candidate algorithms.

3.3 Secondary objective 2

The second secondary object of this study is to investigate the changes in the trans-impedance, intra-cochlear voltage and the NRT thresholds over time to identify the stability of the measurements. The contact points for the patients in the study are surgery, activation (4

weeks after surgery) and 3 months post-operative. Investigating how the measurements change throughout time will involve:

- Excluding patients that have been confirmed (by imaging) to have an anomalous position.
- Characterising the measurements at each contact point
- Charactering the changes in measurements between the contact points

Characterising the measurements at each contact point involves:

- Visualising the distributions (e.g. scatter plots, frequency histograms, boxplots, etc.)
- Calculating the distribution statistics (e.g. normality, mean, median, standard deviation, confidence intervals around mean and medians)
- Parametric or non-parametric ANOVA between the distributions of each contact point

Charactering the changes in measurements between the contact points involves:

- For each patient calculating the difference between the measurements at each time point to create difference distributions
- Visualising the difference distributions (e.g. scatter plots, frequency histograms, boxplots, etc.)
- Calculating the distribution statistics (e.g. normality, mean, median, standard deviation, confidence intervals around mean and medians)

A parametric or non-parametric (Friedman) ANOVA will be conducted on the difference distributions of the periods between each contact point to establish the significance of the time factor.

This information will provide guidance on the time from surgery during which the findings of candidate algorithms can be considered valid.

3.4 Secondary objective 3

To ensure the correct handling and storage of patient data for the study, study measurements and metadata will be uploaded to a clinical study database. The database will be structure accordingly:

- A section storing patient information such as:
    - Identifier
    - Implant
    - Demographics
    - Medical history
    - Adverse events
    - Medical professionals' notes and comments
    - Surgical variables
    - Nature of insertion

- A section storing CT images, the operating medical professionals notes and electrode positioning
- A section storing trans-impedance measurements, the measurement parameters and the point in time the measurement was taken
- A section storing intra-cochlear voltage measurements, the measurement parameters and the point in time the measurement was taken
- A section storing NRT measurements, the measurement parameters and the point in time the measurement was taken
- A section storing Cochlear Implant Fitting Software CDX files

An independent researcher (i.e. not the person who developed the database) will determine whether the information has been correctly uploaded to the database, the integrity of the data in the database and whether patient information has been sufficiently de-identified.

4 Quality control of statistical analysis

The statistical analysis plan will be undertaken primarily using the Python programming language with the Numpy and Scipy libraries. Other programming languages or software packages such as Matlab, SAS or Excel may also be used. The analysis plan will be:

1. Statistical analysis will take place once all of the data has been collected
2. Programmed to retrieve the objective outputs
3. The code and the outputs will be reviewed to determine whether or not what was stipulated in the objectives have been met and whether or not the outputs are in line with what domain experts' expectations

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

If a patient has missing or spurious data, the data that is remaining of quality will be used in the clinical investigation. Patients that drop-out will cause only the data that has been collected to be analysed.

5 Derived data

All derived data as described in the conduction of the statistical analysis will be described in full detail with the steps taken to calculate or produce the derived data.

6 Analysis populations

All patients will be included in the analysis unless a patient has been confirmed to have an anomalous position.

7 Bibliography

[1] Zuniga, MG, Rivas, A, Hedley-Williams, A, Gifford, RH, Dwyer, R, Dawant, BM, Sunderhaus, LW, Hovis, KL, Wanna, GB, Noble, JH, & Labadie, RF 2017, 'Tip Fold-over in Cochlear Implantation: Case Series', Otology & Neurotology: Official

Publication Of The American Otological Society, American Neurotology Society [And] European Academy Of Otology And Neurotology, vol. 38, no. 2, pp. 199-206. Available from: 10.1097/MAO.0000000000001283. [15 March 2017]

[2] Grolman, W., Maat, A., Verdam, F., Simis, Y., Carelsen, B., Freling, N., & Tange, R. A. (2009). Spread of excitation measurements for the detection of electrode array foldovers: a prospective study comparing 3-dimensional rotational x-ray and intraoperative spread of excitation measurements. Otology & Neurotology, 30(1), 27-33.

Appendix A

```
# -*- coding: utf-8 -*-
"""
Created on Fri May 13 14:21:16 2016

[redacted]
"""



# Script was run using Ananconda Scientific Python 2.7 64 bit
# https://www.continuum.io/downloads

# This script is used to:
# i) Calculate the perimodiolar weighted average tip fold-over rate between
# Zuniga et al. (2017) and Grolman et al. (2009).

# ii) For a given sample size, estimate the number of tip fold-overs for a
# binomial cumulative percentage equivalent to 90%.

# iii) With a given sample size minus the number of tip-fold-overs estimated,
# estimated the lower bound of the 90% binomial confidence interval when the
# the probability is equal to 100%. This lower bound value must be equivalent
# to the target specificity of 98%.

# iv) Calculate the postivite predivtive value with a target specificity of 98%
# and target sensitivity of 90%.

# Notes:
# Assuming a 90% confidence interval
# copy.deepcopy() is used to mitigate Python's memory management protocols.
```

```
# Import anaconda/python libraries
import numpy as np
import scipy.stats as sp
import copy
```

```
# Defining functions
# Clopper-Pearson exact binomial confidence interval
# Clopper, C. J., & Pearson, E. S. (1934). The use of confidence or fiducial
# limits illustrated in the case of the binomial. Biometrika, 26(4), 404-413.
# Method detailed on page 407 of Clopper & Peason (1934).
def two_tailed_exact_binomial_CI(p,n,x,alpha,fidelity):
    # p: recorded probability of success
    # n: number of trials performed
    # x: recorded number of successes
    # alpha: confidence interval
    # fidelity: number of probability steps (step size indication) for runs 0
    # to p (lower bound) and from p to 1 (upper bound)


    # Two-tailed
    alpha_2 = alpha/float(2)


    # Finding upper bound probability
    p_upper = 1
```

```
    for k in np.linspace(p,1,fidelity):
        # Resetting summate
        summate = 0

        # Select probability
        input_prob = copy.deepcopy(k)

        # Calculating sum loop
        for m in np.arange(x+1):
            summate = summate + sp.binom.pmf(m, n, input_prob)

        # Determining
        if summate<(alpha_2+0.0020) and summate>(alpha_2-0.0020):
            p_upper = input_prob


    # Finding lower boung probability
    p_lower = 0
    for k in np.linspace(0,p,fidelity):
        # Resetting summate
        summate = 0

        # Select probability
        input_prob = copy.deepcopy(k)

        # Calculating sum loop
        for m in np.arange(x,n+1):
            summate = summate + sp.binom.pmf(m, n, input_prob)

        # Determining
        if summate<(alpha_2+0.0020) and summate>(alpha_2-0.0020):
            p_lower = input_prob


    # Returning
    return [p_lower,p_upper]
```

```
# Main program
if __name__ == '__main__':
    # Selecting sample size for analysis
    n = 154


    # PART (i)
    # Calculating weighted average between the two papers
    # Zuniga et al. (2017) - 5 tip fold-overs out of 0.48*303 perimodiolar
    # electrode arrays
    p_z = 5 / (0.48 * 303)
    # Grolman et al. (2009) - 4 tip fold-overs out of 72 perimodiolar electrode
    # arrays
    p_g = 4 / float(72)
    # Weighted tip fold-over rate
    p_w = ((0.48 * 303)*p_z + (72)*p_g)/float((0.48 * 303)+72)


    # Prior rate of tip fold-over - Using the weighted tip fold-over rate
    p_ETF = copy.deepcopy(p_w)


    # PART (ii)
    # Calculating the binomial probability mass distribution up to and including
    # 15 tip fold-overs
    max_ETF = 15
    record_n_ETF = np.zeros(max_ETF+1)
    record_p_n_ETF = np.zeros(max_ETF+1)
    for k in np.arange(0,max_ETF+1):
        # Number of tip foldovers
        n_ETF = k
```

```
        # Estimating probability of n_ETF
        p_n_ETF = sp.binom.pmf(n_ETF, n, p_ETF)

        # Updating
        record_n_ETF[k] = n_ETF
        record_p_n_ETF[k] = p_n_ETF


    # Integrating the binomial probability mass function and fiding the number
    # of tip fold-overs at a cumulative probability equivalent to 90%.
    # Expected number of tip fold-overs
    expectation = np.ceil(n*p_ETF)
    # Variables to contain results
    max_n_ETF_single = 0
    max_n_ETF_single_cumulative = 0
    # Variable to contain integral
    summation = 0
    # Integrating
    for k in np.arange(len(record_n_ETF)):
        # Recording cumulative percentage
        summation = summation + record_p_n_ETF[k]
        # Determining if the cumulative percentage (summation) is equivalent to
        # 90%.
        if      summation<=(0.90+0.03)      and      summation>=(0.90-0.03)      and
record_n_ETF[k]>=expectation:
            # Recording number of tip fold-overs
            max_n_ETF_single = record_n_ETF[k]
            max_n_ETF_single_cumulative = copy.deepcopy(summation)

    # Finding new sample size
    n_new = n - max_n_ETF_single
    print('Cumulative percentage')
    print(max_n_ETF_single_cumulative)
    print('Resultant sample size')
    print(n_new)


    # PART (iii)
```

```
    # Calculating confindence interval for true negative rate of 100% of
hypothetical
    # algorithm given n_new using the Clopper-Pearson exact method.
    # Probability
    p = 1
    # Recorded number of successes
    x = copy.deepcopy(n_new)
    # 1 - confidence interval
    alpha = 0.10
    # Search fidelity - increases the number of probabilities that are analysed
    fidelity = 8000
    # Clopper-Pearson function
    output = two_tailed_exact_binomial_CI(p,n_new,x,alpha,fidelity)
    p_lower = output[0]
    p_upper = output[1]
    del(output)
    print('Lower bound of binomial confidence interval')
    print(p_lower)



    # PART (iv)
    # Positive predictive value
    # PPV = (prevalence * sensitivity) / ((prevalence*sensitivity)+((1-
specificity)*(1-prevalence)))
    sensitivity = 0.90
    specificity = 0.98
    prevalence = p_w
    PPV = (prevalence*sensitivity) / ((prevalence*sensitivity)+((1-
specificity)*(1-prevalence)))
    print('Positive predictive value')
    print(PPV)
```