PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Internet-based cognitive-behavioral therapy for tinnitus: Secondary analysis to examine predictors of outcomes
AUTHORS	Rodrigo, Hansapani; Beukes, Eldré; Andersson, Gerhard; Manchaiah, Vinaya

VERSION 1 – REVIEW

REVIEWER	Djalilian, Hamid R. University of California Irvine
REVIEW RETURNED	10-Feb-2021

GENERAL COMMENTS	The authors have written a nice study identifying predictors of significant reduction in tinnitus severity for patients undergoing internet-based cognitive behavioral therapy. There are some issues which I have outlined below. Minor comments: -The abstract could benefit from more detailed and numeric results (e.g., odds ratios) -The conclusions mention that advanced AI and machine learning techniques could possibly find predictive factors in the future via complex computations, without further discussing their potential benefits in the paper. I would caution against this statement as more complex methodologies are not always the answer; if there are no straightforward predictors with the current statistical methods, maybe there are no predictors after all (or we need larger cohorts with more comprehensive data in the future, instead of more complex computational analyses). -There are some statements that would benefit from citation, for instance “Most individuals with tinnitus are not much bothered by the sounds, but a proportion (2/10) find tinnitus much distressing and need help to reduce the negative effects of their tinnitus.” in the introduction. -Could you clarify if all the variables were analyzed in univariate analysis, and then all the statistically significant variables on univariate analysis (plus baseline tinnitus severity) were analyzed all-together (to control for their confounding effects) in the multivariate analysis? A table that includes both univariate and multivariate analysis would further simplify this. -Discussion second paragraph: The authors hypothesized that patients that work less due to tinnitus were at lower odds of having successful outcomes possibly because they are likely to have

	higher tinnitus severity according to the literature. However, wasn't tinnitus severity accounted for on the multivariate model and this variable (working less) remained significant? In paragraph 5 of the discussion, the authors similarly hypothesized that patients with depression or anxiety were more likely to have experienced improved outcomes because of their historically higher baseline tinnitus severity according to the literature. But again, was this not controlled for in the multivariate analysis (making depression/insomnia influential independently of baseline tinnitus severity)? Major comments: -The abstract conclusion reports that no strong predictors were identified other than baseline tinnitus severity (which was previously reported by Reference #11, as mentioned by the authors in the introduction). It would be beneficial to clearly state what conclusions this paper had that were different from previous literature. -Methods under data analysis subheading: "post-TFI scores of 38 subjects were missing and were imputed using means of their corresponding trials." Why were these values imputed instead of excluding these patients? -The overall statistical methodologies seem a bit complicated with lengthy results. I would have found a standard univariate/multivariate regression modeling with all the patients included (instead of 80% training and 20% testing) just as plausible for finding independent predictors within the scope of the study. -The last line of methods: "All tests were two sided and threshold at 10% level of significance due to the exploratory nature of the study to account for the medium sample size and control for both Type I and type II errors effectively" Does the exploratory nature really justify a p=0.01 for significance? With the large number of variables that were tested in univariate analysis, what is the benefits of p=0.01 instead of methods such as Bonferroni correction which can lead to potentially smaller (not larger) thresholds for significance? -I think that the user's level of engagement/compliance (and motivation) are among the most important factors for CBT success. The authors mention these in the limitations, but if these data are in fact not available, then they should be adequately explored and discussed based on the available literature on tinnitus or other chronic conditions treated with internet-based CBT.

REVIEWER	Prabhu, Prashanth All India Institute of Speech and Hearing
REVIEW RETURNED	16-Feb-2021

GENERAL COMMENTS	The authors have attempted to identify the predictors of ICBT intervention in individuals with tinnitus. The authors have identified that baseline tinnitus severity is an important predictor of the outcome. The article is well-written with a comprehensive analysis of the data. However, there are some concerns listed below that need to be addressed:

	• Page 11 – Line 17 – The authors should specify the inferential statistical test used to compare pre and post-intervention TFI scores and provide the test statistic value. • Reference to be added for this statement – “Most individuals with tinnitus are not much bothered by the sounds, but a proportion (2/10) find tinnitus much distressing and need help to reduce the negative effects of their tinnitus” • Add reference for “In addition, the improvements noted from ICBT have been maintained for 1-year post-intervention”. • Page 6, Line 52 – ‘from’ to be added: “……collected from…” • There are several typographical and spacing errors throughout the manuscript which has to be corrected. For E.g. ‘examine’, ‘study’ ‘clinical’ etc.

REVIEWER	Haile, Sarah University of Zurich, Epidemiology, Biostatistics and Prevention Institute
REVIEW RETURNED	22-Mar-2021

GENERAL COMMENTS	This was an interesting manuscript that feels a gap in the medical literature. I will restrict my comments to statistical aspects of the manuscript. Were all subjects treated with CBT? In the abstract, the paragraph on results appears to focus on statistics first and the story second. The abstract would be overall stronger if that paragraph focused first on the meaning of the results, and 2nd on reporting the actual statistical results. Table 2 makes me also question some of the causal pathways here. For example, there is a strong negative association between tinnitus severity and disability allowance. The formulation of the model seems to have disability as a predictor for disability allowance, but wouldn't it be more likely that people have disability allowance because they have severe tinnitus?? Could you expand on this in the manuscript? In the methods section, it's noted that mean imputation was used to account for missing values. Unfortunately, any kind of "Single imputation of missing values usually causes standard errors to be too small, since it fails to account for the fact that we are uncertain about the missing values." (Sterne et al 2009 https://www.bmj.com/content/338/bmj.b2393, also noted by other authors e.g. White Royston and Wood 2010 https://onlinelibrary.wiley.com/doi/full/10.1002/sim.4067). Please use another method for multiple imputation. That said, unless MI is combined with e.g. cross-validation (Wahl et al 2016, Schomaker and Heumann 2018), it is difficult to combine MI and model comparison. For example, while it may be possible to compute AIC after MI, these AIC results cannot be pooled. There's a nice overview of this topic by Marshall et al 2009 (https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-9-57). Please reconsider the strategies used here for both missing data and model comparison. Regarding the multivariable models: - The term "multivariable" not "multivariate" should be used (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3518362/)

	- How did you decide which variables were included in the multivariable models? - Have you considered proper scores (like the Brier score) to compare models? In the results, it would be more clinically meaningful to report coefficients and their 95% confidence intervals, rather than just p-values. There are however a lot of results summarized in these paragraphs. It would be clearer, I think, to focus on the story the results tell, perhaps giving some key numeric results, and refer readers to the table for full details. I see that R-square results have been reported. These are not noted in the methods section, but I would recommend removing them entirely. See e.g. http://www.stat.cmu.edu/~cshalizi/mreg/15/lectures/10/lecture-10.pdf and https://data.library.virginia.edu/is-r-squared-useless/ Since the analysis combines data from 3 studies, it might be appropriate to stratify table 1 by center. Table 2 should show CI not SE. Finally, the 13-point cut off that serves as the basis for the logistic regression model seems somewhat random. Therefore, I question the need for the logistic model.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

COMMENT: The authors have written a nice study identifying predictors of significant reduction in tinnitus severity for patients undergoing internet-based cognitive behavioral therapy. There are some issues which I have outlined below.

Response: Thank you for your positive feedback.

COMMENT: The abstract could benefit from more detailed and numeric results (e.g., odds ratios)

Response: We have now added the details related to the odds ratio, 95% Confidence intervals and p-values into the Abstract (Page 2, Line 40) as follows:

Abstract: ….As per the Chi-square univariate analysis, participants with a master's degree or above had the highest odds of having a larger severity change score (OR=3.47; 95% C.I: 1.32, 12.51), compared to the participants who had education only up to high school or less. Additionally, the baseline tinnitus severity was found to be a significant variable (p<0.001, OR :1.05, 95%, C.I.: 1.03, 1.07) contributing to a successful outcome with the intervention. Both linear and logistic regression models have identified education level (linear: p=0.01, logistic: p=0.00), and baseline tinnitus severity (linear: p<0.001, logistic: p<0.001), to be significant predictor variables contributing to reduction in tinnitus severity post-ICBT intervention. As per linear regression model, participants who had received disability allowance had shown 25.30-point (95% C.I: -46.35,-4.24) lower TFI reduction compared to those who didn't have to work less due to tinnitus after adjusting for baseline tinnitus severity and their education level.

COMMENT: The conclusions mention that advanced AI and machine learning techniques could

possibly find predictive factors in the future via complex computations, without further discussing their potential benefits in the paper. I would caution against this statement as more complex methodologies are not always the answer; if there are no straightforward predictors with the current statistical methods, maybe there are no predictors after all (or we need larger cohorts with more comprehensive data in the future, instead of more complex computational analyses).

Response: Thank you for your important comment. We agree with your comment. However, here, we are not making any conclusions that the AI and machine learning models are better than linear/logistic regression. Rather, we want to highlight the fact there can be non-linear dependencies between predictors and the responses which might have not been detected by linear regression models and hence in future we wanted to explore the neural network models to see whether there exist any such predictors which yield better predictive model.

Accordingly, we have changed the conclusion section of the Abstract as follows. Page 3, Line 53

Conclusions: ……As no strong predictors were identified other than the baseline tinnitus severity and the education level, future studies should consider including a heterogeneous group of participants as well as other predictor variables that might have not included in the current study.

However, we have added the following statement to the “Study Limitations and Future Research” section as follows. Page 18, Lines 408

Study Limitations and Future Research: …..Future studies may benefit from examining more relevant predictor variables and also using artificial intelligence and machine learning techniques to examine the non-linear relationship between the variables in predicting the ICBT outcomes.

COMMENT: There are some statements that would benefit from citation, for instance “Most individuals with tinnitus are not much bothered by the sounds, but a proportion (2/10) find tinnitus much distressing and need help to reduce the negative effects of their tinnitus.” in the introduction.

Response:
We have altered this to include the appropriate citation as follows: Page 4, Lines 74

The National Study of Hearing in England found that of the general population surveyed (N = 48, 313), 10.1% reported any tinnitus, 2.8% reported moderately annoying tinnitus, 1.6% reported severely annoying tinnitus, and 0.5% were unable to lead a normal life due to the severity of the tinnitus.1

(Reference: Davis, A., & Refaie, A. E. (2020). The epidemiology of tinnitus. In R. Tyler (Ed.), The Handbook of Tinnitus (pp. 1–23). Singular.)

COMMENT: Could you clarify if all the variables were analyzed in univariate analysis, and then all the statistically significant variables on univariate analysis (plus baseline tinnitus severity) were analyzed all-together (to control for their confounding effects) in the multivariate analysis? A table that includes both univariate and multivariate analysis would further simplify this.

Response: All the variables were analyzed using the univariate analysis based on the Chi Square/ Fishers Exact tests (see Tables 3.2-3.5 in Appendix 3). In all multivariable analyses, we started off with the full model, including all the predictor variables and used backward elimination based on Akaike Information criterion to select the final model (see Tables 2-4).

We have clarified this as follows.

The univariate analysis was performed using Chi-square or Fisher's exact test to examine the effect of single variables on the ICBT outcome using all the variables (Page 10, Line: 228).

During multivariable analysis, we started off with the full model, including all the predictor variables and used backward elimination based on Akaike Information criterion to select the final model. Page 11, Line 238

COMMENT: Discussion second paragraph: The authors hypothesized that patients that work less due to tinnitus were at lower odds of having successful outcomes possibly because they are likely to have higher tinnitus severity according to the literature. However, wasn't tinnitus severity accounted for on the multivariate model and this variable (working less) remained significant? In paragraph 5 of the discussion, the authors similarly hypothesized that patients with depression or anxiety were more likely to have experienced improved outcomes because of their historically higher baseline tinnitus severity according to the literature. But again, was this not controlled for in the multivariate analysis (making depression/insomnia influential independently of baseline tinnitus severity)?

Response: Thank you for pointing this out, they have been altered as follows:

Paragraph 2 in the Discussion (Page 16 Line 348)

Additionally, those reported to be working less due to tinnitus were at a lower odds of having a successful outcome. This finding needs further exploration in future studies. Working, may for instance provide some distraction from tinnitus as supported by reports during the 2020 COVID-19 pandemic that tinnitus was more bothersome for some individuals as they did not have the distractions from commuting and sounds at work.38 Closely monitoring the effects of tinnitus is important to ensure that tinnitus can be managed so that individuals are still able to work effectively.

Paragraph 5 in the Discussion (Page 17 Line 373)

Regarding studying the clinical factors, those with higher levels of depression were found to have higher reduction in the TFI score. However, the participants with insomnia showed lower odds of success. Interestingly, other clinical factors including anxiety, hyperacusis, hearing disability as well as cognitive functioning were not significant predictors of ICBT in the current study. Further studies and models are required to verify these results.

COMMENT: The abstract conclusion reports that no strong predictors were identified other than baseline tinnitus severity (which was previously reported by Reference #11, as mentioned by the authors in the introduction). It would be beneficial to clearly state what conclusions this paper had that were different from previous literature.

Response:
Thank you for pointing this out. We have revised the abstract to be clearer regarding our findings as follows:

Results: Of the 228 subjects who were included in the study, 65% had a successful outcome of the treatment. As per the Chi-square univariate analysis, participants with a master's degree or above had the highest odds of having a larger severity change score (OR=3.47; 95% C.I: 1.32, 12.51), compared to the participants who had education only up to high school or less. Additionally, the baseline tinnitus severity was found to be a significant variable (p<0.001, OR :1.05, 95%, C.I.: 1.03, 1.07) contributing to a successful outcome with the intervention. Both linear and logistic regression

models have identified education level (linear: p=0.01, logistic: p=0.00), and baseline tinnitus severity (linear: p<0.001, logistic: p<0.001), to be significant predictor variables contributing to reduction in tinnitus severity post-ICBT intervention. As per linear regression model, participants who had received disability allowance had shown 25.30-point (95% C.I: -46.35,-4.24) lower TFI reduction compared to those who didn't have to work less due to tinnitus after adjusting for baseline tinnitus severity and their education level. Page 2, Line 39

Conclusions: Predictors of intervention outcome can be used as a means of triaging patients to the most suited form of treatment to achieve optimal outcomes and to make healthcare savings. As no strong predictors were identified other than the baseline tinnitus severity and the education level, future studies should consider including a heterogeneous group of participants as well as other predictor variables that might have not included in the current study. Page 3, Line 51

COMMENT: Methods under data analysis subheading: "post-TFI scores of 38 subjects were missing and were imputed using means of their corresponding trials." Why were these values imputed instead of excluding these patients?

Response: Thank you for your comment. These participants had details for all predictor variables except their post-TFI score which were missing at random. If we were to exclude them from our analysis, it will reduce the power of the study (Loder, 2013). Nevertheless, we have analyzed the same data with (Predictive Mean Matching) and without any data imputation. Similar conclusions were observed under both conditions. Therefore, we have decided to present the results after the data imputation.

This has been clarified as follows in the manuscript Page 10, Line 222:

With the intention of preserving the power of the analysis, we have retained those subjects in the analysis after applying the predictive mean matching (PMM) data imputation.34 Data imputation with PMM has been identified to be less vulnerable to model misspecification as, there is no need to define an explicit model for the distribution of the missing values35 with that.

References:
34. Lodder, P. (2013). To impute or not impute: That's the question. Advising on research methods: Selected topics. Huizen: Johannes van Kessel Publishing.

35. 35. Little, R. J. A., and D. B. Rubin. 2002. Statistical Analysis with Missing Data. New York: John Wiley & Sons.)

COMMENT: The overall statistical methodologies seem a bit complicated with lengthy results. I would have found a standard univariate/multivariate regression modeling with all the patients included (instead of 80% training and 20% testing) just as plausible for finding independent predictors within the scope of the study.

Response: Thank you for these comments. The purpose of allocating 80% data for training and 20% for testing is to make a fair comparison among all the predictive models that we build based on this data. This is usually a common strategy which has been used in the literature for predictive model comparisons. By preserving this same proportion, we expect that it would help us to identify the best predictive model which is one of the intended goals of the extended analyses following to this. Nevertheless, we have trained the models using full data, and resulted with similar conclusions.

When presenting the results, we have now used additional tables (Table 4) for a better summarization

of the information and had limited our discussion to the most important results within the manuscript body. Pages 13 and 29

We have added this to the manuscript as follows: Pages 11, Line 232

the full data set was divided into the training (80%, n = 183) and testing (20%, n = 45) to make a fair comparison among all the predictive models.

COMMENT: The last line of methods: “All tests were two sided and threshold at 10% level of significance due to the exploratory nature of the study to account for the medium sample size and control for both Type I and type II errors effectively” Does the exploratory nature really justify a p=0.01 for significance? With the large number of variables that were tested in univariate analysis, what is the benefits of p=0.01 instead of methods such as Bonferroni correction which can lead to potentially smaller (not larger) thresholds for significance?

Response: Thank you for valuable comment. We have taken the multiple comparison issue into consideration and in the revised analysis, we have used 5% level of significance. Page 11, Line 250

COMMENT: I think that the user’s level of engagement/compliance (and motivation) are among the most important factors for CBT success. The authors mention these in the limitations, but if these data are in fact not available, then they should be adequately explored and discussed based on the available literature on tinnitus or other chronic conditions treated with internet-based CBT.

Response:
Thank you for this valid comment. We have included this as follows: Page 18, Line 398

These factors were not investigated for this study. As they have been found to contribute to outcomes45, they should be included in future studies.

Reviewer 2

COMMENT: The authors have attempted to identify the predictors of ICBT intervention in individuals with tinnitus. The authors have identified that baseline tinnitus severity is an important predictor of the outcome. The article is well-written with a comprehensive analysis of the data.

Response: Thank you for your positive feedback.

COMMENT: Page 11 – Line 17 – The authors should specify the inferential statistical test used to compare pre and post-intervention TFI scores and provide the test statistic value.

Response: We have included the details of the inferential statistical tests that was used to compare the pre and post intervention TFI scores in the revised manuscript (Page 12, Line 261) as follows:

Figure 1 presents the pre-and post-intervention tinnitus severity (TFI) score variation, indicating statistically significant differences between these scores ($p < 0.001$) with the paired t-test.

COMMENT: Reference to be added for this statement – “Most individuals with tinnitus are not much bothered by the sounds, but a proportion (2/10) find tinnitus much distressing and need help to reduce the negative effects of their tinnitus”

Response: We have altered this to include the appropriate citation as follows:

The National Study of Hearing in England found that of the general population surveyed (N = 48, 313), 10.1% reported any tinnitus, 2.8% reported moderately annoying tinnitus, 1.6% reported severely annoying tinnitus, and 0.5% were unable to lead a normal life due to the severity of the tinnitus.1 Page 4, Line 74

Reference: Davis, A., & Refaie, A. E. (2020). The epidemiology of tinnitus. In R. Tyler (Ed.), The Handbook of Tinnitus (pp. 1–23). Singular.

COMMENT: Add reference for “In addition, the improvements noted from ICBT have been maintained for 1-year post-intervention.11”

Response: We have added the corresponding reference to the Reference list in the revised manuscript. Page 5, Line 97

Beukes, E. W., Allen, P. M., Baguley, D. M., Manchaiah, V., & Andersson, G. (2018). Long-Term Efficacy of Audiologist-Guided Internet-Based Cognitive Behavior Therapy for Tinnitus. American journal of audiology, 27(3S), 431–447. https://doi.org/10.1044/2018_AJA-IMIA3-18-0004

COMMENT: Page 6, Line 52 – ‘from’ to be added: “……collected from…”

Response: We have corrected this in the revised manuscript. Page 6, Line 127

COMMENT: There are several typographical and spacing errors throughout the manuscript which has to be corrected. For E.g. ‘examine’, ‘study’ ‘clinical’ etc.

Response: We have carefully reviewed the manuscript for grammatical errors and have made revisions to correct them including the above.

Reviewer 3

COMMENT: This was an interesting manuscript that feels a gap in the medical literature. I will restrict my comments to statistical aspects of the manuscript.

Response: Thank you for your positive feedback.

COMMENT: Were all subjects treated with CBT?

Response: Yes, CBT was delivered via the internet (ICBT)

This has been clarified on Page 9 Line 186:

As a secondary analysis, no patients were involved in these studies. The data originates for individuals with tinnitus who had previously undertaken a tinnitus ICBT intervention and received CBT delivered via the internet.

COMMENT: In the abstract, the paragraph on results appears to focus on statistics first and the story second. The abstract would be overall stronger if that paragraph focused first on the meaning of the results, and 2nd on reporting the actual statistical results.

Response: We have now modified the Abstract highlighting the meaning of the results, with statistics in brackets as to focus on the study findings as follows: Page 2 Line 39

Results: Of the 228 subjects who were included in the study, 65% had a successful ICBT outcome. As per the Chi-square univariate analysis, participants with a master's degree or above had the highest odds of having a larger severity change score (OR=3.47; 95% C.I: 1.32, 12.51), compared to the participants who had education only up to high school or less. Additionally, the baseline tinnitus severity was found to be a significant variable (p<0.001, OR :1.05, 95%, C.I.: 1.03, 1.07) contributing to a successful outcome with the intervention. Both linear and logistic regression models have identified education level (linear: p=0.01, logistic: p=0.00), and baseline tinnitus severity (linear: p<0.001, logistic: p<0.001), to be significant predictor variables contributing to reduction in tinnitus severity post-ICBT intervention. As per linear regression model, participants who had received disability allowance had shown 25.30-point (95% C.I: -46.35,-4.24) lower TFI reduction compared to those who didn't have to work less due to tinnitus after adjusting for baseline tinnitus severity and their education level.

COMMENT: Table 2 makes me also question some of the causal pathways here. For example, there is a strong negative association between tinnitus severity and disability allowance. The formulation of the model seems to have disability as a predictor for disability allowance, but wouldn't it be more likely that people have disability allowance because they have severe tinnitus?? Could you expand on this in the manuscript?

Response: Thank you for pointing this out, yes it was those on disability allowance due to severe tinnitus. We have clarified this as follows on Page 14 Line 300:

Those who received disability allowance due to having severe tinnitus and being unable to work, had shown 25.30-point (95% C.I: -46.35,-4.24) lower TFI reduction compared to those who did not have to work less due to tinnitus.

COMMENT: In the methods section, it's noted that mean imputation was used to account for missing values. Unfortunately, any kind of "Single imputation of missing values usually causes standard errors to be too small, since it fails to account for the fact that we are uncertain about the missing values." Please use another method for multiple imputation.

That said, unless MI is combined with e.g. cross-validation (Wahl et al 2016, Schomaker and Heumann 2018), it is difficult to combine MI and model comparison. For example, while it may be possible to compute AIC after MI, these AIC results cannot be pooled. There's a nice overview of this topic by Marshall et al 2009 (https://nam10.safelinks.protection.outlook.com/?url=https%3A%2F%2Fbmcmedresmethodol.biomedcentral.com%2Farticles%2F10.1186%2F1471-2288-9-57&data=04%7C01%7Chansapani.rodrigo%40utrgv.edu%7C92c32e153dbd4e4448a008d8fdb1bdfc%7C990436a687df491c91249afa91f88827%7C0%7C1%7C637538287150103897%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C1000&sdata=xUo53OLY%2Bt7A7xYXqdxCU5R1OIj8Q3VruNNdzRnrZOo%3D&reserved=0). Please reconsider the strategies used here for both missing data and model comparison.

Response: We have now adopted the Predictive Mean Matching data imputation and revised our analysis, accordingly. Moreover, the results with predictive mean matching data imputation and without any data imputation has been checked and had led to similar conclusions. Use of data imputation over discarding participants who had all their predictor variables except their post TFI score, is important as to preserve the power of the study. The details of these are included on Page

10, Lines 221 as follows.

There were 98 subjects who had all their predictive variables except their post TFI scores. With the intention of preserving the power of the analysis, we have retained those subjects in the analysis after applying the predictive mean matching (PMM) data imputation.

Although, we have not included any model comparison details here in this paper, for future analyses, we had trained the model using a training data set and had make predictions using the testing data. This is usually a common strategy used in MI analyses.

COMMENT: Regarding the multivariable models: The term "multivariable" not "multivariate" should be used.

Response: Thank you for your comment. We have now used the term “multivariable” throughout the manuscript.

COMMENT: How did you decide which variables were included in the multivariable models?

Response: The 32 predictor variables were selected based on the previous literature associated with Tinnitus research. We have used the backward elimination based on the Akaike Information criterion when selecting the best multivariable model.

We have clarified this on Page 9, Lines 198 as follows:

Predictor variables were selected, based on clinical reasoning and findings from previous studies by Beukes et al.11 (see Appendix 2 for details). Thirty-two variables were selected as potential predictor (independent) variables and included demographic, tinnitus and hearing-related variables, tinnitus treatment related variables, clinical factors as follows:

COMMENT: Have you considered proper scores (like the Brier score) to compare models?

Response: Thank you for the comment. In the current study, we are not comparing the linear or logistic regression models. However, for our extended analyses in future, we will make sure to use both area under the receiver operating characteristic curve (AUC) and the Brier scores, as you have suggested. We have added this to the limitation section as follows: Page 18, Lines 410

Brier scores should also be used to compare models.

COMMENT: In the results, it would be more clinically meaningful to report coefficients and their 95% confidence intervals, rather than just p-values. There are however a lot of results summarized in these paragraphs. It would be clearer, I think, to focus on the story the results tell, perhaps giving some key numeric results, and refer readers to the table for full details.

Response: We have now added the 95% confidence intervals for the model parameters and corresponding p-values into the results section. Additionally, we have included a separate table (Table 4, Page 29) summarizing all the non-significant variables with their p-values. Pages 14, Line 312.

We have revised the results section to tell the story in a clearer manner.

COMMENT: I see that R-square results have been reported. These are not noted in the methods section, but I would recommend removing them entirely. See e.g.

https://nam10.safelinks.protection.outlook.com/?url=http:%2F%2Fwww.stat.cmu.edu%2F~cshalizi%2Fmreg%2F15%2Flectures%2F10%2Flecture-10.pdf&data=04%7C01%7Chansapani.rodrigo%40utrgv.edu%7C92c32e153dbd4e4448a008d8fdb1bdfc%7C990436a687df491c91249afa91f88827%7C0%7C1%7C637538287150103897%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C1000&sdata=SQwNem5i8WT6Cg4i3dzW6I%2F8kJq58cagAHaS4cu4eT0%3D&reserved=0 and https://nam10.safelinks.protection.outlook.com/?url=https%3A%2F%2Fdata.library.virginia.edu%2Fis-r-squared-useless%2F&data=04%7C01%7Chansapani.rodrigo%40utrgv.edu%7C92c32e153dbd4e4448a008d8fdb1bdfc%7C990436a687df491c91249afa91f88827%7C0%7C1%7C637538287150113889%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C1000&sdata=fvGLA89sbj0qTOMYq52JXJfa6HeFFPnUL2M73f%2BcnLY%3D&reserved=0

Response: Thank you for your comment. Yes, we agree that neither R squared, nor Adj. R squared can be used to compare the models or the predictive power of the models. Nevertheless, it is usually used as a statistical measure of fit that indicates how much variation of the outcome is explained by the predictor variable(s) in the linear regression model. Therefore, based on the revised analysis with predictive mean matching data imputation, we still mention the R squared and Adj. R squared values for transparency as has been done by other studies. Additionally, we have also reported the mean squared error as it is a better measure of prediction accuracy using training data. Please see Page 11, Line 240.

R squared and Adj. R squared values has been reported as they are statistical measures of fit that indicates how much variation of the outcome is explained by the predictor variable(s) in a linear regression model.37 We also reported the mean squared error as it is a better measure of prediction accuracy.

Reference:
37. Montgomery, D. C., Peck, E. A. & Vinning, G. G. (2012), Introduction to Linear Regression Analysis, Wiley Series in Probability and Statistics)

COMMENT: Since the analysis combines data from 3 studies, it might be appropriate to stratify table 1 by center.

Response: Although 3 studies were combined, the same protocol was followed for each study and the same intervention received. The study center was the same for all studies ie and internet treatment. Hence stratification is not required for purposes of this study.

We have added the following for clarity Page 9 :Line 188: As the same protocol was followed for all study participants and the all received the same intervention, merging this data was possible.

COMMENT: Table 2 should show CI not SE.

Response: We have revised the table 2 and have included the CI. Page 28

COMMENT: Finally, the 13-point cut off that serves as the basis for the logistic regression model seems somewhat random. Therefore, I question the need for the logistic model.

Response: The 13 points cut off was taken as it was identified as a clinically significant change in the score during the TFI validation (Meikle et al., 2012). This is clarified in the methods section as follows: Page 9, Line 193

The 13-point change in TFI scores, identified as a clinically meaningful (or significant) change by the original authors26 was used to define a clinically significant intervention outcome.

(Reference: Meikle, M. B., Henry, J. A., Griest, S. E., Stewart, B. J., Abrams, H. B., McArdle, R., Myers, P. J., Newman, C. W., Sandridge, S., Turk, D. C., Folmer, R. L., Frederick, E. J., House, J. W., Jacobson, G. P., Kinney, S. E., Martin, W. H., Nagler, S. M., Reich, G. E., Searchfield, G., … Vernon, J. A. (2012). The Tinnitus Functional Index. Ear & Hearing, 33(2), 153–176. https://doi.org/10.1097/aud.0b013e31822f67c0.)

Regarding the selection of the logistic model, while a linear regression model can be used to identify the factors affecting a significant TFI score change, the logistic model can be used to evaluate the factors which specifically effects outcomes and was thus selected.

VERSION 2 – REVIEW

REVIEWER	Prabhu, Prashanth All India Institute of Speech and Hearing
REVIEW RETURNED	17-Jun-2021

GENERAL COMMENTS	The authors have incorporated all the corrections that were suggested.

REVIEWER	Haile, Sarah University of Zurich, Epidemiology, Biostatistics and Prevention Institute
REVIEW RETURNED	14-Jun-2021

GENERAL COMMENTS	Thank you for this revision. I notice that other reviewers have commented on the point about AI and machine learning techniques. While such techniques may be of use here, it should be noted that there are other approaches to account for non-linear relationships here, for example, including polynomial or spline terms. It might also be of interest here to use generalized additive models (GAMs) or classification trees, which I would not necessarily consider to be AI or ML, although they are certainly not one of the traditional statistical techniques. The sentences starting with "future studies" and "analyses should be extended" are currently broken up by the point about Brier scores. Would it be possible to combine these two sentences, pointing out a) specific statistical approaches, b) specific other predictors, and c) which relationships are perhaps non-linear? From the figures and tables, it's unclear which if any relationships are maybe non-linear. Providing this information, even in the supplementary material, would be helpful for future studies. Will the data be made available to other researchers?

VERSION 2 – AUTHOR RESPONSE

Reviewer 2

COMMENT: The authors have incorporated all the corrections that were suggested.

Response: Thank you for your positive feedback.

Reviewer 3

COMMENT: I notice that other reviewers have commented on the point about AI and machine learning techniques. While such techniques may be of use here, it should be noted that there are other approaches to account for non-linear relationships here, for example, including polynomial or spline terms. It might also be of interest here to use generalized additive models (GAMs) or classification trees, which I would not necessarily consider to be AI or ML, although they are certainly not one of the traditional statistical techniques. The sentences starting with "future studies" and "analyses should be extended" are currently broken up by the point about Brier scores. Would it be possible to combine these two sentences, pointing out a) specific statistical approaches, b) specific other predictors, and c) which relationships are perhaps non-linear? From the figures and tables, it's unclear which if any relationships are maybe non-linear. Providing this information, even in the supplementary material, would be helpful for future studies.

Response: Thank you for your helpful feedback. We have provided additional details about non-linear relationships as well as clarified the text in this section as suggested.

COMMENT: Will the data be made available to other researchers?

Response: Yes, the de-identified data are available upon reasonable request. We have added a data availability statement for this.