Placebo effects in hearing aid trials are reliable

Journal:	*International Journal of Audiology*
Manuscript ID:	TIJA-2012-11-0257.R1
Manuscript Type:	Original Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Dawes, Piers; University of Manchester, Psychological Sciences; Hopkins, Rachel; University of Manchester, School of Psychological Sciences Munro, Kevin; University of Manchester, School of Psychological Sciences
Keywords:	Hearing Aids, Behavioral Measures, Hearing Aid Satisfaction, Psycho-social/Emotional

Placebo effects in hearing aid trials are reliable

Piers Dawes,* PhD

Rachel Hopkins, MSc

Kevin J Munro, PhD

Audiology and Deafness Research Group,

School of Psychological Sciences

University of Manchester

Manchester, UK

Key words: placebo effect, reliability, hearing aid trial

Abbreviations: FAAF, Four Alternative Auditory Feature test, NAL-NL1; National Acoustics Laboratories-Non-Linear 1

*Corresponding author. Present address: Audiology and Deafness Research Group, School of Psychological Sciences, University of Manchester, Oxford Road, Manchester, UK, M13 9PL.

Tel: +44 161 3061758, Fax: +44 161 275 3373

Email address: piers.dawes@manchester.ac.uk

Abstract

Objective: A recent study suggested that placebo effects are a source of bias in non-blinded hearing aid trials. Given the potential impact of this finding on the interpretation of non-blinded trials and design of future research trials, the objective of the present study was to investigate the reliability of this effect.

Design: Using the same procedure as an earlier study, participants were told that they were taking part in a trial of new hearing aid technology. Participants compared two devices that were acoustically identical, except one was described as "new" and the other as "conventional". Participants completed a speech-in-noise test, sound quality ratings and rated overall personal preference for both hearing aids.

Study sample: Sixteen adult hearing aid users.

Results: Participants had significantly better mean speech-in-noise performance (70.9% versus 66.8%,Z=2.30, p=0.02, effect size Pearson's r=0.15) and sound quality ratings for the "new" hearing aid (8.1 versus 7.4, Z=-2.99, p=0.003, r=0.28). A significant proportion of participants (75%) expressed an overall preference for the "new" hearing aid (p=0.001, effect size ϕ_c=0.66).

Conclusion: Placebo effects reliably impact on hearing aid trials. In order to control for placebo effects, double-blind methodology is optimal. However, when double-blinding is not possible other strategies may be appropriate.

Placebo effects are clinical responses associated with the expectation surrounding treatment, rather than with any intrinsic property of the treatment. Placebo effects can occur in conjunction with active treatments, and either facilitate or inhibit them (Luparello et al., 1968; Colloca et al., 2004; Benedetti et al., 2007). In relation to clinical medicine, it is suggested that placebo effects could be (and are) used to optimise outcomes for patients. Such positive effects might be encouraged by the use of suggestion or through positive interaction between the clinician and the patient (Thompson, 2000; Benedetti, 2011; Finniss et al., 2011). Placebo effects are also widely recognised in medical research, and control conditions with double-blind methodology[1] are routinely incorporated into clinical trials (Thompson, 2000; Price et al., 2008). However, in clinical audiology and audiological research, placebo effects are not typically considered. There is some research evidence to suggest that they may have relevance for audiology, as follows.

Around the time that digital hearing aids were being first introduced, Bentler and colleagues (2003), investigated whether labelling hearing aids as 'digital' impacted on measures of hearing aid benefit and overall preference. They provided participants with two identical hearing aids that were labelled as 'digital' (versus 'conventional'). Participants used each hearing aid for one month. They then completed a battery of hearing aid benefit measures, involving speech perception and self-report measures. Effects were generally small; the effect of labelling accounted for 2 to 32 % of variance of the outcome measures. Differences between labelling conditions were not statistically significant individually (with the exception of some subscales from the Abbreviated Profile of Hearing Aid Benefit; Cox & Alexander, 1995). Overall, a statistically significant majority of participants (33 of 40)

[1] Double-blind studies are those where neither the participant nor the experimenter is aware which the experimental condition is and thus control for the possible affect of expectation.

reported that they preferred the 'digital' hearing aid. Bentler et al concluded that the participants' expectation that 'digital' hearing aids must be better had affected both performance of measures of hearing aid benefit and their overall preference.

Dawes et al (2011) went on to investigate the impact of patient expectation on the outcome of a trial of two hearing aids. The aim of Dawes et al's study was to test whether the effect of expectation observed by Bentler and colleagues was specific to the 'digital' label, or whether a more general label of 'new technology' could set up positive expectations about performance, with a consequent impact on outcome measures. If this were the case, it would relevant for any trial of new hearing aid technology. Twenty experienced adult hearing aid users were told that they were taking part in a trial of new hearing aid technology. In a single test session, users completed a speech-in-noise test, sound quality ratings and overall personal preference. Two hearing aids were compared, one being described as 'new' and the other 'conventional'. In reality, both hearing aids were identical and were programmed to the same prescription target for a typical age-related hearing loss. SIN performance was marginally better (by around 2%) for the 'new' hearing aid for the speech-in-noise test, sound quality ratings were statistically significantly higher for the 'new' hearing aid, and 15 out of 20 participants expressed an overall personal preference for the 'new' hearing aid (with the remainder expressing no preference). The interpretation was that describing one hearing aid as being 'new' had set up an expectation in the minds of the participants that performance would be better for the 'new' hearing aid, and that participants' expectation had impacted upon performance of the test measures. On the basis of Bentler et al's (2003) study and Dawes et al's (2011) findings, Dawes et al suggested

a need to control for placebo effects and advised caution in interpreting the results of trials that did not control for placebo effects.

One area of uncertainty is whether expectation does have a reliable impact on performance and measures of hearing aid benefit. In Bentler et al's study, effects on individual tests (with the exception of some subscales for the self-report questionnaire) were generally small and statistically non-significant (although there was a significant effect overall). In Dawes et al's study, the difference in performance on the speech in noise test was small and statistically non-significant (although the authors argued that this small difference was significant in the context of positive findings for the other measures and in hearing aid trials generally, where small effects are typically sought).

If placebo effects are reliable in the context of aided performance and measures of hearing aid benefit, there would be an obvious need to account for placebo effects in the design of future hearing aid trials. Such effects may also have relevance for clinical audiology, if for example participant expectation might be utilised to increase the benefit of audiological treatments. Thus, the objective of the present study was to investigate the reliability of the effect observed by Dawes et al (2011) by replicating that study in a different group of participants and with a different experimenter administering the test measures. The hypothesis was that if placebo effects on hearing aid trials are reliable, a similar pattern of results to that observed by Dawes et al (2011) would be obtained. This would add weight to the suggestion that placebo effects should be controlled for in trials of hearing aid (or similar) technology.

Method

Methodology follows that used by Dawes et al (2011) but is summarised below.

Participants

A sample size was chosen to be comparable with Dawes et al's (2011) study and was intended to be similar to a typical of hearing aid trial, which use small to medium sample sizes that are sufficient to detect effects that are large enough to be of clinical relevance. Accordingly, sixteen participants aged between 61 to 86 yr (M = 76 yr, SD = 7 yr) were recruited from a local hospital-based audiology clinic. Inclusion criteria were: i) at least 12 months daily hearing aid use, ii) English as a first language, iii) symmetrical, mild-to-moderate, sloping high frequency sensorineural hearing loss of at least 45 dB HL at 2 – 6 kHz, iv) ≤5 dB difference between the ears at two or more adjacent frequencies between 0.25 and 8kHz and v) normal middle ear function. Pure-tone audiometry was performed in a sound-treated room using a calibrated Kamplex KLD 21 audiometer with TDH-39 headphones and a B-71 bone vibrator. Normal middle ear function was confirmed with a GSI 38 Auto Tymp. All participants were unilateral hearing aid users. The reason for this was that unilateral hearing aids are typically prescribed by the audiology clinic from which participants were recruited. When invited to take part, participants were told that the purpose of the study was to evaluate new hearing aid technology. On completion of the study, participants were informed about the true purpose of the study. Ethical approval was obtained from the NHS Central Manchester Research Ethics Committee. Written consent was obtained from all participants. Participants were reimbursed travel costs but no other compensation was provided.

Test hearing aids

Test hearing aids were two Starkey A312 Strata behind-the-ear digital aids with seven band, three channel wide dynamic range compression with a noise reduction algorithm. One hearing aid was to be referred to as the 'new' one while the other was referred to as the 'conventional' aid. The 'new' hearing aid had a yellow case while the 'conventional' aid had a beige case. The hearing aids were programmed to the same NAL-NL1 prescription target based on an audiogram for a typical age-related hearing loss (35 dB HL at 500Hz, 40 dB HL at 1000Hz, 50 dB HL at 2000 Hz, 60 dB HL at 4 kHz, and 80 dB HL at 8 kHz). Casings were switched between hearing aids so that for half the group, one hearing aid had a yellow case and was the 'new' hearing aid, while the other had a beige case and was the 'conventional' aid. For the other half of the group, casings and designation (new/conventional) was switched. This procedure was used to control for any actual acoustic differences between hearing aids.

In order to ensure that the two hearing aids produced identical amounts of gain, repeated coupler measurements were obtained after initial programming, after switching cases and after completion of testing. All measurements for the two hearing aids at all frequencies were consistently within 1dB. Coupler gain targets and repeated gain measurements are shown in Table 1. Listening tests confirmed the similarity of the hearing aids. All participants used their current ear mould (all hard acrylic with a pressure equalization vent) with the test hearing aids. Test hearing aids were thus trialled with the ear that was normally fitted with the hearing aid.

(Table 1 here)

Outcome measures

Participants were seated in a comfortable chair 1.5m from a loudspeaker at 0 degrees azimuth. Sound levels for each stimulus refer to the level measured at the reference point, defined as the centre of the participants head with the participant absent.

Speech in noise test

The Four Alternative Auditory Feature test (FAAF; Foster and Haggard, 1987) is a computerised 80 item single-syllable, closed set word recognition test. Participants hear the sentence 'Can you hear X clearly?' and are required to choose from a selection of four words on a screen using a mouse click to identify the word that they heard. The test was administered with the sentences at 65 dB (A) in the presence of speech-shaped noise at +2 dB SNR. A practice list of 12 words was used to familiarise participants with the test procedure with participants using their own hearing aid. Participants completed separate runs of the FAAF with both the new and the conventional aid, with order of testing counter-balanced across participants.

Sound quality rating test

Participants listened to six sound samples and rated them on clarity, comfort and overall impression using a 10 point visual analogue scale (based on Arlingler et al, 1998). Sound samples were Bamford-Kowal-Bench sentences (BKB; Bench et al, 1979) spoken by male and female voices in quiet and in noise, music and an environmental sound (robin song). All samples were 10 seconds long, digitized at 44 kHz. Samples were equalised to have the same long-term RMS power and presented at 65 dB (A). BKB sentences in noise were presented at +2 dB SNR in broadband noise at the same overall presentation level for the sentences in isolation (65 dBA).

Personal preference

At the end of the test session, participants were asked to indicate if they had an overall personal preference for either hearing aid by choosing one of three categories: 'the new hearing aid is best', 'the conventional hearing aid is best' or 'I cannot tell any difference'.

Procedure

At the beginning of the session, participants were given an explanation of the (false) aim of the study, i.e. to evaluate new hearing aid technology and were shown the two hearing aids. The hearing aid with a yellow case was introduced as the one containing 'new technology', while the hearing aid with a beige case was introduced as the 'conventional' hearing aid. No negative comments were made about the 'conventional' hearing aid. Otoscopy, tympanometry and pure tone audiometry was carried out first, followed by FAAF test, sound quality ratings and personal preference. The order of testing was counterbalanced so that half the participants performed the tasks with the 'new' hearing aid first, while the other half performed them with the 'conventional' aid first. All measurements were made within a single test session of around one hour's duration.

Statistical analysis

Non parametric tests (Wilcoxon signed-rank test) were selected for FAAF test because of non-normal distribution of data (positive skew) and to allow comparability with Dawes et al (2011). Performance was above chance levels for all participants, and because relative differences between two conditions are of interest rather than absolute performance, data from all participants was retained and appropriate non-parametric tests applied. Non-

parametric tests were also applied for sound quality ratings, as appropriate for ordinal data. To test for an effect of the order of testing (i.e. new or conventional first), Mann-Whitney U test was applied for both FAAF and sound quality ratings. A Chi-squared test was used to test whether the proportion who reported a preference for either hearing aid was different to that expected by chance. In order to estimate the magnitude of the effect of expectation and to facilitate comparison with effects observed by Dawes et al (2011) and with experimental effects sought in clinical hearing aid trials, effect sizes were calculated using Pearson's r statistic[2] using the means and standard deviations for the respective variables. Cramér's V[3] (denoted as ϕ_c) was calculated for the Chi-squared test.

Results

There were no significant effects of test order (new or conventional first) on FAAF score or sound quality ratings; *p*'s 0.21 to 0.87.

FAAF test

Table 2 shows mean performance for the 'new' and the 'conventional' hearing aids in the present study and in Dawes et al (2011). In the current study, performance was statistically significantly higher for the 'new' hearing aid (compared to the 'conventional' hearing aid, with the size of the difference approximately double that reported in Dawes et al (2011). In the current study, twelve participants (75%) demonstrated a better performance in 'new' hearing aid condition compared to the 'conventional' hearing aid condition. One participant

[2] The absolute magnitude of Pearson's *r* varies between 0 and 1, with 1 indicating a perfect relation between the two variables and 0 indicating no relation. As a guide, effects of 0.2 are considered 'small', 0.5 'medium' and 0.8 'large'.

[3] Cramér's V is applied to goodness-of-fit chi-squared models when there is a 1xk table (i.e. degrees of freedom are greater than 1). Like the Pearson r statistic, it provides a measure of association between two variables that varies in size between 0 and 1.

(6%) performed equally in both conditions and three participants (19%) performed worse in the 'new' condition than the 'conventional' condition.

(Table 2 here)

Sound quality rating test

Median scores for each of the six different sound samples played to the participants for each dimension of clarity, comfort and overall impression were obtained for each participant. Both hearing aids scored well, although the 'new' hearing aid scored statistically significantly higher than the 'conventional' hearing aid. Overall sound quality rating (median of all subjective ratings) was also significantly higher for the 'new' hearing aid compared to the 'conventional' hearing aid and this was statistically significant. Table 3 details mean ratings for each dimension and results of statistical comparisons for the present study and for Dawes et al (2011). The pattern of better ratings for the 'new' hearing aid is common to both studies.

(Table 3 here)

Overall preference

Twelve participants stated a preference for the 'new' hearing aid, while four reported that they could not tell the difference between the two hearing aids. None preferred the 'conventional' hearing aid. This difference is statistically significant ($\chi^2(2) = 14.0$, $p = 0.001$, $\phi_c = 0.66$). The proportion of those who preferred the 'new' hearing aid is the same as that reported by Dawes et al (2011); 75%.

Discussion

Overall, a similar pattern of findings was obtained in the current study compared to Dawes et al (2011); participants performed better on the speech in noise test and made more favourable sound quality ratings for the 'new' hearing aid than the 'conventional' one. A significantly higher proportion of participants preferred the 'new' hearing aid. Effect sizes were larger for sound quality ratings and overall preference (measures that rely on participant self-report) over the speech in noise test. As with Dawes et al (2011), the interpretation of these findings is that describing one hearing aid as 'new' set up expectations in the participants that this hearing aid must perform better than the 'conventional' one, and that this expectation impacted upon participant's performance of the experimental tasks. Such an expectation may have resulted in improved performance with the 'new' hearing aid or worse performance with the 'conventional' one (consistent with a 'nocebo' effect; Benedetti et al, 2007) or both. Thus, participant expectation appears to have a reliable effect on performance.

In the current study, FAAF scores tended to be higher while sound quality ratings tended to be lower than in Dawes et al's (2011) study. This difference is unlikely to be due to difference in procedure; hearing aids, tests and procedure were identical. The primary differences between the two studies are i) a different group of participants and ii) a different experimenter. It seems unlikely that the influence of the experimenter could be a primary reason for this difference. If there was a systematic effect of the experimenter, the direction of the effect should be the same for both measures. This difference between studies is more likely to be due to uncontrolled differences between the participants in the two studies, such as the average level of hearing loss.

As with Dawes et al (2011), because of time and ethical constraints, outcome measures were only made within a single test session in the current study. The duration of effects seen in current study and in Dawes et al (2011) are therefore not known. Placebo effects in clinical medicine can be very long lasting (effects of up to a year have been recorded; Dimmons et al. 1960). Bentler et al (2003) found that the influence of labelling a hearing aid as 'digital' persisted after one month of hearing aid use, and so it is possible that the effects of expectation and labelling on hearing aid benefit may be long lasting. This is a topic for future research (see also the section on the role of placebo effects in clinical audiology below).

If placebo effects do have a reliable impact on trials of hearing aids (or similar technology), then trials should include controls for placebo effects. This is necessary so that one may have confidence that any benefit associated with the experimental condition is not at least partly due to a placebo effect. But how should one control for placebo effects in hearing aid trials? A randomised controlled trial is the gold standard methodology for clinical trials (Benedetti, 2009). In a randomised controlled trial, participants are randomly allocated to test and control conditions. However, this does not eliminate placebo effects because if participants or experimenters are aware of the identity of the groups, there is the potential for the expectation of the participant and/or the experimenter to impact on the outcome (Gracely et al, 1985). The gold standard control for placebo effects is a double-blinded design in which neither participants nor experimenters are aware of the identity of the control or experimental condition. A double-blinded design means that neither the participant's nor the experimenter's expectations should influence the outcome of the study. This design may be applicable for many instances in audiological research. For

example, two signal processing schema could be compared in a double-blinded design using identical hardware such as a programmable digital hearing aid or via headphone simulation. Researchers wishing to apply the most rigorous methodology should consider trials of new technology that would be amenable to double-blind designs.

However, an experimental hearing aid may look physically different to the comparison one, in which case a double-blind design may not be possible. In this case, it would be desirable to establish an alternative to double-blind methodology. One method of minimising placebo effects may be employing an experimenter who would provoke minimal labelling effects. Clinician characteristics such as warmth, empathy, prestige and friendliness have been demonstrated to impact on outcome (Price, Finniss et al., 2008), and a similar situation is likely in the context of a clinical trial. However it seems problematic to reliability identify the relevant features of the experimenter and how they would be perceived by individual participants. Additionally, it seems undesirable that a hearing scientist should strive to minimise warmth, empathy or friendliness towards participants, or take steps to minimise his or her prestige in the eyes of the participant. Besides lowering response rates and increasing dropout rates, this strategy may have the unintended consequence of having an adverse impact on the trial via negative expectation (i.e. by invoking a nocebo effect).

It may be helpful to formulate instructions and procedures so as to minimise the effect of expectation. However, this is not a perfect solution. In the present study, the only information provided to participants was that the aim of the study was to evaluate hearing aids with 'new technology', and no negative comments were made about the conventional hearing aid. It seems that even this minimal information was sufficient to set up an expectation in participants that the 'new' hearing aid must be better, and to influence the

performance of outcome measures. In actual hearing aid trials, participants are ethically required to be given information about the test hearing aid and the hypotheses for the trial. This may have a powerful effect on participant expectation (Benedetti, 2009), so that placebo effects in real hearing aid trials may be larger than those reported here.

Dawes et al (2011) suggested that a potential method of mitigating placebo effects in hearing aid trials may be to estimate the size of placebo effect, with additional effects likely to be due to the experimental manipulation in question (Glasziou et al., 2007). Dawes et al cautioned that this approach is problematic for two reasons: First, because placebo effects are known to fluctuate in size depending on various situational variables (such as experimenter characteristics) they are difficult predict. Second, the placebo effects observed in this study and in Dawes et al (2011) are similar in size to experimental effects typically sought in clinical hearing aid trials. It is unlikely that the experimental effect in a clinical hearing aid trial would be substantially larger than that which may be due to a placebo effect. Compared with unaided listening, hearing aids provide a substantial benefit. Innovations in hearing aid technology are not expected to substantially increase performance, and so improvements in speech recognition of a few per cent are realistic goals for hearing aid researchers. For example, Wood and Lutman's (2004) hearing aid trial reported statistically significantly better speech recognition of between 1 to 4 % for an analogue compared to a digital hearing aid, while Valente et al's (1998) study reported a 1.6% advantage for the test hearing aid. In comparison, the size of the placebo effect in the present study and in Dawes et al (2011) was 2-4%. Neither Wood and Lutman's nor Valente et al's study used double-blinded methodology, so the apparent benefits reported in each study could potentially be wholly accounted for by a placebo effect.

A further method of controlling for placebo effects centres on identifying and removing individual placebo responders either before commencing a trial or during a 'run-in' phase. In a placebo run-in, all eligible participants are given a placebo treatment. Those who respond to the placebo treatment are then withdrawn from the study prior to random allocation to treatment condition (For example, Gong et al, 1996). Alternatively, investigators may attempt to exclude participants based on psychological predictors of the placebo response, such as motivation (Geers et al., 2005), suggestibility (De Pascalis et al., 2002) or social acquiescence (McNair et al., 1979). Other participant characteristics might also be used to exclude placebo responders, such as those with a particular profile of symptoms or those who have not previously had any experience with a particular type of treatment (Newcorn et al., 2009). These strategies depend on the assumptions that there is a group of participants who are reliable placebo responders across trials, and that these placebo responders may be identified by certain psychological traits or demographic characteristics. Neither of these assumptions seems to be supported. First, participants vary in the extent to which they show a placebo response depending on a wide range of contextual and personal factors (Price, Finniss et al., 2008) such that a participant may not exhibit a placebo response in one situation but exhibit a placebo response in another similar situation (Beecher, 1955; Liberman, 1964). Thompson (2000) concluded that because of this inconsistency, there is no evidence for the effectiveness of the 'placebo run-in' as a method for controlling for placebo effects. Second, the defining traits of placebo responders vary between studies, and reviewers have concluded that participant characteristics are not good predictors of placebo responses (Beecher, 1955; Liberman, 1964; Shapiro & Shapiro, 1984; Thompson, 2000; Price, Finniss et al., 2008). In relation to hearing aid trials, there is no evidence to suggest whether a strategy of identifying and excluding placebo responders

would be feasible or not. There are several questions that must be answered. It is currently unknown whether placebo effects in hearing aid trials are reliable at an individual level and if there are any features that reliably characterise placebo responders that would allow identification and exclusion. Even if it were possible to identify and exclude placebo responders from hearing aid trials, this seems potentially problematic if the consequence was the introduction of a selection bias. Excluding a specific portion of participants may reduce the generalizability of the result.

Finally, in clinical medicine placebo effects could be (and are) ethically utilised to optimise the outcome for patients (Turner et al., 1994; Thompson, 2000; Price, Finniss et al., 2008). Placebo effects may also be utilised in clinical audiological settings to increase patient benefit. In relation to tinnitus, Tyler et al (2001) suggested that ensuring that one is perceived by patients as a sympathetic, competent and confident professional may translate into positive patient expectations and a higher likelihood of a successful outcome from treatment. Boosting outcomes via conscious encouragement of placebo effects via supportive and positive clinical interactions is likely to be ethically acceptable. Whether in some circumstances and with appropriate limitations, it may be ethically acceptable to prescribe a treatment primarily to elicit a placebo effect is a more difficult question (Finniss, Kaptchuk et al., 2011). The utility and ethical application of placebo effects in clinical audiological settings is a matter for further enquiry and informed discussion among audiologists and professional bodies.

Conclusions

Placebo effects appear to have a reliable impact on measures of hearing aid benefit such as those typically used in hearing aid trials. Placebo effects need to be controlled for in hearing

aid trials, and given the current state of knowledge there is no satisfactory alternative to a double-blind design. Hearing aid trials that do not include such controls should be interpreted with caution. The influence of expectation is likely to impact on outcomes in clinical audiology practice but this awaits investigation.

Acknowledgements

The authors thank the patients and staff at the Audiology Department of Withington Community Hospital, South Manchester, for their support in this research. The authors also thank Keith Wilbraham for technical support and Starkey Laboratories Ltd (UK) for providing the hearing instruments and funding for participant transport. This research was presented at the British Society of Audiology annual conference, September 2012, Nottingham UK.

References

Arlinger S., Billermark E., Oberg M., Lunner T. & Hellgren J. 1998. Clinical trial of a digital hearing aid. *Scandinavian Audiology*, 27, 51-61.

Beecher H.K. 1955. The powerful placebo. *JAMA*, 159, 1602-1606.

Bench J., Kowal A. & Bamford J. 1979. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *The British Journal of Audiology*, 13, 108-112.

Benedetti F. 2009. *Placebo effects. Understanding the mechanisms in health and disease*. Oxford: Oxford University Press.

Benedetti F. 2011. *The Patient's Brain. The neuroscience behind the doctor patient relationship*. Oxford: Oxford University Press.

Benedetti F., Lanotte M., Lopiano L. & Colloca L. 2007. When words are painful: unravelling the mechanisms of the nocebo effect. *Neuroscience*, 147, 260-271.

Bentler R.A., Neibuhr D.P., Johnson T.A. & Flamme G.A. 2003. Impact of digital labelling on outcome measures. *Ear & Hearing*, 24, 215-224.

Colloca L., Lopiano L., Lanotte M. & Benedetti F. 2004. Overt versus covert treatment for pain, anxiety, and Parkinson's disease. *The Lancet Neurology*, 3, 679-684.

Cox R.M. & Alexander G.C. 1995. The Abbreviated Profile of Hearing Aid Benefit. *Ear & Hearing*, 16.

Dawes P., Powell S. & Munro K.J. 2011. The placebo effect and the influence of participant expectation on hearing aid trials. *Ear & Hearing*, 32, 767-774.

De Pascalis V., Chiaradia C. & Carotenuto E. 2002. The contribution of suggestibility and expectation to placebo analgesia and phenomenon in an experimental setting. *Pain*, 96, 393-402.

Dimons E.G., Kittle C.F. & Crockett J.E. 1960. Comparison of internal mammary ligation and sham operation for angina pectoralis. *American Journal of Cardiology*, 5, 483-486.

Finniss D.G., Kaptchuk T.J., Miller F. & Benedetti F. 2011. Biological, clinical, and ethical advances of placebo effects. *The Lancet*, 375, 686-695.

Foster J.R. & Haggard M.P. 1987. The four alternative auditory feature test of speech perception (FAAF): Linguistic and psychometric properties of the material with normative data in noise. *British Journal of Audiology*, 21, 165-174.

Geers A., Weiland P.E., Kosab K., Landry S.J. & Helfer S.G. 2005. Goal activation, expectations and the placebo effect. *Journal of Personality and Social Psychology*, 89, 143-159.

Glasziou P., Chalmers I., Rawlins M. & McCulloch P. 2007. When are randomised trials unnecessary? Picking signal from noise. *BMJ*, 334(7589), 349-351.

Gong L., Zhang W., Zhu J., Kong D., Page V., et al. 1996. Shanghai trial of nifedipine in the elderly (STONE). *Journal of Hypertension*, 14, 1237-1245.

Gracely R.H., Dubner R., Deeter W.R. & Wolskee P.J. 1985. Clinician's expectations influence placebo analgesia. *Lancet*, 1, 43.

Liberman R. 1964. An experimental study of the placebo response under three different situations of pain. *Journal of Psychiatric Research*, 2, 233-246.

Luparello T., Lyons H.A., Bleecker E.R. & McFadden E.R. 1968. Influences of suggestion on airway reactivity in asthmatic subjects. *Psychosomatic Medicine*, 30, 819-825.

McNair D.M., Gardos G., Haskell D.S. & Fisher S. 1979. Placebo response, placebo effect, and two attributes. *Psychopharmacology*, 63, 245-250.

Newcorn J.H., Sutton V.K., Zhang S., Wilens T., Kratochvil C., et al. 2009. Characteristics of placebo responders in pediatric clinical trials of attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, 48, 1165-1172.

Price D.D., Finniss D.G. & Benedetti F. 2008. A comprehensive review of the placebo effect: Recent advances and current thought. *Annual Review of Psychology*, 59, 565-590.

Rosenthal R. 1994. Parametric measures of effect size. *In:* H. Cooper & L.V. Hedges (eds.) *The Handbook of Research Synthesis*. New York: Russell Sage Foundation, pp. 231-244.

Shapiro A.K. & Shapiro E. 1984. Patient-provider relationships and the placebo effect. *In:* J.D. Matarazzo, S.M. Weiss, J.A. Herd & N.E. Miller (eds.) *Behavioral heath: A handbook of health enhancement*. New York, NY: Wiley-Interscience, pp. 371-383.

Thompson W.G. 2000. Placebos: A review of the placebo response. *The American Journal of Gastroenterology*, 95, 1637-1614.

Turner J.A., Deyo R.A., Loeser J.D., Von Korff M. & Fordyce W.E. 1994. The importance of placebo effects in pain treatment and research. *JAMA*, 271, 1609-1614.

Tyler R.S., Haskell G.B., Preece J. & Bergan B.A. 2001. Nurturing patient expectations to enhance the treatment of tinnitus. *The Hearing Journal*, 22, 15-21.

Valente M., Fabry D.A., Potts L.G. & Sandlin R.E. 1998. Comparing the performance of the Widex SENSE digital hearing aid with analog hearing aids. *Journal of the American Academy of Audiology*, 9, 342-360.

Wood S.A. & Lutman M.E. 2004. Relative benefits of linear analogue and advanced digital hearing aids. *International Journal of Audiology*, 43, 144-155.

Table 1. Coupler gain targets and measured coupler gain (in decibels) for the test hearing aids

	Frequency (Hz)					
	500	750	1000	2000	3000	4000
NAL-NL1 gain targets at 65 dB SPL	6	12	16	28	32	30
Hearing aid 1 mean gain (SD)*	6 (0.0)	9 (1.0)	14 (0.0)	23 (0.6)	32 (0.0)	34 (0.6)
Hearing aid 2 mean gain (SD)*	6 (0.0)	9 (0.6)	14 (0.0)	23 (0.6)	32 (0.0)	34 (0.6)

* Mean gain and standard deviation is the product of three repeated coupler measures

Table 2. Four Alternative Auditory Feature (FAAF) test results

	Mean FAAF % correct 'New' (SD)	Mean FAAF % correct 'Conventional' (SD)	Difference (New – Conventional)	*Z**	*p*	Effect size (*r*)
Current study	70.9 (12.6)	66.8 (14.1)	4.1% (95% CI: 0.6 to 7.6)	2.30	0.02	0.15
Dawes et al 2011	62.3 (10.4)	60.7 (9.0)	1.6% (95% CI: -1.0 to 4.2)	1.84	0.06	0.08

**Wilcoxon signed-rank test statistic.*

Table 3. Sound Quality Ratings

		Comfort	Clarity	Overall Impression	Overall sound quality rating
Current study	Mean rating "New" (SD)	7.98 (1.76)	8.77 (0.93)	7.84 (1.86)	8.20 (1.39)
	Mean rating "Conventional" (SD)	7.22 (1.62)	7.76 (1.27)	7.16 (1.62)	7.38 (1.37)
	Z*	-2.9	-3.14	-2.36	-2.99
	p (two tailed)	0.001	0.004	0.018	0.003
	Effect size (*r*)	0.22	0.41	0.19	0.28
Dawes et al 2011	Mean rating 'new' (SD)	8.95 (1.12)	9.28 (1.15)	9.00 (1.11)	9.12 (1.02)
	Mean rating 'conventional' (SD)	8.40 (1.21)	8.61 (1.28)	8.1 (1.47)	8.35 (1.17)
	Z*	-1.94	-2.77	-2.98	-2.88
	p (two tailed)	0.053	0.006	0.003	0.004
	Effect size (*r*)	0.23	0.27	0.33	0.33

**Wilcoxon signed-rank test statistic.*