Self-report outcome in new hearing-aid users: Longitudinal trends and relationships between subjective measures of benefit and satisfaction

Martin D. Vestergaard
Ørsted-DTU Acoustic Technology, Technical University of Denmark Lyngby, Denmark, and Research Centre Eriksholm, Snekkersten, Denmark

Abstract

This study focussed on self-report outcome in new hearing-aid users. The objectives of the experiment were changes in self-report outcome over time, and relationships between different subjective measures of benefit and satisfaction. Four outcome inventories and a questionnaire on auditory lifestyle were administered to 25 hearing-aid users repeatedly after hearing-aid fitting, and assessments took place one week, four weeks, and 13 weeks after hearing-aid provision. The results showed that, for first-time users who used their hearing aids more than four hours per day, self-reported outcome increased over 13 weeks in some scales, although there was no change in amplification during this time. Furthermore, it was found that, for data collected immediately post-fitting, some subscales were much less face valid than for data collected later. This result indicates that the way in which hearing-aid users assess outcome changes over time. The practical consequence of the results is that early self-report outcome assessment may be misleading for some self-report outcome schemes.

Keywords

One aim of self-report outcome assessment is to evaluate the degree to which provision of hearing aids helps individual patients in overcoming their hearing difficulties. While objective outcome measures, such as speech recognition scores, measure the benefit provided by the amplification and other technical features of a hearing aid, self-reporting has the potential to evaluate the entire process of hearing-aid intervention.

Self-report outcome includes elements such as counselling and fine-tuning and takes into account individual needs. Self-reporting schemes therefore address questions about pre-intervention disability and handicap, perceived benefit, residual disability, satisfaction, daily use of hearing aids and health-related quality of life. In clinical practice, it is not uncommon to find a discrepancy between hearing-aid benefit in terms of speech understanding and self-experienced outcome, and for that reason there has been considerable interest in investigating relationships among functional outcomes and satisfaction. Much effort has been concerned with developing questionnaires that are sensitive in the domains of self-experienced outcome. In 1999, participants at the second audiological workshop at

Martin D. Vestergaard Centre for the Neural Basis of Hearing, Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3EG, UK. E-mail: mdv23@cam.ac.uk.

Eriksholm discussed issues concerning measuring outcome in audiological rehabilitation (Cox, 2000), and they proposed a short inventory as a supplement to existing outcome schemes (Cox et al, 2000). For more on self-report outcome measures, see Gatehouse (2001) or Wong et al (2003).

A few studies have considered the (sometimes limited) quality of the various quantitative measures of self-reported benefit and satisfaction that can be derived from questionnaires. Wong et al (2003) concluded that many new hearing-aid users have expectations that are too high, and speculated that if expectations are not met then they are likely to change over time. If this happens, then the relationship between expectation and benefit changes, which could cause satisfaction to go up over time while benefit really stays the same. Wong et al discussed how responses to items that are supposed to measure the quality of hearing healthcare correlate strongly with those measuring the technical performance of the care process. They argue that being satisfied in certain specific situations does not necessarily imply general satisfaction. Furthermore, many self-report benefit measures have been developed by audiologists with many years of experience with hearing-aid rehabilitation. That might entail questions that are not relevant for new users of hearing aids (e.g. items addressing aspects of the care process, which new users have not yet experienced). Therefore the relationships between the quality and validity of self-report benefit measures need to be established.

A common approach is to use factor analysis to establish the number and nature of distinct outcome dimensions (e.g. Cox & Alexander, 2001, 2002; Humes, 1999; Humes et al, 2001; Kramer et al, 2002; Stephens, 2002). For example, SADL (satisfaction with amplification in daily life) has four independent components (Cox & Alexander, 2001), and IOI-HA (international outcome inventory for hearing aids) has two (Cox & Alexander, 2002; Kramer et al, 2002; Stephens, 2002). Humes et al (2001) reported a study in which a large number of outcome measures, objective as well as subjective, were collected. Independent component analysis was used to deduce seven distinct outcome dimensions, four of which were to do with self-experienced outcome (subjective benefit and satisfaction, aid-use, reduction of hearing handicap, and sound quality).

Some studies have shown that hearing-aid users might improve performance over time post-fitting (Cox & Alexander, 1992; Cox et al, 1996; Gatehouse, 1992, 1993; Horwitz & Turner, 1997; Kuk et al, 2003; Munro & Lutman, 2003). Kuk et al (2003) found that improvement in auditory performance evened out after one month, whereas Gatehouse (1992) suggested that improvement might last up to 18 weeks. However, there are also studies that have failed to show improvement over time (Bentler et al, 1993a; Humes & Wilson, 2003; Mulrow et al, 1992; Saunders & Cienkowski, 1997; Surr et al, 1998; Taylor, 1993). Auditory acclimatization is defined as improvement in auditory performance over time by Arlinger et al (1996) who also note that any such improvement must not be linked to task-related learning underlying the measurement procedures (reviews of acclimatization studies can be found in Gatehouse, 1996; Palmer et al, 1998; Turner et al, 1996). The term 'acclimatization' appears to date back to Kapteyn (1977) who asked his patients to self-assess how long they [had] needed to get used to wearing hearing aids, but the concept initially stems from Watson and Knudsen (1940) who reported that one of their participants improved auditory performance over time. It is obvious that auditory performance must be related to hearing-aid outcome, which can be measured either objectively (functional outcome) or subjectively (self-report outcome). While most of the studies cited above measured functional outcome, a few (cited below) also included self-report outcome.

The question of whether self-report outcome changes over time has not been resolved. Cox et al (2004) reported that overall self-report outcome tended to decrease somewhat over

time, and that changes over time were associated with certain aspects of personality. However, 75% of the participants reported stable outcome over time. The question is whether or not auditory acclimatization occurs for self-experienced outcome. While there are a few studies that appear to provide evidence in favour of auditory acclimatization for self-experienced outcome (Cox & Alexander, 1992; Horwitz & Turner, 1997), several studies have failed to demonstrate auditory acclimatization in the self-report domain (Bentler et al, 1993b; Humes & Wilson, 2003; Mulrow et al, 1992; Surr et al, 1998; Taylor, 1993). However, in most studies where no acclimatization was found, the evaluations did not start until three to six weeks post-fitting.

Kapteyn (1977) found that it took first-time users longer to acclimatize than users with previous hearing-aid experience, and that those who took the longest to acclimatize also reported the lowest satisfaction. Munro & Lutman (2004) reported a study in which one of their subject groups showed a small but statistically significant improvement in self-report outcome, while in another group, using a different version of the same questionnaire (GHABP), no such improvement was observed. They suggested that self-reporting is unfit for measuring changes in auditory performance over time because the result will depend on the type of questionnaire used. Kuk et al (2003) postponed initial self-report outcome assessment by one month because self-reporting without any experience might arguably be meaningless. They failed to show evidence of acclimatization in the self-report domain, and concluded that future studies should seek to examine self-report benefit and satisfaction closer to initial fitting.

The present study investigated the extent to which self-report benefit and satisfaction changed over 13 weeks. It also assessed the validity of, and relationships between, various measures of self-report outcome. To unravel some of the divergent findings regarding the existence of auditory acclimatization in the self-report domain, the objectives of the study were achieved by administering *several* self-report schemes repeatedly over time starting *immediately* post fitting. Aspects of validity and reliability of scales and subscales and verbal and intellectual sophistication of the items in the questionnaires are discussed.

Experiment design

Five questionnaires were included in the study: the Glasgow hearing aid benefit profile (GHABP), the international outcome inventory for hearing aids (IOI-HA), the hearing aid performance questionnaire (HAPQ), satisfaction with amplification in daily life (SADL) and auditory lifestyle and demand (ALD). Brief introductions to the questionnaires are provided in the following subsections. The evaluations took place in the framework of a larger-scale study (Vestergaard, 2004) on benefits from amplification in people with precipitous hearing loss.

Questionnaires

GHABP uses six predefined subscales. They address initial disability and handicap (pre-intervention); and aid-use, hearing-aid benefit, residual disability, and satisfaction (post-intervention) in four predefined and four optional user-nominated listening situations. These subscales have been validated by using a paradigm in which sensitivity to alterations in audibility was the main factor determining the selection of predefined listening situations (Gatehouse, 1999b).

IOI-HA is a seven-item inventory for outcome assessment not intended by the inventors to be a substitute for, but rather to supplement, other outcome schemes (Cox et al, 2000). IOI-HA has been validated in the original English version by Cox and Alexander (2002) for a group of subjects in the Americas, and by Stephens (2002) for a group of English-speaking

patients in Wales, and in a Dutch translation by Kramer et al (2002). The authors identified two independent subscales that will here be designated Introspection and Interaction, as suggested by Cox and Alexander (2002).

HAPQ uses a visual analogue scale for 18 items addressing contentment with hearing aids in various specific listening situations. Previous work with this questionnaire at the Research Centre Eriksholm, has indicated that between one and five independent subscales can be derived from the responses to the 18 questions.

SADL consists of 15 items on hearing-aid performance and user satisfaction, and four additional questions concerning hearing-aid experience and use. Cox & Alexander (1999) found that SADL comprised four subscales, which they labelled positive effect, service and cost, negative features, and personal image (Cox & Alexander, 2001). The SADL subscales are largely independent, as is also the case for the IOI-HA subscales, arguably due to the extraction method that in all cases was based on independent component analysis (Cox & Alexander, 2002; Kramer et al, 2002; Stephens, 2002).

ALD is a questionnaire intended to assess the richness of the auditory environments in which individuals are required to function (Gatehouse, 1999a). The questionnaire presents 25 listening situations, and for each situation subjects were asked to indicate how often the situation occurred and how important it was. Gatehouse found a high correlation between ALD scores and the variability of sound levels encountered during the day by listeners wearing a dosimeter (a sound-level logging device).

Danish translations of the five questionnaires were used. GHABP comes in an official Danish version, which is distributed via the website of the Institute of Hearing Research (www.ihr.gla.ac.uk). IOI-HA has also been presented in an official Danish translation (Cox et al, 2002); however neither GHABP nor IOI-HA has yet been validated in the Danish versions. HAPQ and ALD have been used in Danish in several unpublished studies at the Research Centre, Eriksholm. Only SADL was not available in Danish prior to the experiment. SADL was translated into Danish by the author of the present paper and translated back by a native English speaker who had no a priori expertise in SADL, so as to validate the Danish translation, as suggested by Arlinger (2000). The English and Danish wordings are given in Vestergaard (2004).[1] It should be mentioned that while the ALD version used in this study is the questionnaire originally devised by Gatehouse, he left out item 9 in his later studies (Gatehouse, 1999a; Gatehouse et al, 2000).

Subjects

Twenty five subjects were recruited from an audiological clinic at a local hospital. They had steeply-sloping hearing losses with an average (and standard deviation) slope of 55 dB/oct (s.d. 17 dB/oct) and an average corner frequency of 1255 Hz (s.d. 595 Hz). The corner frequency for each ear was determined by visual inspection of a hearing-threshold curve measured with 1/6th oct frequency resolution. At the corner frequency, the average hearing threshold was 14 dB HL[2] (s.d. 10 dB HL).

Figure 1 shows the range of HL values for the participants along with a typical audiogram. Three women and 22 men with an average age of 60.4 years (s.d. 10.8 yrs) took part. Five subjects had previous hearing-aid experience while 20 were first-time users.

---

[1]Also available at www.pdn.cam.ac.uk/cnbh/mdv/arc/
[2]HL (hearing level) is only defined for standard audiometric frequencies; the HL values at other frequencies were determined from thresholds in dB SPL by linear interpolation along logarithmic frequency of the RETSPL (reference equivalent threshold sound pressure level) values for the Senheiser HDA200 headphones.

Protocol

A week before the assessments started, the subjects were bilaterally fitted with new hearing aids, provided at no cost. The subjects filled in the outcome questionnaires three times, one week, four weeks and 13 weeks post-fitting. They were instructed to fill in the questionnaires on their own, the night before each session, but at each session, the experimenter was available to assist the subjects in case of any doubts regarding the meaning of the items. This was done to reduce potential bias of the results caused by the experimenter being involved in the scoring process, while limiting the number of items left unanswered. ALD was administered before HA fitting and repeated 13 weeks post-fitting. This schedule was optimized for the larger-scale study (Vestergaard, 2004) in which the assessments took place.

For GHABP, items concerning initial disability and handicap were filled in during the introduction to the questionnaire. In subsequent presentations of the questionnaire, these answers were carried forward to give the subjects the possibility of recalling their initial references. For SADL, items 14 and 15 were not used because they address cost and reliability of the hearing instruments, issues that were expected to be irrelevant when evaluating a new hearing instrument provided at no cost.

Hearing aids

The hearing aids were Oticon Adapto. It is a programmable automatic multi-band instrument with two-channel compression and no volume control. They were programmed with the manufacturer's fitting software, 'Genie', to a modified version of the proprietary fitting rationale 'SKI'. With the SKI rationale, the cross-over frequency of the two channels is matched to the cutoff frequency of the hearing loss in such a way that a different compression strategy can be programmed for low frequencies, where the user has mild hearing impairment, than for the high frequencies in which the ski-slope hearing loss is characterized by being severe to profound. The SKI rationale prescribes fast compression in the low frequencies and moderate linear amplification in the high frequencies. In the present experiment 10 dB extra gain at high frequencies was prescribed to ensure that a significant increase in high-frequency audibility was established compared to unaided listening. In practice this modification involved prescribing about linear half-gain in the high frequencies, and non-linear gain as recommended by the manufacturer in the low frequencies. If the prescribed amplification caused acoustic feedback, then gain was reduced in one or more 1000 Hz wide high-frequency bands as necessary to eliminate the feedback. The automatic feedback cancellation system of the instrument was made active, whereas the adaptive gain reduction scheme, which is intended to increase comfort in noisy situations, was not made available. The hearing aids were fine-tuned at the initial fitting, but the subjects were not offered fine-tuning during the later course of the study. While it is recognized that this would be unusual in clinical practice, it was done to keep constant the amount of audibility provided by the hearing aids during the course of the experiment.

Shell style was chosen according to normal audiological practice, in which gain requirements and cosmetic preferences are considered. Vent size was prescribed by the fitting software according to the manufacturer's Open Ear Acoustics philosophy.

Analysis method

The categorical answers to the outcome questionnaires were transformed to numbers, normalized and reversed in such a way that values from zero to one represented the response categories from poorest to best, respectively. Measures of disability and handicap were reversed too, so that large values indicate little disability and handicap.

For ALD, the answers were transformed to values from 0–100 arbitrary units. The score for a particular listening situation was the frequency score weighted by the importance score, and the total ALD score was the weighted scores averaged across the 25 situations. In this way larger values represent a richer auditory lifestyle.

All subjects but one completed the three sessions. One subject did not complete the final session, and in subsequent data analyses, this subject was excluded or his missing data were replaced with means, as appropriate for the various statistical methods.

Identifying subscales

The data from the outcome questionnaires were subjected to factor analysis using principal component extraction and varimax rotation (Kaiser, 1958), and the number of factors to consider was chosen according to the Kaiser criterion (Kaiser, 1960). The purposes were first to investigate whether subscales identified by other authors were valid, and second to look into the reliability over time of the underlying factor structures. Since the questionnaires were filled in three times, each dataset was analysed independently to identify how outcome dimensions change over time.

For GHABP, the performance subscales are predefined: benefit, residual disability, and satisfaction. Another meaningful predefined subscale was daily hearing-aid use, which in this study was defined as the average value of the normalized items: GHABP item 3, IOI-HA item 1, and SADL additional item 3, which all address hearing-aid use and hence would be expected to provide a valid subscale for hearing-aid use (per day).

Validity of subscales

A subscale is a valid measure if it measures what it claims to. Of particular interest here are three kinds of validity: 1) Convergence refers to the extent to which a set of items measures a single one-dimensional underlying construct, 2) Content validity refers to the extent to which the items represent the domain, and 3) Face validity refers to the extent to which scales immediately appear valid (AERA, 1999). Validity of scales and subscales was assessed by looking at item-total correlation and average inter-item correlation for the contributing items and Cronbach's (1951) alpha coefficient. While the item-total correlation indicates the relationship between the item and the *total scale score*, the inter-item correlation indicates the *average* relationship between the item and the other items. Cronbach's alpha coefficient tells how well the scale measures a single underlying construct.

Longitudinal effects

Changes over time in outcome were identified in two ways. First, the factors derived from the repeated factor analyses reveal possible changes over time in the ways in which the subjects used the questionnaires. Second, between-session effects were assessed by a repeated-measures analysis of variance (ANOVA). In this ANOVA, session number and subscale were within-subject factors while previous hearing-aid experience was a between-subjects factor. For GHABP, changes in the correlations between subscales were tested with Fisher's Z transformation, which converts Pearson correlation coefficients to normally distributed values.

Results

Of the 141 items per subject (47 items, three times), on average 1.25 items were unanswered. That corresponds to a response proportion of more than 0.99. No single item accounted primarily for the missing answers, but many answers were missed by the same subject, who accounted for 12 of the missing answers (40%).

In the tables presenting results from the factor analyses, only loadings greater than 0.6 are tabulated, except where an item did not contribute to any factor by at least 0.6, in which case the strongest loading from that item is tabulated also. While it is recognized that this criterion is rather arbitrary, the chosen format is intended to allow a ready appreciation of the loading of different items by identification of non-empty cells in the tables.

GHABP

Table 1 shows the Pearson correlation, ρ, between the GHABP subscales. Initial disability is strongly correlated with handicap, $\rho = 0.80$, $p = 0.001$, and the subscales, benefit, residual disability, and satisfaction, are also correlated with one another, $0.60 \leq \rho \leq 0.90$, $p < 0.01$; the strongest correlation is between satisfaction and benefit. Note that the correlations are all positive because the answers to residual disability and handicap were reversed. The correlations between initial disability and residual disability and between aid use and benefit were only significant at session three ($\rho = 0.50$, $p < 0.05$). Even though the GHABP subscales have never been claimed to be independent, the strong between-subscales correlations might lead one to challenge the necessity for all three performance subscales. This is further discussed later.

There were some significant changes over time in the correlations from sessions one to three. Between benefit and satisfaction the correlation went from 0.90 to 0.75 ($p < 0.01$). However, there were also significant changes in the otherwise non-significant correlation coefficients. For initial disability, a significant change in correlation happened with benefit ($p < 0.05$), residual disability ($p < 0.05$), and satisfaction ($p < 0.05$); and for handicap, it happened with aid use ($p < 0.05$) and satisfaction ($p < 0.05$). These results indicate that even in situations where no significant relationship could be established, the direction in which it changed was statistically significant.

IOI-HA

For IOI-HA, Cronbach's alpha was 0.67, which in spite of the small number of items is not sufficient to suggest that the seven items measure the same underlying attribute. However, for item 5 (residual disability) the item-total correlation was negative, and deletion of item 5 made alpha increase to 0.83. Negative item-total correlation should not occur because of the way response values were reversed as explained earlier. Thus, the result indicates that the subjects in this study had trouble assessing residual disability. Cox and Alexander (2002) also observed an inconsistency with item 5. Combined with the present results this may raise doubt about the content validity of this particular item. The general application of this concern is taken up in the Discussion section.

Table 2 shows the results from the factor analyses. For items normalized as described earlier, the sign of the factor loadings should be the same within factors in order for the items to contribute to the factor in a meaningful manner. Except for item 5, the two factors were comparable to those identified by Cox and Alexander (2002), Kramer (2002) and Stephens (2002). Sixty-five percent of the response variance could be explained by the two factors.

The results suggest that the subscales introspection (items 1, 2, 4, 7) and interaction (items 3, 5, 6), reported by Cox and Alexander (2002), are reasonable dimensions for the results of this study. For introspection, Cronbach's alpha was 0.85, suggesting high convergence. By contrast, for the subscale interaction Cronbach's alpha was 0.28 and deletion of item 5, which in the former paragraph was identified as a potential problem, increased alpha to 0.64. The results corroborate the results from the factor analyses in that deletion of item 5 increases the validity of the subscale Interaction.

HAPQ

For HAPQ, convergence was very high (a =0.95). However, a high Cronbach coefficient is an intrinsic consequence of a large number of items. The results of the factor analyses are given in Table 3. Note that there is a change over time in the amount of variance explained by items 10 and 11 that both address the sound quality of the aid user's own voice: at session 2, these items load heavily on a factor that explains 22% of the variance, while 9 weeks later at session 3 a comparable factor only accounts for 10% of the variance.

There is no clear picture with regard to the distinctness of the underlying factor structures. A weak tendency is that items 1–3 loaded on the same factor, as did items 10–11 and items 16–18. Yet, for the first session the meaning behind the factors is ambiguous. By contrast, for session 3 the five factors have high face validity; the session-3 factors could reasonably be labelled speech and music (items 1, 2, 3, 4, 5, 12, 15); noise (items 6, 7, 9, 15), soft sounds (items 8, 12, 16, 17, 18); driving car (items 13, 14); and own voice (items 10, 11). While the subscales for soft sounds and own voice exist in sessions 1 and 2, respectively, there is no general agreement between session-1, session-2 and session-3 factors. For session 1, the sound of own voice loads on the same factor as noisy situations, but other factors have virtually no face validity for sessions 1 and 2.

SADL

For SADL, the results from the factor analyses are shown in Table 4. Three factors were identified and bore resemblance to three of the four subscales previously reported for SADL (Cox & Alexander, 2001). Because items 14 and 15 were not used, the subscale service and cost (items 12, 14, 15) is represented by only one item in the present study, and therefore it will not be discussed further. Furthermore, since all the subjects had 'ski-slope' hearing losses, unaided conversation on the telephone was often possible, and therefore item 11 was often technically unanswered (if respondents could hear well on the telephone without hearing aids, they could tick a box and did then not need to answer the question about how helpful the hearing aids are on the telephone). Because of insufficient data, item 11 had to be left out of the analysis.

The validity of the subscales positive effect (items 1, 3, 5, 6, 9, 10), negative features (item 2, 7, 11) and personal image (items 4, 8, 13) was as follows: For positive effect, Cronbach's alpha was high (a =0.73), but not as high as found by Cox & Alexander (2001) who reported (a =0.88. By contrast, the subscales negative features (a =0.09) and personal image (a =0.44) had very low validity. For negative features, there were generally problems with the three contributing items, whereas for personal image the problem seemed confined to item 8. The implication of this result is discussed in more detail later.

ALD

Test and retest results for ALD are shown in Figure 2. The regression line and confidence intervals were estimated on the basis of orthogonal regression (Press et al, 1992). Compared to conventional linear regression, in which one of the variables is assumed error-free, orthogonal regression minimizes errors in both variables orthogonally to the regression line.

The test-retest correlation was ρ =0.61 (p <0.001) and there was no systematic change in scores between the two samples. This result indicates that although some subjects changed the evaluation of their auditory lifestyle during the course of the study, there was no predictable direction of this change from the first report. There were no significant correlations between the individual changes in ALD and self-report outcome. The average ALD score was 49.5 (s.d. 13.4).

Between-subscales results

The following subscales were analysed further: the predefined GHABP performance scales, benefit (B), residual disability (R), and satisfaction (S); the IOI-HA subscales, introspection (IS), and interaction (IA) disregarding item 5; the SADL subscales personal image (PI), and positive effect (PE); HAPQ average; and hearing-aid use. In the following, an abbreviated notation will be used to designate these subscales, as shown in Table 1.

The correlations between subscales and ALD (mean values across tests), and between subscales and subject age are shown in Table 6. Several subscales correlate significantly but moderately, with auditory lifestyle. The relationships generally have the direction of poorer self-reported outcome with richer auditory lifestyle. While at first this result might seem unintuitive, one can easily imagine that people with an active auditory life assess the performance of their hearing aids critically. Only IOI-IA scores correlated significantly with age.

Longitudinal effects and interactions

The repeated-measures ANOVA showed a significant effect of subscale ($F_{8,176}$ =13.5, p 0.001) but there were no significant interactions with time or hearing-aid experience. No significant main effects of time or hearing-aid experience were found either; a result that suggests the subscales are stable regarding sample time and previous hearing-aid experience. However, this result needs further discussion because of the observations in Table 3 that suggested that the distinctness of the outcome space changed over time.

Figure 3 shows the increase from week 1 to week 13 for the main scales of the four questionnaires for experienced and first-time users, respectively. It is calculated as the difference between the scores after one week and after 13 weeks in such a way that a positive value represents an increase in self-reported outcome. However, none of the values in Figure 3 was significantly different from zero, and there were no statistically significant differences between the first-timers and experienced users (which is also obvious in the figure).

If outcome changes over time in response to permanent change in the acoustic information available to the aid user post-fitting (Arlinger et al, 1996), then experienced users who are already used to hearing aids should not be expected to show any change in outcome over time. Similarly, one should not expect new hearing-aid users to show this sort of change if they do not wear the hearing aids. It was therefore tested whether the null-result in Figure 3 was independent of daily hearing-aid use. Using a general linear model, daily hearing-aid use according to SADL was found to be a good predictor of the increase in mainscale scores ($F_{1,22}$ =5.25, p =0.032). Those with more daily hearing-aid use reported improved outcome over time. Of course, this result does not imply causality; that is, we do not know whether the subjects reported increased outcome as a result of sufficient hearing-aid use or whether they did not wear the hearing aids because of poor outcomes. Nevertheless, a posthoc criterion for daily hearing-aid use was tested. The idea was to test those who can be expected to show a change against those who cannot.

The first-time users who used their hearing aids more than four hours per day were compared against a control group consisting of the experienced users and those with less than four hours of daily aid use. The results are shown in Figure 4. With this re-stratification, an ANOVA with the same design as above was run. There was now a significant difference in the increase in outcome between the first-timers with more than fours hours of daily hearing-aid use and the control group ($F_{3,69}$ =3.13, p =0.031). A contrast analyses revealed that this effect was confined to GHABP ($F_{1,23}$ =5.95, p =0.023) and IOI-HA ($F_{1,23}$ =5.13, p =0.033). There were no significant effects for HAPQ ($F_{1,23}$

=0.54, p =0.472) and SADL ($F_{1,23}$ =3.86, p =0.062). Notice that the increase was significantly below zero in the case of GHABP for the control group. These results should be compared with the results from above that showed that the subjects changed the way they used HAPQ over time, while GHABP, IOI-HA, and SADL subscales were more stable over time. This will be discussed in the next section.

Discussion

This section focuses on three distinct aspects of the results, namely the validity of, the relations between, and acclimatization effects in, the different measures of self-report benefit and satisfaction.

Validity of subscales

When an item contributes inconsistently to a (sub-) scale, this can happen because the wording is not good (low convergence or content validity), or because the item addresses an attribute that the hearing-aid users are not able to appreciate (no face validity).

IOI-HA item 5 was found to contribute inconsistently to the subscale IOI-IA. The attribute addressed in this question belongs in the category residual disability (as in GHABP-R), in which the hearing-aid users are expected to quantify remaining disability when wearing hearing aids compared to not having hearing difficulties. It is argued here that this may not be a viable task because the reference for residual disability is normal hearing. This baseline may not be available to users whose hearing loss has progressed over a substantial number of years, which is often the case in presbyacusic patients. While this problem concerns content validity, another problem with IOI-HA item 5 is the complexity of the wording. The item asks about the degree to which hearing difficulties affect the things the hearing-aid user can do, but in the Danish translation the words 'over the past two weeks with your present hearing aids' form a subordinate clause that separates semantic subject from object in the main clause. This might cause some respondents to accidentally switch the semantic subject of the question. Thus, there is a risk that the item is misunderstood to be asking the degree to which the *hearing aids* affect the things the user can do. If this happens (and if the hearing aids are any good), the scale orientation is reversed and that could explain the negative item-total correlation.

SADL item 7 was not understood by most respondents. The item addresses the compromise between avoiding acoustic feedback and obtaining adequate gain. All audiologists know that this compromise is pertinent for some users. However, if the hearing aids are well fitted and the ear-mould acoustics do not cause problems with feedback, then obviously the user is not bothered by feedback. For the item to be a sensitive measure of the amount of bother caused by feedback, the respondent must know that there is a trading relationship between gain and whistling, which is not necessarily the case for users of instruments without a volume control. Thus, the item addresses a fine-tuning matter, not necessarily well understood by the average user of an automatic hearing aid.

The above illustrates the difficulty of compiling items of universal applicability. If the question was simply 'are you bothered by feedback?', then some users might answer 'no' ['but I have to turn down the volume so I can't hear']. In that case, we do not capture the true amount of bother caused by potential feedback. When the question is refined to avoid that situation, it becomes so complex that people do not understand what it is about.

Relationships between measures

While the subscales in IOI-HA and SADL were independent and stable over time, those in HAPQ and GHABP were less so. Although it could be argued that there is no obvious

reason why hearing-aid outcome subscales should be independent, if a significant and strong correlation exists between any two subscales then there is no need to keep both, because they will convey the same information. This is not to say, however, that outcome scales should be free from between-item correlation. If the number of items is reduced, so as to handle a given attribute through fewer questions, the risk is an increase of verbal sophistication in the remaining items that may in turn decrease the sensitivity to the attribute in question, as illustrated above for SADL item 7. As such, there is a trading relation between achieving scale sensitivity and having an efficient scale structure of independent subscales, in that from a clinical practice point of view, it is desirable to avoid repeating essentially the same question because it makes the inventory shorter.

For HAPQ, the salience of the own-voice factor decreased during the first 3 months post-fitting. Clinical experience suggests that while first-time users often initially find it particularly troublesome that hearing aids change the sound of their own voice (echo, occlusion effect, loudness or timbre changes), it becomes less of an issue after the benefits of using the hearing aids have been experienced.

Overall, the factors identified in HAPQ did not all have high face validity. Before a subscale can be defined – let alone labelled – the contributing items must address issues that have something in common, which was not the case for the factors derived from sessions 1 and 2. This suggests that early outcome assessment might be spurious if a questionnaire uses many distinct dimensions. The results support the notion that early self-report outcome assessment is only valid if the questions have face value and do not address attributes that new users cannot be expected to handle.

While between-subscales correlations were unexpectedly high for the GHABP, it is noteworthy that they changed with time. Thus, between benefit and satisfaction, the correlation went down during the first three months, while the correlations went up between initial disability and benefit, and residual disability and satisfaction, and changed sign between handicap and aid use and satisfaction. That result also supports the notion that patients become better over time at using the outcome space at hand, albeit not in terms of independent subscale scores in the case of GHABP. Humes et al (2001) found that all GHABP subscales except aid use loaded on the same factor, which the authors simply named subjective benefit.

Auditory acclimatization

The results showed an improvement in self-reported outcome over time in some first-time hearing-aid users. There was a significant difference in the improvement between first-time users with more than four hours of daily aid use and a control group, but only for GHABP and IOI-HA. Auditory acclimatization, as defined by Arlinger et al (1996), involves an improvement in auditory performance with hearing aids over time. The present results show that the existence of auditory acclimatization in self-report outcome is dependent on the amount of daily hearing-aid use. However, a causal relationship could not be established with regard to the correlation between aid use and improvement in outcome over time.

There was another longitudinal effect, in that the subjects became better over time in using the outcome space that was made available through the questionnaires. This was evidenced by factors reaching higher face validity over time in HAPQ, and by some between-subscales correlations decreasing over time in GHABP. One explanation for these results is that the users need enough experience with the different listening situations addressed in the questionnaires before they can evaluate the performance of their hearing aids independently in those situations. While HAPQ and GHABP address hearing-aid performance in different

specific listening environments, SADL and IOI-HA address performance at a more general level.

The subjects filled in the ALD questionnaire twice. Because of the provision of hearing aids between the two tests, the ALD scores might be expected to change over time. If the subjects perceived an increase in auditory competence following the provision of hearing aids, some listening situations might in turn have become more frequent post-fitting. However, no such change was observed. On average, the participants rated the richness of their auditory life the same for pre-fitting and for three months post-fitting.

Consistent correlations were found between auditory lifestyle and self-report outcome in such a way that a higher ALD score was generally associated with lower self-report outcome. This relationship can be interpreted as follows: those subjects who express the most varied need for auditory function assess hearing-aid outcome most critically, leading to a lower average assessment than for subjects with lesser needs for auditory function. It is well known that the correlation between objective and subjective outcome is limited (see, e.g. Vestergaard, 2004), and it is possible that the influence of personal expectations and needs is so strong that it prevents objective outcomes being externalized in subjective-outcome scores.

Since a high ALD score reflects high expectations and needs, perhaps a question addressing pre-fitting expectations would be a useful supplement to some existing inventories for measuring self-experienced outcome. Wong et al (2004) also considered the relation between expectation and perceived benefit, and called deviation between the two disconfirmation. From their preliminary results, they observed that while perceived benefit did not always predict satisfaction, greater satisfaction was associated with fulfilment of expectations in terms of hearing aids performing as well as or better than expected. While this result is neither very surprising nor particularly original, it lends support to the notion that pre-fitting expectations should be considered in studies of self-report outcome.

In the present study, the strongest predictor of outcome was ALD, which explained 14% of the variance in the GHABP subscale Satisfaction. Wong et al (2003) reviewed 45 studies on self-report outcome and concluded that, other than aid use, no intrinsic factors influence benefit and satisfaction. They also concluded that there was no correlation between age and satisfaction. In the present study, however, a small but statistically significant correlation was found between IOI-IA outcome and age.

Conclusions

First-time users who used their hearing aids more than four hours per day reported improved outcome over time compared to users with previous hearing aid experience or users with less than four hours of daily hearing-aid use. However, this result was restricted to GHABP and IOI-HA, which address *general* aspects of hearing-aid benefit and satisfaction. For SADL and HAPQ, which address hearing-aid performance in specific situations, there was no such change over time.

Subjects who report the largest need for auditory function on the ALD scale assess hearing-aid outcome more critically than those with low demand for auditory function. This relationship means that a high ALD score is statistically associated with low self-report outcome. This result leads to the proposal that auditory expectations be included in inventories for measuring self-report outcome.

Early outcome assessment is only valid if the questions have face value for new hearing-aid users. While many of the problems with the validity of hearing-aid benefit and satisfaction

measures can be solved by rigid administration of, and thorough introduction to, the questionnaires as well as careful data analyses, the respondents must be able to appreciate the attributes addressed in the questions. This means that a differentiated assessment of subjective hearing-aid outcome in different listening environments is not feasible from the first day post-fitting. Self-assessment tools that address general aspects of hearing-aid performance are more stable over time, and they can more reliably be used immediately post-fitting.

Acknowledgments

Acronyms

ALD	Auditory lifestyle and demand
GHABP	Glasgow hearing aid benefit profile
HAPQ	Hearing aids performance questionnaire
IOI-HA	International outcome inventory for hearing aids
SADL	Satisfaction with amplification in daily life

References

AERA. American Educational Research Association. Standards for Educational and Psychological Testing. Washington, DC: AERA; 1999.

Arlinger S. Can we establish internationally equivalent outcome measures in audiological rehabilitation? Ear Hear. 2000; 21:97S–99S. [PubMed: 10981599]

Arlinger S, Gatehouse S, Bentler RA, Byrne D, Cox RM, et al. Report of the Eriksholm Workshop on auditory deprivation and acclimatization. Ear Hear. 1996; 17:87S–98S. [PubMed: 8807279]

Bentler RA, Niebuhr DP, Getta JP, Anderson CV. Longitudinal study of hearing aid effectiveness. I: Objective measures. J Speech Hear Res. 1993a; 36:808–819. [PubMed: 8377493]

Bentler RA, Niebuhr DP, Getta JP, Anderson CV. Longitudinal study of hearing aid effectiveness. II: Subjective measures. J Speech Hear Res. 1993b; 36:820–831. [PubMed: 8377494]

Cox RM. Proceedings of a workshop on self-reported outcome measures in audiological rehabilitation including hearing aids. Eriksholm, Denmark, September 1–3, 1999. Ear Hear. 2000; 21(4 Suppl): 1S–115S. [PubMed: 11203477]

Cox, RM.; Alexander, G.; Gray, G. Hearing aid fitting: self-report outcomes over six months; Paper presented at American Auditory Society 2004 Meeting; Scottsdale, AZ. March 7–9; 2004.

Cox RM, Alexander GC. Maturation of hearing aid benefit: objective and subjective measurements. Ear Hear. 1992; 13:131–41. [PubMed: 1397752]

Cox RM, Alexander GC. Measuring satisfaction with amplification in daily life: The SADL scale. Ear Hear. 1999; 20:306–320. [PubMed: 10466567]

Cox RM, Alexander GC. Validation of the SADL questionnaire. Ear Hear. 2001; 22:151–160. [PubMed: 11324844]

Cox RM, Alexander GC. The international outcome inventory for hearing aids (IOI-HA): Psychometric properties of the English version. Int J Audiol. 2002; 41:30–35. [PubMed: 12467367]

Cox RM, Alexander GC, Taylor IM, Gray GA. Benefit acclimatization in elderly hearing aid users. J Am Acad Audiol. 1996; 7:428–441. [PubMed: 8972444]

Cox RM, Hyde M, Gatehouse S, Noble W, Dillon H, et al. Optimal outcome measures, research priorities, and international cooperation. Ear Hear. 2000; 21:106S–115S. [PubMed: 10981601]

Cox RM, Stephens D, Kramer SE. Translations of the international outcome inventory for hearing aids (IOI-HA). Int J Audiol. 2002; 41:3–26. [PubMed: 12467365]

Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951; 16:297–334.

Gatehouse S. The time course and magnitude of perceptual acclimatization to frequency responses: Evidence from monaural fitting of hearing aids. J Acoust Soc Am. 1992; 92:1258–1268. [PubMed: 1401514]

Gatehouse S. Role of perceptual acclimatization in the selection of frequency responses for hearing aids. J Am Acad Audiol. 1993; 4:296–306. [PubMed: 8219296]

Gatehouse, S. Aspects of auditory ecology and psychoacoustic function as determinants of benefits from, and candidature for nonlinear processing in, hearing aids. In: Rasmussen, AN.; Osterhammel, PA.; Andersen, T.; Poulsen, T., editors. Auditory Models and Non-Linear Hearing Instruments. Denmark: Holmens Trykkeri; 1999a. p. 221-233.

Gatehouse S. Glasgow hearing aid benefit profile: derivation and validation of a client-centered outcome measure for hearing aid services. J Am Acad Audiol. 1999b; 10:80–103.

Gatehouse S. Self-report outcome measures for adult hearing aid services: Some uses, users and options. Trends Amplif. 2001; 5:91–110.

Gatehouse, S.; Elberling, C.; Naylor, G. Aspects of auditory ecology and pyschoacoustic function as determinants of benefits from, and candidature for non-linear processing in, hearing aids; Paper presented at IHCON; Lake Tahoe, CA, USA. 2000.

Horwitz AR, Turner CW. The time course of hearing aid benefit. Ear Hear. 1997; 18:1–11. [PubMed: 9058033]

Humes LE. Dimensions of hearing aid outcome. J Am Acad Audiol. 1999; 10:26–39. [PubMed: 9921723]

Humes LE, Garner CB, Wilson DL, Barlow NN. Hearing-aid outcome measured following one month of hearing aid use by the elderly. J Speech Lang Hear Res. 2001; 44:469–486. [PubMed: 11407554]

Humes LE, Wilson DL. An examination of changes in hearing-aid performance and benefit in the elderly over a three-year period of hearing-aid use. J Speech Lang Hear Res. 2003; 46:137–145. [PubMed: 12647894]

Kaiser HF. The varimax criterion for analytic rotation in factor analysis. Psychometrika. 1958; 23:187–200.

Kaiser HF. The application of electronic computers to factor analysis. Educ Psychol Meas. 1960; 20:141–151.

Kapteyn TS. Satisfaction with fitted hearing aids. I: An analysis of technical information. Scand Audiol. 1977; 6:147–156. [PubMed: 929092]

Kramer SE, Goverts ST, Dreschler WA, Boymans M, Festen JM. International outcome inventory for hearing aids (IOIHA): Results from The Netherlands. Int J Audiol. 2002; 41:36–41. [PubMed: 12467368]

Kuk FK, Potts L, Valente M, Lee L, Picirrillo J. Evidence of acclimatization in persons with severe-to-profound hearing loss. J Am Acad Audiol. 2003; 14:84–99. [PubMed: 12830844]

Mulrow CD, Tuley MR, Aguilar C. Sustained benefits of hearing aids. J Speech Hear Res. 1992; 35:1402–1405. [PubMed: 1494282]

Munro KJ, Lutman ME. The effect of speech presentation level on measurement of auditory acclimatization to amplified speech. J Acoust Soc Am. 2003; 114:484–495. [PubMed: 12880059]

Munro KJ, Lutman ME. Self-reported outcome in new hearing aid users over a 24-week post-fitting period. Int J Audiol. 2004; 43:555–562. [PubMed: 15724519]

Palmer CV, Nelson CT, Lindley GA. The functionally and physiologically plastic adult auditory system. J Acoust Soc Am. 1998; 103:1705–1721. [PubMed: 9566316]

Press, WH.; Teukolsky, SA.; Vetterling, WT.; Flannery, BP. Numerical Recipes in C. Cambridge: Cambridge University Press; 1992.

Saunders GH, Cienkowski KM. Acclimatization to hearing aids. Ear Hear. 1997; 18:129–139. [PubMed: 9099562]

Stephens D. The international outcome inventory for hearing aids (IOI-HA) and its relationship to the client-oriented scale of improvement (COSI). Int J Audiol. 2002; 41:42–47. [PubMed: 12467369]

Surr RK, Cord MT, Walden BE. Long-term versus short-term hearing aid benefit. J Am Acad Audiol. 1998; 9:165–171. [PubMed: 9644613]

Taylor KS. Self-perceived and audiometric evaluations of hearing aid benefit in the elderly. Ear Hear. 1993; 14:390–394. [PubMed: 8307242]

Turner CW, Humes LE, Bentler RA, Cox RM. A review of the literature on hearing aid benefit as a function of time. Ear Hear. 1996; 17:14S–28S. [PubMed: 8807271]

Vestergaard, MD. Benefit from amplification of high frequencies in hearing impaired: aspects of cochlear dead regions and auditory acclimatization. The Technical University of Denmark, Acoustic Technology; Lyngby: 2004. PhD thesis

Watson NA, Knudsen VO. Selective amplification in hearing aids. J Acoust Soc Am. 1940; 11:406–419.

Wong LL, Hickson L, McPherson B. Hearing aid satisfaction: What does research from the past 20 years say? Trends Amplif. 2003; 7:117–161. [PubMed: 15004650]

Wong, LLN.; Hickson, L.; McPherson, B. Hearing aid satisfaction and its relationship to expectation, performance, benefit and disconfirmation; Paper presented at American Auditory Society 2004 Meeting; Scottsdale, AZ. March 7–9; 2004.
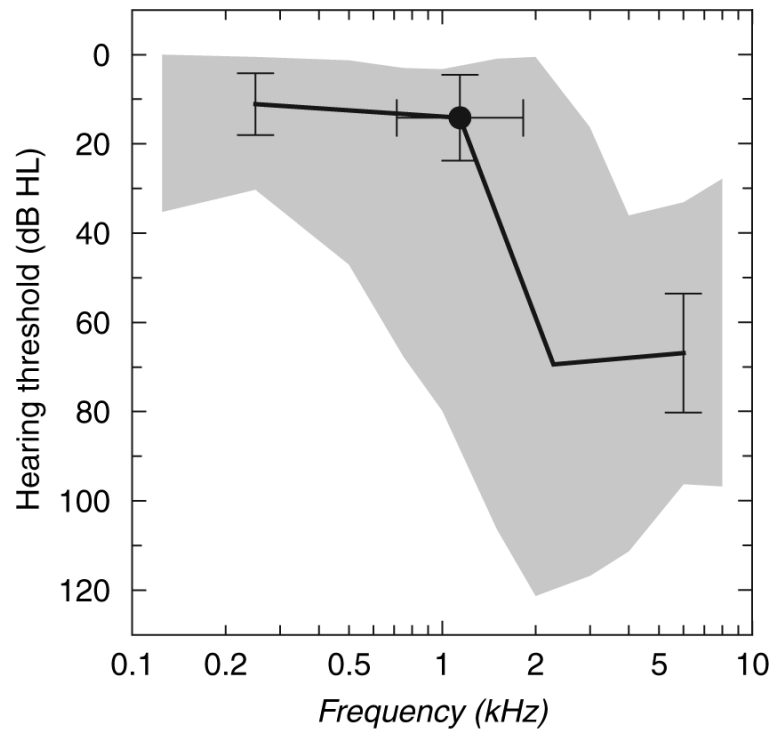
Figure 1.
Range of hearing threshold levels of the subjects and key audiometric values with standard deviation bars (±s.d.). The typical audiogram is composed by connecting with straight lines the following key audiometric points through the average values: Hearing level (HL) at 250 Hz, HL at the corner frequency, HL one octave above the corner frequency and HL at 8000 Hz.
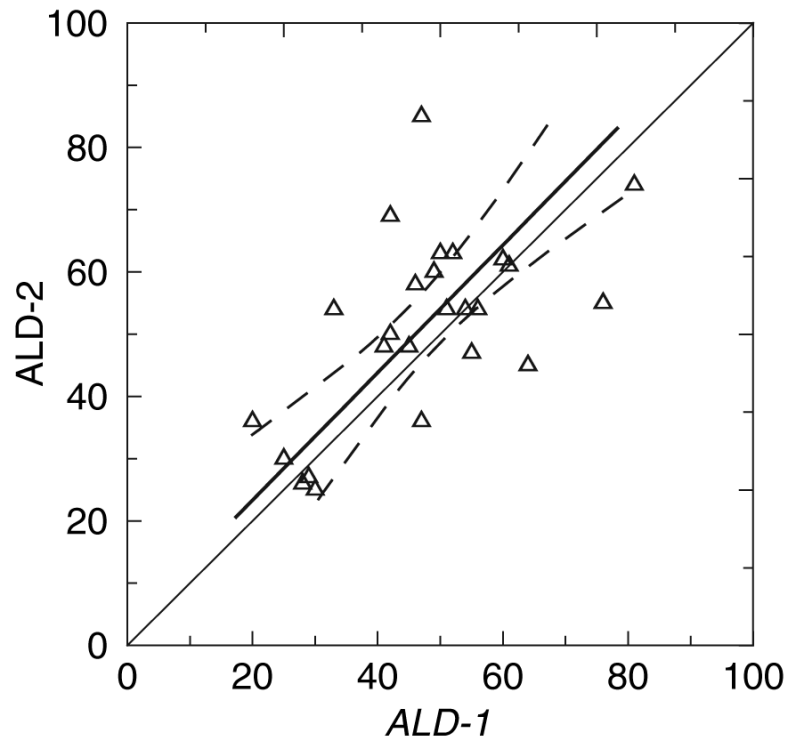
Figure 2.
Test-retest scores for Auditory Lifestyle and Demand (ALD). Regression line (solid line)
and 95% confidence interval for the regression line (dashed lines) are also shown.
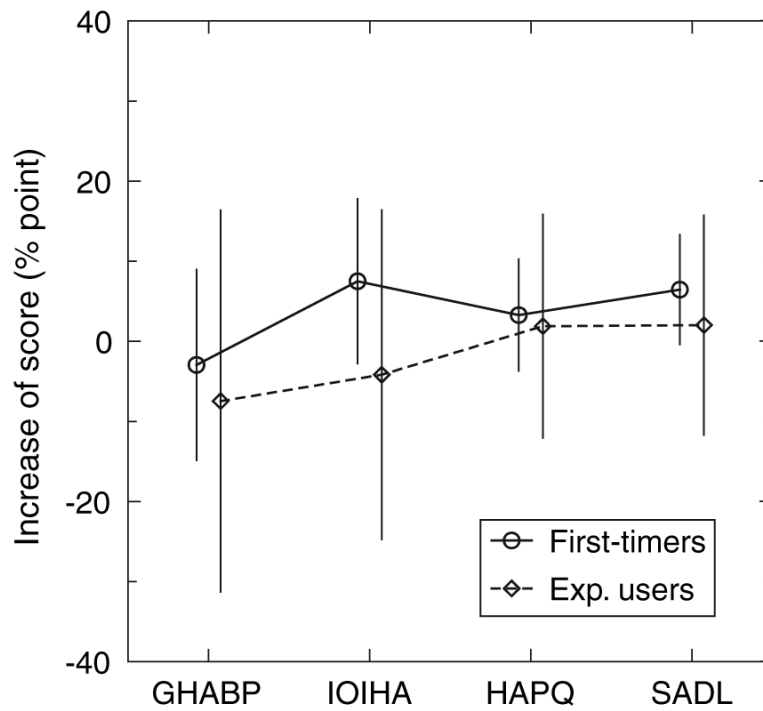
Figure 3.
Increase in self-report outcome from session 1 to session 3 of the main scales, GHABP benefit, IOI-HA global score, HAPQ average, and SADL global score. The increase is shown by circles for first-time users, and by diamonds for the experienced users. The vertical bars denote 95% confidence intervals.

Figure 4.
As Figure 3 for first-time users with more than four hours of daily hearing-aid use,
compared to a control group of experienced users, and aid users with less than four hours of
daily hearing-aid use.

Table 1

Pearson correlation ρ for the GHABP subscales. See text for details.

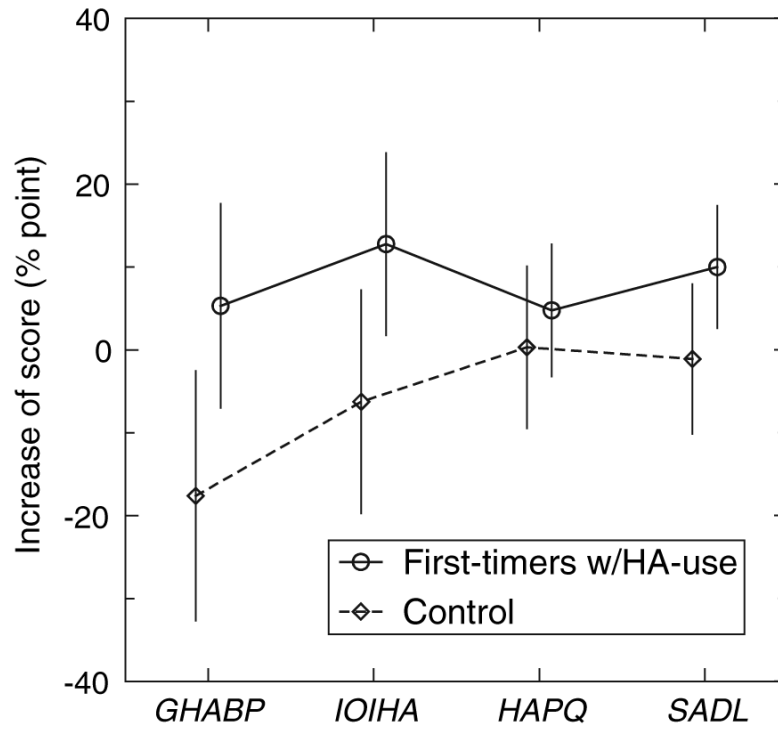| GHABP	Handicap	Aid Use	Benefit	Residual Disability	Satisfaction
Session 1					
Initial Disability	0.80 ***	−0.28	−0.06	0.06	−0.10
Handicap		−0.29	−0.04	0.02	−0.16
Aid Use			0.33	0.16	0.22
Benefit				0.71 ***	0.90 ***
Residual Disability					0.77 ***
Session 2					
Initial Disability	0.80 ***	−0.19	0.03	0.42 *	0.20
Handicap		−0.18	0.00	0.21	0.03
Aid Use			0.48 *	0.01	0.43 *
Benefit				0.60 **	0.84 ***
Residual Disability					0.84 ***
Session 3					
Initial Disability	0.80 ***	0.04	0.34	0.50	0.33
Handicap		0.10	0.25	0.30	0.20
Aid Use			0.50 *	0.00	0.15
Benefit				0.62 **	0.75 ***
Residual Disability					0.85 ***

The statistical significance of ρ is indicated like this:

Entries in italics indicate that the correlation coefficient changed significantly between session 1 and session 2 or 3.

*
$p < 0.05$

**
$p < 0.01$

***
$p < 0.001$.

Table 2

IOI-HA factor loadings.

IOI-HA	Session 1		Session 2		Session 3	
	$F1_1$	$F2_1$	$F1_2$	$F2_2$	$F1_3$	$F2_3$
Item 1	0.59		0.65			−0.74
Item 2	0.68		0.69		0.82	
Item 3		0.79		0.82	0.71	
Item 4	0.76		0.85		0.87	
Item 5	−0.79		−0.63			0.62
Item 6		0.84		0.88		0.61
Item 7	0.94		0.89		0.86	
Explained variance	2.94	1.91	2.91	1.86	3.15	1.42
Proportion total	0.42	0.27	0.42	0.27	0.45	0.20

Table 3

HAPQ factor loadings.

HAPQ	Session 1			Session 2			Session 3				
	$F1_1$	$F2_1$	$F3_1$	$F1_2$	$F2_2$	$F3_2$	$F1_3$	$F2_3$	$F3_3$	$F4_3$	$F5_3$
Item 1	0.77			0.89			0.89				
Item 2	0.70			0.80			0.75				
Item 3	0.80			0.55			0.74				
Item 4			0.75	0.71			0.70				
Item 5		0.62			0.55		0.79				
Item 6		0.80			0.64			0.81			
Item 7		0.87				0.69		0.83			
Item 8			0.71						0.66		
Item 9		0.82		0.57		0.75		0.83			
Item 10		0.78			0.90						0.72
Item 11		0.79			0.87						0.82
Item 12			0.87	0.79			0.61		0.64		
Item 13		0.72				0.70				0.91	
Item 14	0.58					0.73				0.90	
Item 15		0.59				0.75	0.58				
Item 16			0.58			0.72			0.81		
Item 17			0.74	0.63					0.74		
Item 18			0.80			0.63			0.78		
Explained variance	3.35	5.10	4.33	4.79	3.90	5.25	4.46	2.90	3.60	2.19	1.84
Proportion total	0.19	0.28	0.24	0.27	0.22	0.29	0.25	0.16	0.20	0.12	0.10

Table 4

SADL factor loadings.

SADL	Session 1			Session 2			Session 3		
	$F1_1$	$F2_1$	$F3_1$	$F1_2$	$F2_2$	$F3_2$	$F1_3$	$F2_3$	$F3_3$
Item 1	0.60			0.88			0.67		
Item 2		0.59				−0.54	−0.56		
Item 3			0.63	0.89			0.78		
Item 4		0.75			0.81			0.58	
Item 5			0.50			−0.43		−0.38	
Item 6			0.75		0.40			0.64	
Item 7			0.45						0.66
Item 8	0.75					0.82			0.81
Item 9	0.79			0.74			0.72		
Item 10	0.74			0.68					0.78
Item 12	0.70				0.56			0.75	
Item 13		0.76			0.79			0.84	
Explained variance	3.14	1.84	1.91	3.48	2.08	1.77	2.45	2.34	2.08
Proportion total	0.26	0.15	0.16	0.29	0.17	0.15	0.20	0.19	0.17

Table 5

Subscales with contributing items. The scores are calculated as the average of the response values for the contributing items (3rd column), and the subscale is designated in the text by its short notation (2nd column)

Subscale	*Short notation*	*Items*
GHABP		
Benefit	GHABP-B	Benefit[1]
Residual Disability	GHABP-R	Residual Disability[1]
Satisfaction	GHABP-D	Satisfaction[1]
IOI-HA		
Introspection	IOI-IS	1, 2, 4, 7
Interaction	IOI-IA	3, 6
HAPQ		
average	HAPQ	1–18
SADL		
Personal Image	SADL-PI	4, 8, 13
Positive Effect	SADL-PE	1, 3, 5, 6, 9, 10
		GHABP item 3
Hearing aid use	HA-use	IOI-HA item 1
		SADL additional item 3
ALD		
Auditory lifestyle	ALD	1–25 (weighted average)

[1]For GHABP, only answers to listening situations nominated and answered at all three sessions were included in calculating the subscale scores.

Table 6

Pearson correlations, ρ, between self-report outcomes and ALD and age.

	ALD	*Age*
GHABP-B	−0.26**	0.06
GHABP-R	−0.35***	0.10
GHABP-S	−0.38***	0.06
IOI- IS	−0.08	−0.03
IOI- IA	−0.23**	0.21**
HAPQ	−0.33***	−0.05
SADL-PE	−0.24**	0.07
SADL-PI	−0.21**	0.16
HA-use	−0.06	−0.10

*
$p < 0.05$

**
$p < 0.01$

***
$p < 0.001$