

A comparative, controlled, clinical investigation and quality control of a new hearing aid in comparison to the currently marketed device.

Statistical Analysis

Document dated: 2016.11.14

Statistical Plan for the Non-Inferiority Analysis

Christophe Lesimple

14 November 2016

Contents

1. Introduction.....	1
2. Benefit of amplification measured with speech reception threshold	1
3 Non-Inferiority test design.....	4
4. Required sample size.....	6
5.References.....	7

1. Introduction

The goal of the trial is to show that the investigational device performs at least as well as the current marketed device. The platform that supports the investigational device shows promising capabilities in terms of processing power and wireless opportunities. The development of this first generation has the intend to port the main functionalities (amplification, feedback canceler, directional microphones and noise reduction) from the current marketed device on the new platform. For this reason, it is not expected to find that the investigational device outperforms the current marketed device.

A non-inferiority trial aims to demonstrate that the investigational device is at least as good as the current marketed device by more than a pre-specified, small amount. This amount, known as the non-inferiority margin M , has to be determined before the trial start on the main outcome based on clinical and statistical reasoning.

The determination of the non-inferiority margin M has to fulfill following points (D'Agostino et al., 2003):

1. we must have assurance that the current marketed device (Control C) would have been superior to the unaided condition,
2. the non-inferiority active-controlled trial should demonstrate that the investigational device (new treatment T) is within the non-inferiority margin M of the active control C,
3. it is recommended to establish that the investigational device also retains at least a certain amount of the superiority of the active control over the unaided condition.

This statistical plan contains: a meta-analysis that will demonstrate the reference effect of amplification on speech reception threshold, a rationale to compute the non-inferiority margin, a rationale to compute the required sample size.

2. Benefit of amplification measured with speech reception threshold

2.1 Reference study search and summary

The following references give an overview about similar reported trials in order to estimate the benefit of amplification (unaided-aided) for an adaptive speech in noise test. Benefit of amplification with commercially available devices on the 50% speech reception threshold are described regarding test set-up differences:

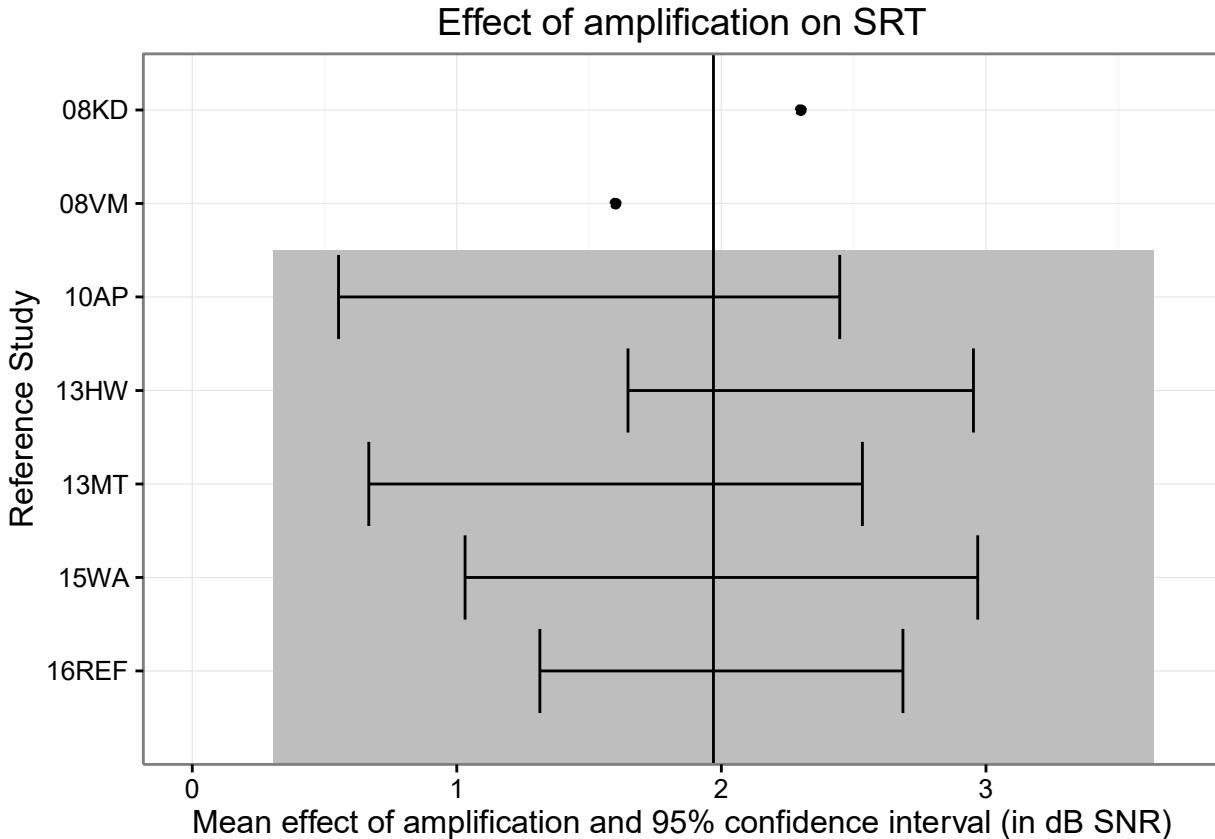
1. 08VM from **Valente & Mispagel (2008)** mean difference of 1.6 dB SNR (sd = na) on 26 subjects with a mild to moderate-severe hearing loss (mean age = 65.6) . Open fitted hearing aids from Vivatone. Speech (HINT) at 65 dB and 8 noise sources.
2. 08KD from **Klemp & Dhar (2008)** mean difference of 2.3 dB SNR (sd = na) on 16 subjects with a

mild to moderate hearing loss (age 50-85). Open fitted hearing aids from Phonak and Widex. Speech at 65 dB A (HINT) and 3 noise sources.

3. 10AP from **Alworth & Plyler (2010)** mean difference of 1.5 dB SNR (sd = 2.3) on 25 subjects with a mild to moderate hearing loss (mean age = 67). Open fitted hearing aids from Bernafon with NL1. Speech at fixed level (HINT) and 1 noise source from the front.
4. 13MT from **Magnusson et al. (2013)** mean difference 1.6 dB SNR (sd = 2.0) on 20 subjects with a mild hearing loss (age 51-64). Open fit with hearing aids from Phonak (own fitting rationale). Speech at 65 dB SPL (Swedish Hagermann) and 4 noise sources.
5. 13HW from **Helbling et al. (2013)** mean difference 2.3 dB SNR (sd = 1.4) on 20 subjects with a moderate hearing loss (age 42-77). Open fit with hearing aids from Sonic Innovation (own fitting rationale). Noise at 65 dB SPL (OLSA) and 1 noise source from the front.
6. 15WA from **Woods et al. (2015)** mean difference 2 dB SNR (sd = 2.3) on 24 subjects with a mild to moderate-severe hearing loss (age 61-81). Open fit with hearing aids from Phonak and GN. Speech at 70 dB SPL (HINT) and ? noise sources.
7. 16REF from unpublished **internal test (2016)** mean difference 2 dB SNR (sd = 2.0) on 35 subjects with a mild to moderate-severe hearing loss (age 37-87). Fitting with domes and custom molds from Bernafon Reference hearing aid (Juna 9). Speech at 65 dB SPL (OLSA) and 4 noise sources.

study	spatial.noise	speech.mat	platform	main.effect	standev	sample.size	ci.low	ci.high
08VM	separated	other	different	1.6	NA	26	NA	NA
08KD	separated	other	different	2.3	NA	16	NA	NA
10AP	colocated	other	same	1.5	2.3	25	0.55	2.4
13MT	separated	other	different	1.6	2.0	20	0.67	2.5
13HW	colocated	olsa	same	2.3	1.4	20	1.65	3.0
15WA	colocated	other	different	2.0	2.3	24	1.03	3.0
16REF	separated	olsa	same	2.0	2.0	35	1.31	2.7

The last reference (16REF) comes from internal test results with the reference marketed device in the same test setup that is planned to be used for the current investigation. The summary of all these studies can be shown in the following graph:



Studies 08KD and 08VM are shown even no confidence interval could be computed based on the reported results. The summary estimation will therefore exclude these two studies.

2.2 Estimating summary effect

The random-effects model is generally preferred because most meta-analyses include studies that are not identical in their methods and/or sample characteristics. Differences in methods and sample characteristics between studies will likely introduce variability among the true effect. The random-effect model considers the studies included in the analysis as a sample from a larger universe of studies that could be conducted. The results from random-effects analyses are generalizable beyond the included set of studies and can be used to infer what would likely happen if a new study were conducted.

The estimation of the main effect was conducted with the **MAd** and **metafor** packages developed for R (Del Re, 2015) on effect size:

```
## Table with effect sizes
# 1. Compute effect size Cohen's "d" (1988)
d.effect.size <- reviewSummary$main.effect/reviewSummary$standev
# 2. Compute Hedges'g from before and J as corrector factor
J <- 1-3/(4*(sample.size-1)-1)
g.effect.size <- round(J*d.effect.size, 3)
# 3. Estimation of the variance of Hedge'g https://www.ncbi.nlm.nih.gov/books/NBK140571/
n <- reviewSummary$sample.size
g.var.estimate <- round(((1/n)*((n-1)/(n-3))*(1+n*g.effect.size^2) - g.effect.size^2)/
                        (J*(n-1))^2, 3)
# 4. Table review.es
```

```
review.es <- data.frame(study, spatial.noise, speech.mat, platform,
                        g.effect.size, g.var.estimate)
```

```
library("knitr")
```

```
kable(tail(review.es, 5), format = "markdown")
```

	study	spatial.noise	speech.mat	platform	g.effect.size	g.var.estimate
3	10AP	colocated	other	same	0.63	0.48
4	13MT	separated	other	different	0.77	0.71
5	13HW	colocated	olsa	same	1.58	2.83
6	15WA	colocated	other	different	0.84	0.82
7	16REF	separated	olsa	same	0.98	1.05

```
# Estimation of the main effect
```

```
library("MAAd")
```

```
library("metafor")
```

```
## Loading required package: Matrix
```

```
## Loading 'metafor' package (version 1.9-9). For an overview
```

```
## and introduction to the package please type: help(metafor).
```

```
library("Matrix")
```

```
reviewSummary <- c(reviewSummary, g.effect.size, g.var.estimate, d.effect.size)
```

```
reviewRegression <- tail(reviewSummary,5)
```

```
#Fit Omnibus regression model
```

```
m0 <- mareg(g.effect.size~1, var = g.var.estimate, method = "REML",
            data = reviewRegression)
```

```
summary(m0)
```

```
##
```

```
## Model Results:
```

```
##
```

```
##      estimate    se      z    ci.l ci.u    p
## intrcpt    0.817 0.408 2.003 0.018 1.615 0.04
```

```
##
```

```
## Heterogeneity & Fit:
```

```
##
```

```
##      QE QE.df   QEp    QM QM.df  QMp
## [1,] 0.305 4.000 0.990 4.014 1.000 0.04
```

Interpreting these summary statistics, we find that the overall effect is $g^+ = .82$ (95% CI = 0.018, 1.615), indicating there was a “large” (based on Cohen’s interpretive guidelines; 1988) and significant treatment effect. The QE statistic does not show any statistically significant heterogeneity between all the different studies (QE = 0.305, $p = 0.989$) and we won’t apply moderators (factors affecting the variability of results) to the model.

We can conclude that the expected benefit of amplification is about +2 dB SNR on SRT measures.

3 Non-Inferiority test design

3.1 Non-Inferiority margin determination

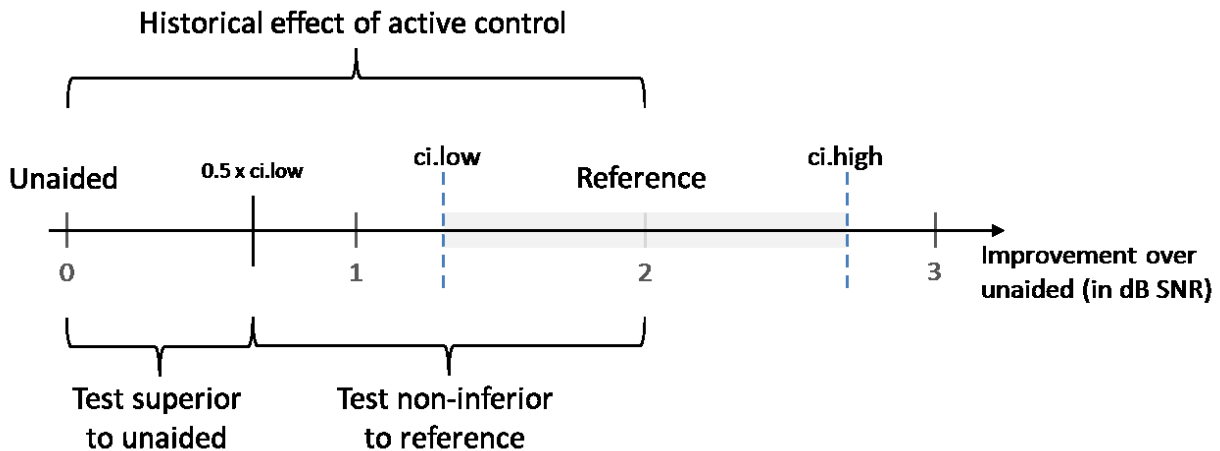
Non inferiority margin set with the following rationale (S. Kaul et al., 2005): “One proposal for selecting the margin is to take one-half of the magnitude of the worst limit of this CI—the so-called “50% rule” or “95-95

method" recommended by the FDA" with unaided condition instead of placebo.

Some considerations:

Assay sensitivity: "the ability to detect differences between treatments if such differences exist." Meta-analysis on effect size (see section 1.2) has shown that there is a positive effect of amplification with speech reception threshold measure (effect = 0.82, SE = 0.41, and p = 0.04).

Constancy assumption: "which assumes that the historical difference between the active control and placebo will be constant in the setting of the current active control trial if a placebo control had been used." Experimental factors like speech material, noise location, and hearing aid type has no significant influence on the amplification effect size. However, we think that the reference test (16REF) presents most similarities with the experimental test design. As the differences are very small regarding the overall effect, we will use the confidence interval of 16REF, for the computation of the non-inferiority margin.



```
n.i.m. <- 2 - reviewSummary$ci.low[reviewSummary$study=="16REF"] / 2
round(n.i.m., 1)
```

```
## [1] 1.3
```

The non-inferiority margin for the trial will be set to 1.3 dB.

3.2 Non-inferiority test hypothesis

Lower SRT indicates a better performance, i.e. a positive difference like between unaided and aided shows an improvement with amplification. Following values definition: The current marketed device is the control : C,

The investigational device is the test : T,

The non-inferiority margin : M.

So if $C - T < 0$ then C is better than T and if $C - T > 0$ then T is better than C.

Test Hypothesis

Under H0: $C - T \leq -M$, i.e. non-inferiority not found

Under H1: $C - T > -M$, i.e. T significantly non-inferior to C

Test analysis

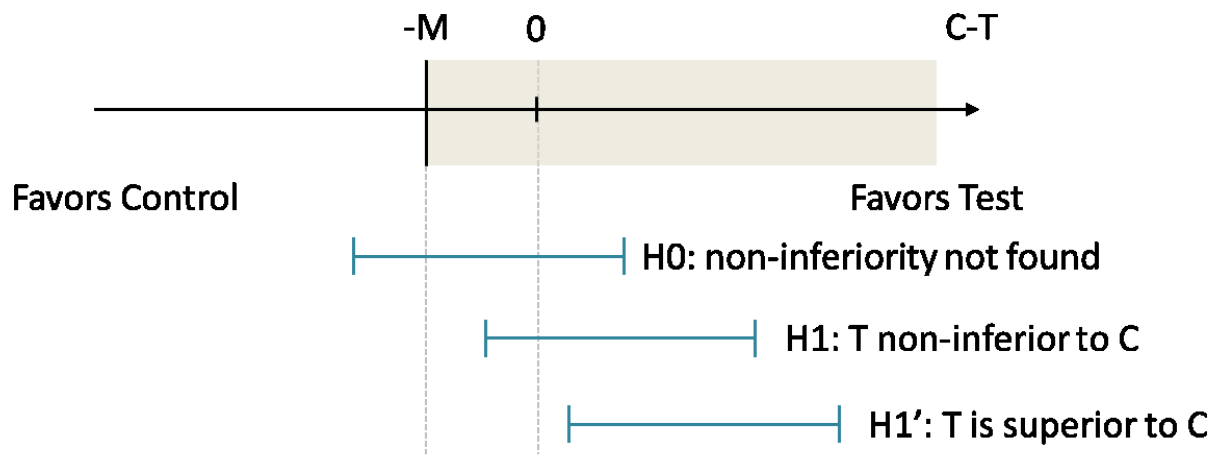
The paired t-test usually tests that the mean differences are zero. The non-inferiority test compares the difference to a non-zero quantity $-M$.

The assumptions of the paired t-test are:

1. The data are continuous (not discrete).
2. The data, i.e., the differences for the matched-pairs, follow a normal probability distribution.
3. The sample of pairs is a simple random sample from its population. Each individual in the population has an equal probability of being selected in the sample.

3.3 Switching to superiority hypothesis

If the confidence interval for the difference between both devices lies entirely above 0, then *“it is acceptable to calculate the P value associated with a test of superiority and to evaluate whether this is sufficiently small to reject convincingly the hypothesis of no difference. There is no multiplicity argument that affects this interpretation because, in statistical terms, it corresponds to a simple closed test procedure. Usually this demonstration of a benefit is sufficient on its own, provided the safety profiles of the new agent and the comparator are similar.”*



Reference: Committee for Proprietary Medicinal Products (CPMP). (2001). Points to consider on switching between superiority and non-inferiority. British Journal of Clinical Pharmacology, 52(3), 223–228.

4. Required sample size

Reference: Chow S, Shao J, Wang H. 2008. Sample Size Calculations in Clinical Research. 2nd Ed. Chapman & Hall/CRC Biostatistics Series. page 52.

With the following requirements: non-inferiority margin: 1.34, standard deviation: 2 dB, α : 0.025, β : 0.1.

Required sample size:

```
delta <- n.i.m.
sd <- reviewSummary$standev[reviewSummary$study=="16REF"]
alpha <- 0.025
beta <- 0.1
```

```
out.sample.size <- (sd*(qnorm(1-alpha)+qnorm(1-beta))/delta)^2  
ceiling(out.sample.size)
```

```
## [1] 24
```

A sample size of 24 will be use to test the non-inferiority hypothesis with a non-inferiority margin of 1.3 dB.

5. References

- Alworth, L. N., Plyler, P. N., Reber, M. B., & Johnstone, P. M. (2010). The effects of receiver placement on probe microphone, performance, and subjective measures with open canal hearing instruments. *Journal of the American Academy of Audiology*, 21(2010), 249–266. <http://doi.org/10.3766/jaaa.21.4.4>
- Chow, S., Shao, J., & Wang, H. (2008). *Sample Size Calculations in Clinical Research*. 2nd Ed. Chapman & Hall/CRC Biostatistics Series. page 52.
- Committee for Proprietary Medicinal Products (CPMP) (2001). Points to consider on switching between superiority and non-inferiority. *British Journal of Clinical Pharmacology*, 52(3), 223–228.
- D’Agostino, R. B., Massaro, J. M., & Sullivan, L. M. (2003). Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. *Statistics in Medicine*, 22(2), 169–86. <http://doi.org/10.1002/sim.1425>
- Del Re, A. C. (2015). A Practical Tutorial on Conducting Meta-Analysis in R. *The Quantitative Methods for Psychology*, 11 (1), 37-50.
- Helbling, T., Vormann, M., & Wagener, K. (2013). Speech priority noise reduction: A digital noise reduction preference study. *Hearing Review*. 2013;10(11):34-43.
- Johnson, B. T., & Huedo-Medina T., B., (2013). Meta-Analytic Statistical Inferences for Continuous Measure Outcomes as a Function of Effect Size Metric and Other Assumptions [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013 Apr. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK140575/>
- Kaul S., Diamond G. A., & Weintraub W. S. (2005). Trials and tribulations of non-inferiority: the ximelagatran experience. *J Am Coll Cardiol*. 2005 Dec 6;46(11):1986-95.
- Klemp, E. J., & Dhar, S. (2008). Speech Perception in Noise Using Directional Microphones in Open-Canal Hearing Aids. *Journal of the American Academy of Audiology*, 19(7), 571–578. <http://doi.org/10.3766/jaaa.19.7.7>
- Magnusson, L., Claesson, A., Persson, M., & Tengstrand, T. (2013). Speech recognition in noise using bilateral open-fit hearing aids: The limited benefit of directional microphones and noise reduction. *International Journal of Audiology*, 52(1), 29–36. <http://doi.org/10.3109/14992027.2012.707335>
- Valente, M., & Mispagel, K. M. (2008). Unaided and aided performance with a directional open-fit hearing aid. *International Journal of Audiology*, 47(6), 329–36. <http://doi.org/10.1080/14992020801894832>
- Woods, D. L., Arbogast, T., Doss, Z., Younus, M., Herron, T. J., & Yund, E. W. (2015). Aided and unaided speech perception by older hearing impaired listeners. *PloS One*, 10(3), e0114922. <http://doi.org/10.1371/journal.pone.0114922>