bernafon
Your hearing · Our passion

A comparative, controlled, clinical investigation of a currently marketed hearing aid programmed with two different fitting methods

Statistical Plan

Document dated: 2017.12. 21

Statistical Plan BF002-1707

version 1.0

Christophe Lesimple

21.12.2017

Contents

1. Test Design **2**
1.1 Research question 2
1.2 Timeline 2
1.3 Random effect 2

2. Outcomes **3**
2.1 Primary outcomes 3
2.2 Secondary outcomes 3
2.3 Explanatory variables 4

3. Sample size and power analysis **5**
3.1 Sample size proposal 5
3.2 Power analysis 6

4. Analysis **8**
4.1 Analysis methods 8
4.2 Analysis tools 8

5. References **8**

This test will compared two fitting procedures and devices with the following label:

- Reference Medical Device (RMD): marketed Bernafon device fitted with standardized clinical procedure,
- Investigational Medical Device (IMD): new fitting procedure and device.

While both devices will have the same electro-acoustical performances, it is suspected that changing fitting procedure might influence the subjective benefit for hearing aid user. Research suggest that placebo effect might impact hearing aid trials (Dawes et al., 2013) and that clinician behavior might also modify the perceived benefit of hearing aids (Naylor et al., 2015). The designed test will investigate the potential effect of the IMD compared to the RMD on first time users.

1. Test Design

As the fitting procedure is part of the investigation, no true blinding can be hold. However, the participants will not be told that there is no electro-acoustical difference between IMD and RMD. A balanced cross-over design will be used to guarantee control for period or learning effect. Repeated measures will be analyzed as it reduces the sample size needed to achieve the same test power. No wash out period is needed as it could not be demonstrated that amplification from hearing aid induces a carry-over effect.

1.1 Research question

The test should answer following research questions:

1. Narrative effect of fitting procedure: are different fitting procedures leading to differences in perceived benefit with hearing aids despite similar electro-acoustical performances? (Questionnaires)
2. Self-administrated audiometry: is AMTAS self-administrated audiometry equivalent to standard clinical procedure? (hearing loss thresholds)

1.2 Timeline

The participants will be randomly allocated into two groups that will have to test the IMD and RMD. Both groups will differ based on test order:

- Group A: IMD for one week and then RMD for one week,
- Group B: RMD for one week and then IMD for one week.

1.3 Random effect

Standardized fitting process is not a homogeneous concept and we expect to have a potential investigator effect that can influence the outcomes. Investigator can be included in the analysis as a fixed or random effect as described by Feaster et al. (2011). As balance between investigators is complicate to hold in practice and limits the generalizability of the results, we will include investigator as a random effect. Feaster et al. (2011) recommends a minimum of 4 investigators trials with site or investigator as random effect. An investigator-by-treatment interaction terms will also be included in the model as recommended by Kraemer and Robinson (2005).

2. Outcomes

2.1 Primary outcomes

The perceived benefit will be measured with:

a. the IOI-HA questionnaire and its official translation into German (Cox et al., 2002)

b. a custom questionnaire based on reference article from Naylor et al. (2015).

These two questionnaires were used in the reference article from Naylor et al. (2015) to measure the effect of narrative embodied in the hearing aid fitting process. Even current research question addresses another aspect of the fitting process, these tools have shown a potential to detect changes in the investigated processes.

IOI-HA questionnaire

Perceived benefit of amplification assessed with 7 questions on a 5 points scale (maximum score 35). Indication from reference study found an effect of fitting process between the preferred and less preferred possibility:

- significant 3.2 scale points and an effect size of 0.68 with experience hearing aid users (n = 24),
- non-significant 0.5 scale points with first time hearing aid users (n = 16).

Custom questionnaire

From the reference study, the questionnaire will directly ask about the preference (much better/slightly better/no difference), the certainty of preference (10 points VAS), and the reason (multiple choice: sound quality, speech intelligibility, listening effort, comfort, loudness, other).

- Fitting procedures: no significant difference between both fitting procedures for first time and experienced hearing aid users,
- Placebo effect: 20 out of 24 indicated a difference with a mean certainty of 7.9/10 for the experienced users group. 14 out of 16 indicated a difference with a mean certainty of 7.0/10 for the first time users group

Order effect

From the reference study, test order might play a different role for first time or experienced users:

- No order effect was found with the experienced hearing aid users,
- A highly significant test order effect in favor for the second tested hearing aid was found for the first time users.

2.2 Secondary outcomes

2.2.1 Audiometry

AMTAS is a self-administrated audiometry procedure that have shown equivalence to manual audiometry (Elkeboom et al., 2013). Across all tested audiometrical frequencies (250 up to 8000 kHz), the mean difference between both procedures (self-administrated vs. manual) was about 0.1 dB HL (min = -3 dB HL at 250 Hz and max = 3.6 dB HL at 8000 Hz). Data analysis will focus on hearing threshold differences with both procedures in a nested model: ear-participant-investigator.

2.2.2 Data logging

The usage of hearing aids will be recorded as another indication of hearing aid use. The data will be averaged from each ear. Following data will be used:

1. Daily usage in hours per day,
2. Acoustical environment in percentage for following listening environments: quiet, speech in quiet, speech in noise, and noise.

2.3 Explanatory variables

2.3.1 Speech intelligibility Freiburger Test

The freiburger sprachtest is a standardized speech test in German speaking areas. Percentages of correctly repeated words are recorded as test outcomes by the investigator for the unaided and aided conditions for each test session.

Differences between unaided and aided (IMD vs. RMD) at 50% speech intelligibility will be used as an explanatory variable for the model.

2.3.2 Gothenburg profile

The perceived handicap associated with hearing impairment might also play a role in perceived benefit from hearing aid. The Gothenburg Profile (GP) is a 20 items questionnaire with a 5 points scale (how often. . . ? never, rarely, sometimes, often, and always). First block of 10 questions addresses speech intelligibility and localization while second block is focused on experienced handicap and the relation to others.

Our hypothesis is that participants that are not often affected in their daily life by their hearing impairment, might report smaller benefit from amplification with a hearing aid. Understanding and quantifying their struggles with principal component analysis could be used to improve the modelling of the results.

3. Sample size and power analysis

Literature search about this tested topic does not provide enough strong references to build a confirmatory analysis. It seems that results might be strongly influenced by the inclusion criteria (first time vs experienced hearing aid users) or test design (cross-over, multi-site, or between group design). There is no article to our knowledge that includes multi-site and investigator effect in the test design when looking at the effect of different fitting procedures. Adding a by investigator and by treatment interaction term seems to be mandatory if this trial wants to be able to generalize the results for a broader environment. There is no reasonable motivation to add the by site effect as the participants will be randomly assigned to one of the four investigator who should be able to work on both test sites.

3.1 Sample size proposal

Sample size cannot be computed on reported effect sizes from previous experiment, however, the following summary shall be used as an indication to determine the sample size for participants per investigator:

- Naylor et al. (2015) about the narrative effect during the fitting with 24 experienced users and 16 first time users. 83% of the experienced HA users had a preference for a given fitting method and 87.5 % for the first time users. Test design was repeated measures and separate analysis on groups based on experience.
- Dawes et al. (2011) about the placebo effect in hearing aid trials with 20 experienced hearing aid users. 75% of the listeners had a preference for the labelled “new” hearing aid (effect size = 0.50),
- Dawes et al. (2013) about the placebo effect in hearing aid trials with 16 experienced hearing aid users. 75% of the listeners had a preference for the labelled “new” hearing aid (effect size = 0.66),
- Humes et al. (2017) about the performances with over the counter hearing aids compared to audiology best practice with 154 participants. Between groups test design with 3 arms (placebo, reference device, and over the counter alternative).

Placebo effect was found with sample sizes about 20 participants in single site study design. Adding an investigator effect in the design might bring new insights in this domain but need to have a minimum participant per investigator to reduce imbalance between them. A minimum of 24 participants will be included in the test with 4investigators and at least 6 participants per investigator.

This number is in line with Naylor et al. (2015) proposal who wanted “to compare the effect sizes generated by the experimental contrast in this study with those generated in typical studies of alternative HA systems, so we chose a sample size (n = 24) commensurate with such studies. Approximately 75% of studies comparing alternative HA systems utilize n ≤ 25 (Naylor 2005). Statistical power calculations based on HHIE data from another study of a non-acoustical intervention (Thorén et al. 2014) indicate that n = 18 would be sufficient to detect a 10-point change in HHIE score in either direction with an alpha of 0.05 and a Power of 0.8.”

3.2 Power analysis

The expected power as a function of the effect size can be computed with G*Power (version 3.1.9.2).

Preference: χ^2 distribution can be used to investigate if the preference distribution is due to chance or to the manipulated effect. We can compute the effect size from the reported literature with Cohen's w (Cohen, 1988):

$$w = \sqrt{\sum_{i=1}^{N} \frac{(p_{0i} - p_{1i})^2}{p_{oi}}}$$

where p_{0i} is the value of the i^{th} cell under H0 and p_{1i} is the value of the i^{th} cell under H1. However, there are two possible analysis options that can be differentiated by the counting method.

1. **with 2 outcomes** i.e. there is an audible difference or not,

Naylor (2015) experiment 1: w = 0.67

Naylor (2015) experiment 2: w = 0.75

Naylor (2015) experiments 1 and 2: w = 0.7

Dawes (2011) experiment 1: w = 0.5

Dawes (2013) experiment 2: w = 0.5

Dawes experiments 1 and 2: w = 0.5

All the experiments together: w = 0.6

With an effect size of 0.6, 1 degree of freedom, 24 participants and α of 0.05, we can achieve a power of 0.84. This analysis method is used by Dawes in his papers because only one device was preferred over the other if the listeners had a preference. The influence of effect size variation on the achieved power is show in figure 1.

This analysis allows to answer the question is there is a difference between both method used in the test. But it does not give an indication about the preferred method.

2. **with 3 outcomes** i.e. preference for A, preference for B, or no preference,

Naylor (2015) experiment 1: w = 0.54

Naylor (2015) experiment 2: w = 0.47

Naylor (2015) experiments 1 and 2: w = 0.5

Dawes (2011) experiment 1: w = 0.94

Dawes (2013) experiment 2: w = 0.94

Dawes experiments 1 and 2: w = 0.94

all the experiments together: w = 0.65

With an effect size of 0.65, 2 degrees of freedom, 24 participants and α of 0.05, we can achieve a power of 0.82. The influence of effect size variation on the achieved power is show in figure 2.

Adding a degree of freedom in the analysis leads to a small loss of power, even if it is always higher than 80%. However, it might give more information about which method will be preferred.

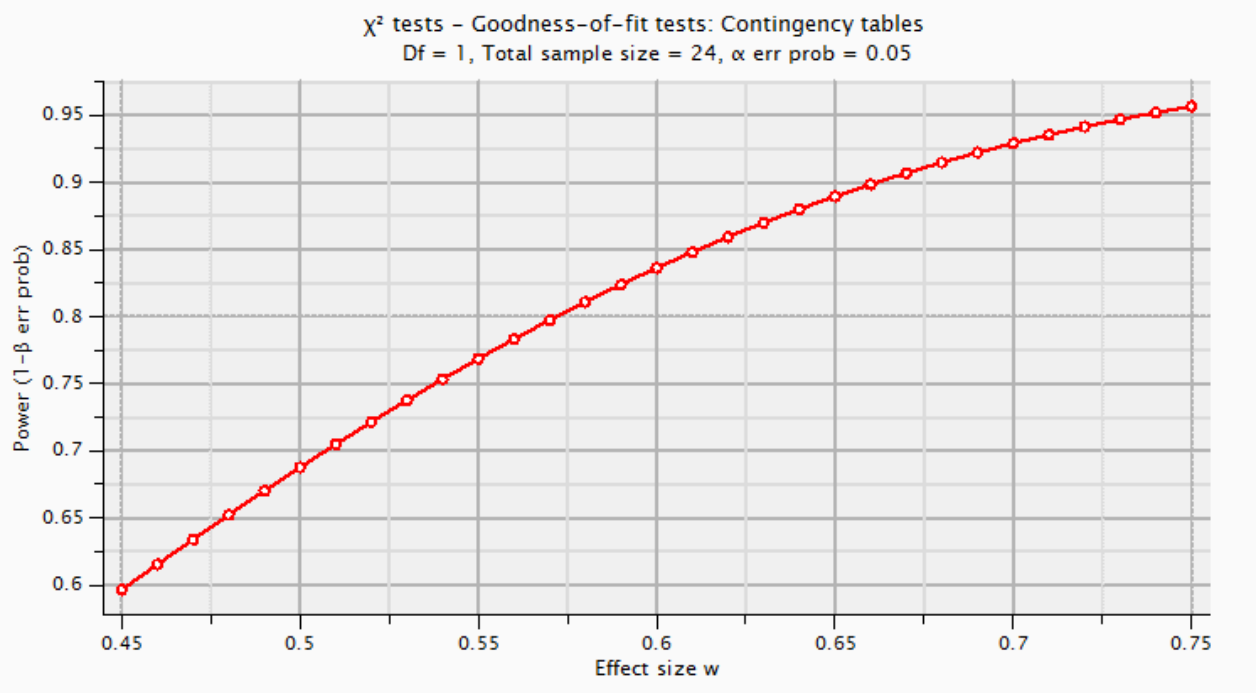

Figure 1: Expected power as a function of the effect size with analysis on 2 levels (audible difference or not)

Figure 2: Expected power as a function of the effect size with analysis on 3 levels (preference for A, preference for B or no preference)

4. Analysis

4.1 Analysis methods

Linear mixed effect regression will be used on the IOI-HA questionnaire, chi2 distribution on the preference test and paired t-test on audiometrical thresholds. Principal component analysis will be made on different outcomes and explanatory variables combinations.

4.2 Analysis tools

R (latest available and validated version) downloaded from official Comprehensive R Archive Network (https://cran.r-project.org/). R-Studio IDE will be used to integrate analysis to the report. R provides adequate packages for descriptive statistics (base, stats, and Rmisc), data vizualisation (ggplot2), mixed effect models (lmer, lmerTest, and nlme), and principal component analysis (FactoMineR).

5. References

Cohen, Jacob (1988). Statistical Power Analysis for the Behavioral Sciences. Routledge.

Cox, R. M., Stephens, D., & Kramer, S. E. (2002). Translations of the International Outcome Inventory for Hearing Aids (IOI-HA). International Journal of Audiology, 41, 3–26.

Dawes, P., Powell, S., & Munro, K. J. (2011). The Placebo Effect and the Influence of Participant Expectation on Hearing Aid Trials. Ear and Hearing, 32(6), 767–774.

Dawes, P., Hopkins, R., & Munro, K. J. (2013). Placebo effects in hearing-aid trials are reliable. International Journal of Audiology, 52(7), 472–477.

Eikelboom, R. H., Swanepoel, D. W., Motakef, S., & Upson, G. S. (2013). Clinical validation of the AMTAS automated audiometer. International Journal of Audiology, 52(5), 342–349.

Feaster, D. J., Mikulich-Gilbertson, S., & Brincks, A. M. (2011). Modeling site effects in the design and analysis of multi-site trials. The American Journal of Drug and Alcohol Abuse, 37(5), 383–91.

Kraemer, H. C., & Robinson, T. N. (2005). Are certain multicenter randomized clinical trial structures misleading clinical and policy decisions? Contemporary Clinical Trials, 26(5), 518–529.

Margolis, R. H., Saly, G. L., Le, C., & Laurence, J. (2007). Qualind: A method for assessing the accuracy of automated tests. Journal of the American Academy of Audiology, 18(1), 78–89.

Naylor, G., Öberg, M., Wänström, G., & Lunner, T. (2015). Exploring the Effects of the Narrative Embodied in the Hearing Aid Fitting Process on Treatment Outcomes. Ear and Hearing, 36(5), 517–526.

Ringdahl, A., Eriksson-Mangold, M., & Andersson, G. (1998). Psychometric evaluation of the Gothenburg Profile for measurement of experienced hearing disability and handicap: applications with new hearing aid candidates and experienced hearing aid users. British Journal of Audiology, 32(6), 375–385.