



Published in final edited form as:

*Int J Audiol.* 2016 May ; 55(5): 285–294. doi:10.3109/14992027.2015.1120892.

## Predicting 3-month and 12-month Post-Fitting Real-World Hearing Aid Outcome using Pre-fitting Acceptable Noise Level (ANL)

Yu-Hsiang Wu<sup>1</sup>, Hsu-Chueh Ho<sup>2,3</sup>, Shih-Hsuan Hsiao<sup>2,3</sup>, Ryan B. Brummet<sup>4</sup>, and Octav Chipara<sup>4</sup>

<sup>1</sup>Department of Communication Sciences and Disorders, The University of Iowa, Iowa City, Iowa, USA

<sup>2</sup>Department of Otolaryngology, Buddhist Dalin Tzu-Chi General Hospital, Chiayi, Taiwan

<sup>3</sup>School of Medicine, Tzu-Chi University, Hualien, Taiwan

<sup>4</sup>Department of Computer Science, The University of Iowa, Iowa City, Iowa, USA

### Abstract

**Objective**—Determine the extent to which pre-fitting acceptable noise level (ANL), with or without other predictors such as hearing aid experience, can predict real-world hearing aid outcomes at 3 and 12 months post-fitting.

**Design**—ANLs were measured before hearing aid fitting. Post-fitting outcome was assessed using the International Outcome Inventory for Hearing Aids (IOI-HA) and a hearing aid use questionnaire. Models that predicted outcomes (successful vs. unsuccessful) were built using logistic regression and several machine learning algorithms, and were evaluated using the cross-validation technique.

**Study sample**—132 adults with hearing impairment.

**Results**—The prediction accuracy of the models ranged from 61% to 68% (IOI-HA) and from 55% to 61% (hearing aid use questionnaire). The models performed more poorly in predicting 12-month than 3-month outcomes. The ANL cutoff between successful and unsuccessful users was higher for experienced (~18 dB) than first-time hearing aid users (~10 dB), indicating that most experienced users will be predicted as successful users regardless of their ANLs.

**Conclusions**—Pre-fitting ANL is more useful in predicting short-term (3 months) hearing aid outcomes for first-time users, as measured by the IOI-HA. The prediction accuracy was lower than the accuracy reported by some previous research that used a cross-sectional design.

---

Correspondence: Hsu-Chueh Ho, No.2, Min-Sheng Road, Dalin Town, Chia-Yi, Taiwan, R.O.C., Telephone number: 886-5-2648000 ext 5240, Fax number: 886-5-2648006, L367578@tzuchi.com.tw.

#### DECLARATION OF INTEREST

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

## Keywords

Acceptable noise level; Hearing aid; Outcome; International Outcome Inventory for Hearing Aids (IOI-HA); Machine learning

---

## INTRODUCTION

Acceptable noise level (ANL) is a measure that quantifies an individual's willingness to accept background noise while listening to speech (Nabelek et al, 1991; Nabelek et al, 2006). In a series of early studies, Nabelek and her colleagues first demonstrated the association between ANL and real-world hearing aid outcomes (Nabelek et al, 2006; Nabelek et al, 2004; Nabelek et al, 1991). For example, Nabelek et al (2006) investigated the relationship between ANL and the pattern of hearing aid use for 191 adults with hearing impairment. Most of the users had between three months to three years hearing aid experience. To assess the hearing aid use pattern, a questionnaire (referred to as the HA-Use in this article) that classified respondents into full-time, part-time, and non-users was employed. Nabelek et al found that the mean ANL of full-time users (7.7 dB) was lower than part-time users (13.5 dB) and non-users (14.4 dB). Nabelek et al (2006) further grouped the participants as successful users (full-time users) and unsuccessful users (part-time and non-users), and used a logistic regression analysis to examine the relationship between ANL and the probability of success. The results indicated that ANL was significantly associated with the probability of success. The classification accuracy of the logistic regression model was as high as 85%.

Since Nabelek's works were published, several studies have been conducted to investigate the relationship between ANL and real-world hearing aid outcomes (Table 1). For example, using HA-Use as the outcome measure, Freyaldenhoven et al (2008b) replicated the results of Nabelek et al (2006) and reported 68% classification accuracy. Using the same subject group as Nabelek et al (2006), Freyaldenhoven et al (2008a) further demonstrated that when combining ANL and the unaided Abbreviated Profile of Hearing Aid Benefit questionnaire (APHAB; Cox & Alexander, 1995), classification accuracy increased to 91%. Note that these studies (Freyaldenhoven et al, 2008a, b; Nabelek et al, 2006) used a cross-sectional design in which research participants completed the ANL test and used questionnaires to report their recent experience with hearing aids.

The association between ANL and outcomes measured using standardized questionnaires has also been investigated in prospective studies. Ho et al (2013b) used the Chinese version (Cox et al, 2002) of the International Outcome Inventory for Hearing Aids questionnaire (IOI-HA; Cox & Alexander, 2002) to measure the outcomes for 80 adults three months post hearing aid fitting. Most participants (77.5%) were first-time hearing aid users. The results indicated that users with lower unaided ANLs, which were measured before the hearing aid fitting, tended to report better outcomes at three months post-fitting. ANL significantly explained 16.2% of the variance of the IOI-HA score. A logistic regression analysis further indicated that the classification accuracy for hearing aid success defined by the IOI-HA was 67.5%, which was very close to the 68% accuracy reported by Freyaldenhoven et al (2008b).

Consistent with Ho et al (2013b), Taylor (2008) tested 27 first-time hearing aid users and found that pre-fitting ANL explained 16.8% of the variance in IOI-HA outcomes measured at 30 days post-fitting.

However, several studies (Olsen et al, 2012; Schwartz & Cox, 2012; Walravens et al, 2014; Table 1) did not demonstrate a clear association between ANL and hearing aid outcome. In a cross-sectional study, Olsen et al (2012) recruited 63 adults whose mean hearing aid experience was 11 years. Consistent with the trend found by Nabelek et al (2006), the mean ANLs of full-time users were approximately 2 to 6 dB lower (better) than ANLs of part-time users and non-users. However, because most participants were full-time users (90.5%), no statistical analysis was conducted to examine the relationship between ANL and hearing aid use pattern. Olsen et al (2012) further indicated that, contrary to Ho et al (2013b) and Taylor (2008), there was no clear association between ANL and IOI-HA outcome.

In another cross-sectional study (Walravens et al, 2014), ANL was measured for 96 hearing aid owners. Among these participants, 48% and 42% owned hearing aids for one to five years and for more than five years, respectively. The results of the HA-Use revealed that the ANL of full-time users (7.5 dB) was higher (poorer) than that of part-time (4.9 dB) and non-users (4.5 dB). Although the difference was not statistically significant, the trend of the findings by Walravens et al (2014) was contrary to Nabelek et al (2006).

Despite the growing body of ANL literature, the usefulness of using ANL to predict real-world hearing aid outcomes remains unclear for several reasons. First, most previous studies used a cross-sectional design. It is unknown to what extent the results of these studies can generalize to the condition wherein ANL is used to predict future hearing aid outcome.

Second, although the prospective studies by Taylor (2008) and Ho et al (2013b) demonstrated the association between pre-fitting ANL and short-term post-fitting outcome (one and three months post-fitting, respectively), it is unknown if ANL can predict longer-term outcomes. This is because the ability for ANL to predict outcome may decrease over time after hearing aid fitting. More specifically, one possible reason for individuals with higher ANLs tending to report poorer outcomes is that they are less willing to accept the noise generated or amplified by hearing aids. These individuals may eventually acclimatize to the noise or benefit from hearing aids' noise reduction technologies after a longer period of time. As a result, long-term outcomes might more likely be affected by factors other than noise acceptance and therefore might not be predictable by ANL. This hypothesis is supported by Walravens et al (2014), who suggested that the non-significant relationship between ANL and hearing aid use was due to their participants' longer hearing aid experience.

The third reason why the usefulness of ANL has not been fully supported is that the prediction accuracy reported in the literature might be overestimated. In previous research the performance of the prediction model was often evaluated by the dataset that was utilized to build the model (e.g., Freyaldenhoven et al, 2008a, b; Ho et al, 2013b; Nabelek et al, 2006). As indicated by Nabelek et al (2006), using the same dataset to build and evaluate the prediction model would overestimate the model's performance. Currently, only one study

(Schwartz & Cox, 2012) has tried to use a new dataset to evaluate the prediction model established by other research. Schwartz and Cox recruited 50 adults who had at least six months of bilateral hearing aid experience. ANL was used to predict hearing aid success, such that the participants who had ANLs equal to or smaller than 7 dB were predicted to be successful users (Nabelek et al, 2006). Four standardized questionnaires were used to measure outcomes, including the Satisfaction with Amplification in Daily Life (SADL; Cox & Alexander, 1999) and APHAB. For each outcome measure, the participants were classified as successful or unsuccessful users based on somewhat arbitrary criteria. For example, the participants who had a SADL score less than 5 points were defined as unsuccessful users (SADL scores range from 0 to 7). The results revealed that the accuracy of the prediction made by ANL ranged from 52% to 64%, which was much lower than the 85% accuracy reported by Nabelek et al (2006). However, the discrepancy in prediction accuracy between these two studies could be due to Nabelek et al (2006) establishing the ANL criteria based on hearing aid use pattern, while Schwartz and Cox (2012) used questionnaires other than the HA-Use to measure outcomes.

In short, the gap in the literature regarding the usefulness of using ANL to predict hearing aid outcomes stems from (1) the cross-sectional design of previous research, (2) the lack of long-term evaluation in prospective studies that take into account the effect of hearing aid experience, and (3) the limitation of how prediction models have been built and evaluated. To fill the gap, the objective of the current study was to investigate the extent to which pre-fitting ANL, with or without other predictors such as hearing aid experience, could predict short-term (3 months post-fitting) and long-term (12 months) outcomes. To achieve this objective, prediction models were built using logistic regression and several machine learning classifiers (e.g., decision tree). The performance of the prediction models was evaluated and compared using the cross-validation technique. The current study was part of a larger study and the 3-month outcome results of the first 80 participants of the current study have been reported in Ho et al (2013b).

## METHODS

### Participants

Participants were recruited from the Hearing Aid Clinic in the Buddhist Dalin Tzu-Chi General Hospital, Taiwan. Participants were eligible for inclusion in this study if they (1) were older than 20 years of age, (2) spoke Taiwanese as their primary language, and (3) decided to purchase hearing aids. In agreement with Nabelek et al (2006), once enrolled, if a participant's rationale for hearing aid disuse was not related to instrument performance, participation was terminated (severe illness,  $n = 2$ ; device lost,  $n = 1$ ). In total, 132 adults participated in the study and completed at least one outcome measure (see below).

Table 2 summarizes the participants' demographic, audiometric, and hearing aid fitting data. Hearing loss was defined as a mixed hearing loss if the mean air-bone gap across 0.5, 1, 2, and 4 kHz was greater than 10 dB. Although approximately half of the participants' hearing aids were fit unilaterally, none of the participants had a unilateral hearing loss. Table 2 also summarizes the results of the Chinese version (Chang et al, 2009) of the Hearing Handicap Inventory for Elderly-Screening questionnaire (HHIE-S; Ventry & Weinstein, 1983). The

HHIE-S is a 10-item questionnaire that was designed to screen self-reported emotional and social consequences of hearing loss. The HHIE-S scores range from 0 to 40, with higher scores representing more negative impacts.

The participants' hearing aids were fit by audiologists who were independent of the study. The choice of hearing aid model, style, features, and bilateral/unilateral fitting was not controlled in the study and was determined on an individual basis by the audiologists and study participants. As reported by Ho et al (2013b), the gain/output and setting of the features were determined or guided by the manufacturer's fitting software and no real-ear measures were conducted.

### **ANL test**

ANL was measured using Taiwanese speech material. The material was a story taken from a Chinese children's book and read by a male Taiwanese adult at a normal conversational effort and speed. The twelve-talker babble from the official ANL CD (Cosmos Dist. Inc.) was used as the noise signal. The details of the development of the speech material, the selection of the babble, and the validation of the Taiwanese ANL are described in Ho et al (2013a).

The standard procedures described in the official ANL CD manual (Cosmos Dist. Inc.) were used to measure ANL. In brief, to measure the most comfortable level (MCL), the participants used hand signals to adjust the speech level. The speech signal was initially presented at 30 dB HL (American National Standards Institute, 2010). The participants first signaled to increase the speech level until it was too loud and then signaled to decrease the level until it was too soft in 5 dB steps. The speech level was then adjusted in 2 dB steps to the level that was most comfortable for listeners. Once the MCL had been established, the noise was added to find the maximum background noise level (BNL). The noise was initially presented at 16 dB below the MCL. As with the MCL, the participants increased the noise until it was too loud, and then decreased the noise until it became too soft in 5 dB steps. Finally, the participants were asked to find the maximum level that they could accept or put up with while listening to the speech. The background noise was adjusted in 2 dB steps. The ANL was calculated by subtracting the BNL from the MCL.

Before testing, verbal and written instructions were provided to participants. The instructions were translated from the English version included with the ANL CD. Special care was taken to ensure that the phrases "accept" and "put up with" were accurately translated (Ho et al, 2013a). Before the commencement of the formal measure, several (typically one to two) ANL practices were taken until the participants fully understood the procedures. For the first 105 participants, ANL was measured once. ANL was measured twice consecutively for the rest of 27 participants. For these 27 participants, the two ANLs were averaged and used in analyses. Because only 2 out of the 27 participants had differences between the two ANL measures larger than 2 dB (4 dB,  $n = 1$ ; 6 dB,  $n = 1$ ), it is likely that for the first 105 participants one measurement was able to assess ANL with reasonable accuracy.

The participants' ANLs were measured binaurally in a sound-treated booth without wearing hearing aids. The speech and noise stimuli were generated by a computer and a sound interface, routed to a GSI-61 audiometer, and then presented to the listener at 0° azimuth and 0° elevation from a Grason-Stadler loudspeaker. The loudspeaker was located in a corner of the booth. The distance between the loudspeaker and the listener was 1.2 m. The audiometer and sound field were calibrated according to American National Standards Institute S3.6-2010.

### Hearing aid outcome measure

Hearing aid outcome was assessed using two self-report inventories. The first inventory was the Chinese version of the IOI-HA (Cox et al, 2002). This inventory is a seven-item questionnaire designed to evaluate the effectiveness of hearing aid interventions. Each of the seven items assesses one of the outcome domains that are important to the overall success of hearing aid: (1) daily use, (2) benefit, (3) residual activity limitation, (4) satisfaction, (5) residual participation restriction, (6) impact on others, and (7) quality of life. Possible scores for each item range from 1 to 5, with higher scores suggesting better outcomes. The global score, which is the sum of the scores of the seven items (ranging from 7 to 35), was used to quantify overall hearing aid outcome. In the current study, participants who had global scores higher than 26.3, which is the mean norm score of the Chinese IOI-HA reported by Liu et al (2011), were defined as successful users.

The second inventory was the Chinese translation of the HA-Use (Nabelek et al, 2006). The HA-Use has only one question ("How do you use your hearing aids?") with three possible responses: (1) wearing hearing aids whenever needed, (2) occasionally, and (3) not wearing hearing aids. In accordance with Nabelek et al (2006), participants who wore hearing aids whenever needed (full-time users) were defined as successful users while those who wore hearing aid occasionally (part-time users) or did not wear hearing aids (non-users) were defined as unsuccessful users.

### Procedures

All participants read and signed a statement of informed consent approved by the Institutional Review Board at the Buddhist Dalin Tzu-Chi General Hospital. After agreeing to participate in the study, ANL was measured. Three and twelve months after hearing aid fitting, a research assistant called the participants on the phone to administer the IOI-HA and HA-Use. The assistant read the questions and available responses to the participants and then recorded their responses. If the participant had difficulty understanding the assistant on the phone, a participant's family member was asked to serve as a liaison to assist communication between the participant and assistant. For various reasons such as loss of contact with participants and participants' unwillingness to complete the longer IOI-HA, outcome data were not collected from all participants: the numbers of completed 3- and 12-month IOI-HA and 3- and 12-month HA-Use were 130, 123, 131, and 128, respectively. Because the current study was an observational study, the assistant did not encourage the participants who reported poorer outcomes to return to the clinic.

## Prediction model

To examine the extent to which pre-fitting ANL, with or without other predictors, could predict hearing aid outcomes (successful vs. unsuccessful), logistic regression that takes a linear combination of predictors to compute the probability of a class of the categorical dependent variable was used. Logistic regression was selected because it has been used in previous ANL research (Freyaldenhoven et al, 2008a, b; Ho et al, 2013b; Nabelek et al, 2006).

In addition to logistic regression, five machine learning algorithms, or classifiers, were included in the current study. These classifiers were selected due to their popularity in the machine learning literature. The reason for including other classifiers is that logistic regression is an inherently simple classifier that assumes the classes of the categorical dependent variable are linearly separable. The alternative classifiers selected make different assumptions about the relationship between variables and use various mechanisms to make predictions and, therefore, might outperform logistic regression.

The first algorithm was a naïve Bayes classifier, assuming independence between predictors. The second algorithm was a 3-nearest-neighbors instance based classifier, which uses the characteristics of the three data points that are closest to a given instance using Euclidean distance in the data space to predict the categorical class. A decision tree created using the C4.5 algorithm, which uses the most informative predictors to split classes, was the third classifier. The fourth algorithm was a multilayer perceptron classifier, which is an artificial neural network classification algorithm. The fifth algorithm was a sequential minimal optimization support vector machine, which can perform a non-linear classification by finding a maximum margin hyper-plane that separates the categorical classes. For detailed information about these algorithms, see Witten and Frank (2005).

For each of the four hearing aid outcomes (3- and 12-month IOI-HA and HA-Use) and each of the six classifiers (logistic regression plus the five machine learning algorithms), three prediction models were built (i.e., trained) using the software Weka 3.6.12 (Hall et al, 2009). The first model used ANL as the sole predictor. The second model employed eight patient-centered variables available in the current study as predictors: ANL, age, gender (male or female), pure tone average across ears, hearing loss type (mixed or sensorineural), hearing aid experience (first-time or experienced user), unilateral/bilateral fitting, and HHIE-S score. These variables were used because they are typically available to audiologists before or at the time of hearing aid fitting and might be useful for hearing aid success prediction. The third model used ANL and hearing aid experience as predictors. Weka's correlation-based feature selection algorithm indicated that ANL and hearing aid experience were the predictors most relevant to hearing aid outcome prediction.

In addition to the above-mentioned prediction models, a simple classifier called ZeroR was included in the current study. ZeroR is the simplest classifier which ignores all predictors and predicts the majority category. For example, ZeroR will be trained to predict all hearing aid users as successful users if most users in the dataset utilized to train this classifier were successful users. ZeroR has no predictability power; it is often used to determine a baseline performance and serves as a benchmark for other prediction methods.

The prediction models, including ZeroR, were evaluated using ten iterations of ten-fold cross-validation. Specifically, the dataset was randomly partitioned into ten equal size subsets. Nine of the subsets were used to train the model (i.e., the training set) and the remaining subset was utilized to evaluate the model (i.e., the test set). After each evaluation, several metrics such as prediction accuracy, area under the receiver operating characteristic curve (AUC), and true positive and negative rates were computed. This evaluation process was then repeated ten times (the folds), with each of the ten subsets used exactly once as the test data set. The ten-fold cross-validation was repeated ten times (the iterations), resulting in 100 test results for each model. The cross-validation process was conducted using the Weka Experimenter interface. In the current study the overall performance of the prediction model was evaluated using the AUC.

## RESULTS

### Hearing aid outcome

Recall that for the IOI-HA (scores ranging from 7 to 35), a participant was a successful user if his/her IOI-HA global score was higher than 26.3 (the mean norm score of the Chinese IOI-HA). For the HA-Use, a participant who wore hearing aids whenever needed (full-time user) was a successful user. The first column of Table 3 shows the mean global score of the IOI-HA and the numbers of full-time, part-time, and non-users defined by the HA-Use. The mean IOI-HA scores (27.3 and 28.1) were close to, but slightly higher than, the mean score (26.3) of the norm reported by Liu et al (2011). For 3- and 12-month IOI-HA outcomes, 63.1% and 73.2% of the participants, respectively, were successful users (the second column of Table 3). For the HA-Use, approximately 75% of the participants were successful users, which is higher than the 36% reported by Nabelek et al (2006) but lower than the 91% reported by Olsen et al (2012).

To examine the relationship between hearing aid success defined by the IOI-HA and HA-Use, chi-square tests were conducted. The results indicated that the two types of hearing aid success were associated ( $p < 0.001$  and  $\phi = 0.43$  for 3 months;  $p < 0.001$  and  $\phi = 0.55$  for 12-months). The significant but moderate associations suggested that the IOI-HA and HA-Use measured similar but different aspects of outcome. Because the IOI-HA's first item assessed the degree of hearing aid daily use, this item should generate consistent results with the HA-Use (Nabelek et al, 2006). The mean scores of the IOI-HA's first item (ranging from 1 to 5) for full-time, part-time, and non-users were 4.7, 3.2, and 1.0, respectively (3- and 12-month data combined). T-tests with Bonferroni correction indicated that all differences in the item score between the three user groups were significant.

Figure 1 shows the relationship between 3- and 12-month IOI-HA global scores. The significant correlation ( $r = 0.79$ ,  $p < 0.001$ ) indicated that in general the IOI-HA outcomes were stable across time. For hearing aid success defined by the IOI-HA, 11.5% ( $n = 14$ ) of the participants who completed this measure at both 3 and 12 months changed from unsuccessful to successful users and 4.1% ( $n = 5$ ) reported the opposite. For the HA-Use, 3.9% ( $n = 5$ ) of the participants changed from unsuccessful to successful users and 7.9% ( $n = 10$ ) reported the opposite.

### Pre-fitting ANL and hearing aid success

The third and fourth columns of Table 3 show the mean ANLs for successful and unsuccessful users defined by each outcome measure. Although the differences were not large, successful users generally had lower (better) pre-fitting ANLs. Four separate t-tests were conducted for each outcome to determine if ANL was different for successful and unsuccessful users. Bonferroni correction was applied to adjust multiple comparisons. The results indicated that the ANL of successful users was lower than that of unsuccessful users for 3-month IOI-HA (unadjusted  $p < 0.001$ ), 12-month IOI-HA (unadjusted  $p = 0.001$ ), and 3-month HA-Use (unadjusted  $p = 0.004$ ). However, the difference was not significant for 12-month HA-Use (unadjusted  $p = 0.033$ ). Figure 2 shows IOI-HA score as a function of ANL. In general, IOI-HA score decreased as ANL increased.

### Predicting hearing aid success

Figure 3 shows the mean AUC averaged across 100 cross-validation results of each prediction model when the model used ANL (Figure 3A), all patient-centered variables (3B), and ANL plus hearing aid experience (3C) to predict hearing aid success. An AUC value of 1 represents a perfect prediction model while a value of 0.5 represents a worthless model. In general, a model with an AUC value lower than 0.7 is considered to be a poor model (Masegosa, 2013). A series of paired t-tests were conducted to examine the difference in AUC between the logistic regression and each of the remaining classifiers (including ZeroR), corrected for multiple comparisons. The difference that reached the significance level is labeled by an asterisk in Figure 3. The results first indicated that, in most cases, logistic regression had higher AUCs (i.e., better performance) than ZeroR, which was a worthless classifier and had an AUC of 0.5. However, the logistic regression model that used ANL as the sole predictor to predict the 12-month HA-Use outcome did not outperform ZeroR. The results further indicated that the logistic regression models' AUCs were significantly higher than several classifiers and were not lower than any of the classifiers evaluated in the current study. Therefore, the rest of the paper will focus on logistic regression.

The top half of Table 4 shows the mean prediction accuracy and AUC averaged across the 100 cross-validation results of each logistic regression model that predicted hearing aid success. Prediction accuracy represents the probability for the model to correctly identify successful and unsuccessful users from all individuals in the dataset. Table 4 also shows the true positive and negative rates (TPR and TNR, respectively) of each model. The TPR represents the probability for the model to identify unsuccessful users from those who were truly unsuccessful with hearing aids, while the TNR reflects the probability to identify successful users from those who were truly successful with hearing aids.

Table 4 indicates that many models, especially those that predicted HA-Use outcomes, while having relatively high prediction accuracy (~70% to 78%), had relatively low AUCs (~0.6 to 0.65). Furthermore, most models had high TNRs but low TPRs. These results were due to the imbalance between the numbers of successful and unsuccessful users in the dataset. Specifically, the ratio of successful to unsuccessful users was approximately 3 to 1 for the HA-Use outcome (Table 3). Because successful users outnumbered unsuccessful users, the

logistic regression models were trained to predict most participants as successful users so that the prediction accuracy could be maximized. As a result, although the models could achieve high accuracy and could correctly predict most successful users (i.e., high TNRs shown in Table 4), only a small portion of unsuccessful users could be identified (i.e., low TPRs). Because the predictions made by these models were similar to the ZeroR that predicted all participants were successful users, the models had low AUCs and their high prediction accuracy was misleading.

To remedy the data imbalance problem, the logistic regression models were re-trained using cost-sensitive machine learning algorithms. In short, standard learning algorithms, such as logistic regression, compute the probability of a given category (e.g., successful user category) for a given instance (e.g., a patient). A probability threshold (typically 50%) is then used to transform the probability into nominal predictions. Cost-sensitive learning is an approach that changes the probability threshold without explicitly doing so by specifying the “costs” of different misclassifications. In the current study the misclassification cost was determined by the ratio of the number of successful users to the number of unsuccessful users in the dataset. For example, if the ratio of successful to unsuccessful users in the dataset is 3 to 1, the cost of misclassifying an unsuccessful user as a successful user will be set to three times of the cost of doing the opposite. This misclassification cost would ensure that the TPR of a given model was roughly equal to its TNR. The rationale for equating the TPR and TNR is based on the assumption that it is equally important to identify successful and unsuccessful users. The cost-sensitive learning and the selection of the cost of the current study was conceptually similar to Freyaldenhoven et al (2008a, b), which used the ratio of the number of successful users to the number of all users in the dataset to determine the probability threshold of the logistic regression model.

The results of the cost-sensitive logistic regression models are shown in the bottom half of Table 4. These models had the same AUCs as the original models because (1) the receiver operating characteristic curve is created by varying the probability threshold of a model and (2) the cost-sensitive model was developed from the original model by changing its probability threshold. The TPR and TNR of the cost-sensitive models were roughly equal, indicating that these models were equally good at identifying successful and unsuccessful users. Compared to the original models, the cost-sensitive models had lower prediction accuracy, ranging from 61% to 68% (IOI-HA) and from 55% to 61% (HA-Use). These accuracy values were more consistent with the AUCs and more reasonably reflected the extent to which ANL could predict outcomes. Therefore, the rest of the paper will focus on the results of the cost-sensitive models.

To examine the effect of time (3-month/12-month), outcome measure (IOI-HA/HA-Use), and predictor (ANL/all variables/ANL plus hearing aid experience) on the overall performance of the prediction model, a three-way analysis of variance (ANOVA) was conducted. The dependent variable was AUC (obtained from 100 cross-validation tests). The results revealed that all the main effects were significant (time:  $F_{1, 396} = 14.3, p < 0.001$ ; outcome:  $F_{1, 396} = 39.6, p < 0.001$ ; predictor:  $F_{2, 792} = 11.1, p < 0.001$ ). Follow-up analyses further indicated that the AUC of the ANL-plus-hearing aid experience models did not differ from the models that included all variables, while it was larger than the AUC of the models

that used ANL as the sole predictor. No interaction was significant. These results indicated that the prediction models performed more poorly in predicting the 12 months than 3 months post-fitting outcomes, in predicting the HA-Use outcomes than IOI-HA outcomes, and when ANL was used as the sole predictor.

Because the models that used ANL and hearing aid experience as the predictors had similar AUCs, but were simpler than the models that included all variables, they are suitable for clinical use. Table 5 shows the models' ANL cutoff between successful and unsuccessful users. Individuals who have ANLs lower than the cutoff will be predicted as successful users. The cutoffs were higher for the original models than the cost-sensitive models, reflecting that the original models tended to predict most participants to be successful users. The cutoffs were higher for experienced users than first-time users. For first-time users, the ANL cutoffs of all cost-sensitive models were around 9 to 10 dB.

## DISCUSSION

### Hearing aid outcome

The Chinese versions of the IOI-HA and HA-Use were administered on the phone to measure hearing aid outcomes in the current study. The significant associations between these two measures and the strong relationship between the IOI-HA's first item (hearing aid daily use) and the HA-Use supported the validity of the outcome measures used in the current study.

For both the IOI-HA and HA-Use, outcomes were generally stable between 3 and 12 months post-fitting. This is consistent with the literature (Humes et al, 2002).

### Pre-fitting ANL and hearing aid success

The results of the study indicated that, for the IOI-HA, the participants who had lower pre-fitting ANLs tended to report better outcomes at both 3 and 12 months post-fitting. These results are consistent with Ho et al (2013b) and Taylor (2008).

Compared to the IOI-HA, the relationship between pre-fitting ANL and HA-Use was not as strong. Specifically, although the mean pre-fitting ANL of full-time users was lower than that of part-time/non-users, this difference was not statistically significant at 12 months post-fitting. Because the ANL differences (1.5 to 2.1 dB; Table 3) were smaller than those reported by Nabelek et al (2006) and Freyaldenhoven et al (2008b) (> 4 to 5 dB), the current study was unable to replicate the strong relationship between ANL and HA-Use reported by these two studies. This discrepancy could stem from previous studies measuring ANL and outcome at the same time point, while the current research examines the relationship between pre-fitting ANL and post-fitting outcomes. It could also result from the uncontrolled subject number in each user group (full-time, part-time, and non-user) in the current study, while Nabelek et al (2006) and Freyaldenhoven et al (2008b) aimed to have an equal subject number across the three groups and therefore had more non-users.

### Predicting hearing aid success

All logistic regression models, except for the model that used ANL as the sole predictor to predict the 12-month HA-Use outcomes, performed better than ZeroR, which had no predictability power (Figure 3). These results suggested that in most cases ANL, with or without other predictors, provided useful information for outcome prediction. However, except for the models that predicted 3-month IOI-HA outcomes, most models had AUCs lower than 0.7 (Table 4) and can be considered as poor models (Masegosa, 2013). The cost-sensitive models' prediction accuracy ranged from 61% to 68% for the IOI-HA and from 55% to 61% for the HA-Use.

Caution should be taken when comparing the model's prediction accuracy in the current study to previous research. This is because (1) the current study used a prospective design while most of the previous research used a cross-sectional design and (2) previous research often used the same dataset to train and evaluate the prediction model, which can overestimate model performance. Regardless, in terms of using ANL as the sole predictor to predict the HA-Use outcome, the 55% to 60% prediction accuracy of the cost-sensitive models in the current study is somewhat close to the 68% reported by Freyaldenhoven et al (2008b) but much poorer than the 85% reported by Nabelek et al (2006). The 55% to 68% accuracy across all cost-sensitive models of the current study, however, was fairly close to the 52% to 64% accuracy reported by Schwartz and Cox (2012), which used the ANL criteria established by Nabelek et al (2006) based on the HA-Use to predict outcomes measured using questionnaires other than the HA-Use.

### Role of time and hearing aid experience

The current study suggests the effect of time on outcome prediction. Approximately 16% (n = 19) of the participants who completed the IOI-HA at both 3 and 12 months reported changing hearing aid success and most of them (n = 17) were first-time users. For the HA-Use, 12% (n = 15) of the participants changed outcomes and all of them were first-time users. The mean ANL of those who changed from unsuccessful at 3 months to successful users at 12 months was higher than the ANL of those who reported the opposite (12 dB vs. 10 dB, for both IOI-HA and HA-Use). Because a number of participants who had higher ANLs initially reported poorer outcomes at 3 months and then became successful users at 12 months (and vice versa), the prediction models performed more poorly in predicting the 12-month than 3-month outcomes. This result suggests that the role of noise acceptance in determining outcomes may decrease over time after hearing aid fitting. If this is the case, more long-term post-fitting outcomes (e.g., years) may not be predictable from pre-fitting ANL measurements. An alternative explanation is that some participants' noise acceptance (and ANL) could have changed from pre-fitting to 12 months post-fitting, which could in turn cause hearing aid outcome to change. If this is the case, although pre-fitting ANL may not predict long-term post-fitting outcome, the relationship between ANL and outcome still exists. A prospective study that measures both ANL and hearing aid outcomes at same time points across a longer period of time is needed to examine these speculations.

The result that experienced users had higher ANL cutoffs than first-time users (Table 5) suggests the importance of hearing aid experience in outcome prediction. In the clinical

population of the current study, experienced users had better outcomes and most (~90%) of them were successful users regardless of their ANLs. As a result, the prediction models had very high ANL cutoffs for experienced users (Table 5). This result confirms the role of previous hearing aid experience in outcomes (Humes & Humes, 2004). The results also suggest that, for those who have previous hearing aid experience and are willing to come to the clinic and purchase hearing aids again, ANL is less important in predicting their post-fitting outcomes.

For first-time users, the ANL cutoff ranged from 9 to 10 dB (Table 5), which is almost identical to those reported by Nabelek et al (2006) and Freyaldenhoven et al (2008b). Therefore, it seems that an ANL around 9 to 10 dB is a reasonable cutoff between successful and unsuccessful hearing aid users for both IOI-HA and HA-Use.

### Limitations

The current study has several limitations concerning its generalizability. Firstly, because the IOI-HA and HA-Use had different relationships with ANL, the results of the current study may not generalize to other outcome measures. Secondly, the current study's participants were very heterogeneous in terms of age, hearing loss type, and unilateral/bilateral fitting. They also had more severe hearing losses (mean pure tone average across ears = 70 dB HL) compared to previous ANL research and had more males than females. The heterogeneity seemed not to affect the performance of the prediction model because including more patient-centered variables other than hearing aid experience (e.g., gender) into the model did not improve the AUC, i.e., the model's performance. However, it is still unknown if the results of the current study can generalize to other clinical populations in different cultures. Finally, all of the limitations reported in Ho et al (2013b), which include (1) only one ANL measurement was conducted on most participants and (2) hearing aids were fitted without using real-ear measures, apply to the current study. See Ho et al (2013b) for the detailed discussion of these limitations.

### CONCLUSIONS

The current study indicates that pre-fitting ANL, together with other predictors such as hearing aid experience, could predict post-fitting outcomes. The performance of the prediction models was better when predicting the IOI-HA than HA-Use. The prediction accuracy ranged from 61% to 68% for the IOI-HA and from 55% to 61% for the HA-Use. The performance of the prediction models was affected by time, such that the models performed more poorly in predicting the 12-month outcomes than 3-month outcomes. Finally, because most experienced hearing aid users in the clinical population were successful users regardless of their ANLs, ANL played a smaller role in predicting their outcomes.

### Acknowledgments

This research was supported by a research Grant DTCRD100(2)-E-07 from Buddhist Dalin Tzu-Chi General Hospital, Chiayi, Taiwan. The first author was supported by grants R03-DC012551 and R01-DC012769 from the National Institute on Deafness and Other Communication Disorders, USA.

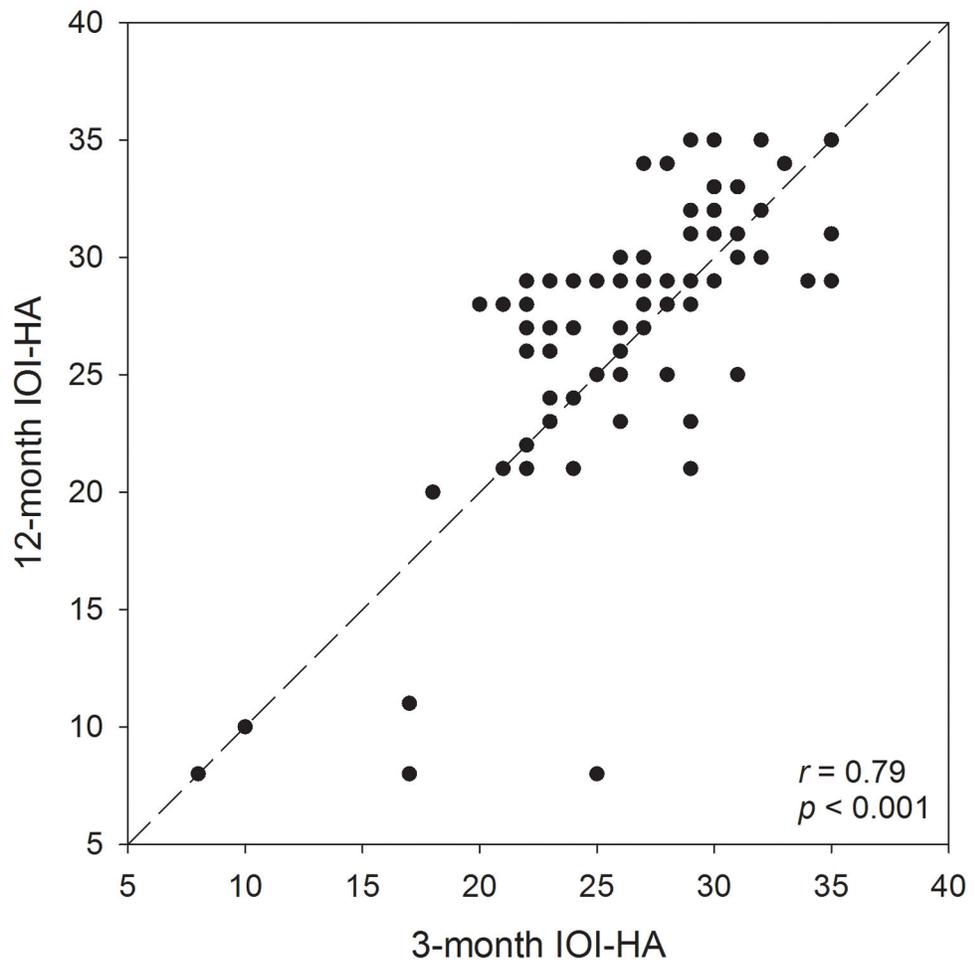
## Acronyms and abbreviations

<b>ANL</b>	acceptable noise level
<b>AUC</b>	area under the receiver operating characteristic curve
<b>BNL</b>	background noise level
<b>HA-Use</b>	the hearing aid use questionnaire developed by Nabelek et al (2006)
<b>IOI-HA</b>	International Outcome Inventory for Hearing Aids
<b>HHIE-S</b>	Hearing Handicap Inventory for Elderly-Screening
<b>MCL</b>	most comfortable level
<b>TNR</b>	true negative rate
<b>TPR</b>	true positive rate

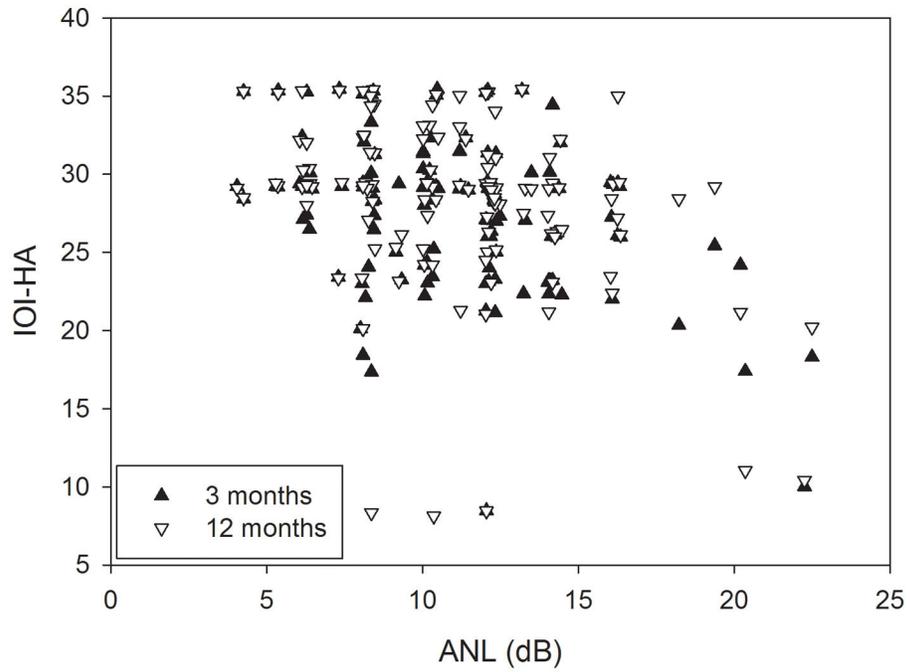
## References

- American National Standards Institute. Specification for Audiometers (ANSI S3.6). New York: American National Standards Institute; 2010.
- Chang HP, Ho CY, Chou P. The factors associated with a self-perceived hearing handicap in elderly people with hearing impairment—results from a community-based study. *Ear Hear.* 2009; 30:576–583. [PubMed: 19633566]
- Cox RM, Alexander GC. The abbreviated profile of hearing aid benefit. *Ear Hear.* 1995; 16:176–186. [PubMed: 7789669]
- Cox RM, Alexander GC. Measuring satisfaction with amplification in daily life: The SADL scale. *Ear Hear.* 1999; 20:306. [PubMed: 10466567]
- 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, Stephens D, Kramer SE. Translations of the International Outcome Inventory for Hearing Aids (IOI-HA). *Int J Audiol.* 2002; 41:3–26. [PubMed: 12467365]
- Freyaldenhoven MC, Nabelek AK, Tampas JW. Relationship between acceptable noise level and the abbreviated profile of hearing aid benefit. *J Speech Lang Hear Res.* 2008a; 51:136–146. [PubMed: 18230861]
- Freyaldenhoven MC, Plyler PN, Thelin JW, Muenchen RA. Acceptance of noise growth patterns in hearing aid users. *J Speech Lang Hear Res.* 2008b; 51:126–135. [PubMed: 18230860]
- Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, et al. The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter.* 2009; 11:10–18.
- Ho HC, Wu YH, Hsiao SH, Stangl E, Lentz EJ, et al. The equivalence of acceptable noise level (ANL) with English, Mandarin, and non-semantic speech: a study across the U.S. and Taiwan. *Int J Audiol.* 2013a; 52:83–91. [PubMed: 23153252]
- Ho HC, Wu YH, Hsiao SH, Zhang X. Acceptable noise level (ANL) and real-world hearing-aid success in Taiwanese listeners. *Int J Audiol.* 2013b; 52:762–770. [PubMed: 24099584]
- Humes LE, Humes LE. Factors affecting long-term hearing aid success. *Seminars in Hearing.* 2004; 25(1):63–72.
- Humes LE, Wilson DL, Barlow NN, Garner C. Changes in hearing-aid benefit following 1 or 2 years of hearing-aid use by older adults. *J Speech Lan Hear Res.* 2002; 45:772–782.
- Liu H, Zhang H, Liu S, Chen X, Han D, et al. International outcome inventory for hearing aids (IOI-HA): Results from the Chinese version. *Int J Audiol.* 2011; 50:673–678. [PubMed: 21916818]
- Masegosa, AD. *Exploring Innovative and Successful Applications of Soft Computing.* IGI Global; 2013.

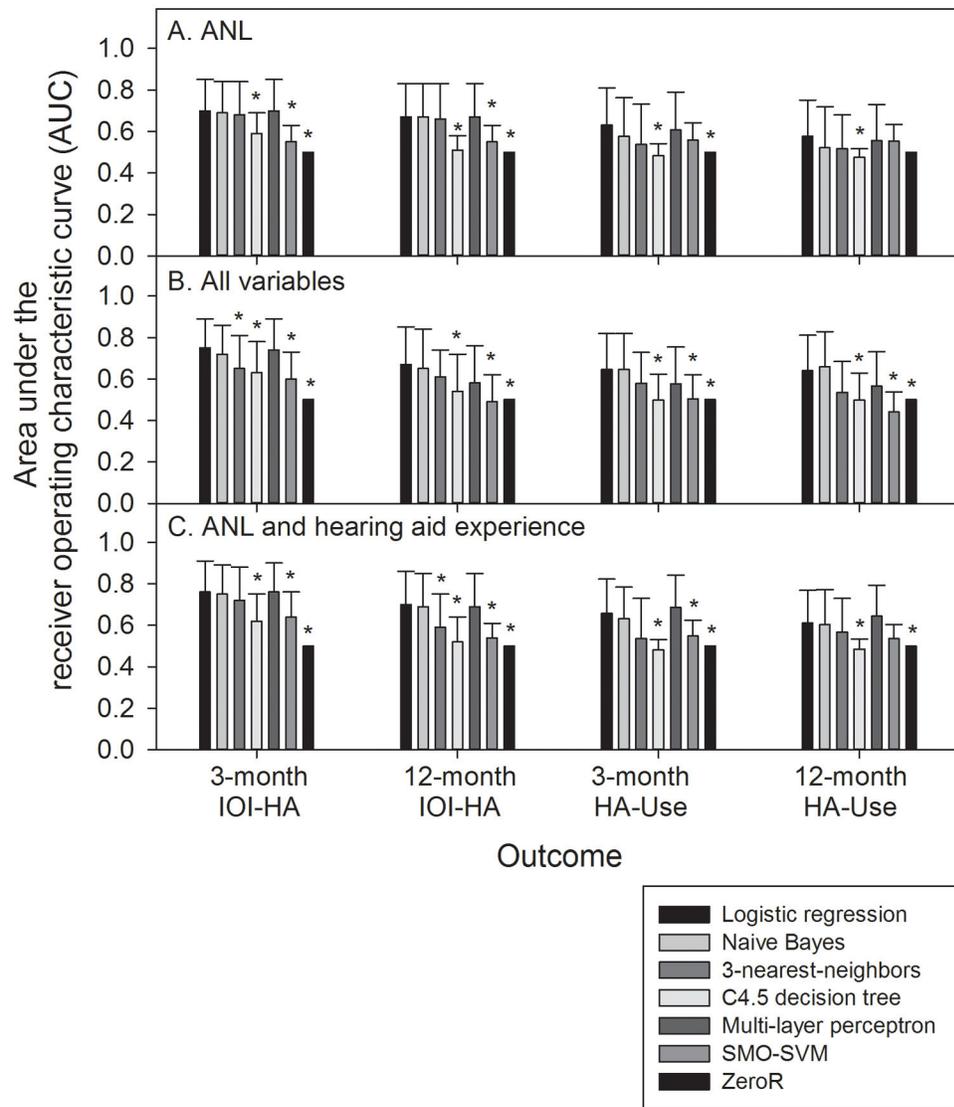
- Nabelek AK, Freyaldenhoven MC, Tampas JW, Burchfiel SB, Muenchen RA. Acceptable noise level as a predictor of hearing aid use. *J Am Acad Audiol*. 2006; 17:626–639. [PubMed: 17039765]
- Nabelek AK, Tampas JW, Burchfield SB. Comparison of speech perception in background noise with acceptance of background noise in aided and unaided conditions. *J Speech Lang Hear Res*. 2004; 47:1001–1011. [PubMed: 15603458]
- Nabelek AK, Tucker FM, Letowski TR. Toleration of background noises: relationship with patterns of hearing aid use by elderly persons. *J Speech Hear Res*. 1991; 34:679–685. [PubMed: 2072693]
- Olsen SO, Lantz J, Nielsen LH, Brannstrom KJ. Acceptable noise level (ANL) with Danish and non-semantic speech materials in adult hearing-aid users. *Int J Audiol*. 2012; 51:678–688. [PubMed: 22731922]
- Schwartz, KS.; Cox, RM. Does Acceptable Noise Level predict hearing aid success?. Poster presented at the Annual Meeting of the American Auditory Society; 2012. Retrieved from [http://www.harlmemphis.org/files/4813/4625/3302/anl2012\\_poster.pdf](http://www.harlmemphis.org/files/4813/4625/3302/anl2012_poster.pdf)
- Taylor B. The Acceptable Noise Level Test as a predictor of real-world hearing aid benefit. *Hear J*. 2008; 61:39–42.
- Ventry IM, Weinstein BE. Identification of elderly people with hearing problems. *Asha*. 1983; 25:37–42. [PubMed: 6626295]
- Walravens E, Keidser G, Hartley D, Hickson L. An Australian version of the acceptable noise level test and its predictive value for successful hearing aid use in an older population. *Int J Audiol*. 2014; 53:S52–S59. [PubMed: 24447234]
- Witten, IH.; Frank, E. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann; 2005.



**FIGURE 1.** Relationship between 3 and 12-month global IOI-HA scores. The dashed line represents perfect match.



**FIGURE 2.** Global IOI-HA score as a function of ANL. The data points were jittered to better illustrate the relationship between IOI-HA and ANL.



**FIGURE 3.** Mean area under the receiver operating characteristic curve (AUC) of the classifiers that used either ANL (3A), all available patient-centered variables (3B), or ANL plus hearing aid experience (3C) to predict outcomes. Error bars = 1 SD. SMO-SVM: sequential minimal optimization support vector machine.

Summary of seven studies evaluating ANL and real-world hearing aid outcomes. The first four studies found significant associations between ANL and outcomes, while this association was less clear in the last three studies. The percentage shown in parentheses in the last column represents the percentage of observed successful users defined by a given outcome measure. Nabelek et al (2006) and Freyaldenhoven et al (2008a) used the same group of subjects.

**Table 1**

	Subject number	Hearing aid experience	ANL delivery	Design	Outcome measure
Nabelek et al (2006)/ Freyaldenhoven et al (2008a)	191	First-time users: 16.2%; Experienced users: 3 months to 3 years experience	Sound field	Cross-sectional	HA-Use (36.1%)
Freyaldenhoven et al (2008b)	69	Information unavailable	Sound field	Cross-sectional	HA-Use (36.2%)
Taylor (2008)	27	All first-time users	Binaural earphones	Prospective (1-month outcome)	IOI-HA
Ho et al (2013b)	80	First-time users: 77.5%	Sound field	Prospective (3-month outcome)	IOI-HA (53.8%) HA-Use (73.8%)
Schwartz and Cox (2012)	50	All experienced users; at least 6 months experience	Sound field	Cross-sectional	Various measures (70% to 88%)
Olsen et al (2012)	63	All experienced users; mean experience = 11 years	Monaural earphone	Cross-sectional	IOI-HA HA-Use (91%)
Walravens et al (2014)	96	All hearing aid owners; < 1 year: 9.4%, 1 to 5 years: 47.9%, > 5 years: 41.7%	Binaural earphones	Cross-sectional	HA-Use (50.9%) Hearing aid daily use

**Table 2**

The participants' demographic, audiometric, and hearing aid fitting data, and scores of the Hearing Handicap Inventory for Elderly-Screening (HHIE-S). Pure tone average (PTA) was the average of hearing thresholds across ears at frequencies 0.5, 1, 2, and 4 kHz.

Variable	Subjects (n = 132)	
Age (years)	Mean	72.3
	SD	9.1
	Range	44 – 87
Gender	Male	n = 84 (63.6%)
	Female	n = 48 (36.4%)
PTA (dB HL)	Mean	70.0
	SD	11.3
	Range	36.9 – 102.5
Hearing loss type	Mixed	n = 39 (30.0%)
	Sensorineural	n = 93 (70.0%)
Hearing aid experience	First-time user	n = 103 (78.0%)
	Experienced user	n = 29 (22.0%)
Unilateral/bilateral fitting	Unilateral	n = 65 (49.2 %)
	Bilateral	n = 67 (50.8%)
HHIE-S	Mean	29.6
	SD	8.0
	Range	8 – 40

**Table 3**

The outcome (first column), percentage of hearing aid success (second column), and ANL of successful and unsuccessful users (third and fourth columns) of each measure. Standard deviations are in parentheses. The three numbers shown at the bottom of the first column indicate the participant numbers of full-time, part-time, and non-users, respectively.

	ANL			
	Outcome	Hearing aid success	Successful	Unsuccessful
IOI-HA				
3 months (n = 130)	27.3 (4.7)	63.1%	9.8 (3.0)	12.5 (3.9)
12 months (n = 123)	28.1 (5.5)	73.2%	10.1 (3.3)	12.6 (4.0)
HA-Use				
3 months (n = 131)	99/28/4	75.6%	10.3 (3.2)	12.4 (4.4)
12 months (n = 128)	94/23/11	73.4%	10.4 (3.2)	11.9 (4.5)

**Table 4**

The mean prediction accuracy, area under the receiver operating characteristic curve (AUC), true positive rate (TPR), and true negative rate (TNR) of the original and cost-sensitive logistic regression models that used either ANL, all available patient-centered variables, or ANL plus hearing aid experience to predict the 3- and 12-month IOI-HA and HA-Use outcomes.

Outcome/Predictor	Logistic regression					
	Accuracy (%)		AUC		TPR/TNR	
	3 months	12 months	3 months	12 months	3 months	12 months
IOI-HA: original model						
ANL	70.8	73.7	0.70	0.67	0.40/0.89	0.12/0.96
All variables	70.4	71.6	0.75	0.67	0.48/0.84	0.24/0.89
ANL and hearing aid experience	71.2	72.6	0.76	0.70	0.57/0.80	0.16/0.93
HA-Use: original model						
ANL	77.8	74.5	0.63	0.58	0.12/0.99	0.05/0.99
All variables	74.1	69.4	0.64	0.64	0.13/0.94	0.12/0.90
ANL and hearing aid experience	76.9	74.5	0.66	0.61	0.13/0.98	0.09/0.98
IOI-HA: cost-sensitive model						
ANL	63.9	61.0			0.65/0.64	0.60/0.61
All variables	68.3	64.5	Identical to the original model		0.68/0.68	0.58/0.67
ANL and hearing aid experience	66.3	63.2			0.71/0.64	0.58/0.65
HA-Use: cost-sensitive model						
ANL	60.3	55.2			0.63/0.61	0.53/0.57
All variables	60.0	61.0	Identical to the original model		0.58/0.60	0.60/0.61
ANL and hearing aid experience	58.2	55.1			0.64/0.57	0.62/0.53

**Table 5**

ANL cutoff between successful and unsuccessful users defined by the IOI-HA and HA-Use for first-time and experienced hearing aid users.

	ANL cutoff (dB)			
	First-time user		Experienced user	
	3 months	12 months	3 months	12 months
IOI-HA, original model	11.8	15.0	18.9	20.7
HA-Use, original model	16.5	17.3	24.4	28.8
IOI-HA, cost-sensitive model	9.4	10.1	17.0	16.0
HA-Use, cost-sensitive model	9.8	9.0	18.6	21.4

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript