**The placebo effect and the influence of participant expectation
on hearing aid trials**

Piers Dawes, Samantha Powell and Kevin J. Munro

School of Psychological Sciences, University of Manchester, UK

Running title: Expectation & hearing aid outcome

Parts of this study were presented at the annual conference of the British Society of
Audiology, Manchester September 2010

Corresponding author: Piers Dawes PhD, School of Psychological Sciences, Ellen
Wilkinson Building, The University of Manchester, Oxford Road, Manchester, M13
9PL, UK

Tel: +44-161-306-1758 Fax: +44-161-275-3373,

Email: piers.dawes@manchester.ac.uk

Abstract

Objective: The aim of the study was to investigate the influence of participant expectation on the outcome of a trial that compared two behind-the-ear hearing aids with identical electroacoustic performance, except that one was called a ‘new’ hearing aid and the other a ‘conventional’ hearing aid.

Design: Twenty experienced adult hearing aid users were told that they were taking part in a trial that compared new and conventional hearing aid technology. They attended a single test session where they were fitted with each hearing aid, in a balanced design, set to the same NAL NL1 prescription target for a typical age-related hearing impairment. Outcome measures were selected to be representative of hearing aids trials and included: i) the Four Alternative Auditory Feature (FAAF) test (presented at 65 dB (A) and signal-to noise ratio of +2 dB), ii) sound quality ratings for six different sound samples (a selection of speech, music and environmental sounds) and iii) overall personal preference.

Results: There was a non-significant trend for better mean performance with the ‘new’ hearing aid on the FAAF test ($M = 62.3\%$, $SD = 10.4$ versus $M = 60.7\%$, $SD = 9.0$; $z = -1.84$, $p = 0.06$, two-tailed, $r = .08$). The ‘new’ hearing aid was also consistently rated more highly on all sound quality ratings and this difference was statistically significant ($M=8.94$, $SD=0.93$ versus $M=8.32$, $SD = 1.07$; $z = -3.30$, $p < .001$, two-tailed, $r = .29$). Fifteen (75%) participants expressed an overall personal preference for the ‘new’ hearing aid with the remainder expressing no preference ($p < .05$, $r = .50$).

Conclusions: These results suggest a need to control for placebo effects in hearing aid trials, as well as a need to interpret cautiously any hearing aid trial that did not control for this effect.

Key words: hearing aid trial, placebo effect, expectation

INTRODUCTION

Placebo effects -clinical responses to inert treatments- are wide-ranging and are recognised in medical research and clinical practice (Price et al. 2008; Thompson 2000). Control conditions for placebo effects are commonly incorporated into clinical trials of medical treatments although these effects are not usually considered in trials of hearing aids. The goal of this study was to examine possible placebo effects in hearing aid trials using experienced adult hearing aid users.

The placebo effect

The placebo effect is “a change in a patient's illness attributable to the symbolic import of a treatment rather than a specific pharmacologic or physiologic property of the treatment” (Brody 1980). Placebo effects have long been recognized, with the first placebo controlled trial conducted in 1801 (de Craen et al. 1999). Placebo effects currently generate a huge amount of research interest. For example, over the period spanning January to October 2010, about 1,650 peer-reviewed publications included the term ‘placebo’ in the title compared to 104 for ‘hearing aid’ based on a Google scholar search. Much of this research is in the area of pain management (Turner et al. 1994), though general conclusions can be drawn that are likely to be of relevance for hearing aid trials.

First, placebo response rates - the proportion of participants who show placebo responses - are variable but tend to be very high. An early review reported that rates of around 35% were typical (Beecher 1955), with a more recent review reporting even higher response rates, with an average of 70% (Roberts et al. 1993). Factors thought to influence the placebo response are discussed below, but depend on aspects of the design of the research study, including the nature of the placebo treatment, the

characteristics of the clinician, and the clinical environment. Effect sizes vary considerably. Vase et al. (2002) reported a small mean effect size (Cohen's $d = 0.15$), though effect sizes ranged from -0.95 to +0.57. Generally, effect sizes were smaller in studies where the placebo was the control condition and larger in studies that examined the placebo mechanism itself.

Placebo responses are not just behavioral biases, but have demonstrable physiological effects. For example, in pain research, placebo-generated analgesic responses are blocked by administration of opioid blockers, suggesting that the placebo response is mediated by endogenous opioid systems (Levine et al. 1978). In addition, sham surgery for angina resulted in better exercise tolerance, fewer electrocardiographic abnormalities, and a need for less medication (Dimons et al. 1960).

The onset of placebo effects may be immediate or delayed, and can mimic those of active treatments. For example, the time course of activity of a placebo drug can mimic that of the active treatment (Lasagna et al. 1958). In a study of Parkinson's disease, a placebo deep brain stimulation procedure resulted in an immediate improvement in motor performance (Pollo 2002). Placebo effects can also be long lasting, with improvements in the symptoms of angina for up to a year after sham surgery (Dimons et al. 1960).

In terms of susceptibility to placebo responses, patient characteristics are not good predictors; placebo responses appear to be independent of demographic, intelligence or other personality factors (Shapiro et al. 1984), and individual responses tend to vary across trials (Beecher 1955; Liberman 1964). A range of clinician characteristics has, however, been shown to impact on placebo effects. The clinician's warmth, friendliness, empathy, positive attitude, sympathy and prestige are all

associated with greater positive placebo effects (Price et al. 2008). The clinician's own expectations may also have an effect on outcome. In a study of dental extractions (Gracely et al. 1985), the placebo response of patients was compared for two groups. In one group, the dentist thought he was administering a narcotic analgesic, a placebo or a narcotic antagonist. In the second group, the dentist thought he was administering either a placebo or a narcotic antagonist. Although the actual range of possible treatments was the same for both groups of patients, the first group reported significantly less pain. The interpretation was that the dentist's knowledge of the range of possible treatments resulted in subtle changes in behavior that influenced the patient's response.

Placebo responses themselves are highly specific and selective. Bennedetti et al. (1999) injected the irritant capsaicin into the four limbs of study participants. Participants experienced a reduction in pain only for those limbs that had been treated with a placebo analgesic.

Several mechanisms have been proposed as underlying specific placebo effects, and two that have particular relevance to hearing aid trials are conditioned learning and expectation (Thompson 2000). A learned response is based on classical conditioning: if a person repeatedly experiences positive benefit with medical treatments, they may come to associate benefit with medical interventions generally. The unconditioned stimulus (the placebo treatment) becomes associated with a conditioned stimulus (e.g. penicillin, analgesics) that reliably relieves symptoms. More generally, subtle conditioning may be experienced through encounters with doctors, medicines and other healing encounters. In terms of expectation, some placebo responses seem to be related to a strong expectation of benefit, or expectation of a particular response to a treatment. For example, a group of pharmacology

students received instruction on sedating and stimulating drugs (Blackwell et al. 1972). They were then given placebo tablets; 1 blue pill or 1 pink pill, or 2 blue pills or 2 pink pills. Effects related to both the color of the pill and the dosage. Those that had taken blue pills experienced sedating effects while those that had taken pink pills experienced stimulating ones. The interpretation was that the color of the pill had set up expectations about the sedating or stimulating nature of the pill. Students that had taken a higher ‘dose’ of placebo pills experienced stronger effects, in line with the student’s expectations.

Placebo effects and hearing aid trials

All of the characteristics of placebos reviewed above might plausibly be relevant to hearing aid trials. First, placebo effects occur across a range of pharmacological, surgical, behavioral treatments, and trials of medical equipment. They also occur in a sizable proportion of people. Effect sizes vary, and these encompass the effect sizes reported in clinical hearing aid trials (as discussed below). Second, placebo effects are ‘real’ in a behavioral and physiological sense, and benefits are seen in specific physiological and behavioral measures. Although physiological measures are not typically used in hearing aid trials, one might expect placebo effects on a range of hearing aid benefit measures. Third, placebo effects could mirror exactly what the participant/client (and the clinician) expects in terms of hearing aid benefit and potentially have long-lasting effects. Fourth, clinician characteristics – sympathy, positive attitude, interest in the patient and the clinician’s own expectations of benefit – can influence the patient’s response, and this may also impact both clinical trials and outcomes in a clinical setting. Fifth, placebo effects are highly specific, and this might be plausibly observed in hearing aid trials. For example, a benefit related to

placebo effects might be seen for localization ability but not speech recognition, if it was expected that a particular hearing aid technology being evaluated would especially improve localization. Finally, in terms of the mechanisms underlying placebo effects, both learned responses to use of medical devices and professional clinical attention as well as an expectation of benefit could plausibly affect hearing aid trial performance.

We are aware of only a single peer-reviewed study that has investigated the influence of expectation on measures of hearing aid benefit. Bentler et al. (2003) provided participants with two identical hearing aids, that were labeled as either ‘conventional’ or ‘digital’. Hearing aids were used for one month, and participants then completed a battery of hearing aid benefit measures including speech perception and self-report measures. Overall, the effect of labeling accounted for 2 to 32% of variance in outcome measures. Effects were generally small. With the exception of some subscales of the Abbreviated Profile of Hearing Aid Benefit (APHAB; Cox et al. 1995), differences in outcome measures were not significant individually. However, overall the authors concluded that the expectation that ‘digital’ hearing aids must be better had affected their performance of measures of hearing aid benefit.

In order to investigate patterns of benefit and how they vary according to study design, we compared the number of statistically significant treatment effects, and the size of effect, according to test type (subjective measures, such as benefit questionnaires and sound quality ratings, or behavioral measures, such as speech in noise tests) across a selection of hearing aid trials that utilised unblinded¹ (Arlinger et al. 1998; Bamford et al. 1999; Boymans et al. 1999; Ricketts et al. 1999; Ringdahl et

¹ Unblinded studies are those where both the participant and the experimenter are aware which is the experimental condition.

al. 1990; Valente et al. 1998), single-blinded² (Bille et al. 1999; Boymans et al. 2000; Wood et al. 2004) or double-blinded³ methodology (Hayes et al. 2000; Larson et al. 2002). Note that the above studies were selected on the basis of a Google scholar search with the key words “hearing aid trial”, and on the availability of the full manuscript. Selection was thus not as rigorous as it should be for a reliable meta-analysis and it is possible that the selection method may have introduced bias. However, the aim was to obtain a feel for general patterns of findings and how they might vary according to study design. The selection method resulted in a small selection of published hearing aid studies that are readily available to clinicians and researchers, and this seems a reasonable point from which to proceed.

Our first impression was that the majority of these hearing aid trials did not utilize double-blind methodology. Double-blind studies may be in the minority because of the extra expense and difficulty of conducting such studies. Until recently with the advent of programmable digital hearing aids, new hearing aids looked quite different from old ones and so were impossible use in blinded studies. Additionally, the possible consequences of using unblinded study methodology may not be widely understood within the audiological community.

These studies all used a similar variety of outcome measures. All included a speech in noise task, and most included either a self report questionnaire about hearing aid benefit, sound quality ratings or personal preference. In the discussion that follows, we draw a distinction between what we saw as two main categories of outcome measure; 1) speech in noise tests, which we refer to as ‘behavioral measures’

² Single-blinded studies are those where the participant is unaware of the treatment condition, but the experimenter is aware of the condition. These studies offer partial control of the effects of expectation. Note that the expectations of the experimenter still have the potential to impact upon outcomes, see Gracely et.al. (1985).

³ Double-blind studies are those where neither the participant nor the experimenter are aware which is the experimental condition, and thus control for possible affects of expectation.

and 2) questionnaires, sound quality ratings and personal preference, which we refer to as ‘subjective measures’. Behavioral measures are those that utilize psychophysical methodology and typically require participants to identify speech in background noise. Examples of this type of test included the Hearing in Noise Test (HINT; Nilsson et al. 1994) and Four Alternative Auditory Feature test (FAAF; Foster et al. 1979). Subjective measures are those that require the participant to make a judgment about their experience, typically using categorical or visual analogue scale-based ratings. Examples of this type of measure included questionnaires (Abbreviated Profile of Hearing Aid Benefit; APHAB; Cox and Alexander 1995), Glasgow Hearing Aid Benefit Profile; GHABP; Gatehouse 1999), ratings of sound quality or indicating a particular personal preference from a selection of alternative hearing aids.

For the unblinded studies cited above, five out of six found a significant benefit for the ‘test’ aid for behavioral measures (speech in noise tests), with effect sizes (r) ranging from 0.15 to 0.49. The mean effect size for behavioral measures was 0.25. Three of these studies included subjective sound quality ratings and all reported significant benefit. Likewise, three studies used self-report questionnaires and all found significant benefits for the ‘test’ aid. For these measures, effect sizes varied widely from 0.1 to 0.99. Three studies also asked the participants for their overall personal preference and all found in favor of the ‘test’ aid, with effect sizes ranging from 0.43 to 0.68. The mean effect size for subjective measures (sound quality ratings, self report questionnaires and personal preference) was 0.5.

For the three single blind studies cited above, two studies found a significant benefit for both behavioral and self-report measures. Effect sizes could be calculated for only one of these studies where the necessary statistics were reported; 0.09 for the behavioral test (speech in noise, FAAF) and 0.45 for personal preference.

For the two double blind studies, neither found any significant differences in performance on speech recognition tests. Larson et al's (2002) study found significant differences on ratings of sound quality (effect sizes 0.16 to 0.47) and self report questionnaire (no statistics reported), while no significant differences were reported for the self report questionnaire utilized in the Hayes and Cormier's (2000) study.

Based on this selection of studies, our overall impression is that unblinded hearing aid trials tend to find significant effects more frequently than double-blind trials. Significant differences appear to be more common for subjective measures than for behavioral ones. Where calculable, effect sizes were generally larger for unblinded versus blinded studies and for subjective measures versus behavioral ones.

In summary, in reviewing the characteristics of placebo effects and taking the findings of Bentler et al. (2003) study together with our brief survey of hearing aid trials, it is plausible that placebo effects may impact upon hearing aid clinical trials. If so, placebo effects would need to be taken into account when designing or interpreting hearing aid trials, and this would be an important consideration for all those involved with clinical hearing aid trials; hearing scientists, hearing aid manufacturers, hearing aid dispensers and hearing aid users. We hypothesized that in a setting designed to reflect a scientific hearing aid trial, participants would perform better on a behavioral (speech and noise) test for a hearing aid described to them as containing new technology compared to an identical hearing aid described to them as being a conventional model. Additionally, the 'new' hearing aid would be rated more highly than the 'conventional' one on sound quality judgments, and that the 'new' aid would be preferred overall. Bentler et al examined the effect of labeling hearing aids as 'digital'. This study extends the work of Bentler et al by examining more broadly the effect of participant expectation in clinical hearing aid trials. Given that a general

pattern in unblinded hearing aid trials was for larger effect sizes for subjective measures, we expected larger effect sizes for the two subjective measures (sound quality ratings and personal preference) compared to the speech in noise test. These hypotheses were examined as follows.

MATERIAL AND METHODS

Participants

With the exception of Larson et al.'s (2000) study, all of the hearing aid trials mentioned above utilized small to medium sample sizes (N 's 12 to 50). We selected a sample size of $N = 20$ (13 males, 7 females) in order to provide similar statistical power to a typical hearing aid trial. Participants were recruited from a local hospital-based audiology clinic and ranged in age from 54 to 80 years ($M = 69$ years, $SD = 6$ years). Inclusion criteria were i) symmetrical mild-to-moderate, sloping high frequency sensorineural hearing impairment (mean of ≥ 45 dB HL at 2-6 kHz), with ≤ 5 dB difference between the ears at two or more adjacent frequencies (between 0.25 and 8 kHz), ii) normal middle ear function based on tympanometry, iii) daily unilateral hearing aid use with minimum of one year experience with a digital non-linear BTE hearing aid and iv) English spoken as a first language. Pure tone audiometry was carried out in a sound-treated room using a calibrated Kamplex KLD 21 audiometer with TDH-39 headphones and B-71 bone vibrator. A GSI 38 Auto Tymp was used to assess middle ear function.

When invited to take part in the study, participants were sent an information sheet that stated that the purpose of the study was to evaluate new hearing aid technology. On completion of testing, they were informed of the true purpose of the study. Ethical approval was obtained from the relevant authorities and written

consent was obtained from all participants. Participants had travel costs reimbursed. No other compensation was provided.

Test hearing aids

The hearing aids used were two Starkey A312 Strata behind-the-ear digital instruments with seven band, three channel wide dynamic range compression with a noise reduction algorithm. The hearing aids had interchangeable yellow and beige covers. All participants used their current ear mould (typically hard acrylic with pressure equalization vent) with the test hearing aids. Both test hearing aids were programmed to the same NAL-NL1 prescription target based on a typical audiogram for age-related sensorineural hearing loss (35 dB HL at 500 Hz, 40 dB HL at 1 kHz, 50 dB HL at 2 kHz, 60 dB HL at 4 kHz, 80 dB HL at 8 kHz). The NAL-NL1 real ear insertion gain prescription target for these pure tone thresholds is shown in the first line of Table 1. Following programming, three repeated coupler measures were made with each hearing aid using an Otometrics Aurical Plus testing system. The mean and standard deviation of these three measures is shown in Table 1. The mean difference in gain between the two test hearing aids was 3.1 dB (between 250 Hz to 4kHz; maximum difference 3.4 dB at 4 kHz). Coupler measurements were also carried out after switching covers between the hearing aids and at the completion of testing to ensure a stable gain response. These measures are shown in Table 1. Overall, differences between repeated coupler measures ranged between 0 and 3 dB, probably largely due to variation in test box position.

Subjective listening checks were carried out by three experienced hearing aid users. BKB sentences (as described below) were played over loudspeaker and out of the listener's view, the hearing aids were either switched or kept the same. After

listening in both conditions, listeners were asked “Does this one sound the same or different to the other hearing aid?” No differences were reported.

Additionally, casings were switched between the hearing aids so that for half the group, one hearing aid had a yellow case and was referred to as the ‘new’ hearing aid while the other was the ‘conventional’ aid with the beige case, while for the other half of the group, the casings and associated designations (‘conventional’ or ‘new’) were switched. This procedure was utilized to control for any actual differences that might exist between the two hearing aids.

Outcome measures

For all measures, participants were seated 1.5 meters from a loudspeaker at zero degrees azimuth. For all the auditory stimuli described below, sound levels refer to the level measured at the reference position, which was defined as the center of the participant’s head with the participant absent.

Speech in noise test

The Four Alternative Auditory Feature (FAAF) test (Foster and Haggard 1979) was used to assess aided speech perception in noise. The FAAF test is a computerized 80 item, single-syllable, closed set word recognition test. Participants are asked to identify the key word embedded in the carrier phrase ‘Can you hear X clearly?’ The participant then selects the word they thought they heard via mouse click from four alternative responses displayed on the computer screen. The target word is presented along with three foils that differ from the target word by either one or two distinctive features. For example, the target word might be “lads” with the foils “land”, “lad” and “lands”. The FAAF test provides an overall percent correct score as well as a

breakdown of errors according to phonetic feature, though only the overall score is reported in this study. The FAAF test was administered with test items at 65 dB (A) in speech-shaped noise at +2 dB SNR. This level of presentation was the same as that used in Wood and Lutman's (2004) hearing aid trial, which also used the FAAF test. This presentation level allows for typical performance accuracy above 50% correct, while remaining challenging enough to be below ceiling. The fixed speech-shaped noise is designed to have the same long-term average spectrum as the test items. Prior to completing the FAAF test, participants familiarized themselves with the task by doing a short run of 12 FAAF test items. The FAAF test was performed with either the 'conventional' or the 'new' hearing aid first, with order of testing counter-balanced across participants.

Sound quality rating test

To assess sound quality, a rating procedure was used based on that used by Arlinger et al. (1998). Participants listened to six different sound samples (described below) and rated them on three dimensions; clarity, comfort and overall impression. The order of testing (i.e. 'new' or 'conventional' hearing aid first) was counter-balanced across participants. Ratings were awarded using a 10-point visual analogue scale (e.g. for overall impression, '1' is 'poor' and '10' is 'excellent') similar to that used by Larson et al (2000). Participants were allowed to listen to each sample as many times as they liked before assigning a rating. Sound samples were; Bamford-Kowal-Bench (BKB) sentences (Bench et al. 1979); for example, "The cat caught a mouse" or "The milk came in a bottle") spoken by male and female voices in quiet and in noise, music (excerpt from Beethoven's symphony number 1) and an environmental sound (robin song). All samples were approximately 10 seconds long, digitized at 44 kHz,

equalized 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 with the same long-term RMS power as the sentences.

Personal preference

At the end of the test session, participants were asked to state their overall personal preference using one of the three categories: ‘the new hearing aid is best’, ‘the conventional hearing aid is best’ or ‘I cannot tell any difference’, as used by Arlinger et al. (1998).

Procedure

At the start of the session, participants were given an explanation about the (false) aims of the study, i.e. to evaluate some new hearing aid technology. Further questions from participants were anticipated at this point, such as “What is the new technology?”, “How does the new technology improve hearing aid function?” and “What sort of benefit does the new technology provide?” Bogus responses to these questions were prepared, for example, in response to the question “What sort of benefits does the new hearing aid provide?” the tester was to answer that the new technology was thought to improve sound quality and speech recognition in noise. However, somewhat surprisingly, no participants had any questions about the trial at this point. Perhaps participants were simply willing to place their trust in the tester and comply with the requirements of testing.

Following the (false) study explanation, participants were shown the two test hearing aids. The hearing aid with the yellow case was introduced to the participant as the ‘hearing aid with new technology’. The reason given for the yellow case was that

this was a prototype hearing aid. The hearing aid with the beige case was introduced as the ‘conventional hearing aid’. Otoscopy, pure tone audiometry and tympanometry were then carried out, followed by the test protocol in the following order: FAAF test, sound quality ratings and personal preference. Test sessions were approximately 1 hour in duration.

RESULTS

Following tests for normality, non-parametric measures for paired data were selected for FAAF test and sound quality ratings.

FAAF test

For the FAAF test, data from a single participant were excluded as responses were at chance level. It transpired that the participant was unable to read the response alternatives on the computer screen and had been guessing responses at random. For the remaining 19 participants, there was a non-significant tendency for better performance with the ‘new’ hearing aid over the ‘conventional’ aid, based on Wilcoxon signed-rank test ($M = 62.3\%$, $SD = 10.4$ versus $M = 60.7\%$, $SD = 9.0$; $z = -1.84$, $p = 0.06$, two-tailed, $r = .08$).

Sound quality rating test

Both hearing aids tended to be rated positively, although the ‘new’ hearing aid was consistently rated more highly than the ‘conventional’ one on the dimensions of comfort, clarity and overall impression (based on based on Wilcoxon signed-rank test). Grand mean sound quality rating (average of all subjective ratings) was

significantly different between aids, with the ‘new’ aid being rated more favourably.

Mean ratings and details of statistical comparisons are shown in Table 2.

Insert Table 2 here

Personal Preference

Fifteen participants preferred the ‘new’ hearing aid while five could not tell the difference. None preferred the ‘conventional’ aid. This difference was statistically significant [$X^2(1) = 5.00, p < .05, r = .50$].

DISCUSSION

The hypothesis that participants would perform better on tests of hearing aid benefit with a hearing aid described as containing new technology was partially supported. Significant differences in favor of the ‘new’ hearing aid were seen on ratings of sound quality and personal preference between two identical hearing aids that were described to participants as ‘new’ and ‘conventional’. However, there was no significant difference on the performance in a speech in noise task (the FAAF test). The inference is that describing one hearing aid as ‘new’, presenting it in the context of a hearing aid trial within an impressive setting (i.e. a University-based sound proof booth fitted with technical testing equipment) set up an expectation in participants that it must be better than a hearing aid labeled as ‘conventional’. This expectation then impacted upon participants’ performance. For sound quality and personal preference, medium to large effect sizes observed. This finding was also consistent with our hypothesis that subjective measures (such as sound quality ratings and personal preference) would be more influenced by expectation than behavioral ones (such as

the FAAF test). If participant's expectation influences performance, then it seems reasonable that subjective measures that rely only on participant's self report or opinion should be highly influenced by expectation. The FAAF test was perhaps less influenced because it utilizes psychophysical methodology with high test-retest reliability (Foster & Haggard, 1979) and is thus reasonably resistant to effects of expectation, presumably mediated by motivation or attention. Note that alternative speech tests utilizing different stimuli and procedures may be more or less susceptible to the effect of expectation on performance depending on how susceptible they are to changes in participant motivation or attention. In actuality, we suspect that the FAAF test was slightly influenced by expectation in this study. Although the difference was small and statistically non-significant ($p = .06$), the difference was in the expected direction and fitted with the general pattern of findings on the other outcome measures. This study was statistically underpowered for the detection of a small effect on FAAF test performance. We advise caution in relation to placebo effects and speech in noise tests; any placebo effect is likely to be small, although this small effect size is not dissimilar to the small effects typically sought in clinical hearing aid trials.

This study sought to investigate the effect of participant expectation on tests of hearing aid benefit by imitating the design of a typical uncontrolled hearing aid trial. This study differs from an actual uncontrolled hearing aid trial in one potentially important respect, however. In an actual uncontrolled hearing aid trial, both the participant and the (hopeful) experimenter would likely have an expectation that the 'new' hearing aid would perform better than the 'conventional' one. In the current study, although participants may have had an expectation of better performance from the 'new' hearing aid, the experimenter was aware that the aids were in fact identical.

Research reviewed in the introduction (Gracely et al., 1985) suggests that the expectation of the experimenter can subtly impact upon participants' expectations. Therefore in the case of the current study, the experimenter's expectations may have encouraged a tendency towards null findings. Placebo effects in actual uncontrolled hearing aid trials may in fact be larger than those reported in this study. Alternatively, given that our hypothesis was that there *would* be differences between the hearing aids, the experimenter's expectation of difference may have had the opposite effect, i.e. to encourage a bias in favor of the 'new' hearing aid. Disentangling the effects of participant and experimenter expectations on outcome would require deception of both the participant and the person collecting the data (i.e. so that the data collector also believed this was a real trial of new hearing aid technology), and this was not possible in the current study.

One of the general conclusions of this study is that placebo effects need to be controlled in order to have confidence in the results of clinical hearing aid trials, ideally by using double-blind methodology in a randomized controlled study. However, the majority of hearing aid trials do not use double-blind designs, perhaps due to practical limitations. For example, the new device may look physically different from the conventional one, or because financial or staffing limitations preclude conducting a more elaborate double-blind study design. How should hearing aid trials that do not use double-blind designs be interpreted? Or, stated another way, do studies that do not use double-blind designs provide reliable data? Glasziou et al. (2007) discussed this issue in relation to clinical medical trials. They concluded that in circumstances where the treatment effect was so large that it was unlikely to be a reflection of bias or factors other than treatment effect, it may be permissible to rely on an uncontrolled study design. Glasziou et al. provide examples where the treatment

effect was around ten times the size of the expected outcome based on the usual natural progression of the condition under examination. This principle is potentially applicable to clinical hearing aid trials. However, the difficulty seems to lie in determining how large an effect could be attributable to treatment. Literature reviewed in the introduction suggests that the effect size attributable to placebo effects is variable according to various design factors. In any case, very large effects are not normally observed in trials of new hearing aid technology, at least with psychophysical outcome measures. The small effect sizes that are typically sought are comparable in size to those observed in this study, which were due to expectation alone. An alternative strategy to control for placebo effects without the use of a double-blind methodology might be to include outcome measures that are not expected to change with the new treatment along with some that were expected to change. Specific changes in particular outcome measures might then be taken as evidence for a ‘real’ treatment effect. However, as reviewed in the introduction, placebo effects can be highly specific and may mimic the pattern of findings expected by researchers and participants. This strategy is thus also potentially open to placebo effects, and is not a reliable research design. A second alternative could be to use single-blind methodology. Single-blind hearing aid trials are preferable to unblinded trials as they do control for the effects of participant expectation. However, expectations of the researcher may still potentially affect results (Gracely et al. 1985). In summary, there do not appear to be any good alternatives to the gold standard double-blind randomized controlled study design if one wishes to rigorously control for placebo effects. Other types of study design should be used (and the findings thereof interpreted) with caution.

The focus of this study was participant expectation in the context of clinical hearing aid trials. One might expect participant expectations to influence satisfaction, benefit and compliance in a clinical setting as well as a research one, and this may be a productive area of future research. Research reviewed in the introduction also suggests that placebo effects occur widely across populations, clinical conditions and treatments. Conclusions of this study may then be relevant to other areas of audiology. Researchers and clinicians may benefit from an awareness of possible placebo effects in other audiological populations and conditions, such as tinnitus or hyperacusis.

Finally, and perhaps most controversially, there is a potential role of placebo effects in clinical audiology. Reviews of the placebo effect in relation to clinical medicine, cited earlier in this manuscript (Price et al. 2008; Thompson 2000; Turner et al. 1994) suggest that placebo effects could be (and are) ethically used to optimize the outcome for patients. Such effects might be fostered by use of suggestion or via positive interactions between the clinician and the patient. In audiology, this could, for example, take the form of emphasizing the empirically proven benefits of hearing aids to a new hearing aid user during an initial fitting. This, combined with a warm, confident and professional attitude on the part of the clinician, could result in greater benefit and compliance on the part of the hearing aid user. Of course, this is a matter for careful research in verifying and quantifying such potential benefits, and for informed debate among audiologists and professional bodies as to how placebo effects might ethically be used in clinical practice. In any case, placebo effects and the manipulation of expectation potentially offer an additional tool to maximize real benefit for audiology patients.

CONCLUSIONS

In this study, measures typical of those used in clinical hearing aid trials were influenced by participant's expectations. This study suggests that there is a need to control for placebo effects in hearing aid trials, as well as need for caution in interpreting findings from trials that did not include such controls.

ACKNOWLEDGEMENTS

Thank you to patients and staff at the Audiology Department of Withington Community Hospital, South Manchester who supported this research. Thank you to Keith Wilbraham for technical support and to Starkey Laboratories Ltd (UK) who provided the hearing instruments and funding for participant transport. The first author (PD) is supported by funding from Starkey Laboratories Inc (US).

REFERENCES

- Arlinger, S., Billermark, E., Oberg, M., et al. (1998). Clinical trial of a digital hearing aid. *Scandinavian Audiology*, 27, 51-61.
- Bamford, J., McCracken, W., Peers, I., et al. (1999). Trial of a two-channel hearing aid (Low-frequency compression-High-frequency linear amplification) with school age children. *Ear & Hearing*, 20.
- 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., Arduino, C., Amanzio, M. (1999). Somatic activation of opioid systems by target-directed expectations of analgesia. *Journal of Neuroscience*, 19, 245-248.
- Bentler, R. A., Neibuhr, D. P., Johnson, T. A., et al. (2003). Impact of digital labelling on outcome measures. *Ear & Hearing*, 24, 215-224.
- Bille, M., Jensen, A., Kjaerbol, E., et al. (1999). Clinical study of a digital vs an analog hearing aid. *Scandinavian Audiology*, 28, 127-135.
- Blackwell, B., Bloomfield, S., Buncher, C. R. (1972). Demonstration to medical students of placebo effects and non-drug factors. *Lancet*, 1, 1279-1282.
- Boymans, M., Dreschler, W. A. (2000). Field trials using a digital hearing aid with active noise reduction and dual-microphone directionality. *Audiology*, 39, 260-268.
- Boymans, M., Dreschler, W. A., Schoneveld, P., et al. (1999). Clinical evaluation of a full-digital in-the-ear hearing instrument. *Audiology*, 38, 99-108.
- Brody, H. (1980). *Placebos and the philosophy of medicine. Clinical, conceptual, and ethical Issues*. Chicago: University of Chicago Press.
- Cox, R. M., Alexander, G. C. (1995). The Abbreviated Profile of Hearing Aid Benefit. *Ear & Hearing*, 16.
- de Craen, A. J., Kaputchuk, T. J., Tijssen, J. G., et al. (1999). Placebos and placebo effects in medicine: Historical overview. *Journal of the Royal Society of Medicine*, 92, 511-515.
- 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.

- Foster, J. R., Haggard, M. P. (1979). FAAF. An effective analytical test of speech perception. *Proceedings of the Institute of Acoustics*, 182, 9-12.
- Gatehouse, S. (1999). Glasgow Hearing Aid Benefit Profile: Derivation and validation of a client-centered outcome measure for hearing aid services. *Journal of the American Academy of Audiology*, 10, 80-103.
- Glasziou, P., Chalmers, I., Rawlins, M., et al. (2007). When are randomised trials unnecessary? Picking signal from noise. *BMJ*, 334(7589), 349-351.
- Gracely, R. H., Dubner, R., Deeter, W. R., et al. (1985). Clinician's expectations influence placebo analgesia. *Lancet*, 1, 43.
- Hayes, D. E., Cormier, K. L. (2000). Double blind comparison of three hearing aid circuits with new hearing aid users. *Scandinavian Audiology*, 29, 10-19.
- Larson, V. D., Williams, D. W., Henderson, W. G., et al. (2002). A multi-center, double blind clinical trial comparing benefit from three commonly used hearing aid circuits. *Ear & Hearing*, 23, 269-276.
- Lasagna, L., Laties, V. G., Dohan, J. L. (1958). Further studies of the 'pharmacology' of placebo administration. *Journal of Clinical Investigation*, 37, 533-537.
- Levene, J. D., Gordon, N. C., Fields, H. L. (1978). The mechanisms of placebo analgesia. *Lancet*, 2, 654-657.
- Liberman, R. (1964). An experimental study of the placebo response under three different situations of pain. *Journal of Psychiatric Research*, 2, 233-246.
- Nilsson, M., Soli, S. D., Sullivan, J. A. (1994). Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America*, 95, 1085-1099.
- Pollo, A. (2002). Expectation modulates the response to subthalamic nucleus stimulation in Parkinsonian patients. *NeuroReport*, 13, 1383-1386.

- 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.
- Ricketts, T., Dhar, S. (1999). Comparison of performance across three directional hearing aids. *Journal of the American Academy of Audiology*, 10, 180-189.
- Ringdahl, A., Erikson-Mangold, M., Israelsson, B., et al. (1990). Clinical trials with a programmable hearing aid set for various listening environments. *British Journal of Audiology*, 24, 235-242.
- Roberts, A. H., Kewman, D. G., Mercier, L., et al. (1993). The power of nonspecific effects in healing: Implications for psychosocial and biological treatments. *Clinical Psychology Review*, 13, 375-391.
- Shapiro, A. K., Shapiro, E. (1984). Patient-provider relationships and the placebo effect. In J. D. Matarazzo, S. M. Weiss, J. A. Herd, et al. (Eds.), *Behavioral health: A handbook of health enhancement* (pp. 371-383). New York, NY: Wiley-Interscience.
- 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., et al. (1994). The importance of placebo effects in pain treatment and research. *JAMA*, 271, 1609-1614.
- Valente, M., Fabry, D. A., Potts, L. G., et al. (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.
- Vase, L., Riley, J. L. I., Price, D. D. (2002). A comparison of placebo effects in clinical analgesic trials versus studies of placebo analgesia. *Pain*, 99, 443-452.

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. Real ear insertion gain targets and measured coupler gain for the test hearing aids

	250 Hz	500 Hz	750 Hz	1000 Hz	2000 Hz	3000 Hz	4000 Hz
NAL-NL1 target (65dB)	4	6	12	17	27	32	31
Hearing aid 1 mean gain (SD)*	3.3 (2.1)	6 (1)	11.3 (1.1)	15 (1.7)	19.3 (1.1)	27.7 (1.1)	27.3 (1.1)
Hearing aid 2 mean gain (SD)*	3.7 (.6)	9.3 (.6)	13.3 (.6)	17.7 (1.1)	22.3 (.6)	31 (0)	30.7 (.6)
Hearing aid 1 gain following case swap	2	8	12	17	21	29	29
Hearing aid 2 gain following case swap	2	8	11	15	21	29	29
Hearing aid 1 gain post-data collection	4	9	13	18	21	29	28
Hearing aid 2 gain post-data collection	4	9	13	19	21	30	28

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

Table 2. Sound quality ratings for ‘new’ and ‘conventional’ hearing aids

	Comfort	Clarity	Overall impression	Overall sound quality rating
Mean rating ‘new’	8.95	9.28	9.00	9.12
(SD)	(1.12)	(1.15)	(1.11)	(1.02)
Mean rating	8.40	8.61	8.1	8.35
‘conventional’ (SD)	(1.21)	(1.28)	(1.47)	(1.17)
Z*	-1.94	-2.77	-2.98	-2.88
p (two tailed)	.053	.006	.003	.004
Effect size (r)	.23	.27	.33	.33

*Wilcoxon signed-rank test statistic