Updated 2021-03-25 21:58:44 PDT

* Note correction on p13: It should be 1533 males, not 1553.

Method

Data analysis [replaces "Statistical analysis", p18]

Data analysis techniques

Two methods of data analysis were used to predict hearing aid purchase from 28 variables of interest (Table 1): logistic regression and a classification tree analysis.

Logistic regression is a commonly used method for analyzing data from experimental designs with multiple predictor variables and a binary outcome variable. A logistic regression model uses a weighted combination of variables to predict an outcome for each participant or case, with the predicted outcome being in the form of log odds. The information obtained from a logistic regression model includes odds ratios that describe how key variables increase or decrease the odds of an outcome occurring.

Similar to logistic regression, a classification tree can be used to analyze data from an experiment with multiple categorical or continuous predictors and a binary outcome. In contrast to logistic regression, a classification tree analysis uses variables to repeatedly classify observations into smaller and "purer" groups of observations, with the goal of finding the sequence of variables (either particular levels of categorical variables or particular cut-off values of continuous variables) that leads to one or the other outcome. The algorithm begins by searching for the predictor that best splits the observations into two groups, with the aim of sorting each type of outcome into its own group. If the predictor is a continuous variable, a cut-off value that best sorts or splits the observations is selected; if the predictor is a categorical value, a particular level or levels may be used to split the observations. A measure such as "information gain" or the Gini index may be selected to calculate how well the splitting rule sorts observations into two outcome groups from one node of the tree to the next. The information obtained from a classification tree models includes the sequence of categories and/or cut-off values of key variables that predicts a particular outcome. The findings from a classification tree model may be more practically useful than the findings from a logistic regression model, as a classification tree can provide specific values from test scales or questionnaire items that can be used to guide clinical decision-making.

The performance of both logistic regression and classification tree models can be measured in terms of accuracy (how well the model correctly predicts each participant's decision to purchase hearing aids based on each participant's predictor variables), as well as sensitivity (participants who were predicted to purchase hearing aids, as a percentage of those who actually purchased hearing aids) and specificity (participants who were predicted not to purchase hearing aids, as a percentage of those who actually did not purchase hearing aids).

The data in the current study were analyzed using R (version 4.0.4, R Core Team, 2021). Logistic regression was performed using the *glm* function, specifying "family = binomial", with 28 predictor variables and a binary outcome variable with hearing aid purchase 'No' coded as 0 and hearing aid purchase 'Yes' coded as 1. After a logistic regression model was constructed with all 28 predictors, a backwards elimination procedure based on the Akaike Information Criterion was conducted using the *step* function, specifying "direction = backward". A final logistic regression model was constructed using only the predictors that improved the model according to the backwards elimination procedure. The rate of hearing aid purchase in the sample was used as the criterion value for determining whether a predicted outcome by the model would be considered a ‘Yes’ or ‘No’ prediction (for example, if the model’s predicted outcome for a case was 0.346 and the proportion of participants who purchased hearing aids was 0.198, the prediction would be considered a ‘Yes’ case; conversely, if the predicted outcome has been 0.146, it would be considered a ‘No’ case).

A classification tree model was built using the *rpart* function from the library *rpart* (version 4.1.15, Therneau & Atkinson, 2019), with the same 28 predictor variables and a binary outcome variable ‘No’ or ‘Yes’. As there were more cases where the outcome was ‘No’ than cases where the outcome was ‘Yes’ (80.2% and 19.8% of all outcomes, respectively), case weights were added such that ‘No’ cases were weighted 0.2466887 (149 ‘Yes’ cases / 604 ‘No’ cases) and ‘Yes’ cases were weighted 1.0, so that the classification tree model would not be biased towards ‘No’ cases. The experimental design was specified using “method = ‘class’” and splitting criterion was specified using “split = ‘gini’”. The growth of the tree model was limited by specifying the maximum depth of the tree using “rpart.control(depth = 2)”, for example. All other parameters were set to default, including a requirement for a minimum of 20 observations in a node before any splitting rule could be implemented. Trees were plotted using the *fancyRPlot* function in the library *rpart.plot* (version 3.0.9; Milborrow, 2020).

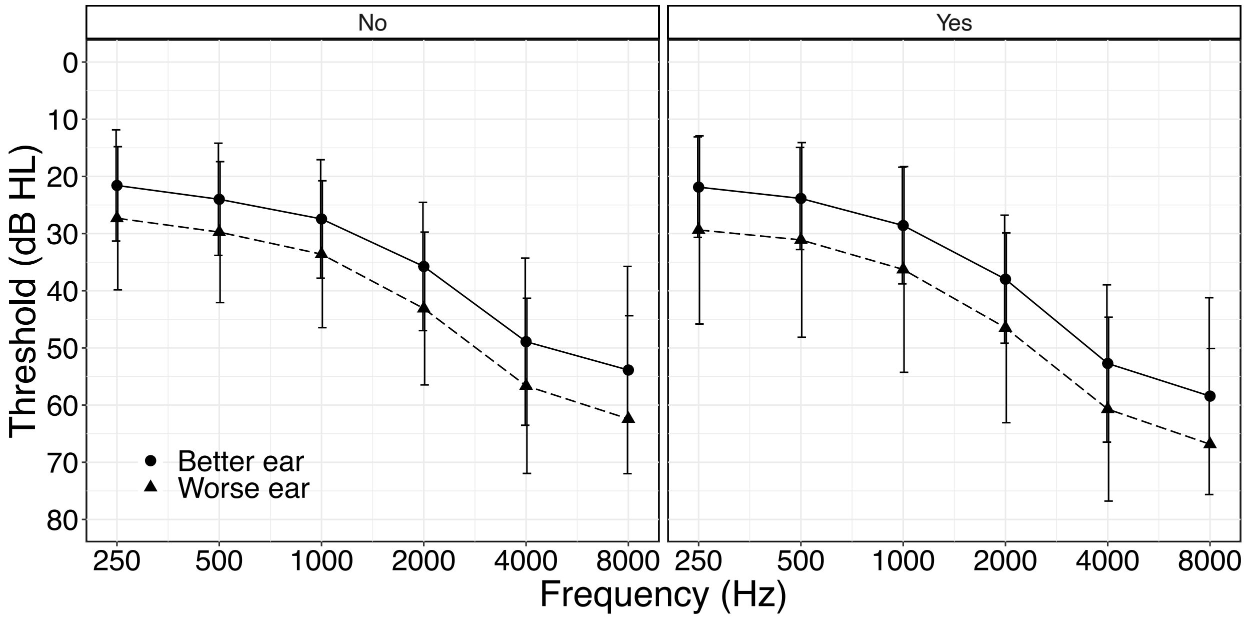
Data inclusion criteria

Participants were included in this analysis if they had a four-frequency pure-tone average in the better ear (PTA-BE) of greater than 25 dB HL, as they would be considered to have at least mild hearing loss according to the WHO categories of hearing loss [ref] and would be potential candidates for a hearing device. Of the 3312 participants who met the basic participation criteria listed in the previous section, 1396 participants met this hearing criterion. All 1396 participants who met the hearing criterion had complete data for the outcome measure of hearing aid purchase. Of these 1396 participants, 753 had no missing observations in 28 predictors, while 643 had between 1 and 16 missing observations in 28 predictors. Logistic regression required complete data across all variables included in the model, and therefore the analyses were limited to the 753 participants with complete data. Table 1 shows the characteristics of participants who did not and who did purchase hearing aids, while Figure 1 shows their audiometric thresholds.

Table 1. Characteristics of 753 participants, grouped by those who did not and those who did purchase hearing aids. Means and standard deviations are shown for continuous variables, while proportions of the level in parentheses are shown for categorical variables.

	Did not purchase	**Did purchase**
Age; years	68.7 (8.9)	71.4 (8.8)
Sex (female)	0.43	0.57
Four-frequency PTA for better ear; dB HL	34.8 (7.6)	36.6 (7.5)
HHIE-S total score	15.0 (9.4)	18.5 (9.6)
Self-reported hearing ability	6.4 (1.7)	6.1 (1.6)
Overall health	3.8 (0.9)	3.9 (0.9)
Overall quality of life	4.1 (0.8)	4.0 (0.9)
Education	3.3 (1.1)	3.3 (1.2)
Married (yes)	0.23	0.77
Accompanied to appointment (yes)	0.84	0.16
Help from neighbors	2.0 (0.9)	2.0 (0.9)
Help with problems	2.2 (0.8)	2.2 (0.8)
Concern from others	3.7 (1.2)	3.6 (1.3)
Frequency of loneliness	1.2 (0.5)	1.2 (0.6)
Subjective age	2.9 (1.0)	2.9 (0.9)
Age stigma	2.4 (0.8)	2.5 (0.8)
Hearing aid stigma	2.9 (1.1)	2.7 (1.1)
Know at least one person who...		
Has a suspected hearing loss	0.14	0.86
Has a known hearing loss	0.15	0.85
Discussed their hearing loss	0.52	0.48
Had a hearing test	0.35	0.65
Obtained hearing aids	0.29	0.71
Sometimes uses hearing aids	0.47	0.53
Regularly uses hearing aids	0.34	0.66
Had a very positive experience with HA	0.41	0.59
Had a somewhat positive experience with HA	0.47	0.53
Had a somewhat negative experience with HA	0.65	0.35
Had a very negative experience with HA	0.83	0.17

Figure 1. Mean pure-tone audiometric thresholds of 753 participants, grouped by those who did not and those who did purchase hearing aids. Standard deviation bars are shown.



Results

Logistic regression analysis

The final logistic regression model consisted of five predictors, of which Age and HHIE were strong predictors, and "know at least one person who has suspected hearing loss", hearing aid stigma, overall health, and "know at least one person who discussed hearing loss" were weak predictors (Table 2). An increase of each year in age led to 4% higher odds of purchasing hearing aids, while an increase of each point on the HHIE led to 5% higher odds of purchasing hearing aids. Compared to Age and HHIE, weak predictors had much wider confidence intervals, indicating high uncertainty about the effects of these variables. The model's accuracy was 65.9%, with 61.1% sensitivity and 67.1% specificity.

Table 2. Coefficients and the associated standard errors and *p*-values of a logistic regression model with five predictor variables and a binary outcome measure of hearing aid purchase, with ‘No’ coded 0 and ‘Yes’ coded 1. Odds ratios were calculated from the coefficients, with 95% confidence intervals. (Know suspected HL = know at least one person who has suspected hearing loss; Know discussed HL = know at least one person who discussed hearing loss)

	Estimate (SE)	*p*	Odds ratio (95% C.I.)
Intercept	-6.21 (1.07)		
Age	0.045 (0.011)	< 0.0001	1.046 (1.024, 1.069)
HHIE	0.052 (0.010)	< 0.0001	1.053 (1.032, 1.075)
Health	0.18 (0.11)	0.088	1.197 (0.976, 1.478)
Hearing aid stigma	-0.17 (0.09)	0.057	0.843 (0.707, 1.004)
Know suspected HL	0.83 (0.34)	0.017	2.282 (1.210, 4.702)
Know discussed HL	-0.32 (0.20)	0.103	0.725 (0.491, 1.066)

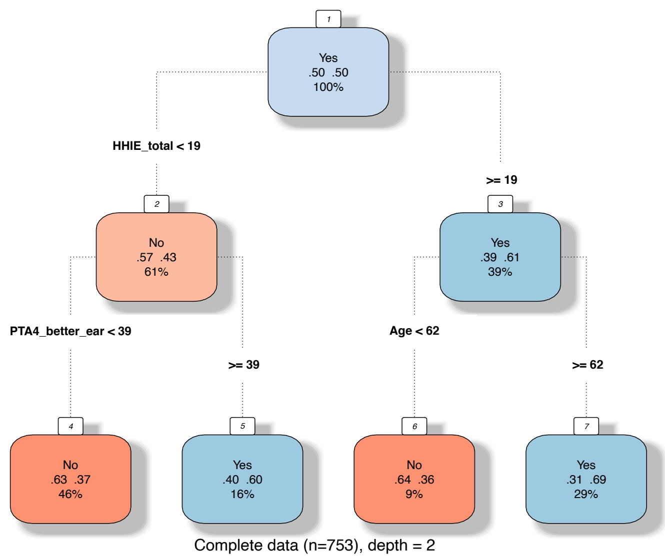
Classification tree analysis

Allowing a classification tree model to grow without restricting the maximum depth of the tree (setting the complexity parameter to 0 and setting all other parameters to default values), the resulting model consisted of 41 splits, and 78.2% accuracy, 92.0% sensitivity, and 74.8% specificity. However, such a model is difficult to interpret, and would likely not generalize to other samples.

Allowing a classification tree model to grow to a maximum depth of 2, the model showed that the predictor that best sorted cases according to hearing aid purchase outcome was HHIE;

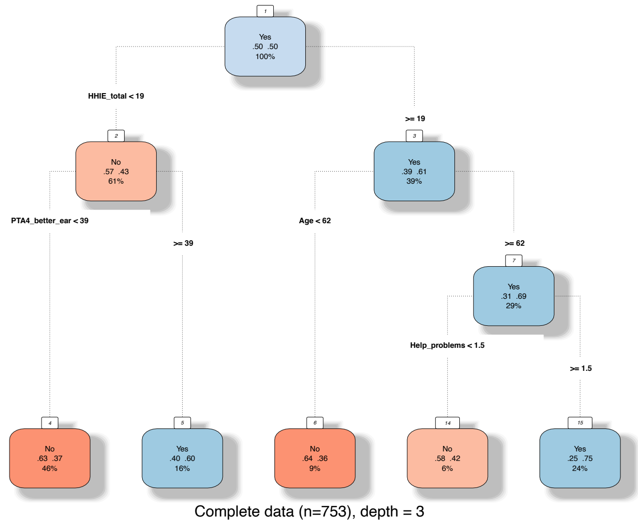
cases with HHIE < 19 were predicted to be ‘No’, while cases with HHIE ≥ 19 were predicted to be ‘Yes’ (Figure 2). Cases from each node were further sorted using the splitting criteria of PTA-BE < 39 dB HL and PTA-BE ≥ 39 dB HL, and Age < 62 and Age ≥ 62, such that a participant who scored less than 19 on the HHIE but had a PTA-BE of 39 dB HL or greater would be predicted to purchase hearing aids, for instance. While this model showed that HHIE, Age and PTA were the predictors that best sorted cases into the two outcomes, a model with just these variables did not sort cases very well, as roughly a third of the cases in each terminal node did not belong to the class indicated by that node. The model’s overall accuracy was 67.2%, with 59.1% sensitivity and 69.2% specificity.

Figure 2. Classification tree model with a depth of 2. Each node is labelled with the node's majority class, with 'No' cases sorted to the left and 'Yes' cases sorted to the right by convention. The proportion of 'No' and 'Yes' cases are shown as a pair of numbers in the middle of each node. The number at the bottom of each node indicates the percentage of all cases in the sample that were sorted into that node, taking into account case weights (e.g., one 'Yes' case would be considered to be the equivalent of about four 'No' cases).



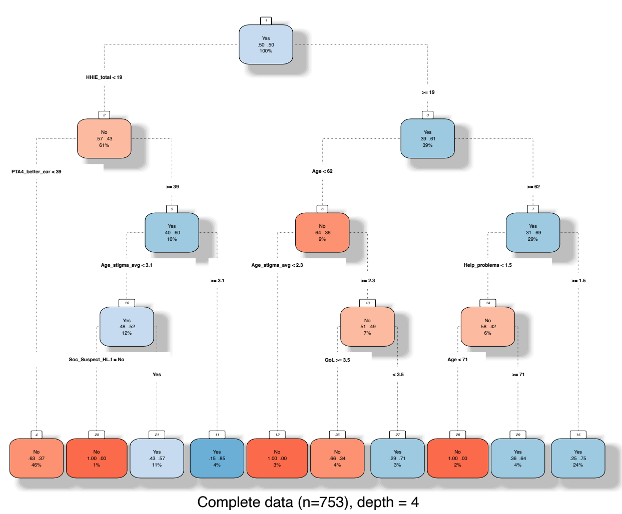
Allowing the model to grow to a maximum depth of 3, an additional social predictor was included ("how many people can you count on to help with serious problems"); this model had slightly better overall accuracy compared to a model with a depth of 2, with 71.5% accuracy, 54.4% sensitivity, and 75.6% specificity.

Figure 3. Classification tree model with a depth of 3. [need to re-draw to improve readability]



Allowing the model to grow to a maximum depth of 4, four additional social predictors were included, and a second cut-off value of Age, but model performance did not change much compared to the model with a depth of 3, with 71.5% accuracy, 63.1% sensitivity, and 73.5% specificity.

Figure 4. Classification tree model with a depth of 4. [need to re-draw to improve readability]



Summary of results

A logistic regression analysis showed that each unit increase of HHIE and Age increased the odds of purchasing hearing aids by 5% and 4%, respectively, while social predictors had relatively weak effects. A classification tree analysis showed similar findings, indicating that HHIE was the most important predictor of hearing aid purchase, followed by Age and PTA-BE, while social predictors had smaller effects. Importantly, the classification tree model indicated that a HHIE score of 19 was the first boundary where a higher score meant that a participant was more likely to purchase hearing aids, especially if they had a PTA-BE of 39 dB HL or greater, or were aged 62 or older.

References

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Terry Therneau and Beth Atkinson (2019). rpart: Recursive Partitioning and Regression Trees. R package version 4.1-15. https://CRAN.R-project.org/package=rpart

Stephen Milborrow (2020). rpart.plot: Plot 'rpart' Models: An Enhanced Version of 'plot.rpart'. R package version 3.0.9. https://CRAN.R-project.org/package=rpart.plot