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)	The reliability of parental assessment of auditory skills in young children: A cross-sectional study in Italian language
AUTHORS	Orzan, Eva; Battelino, Saba; Ciciriello, Elena; Bonifacio, Serena; Pellizzoni, Sandra; SAKSIDA, AMANDA

VERSION 1 – REVIEW

REVIEWER	Berg, Abbey Pace University, Communication Sciences & Disorders
REVIEW RETURNED	18-Aug-2020

GENERAL COMMENTS	This article confirms what other authors have found regarding questionnaires to elicit parental information and auditory. Consider this reference: Fabus, R., Berg, A.L., Serpanos, Y.C., & Hidecker. (2017). Effectiveness of parental questionnaires in the assessment of speech-language and auditory function in children. Folia Phoniatrica et Logopaedia, 69, 261-270. doi: 10.1159/000488054.

REVIEWER	Dillon, Harvey The University of Manchester, Audiology
REVIEW RETURNED	31-Aug-2020

GENERAL COMMENTS	General comments This paper examines the extent to which systematically obtained parental comments can be relied on to detect the presence of hearing loss in children in whom, for various reasons, the loss was not detected during new-born screening. It is an appropriate topic for investigation, and the study has a reasonable sample size, but the paper is lacking in clarity in several places, as detailed below. There are also some inconsistencies in the data that should be attended to. The authors don't seem aware of a previous (English language) questionnaire that uses the same approach. Any broad differences and similarities between their questionnaire and the previous questionnaire should be commented on, particularly for publication in an English-language publication like BMJ Open. Specific comments In the following, page numbers refer to the set at the bottom of the page. The Introduction should comment on how this questionnaire differs from the PEACH (Ching & Hill, 2007; Ching et al, 2010). P2, index 36: There are inconsistencies in the number of participants. Here it is stated to be 297; Table 1 sums to 309; Table 2 sums to 300; and the Methods and Results text refers to 309.

P2, index 56: Given the low sensitivity found in this experiment, I don't see how this conclusion is justified, particularly given that the correlation between threshold and auditory skill ratings was only r=0.36 (Pp, index 47).

P3, index 29 to 31: Similarly I don't see how the results obtained justify this very strongly worded dot point.

P5, index 8 to 12: This statement does not seem accurate given the availability of PEACH.

P5, index 45: In what respect were the responses to the other scales used as control? I can't see that they have been used in that way, other than being "included" in the discriminant analysis, though it is not stated how they were included (P10, index 10 to 12).

P5, index 52 t0 56: Is this section saying that the questionnaire itself sets out the expected behaviours for older children, but not for younger children? If so, that's not how the questionnaire reads.

P7, index 38: Why constrain the Cronbach alpha analysis to the A scale? Was it because of the smaller number of questions in the other scales?

P7, index 42: Add "depending on age" after "2 to 11", for clarity.

P7, index 50 to 52: The questionnaire itself shows that it is not true that "2-9 month-old infants received in total 6 questions on ….". It is true of the age range 2-6 months though.

P8, index 15: What does it mean "Missing data were only accounted for in the analysis of the data"?

P9, index 47 and 56: Values of R-squared cannot be negative. Also, t-values are given in this section, but it is not clear from the description just what t-tests are being carried out.

P10, index 10 to 15: "Positive predictive value" is the probability that subjects with a positive screening test truly have the disease. This sounds different from "children … successfully classified as with or without HI"?

P10, index 17; What does "run over the respective train subgroups" mean?

P10, index 22: What makes 0.5 the optimum cut-off?

P10, index 31: Does "the existing results" mean "completed answers"?

P11, index 31: Again, in what sense were scores on the other scales used as a control?

Table 2: It's not appropriate to express hearing threshold to a precision of three decimal places (i.e. 0.001 dB).

The supplementary material on discriminant analysis is not expressed in a publishable form. Some more explanation is needed.

In the English language translation of the questionnaire, the auditory scale should be headed "A" rather than "U", to match the label used in the paper.

Typos and grammar
P2, Index 36: "fo" should be "for".
P7, index 20 to 22: What is a "put-patient clinic"?
P9, index 15: "con- ductive".
P9, index 42 to 54: I presume that "X^2(2)" refers to Chi-squared with two degrees of freedom. Please replace with appropriate fonts, superscripts and subscripts.
P10, index 36: "run" should be "ran".

References

| | Ching, T. Y. and M. Hill (2007). "The Parents' Evaluation of Aural/Oral Performance of Children (PEACH) scale: normative data." J Am Acad Audiol 18(3): 220-235. |
| | Ching, T. Y., et al. (2010). "Language development and everyday functioning of children with hearing loss assessed at 3 years of age." Int J Speech Lang Pathol 12(2): 124-131. |

VERSION 1 – AUTHOR RESPONSE

Dear Reviewer 1,
Thank you for the suggested literature. The study was created with the aim similar to the two studies already mentioned in the manuscript (Olusanya, 2001; Samelli et al, 2011), namely to create the questionnaire that could complement NHS in identifying HI that progress or have a later onset (comprehension-expression skills were tested on a different data sample). We now refer to this study in the reference list.

Dear Reviewer 2,
Thank you very much for the comprehensive list of constructive comments and suggestions. We believe that the implementation of the changes as suggested by the comments has now significantly improved the overall quality of the text.
The replies to the comments (except for typos and grammar) are written below each comment as to assure that they have all been addressed.

Specific comments
In the following, page numbers refer to the set at the bottom of the page.
The Introduction should comment on how this questionnaire differs from the PEACH (Ching & Hill, 2007; Ching et al, 2010).
Thank you to the reviewer for the suggested literature. Although it was validated in participants with normal hearing, PEACH questionnaire was created predominantly to assess the hearing function in children who already have HI. We now refer to this research and specify the differences between PEACH and QUAC questionnaires..

P2, index 36: There are inconsistencies in the number of participants. Here it is stated to be 297; Table 1 sums to 309; Table 2 sums to 300; and the Methods and Results text refers to 309.
The final sample consists of 309 children for whom there are no missing data for the questions from scale A and who received final diagnosis before december 2019. The counts are now corrected everywhere.
The mistake was the result of the fact that for a subgroup of children (N=12, of which 8 with normal hearing and 4 with HI), diagnostic data were obtained at the very conclusion of the study when the data analysis has already started, and by mistake, the old version of the Table 2 was inserted in the manuscript. We apologize for the mistake and have now inserted the correct version of the Table2.

P2, index 56: Given the low sensitivity found in this experiment, I don't see how this conclusion is justified, particularly given that the correlation between threshold and auditory skill ratings was only r=0.36 (Pp, index 47).
Sensitivity of the questionnaire was not the only measure in the study. Our initial conclusion that "parents of children with verified risk for HI recognise non-typical auditory behavior" was based on the significant differences between the answers for children with and without HI, and on the significant correlations between the scale A responses and the hearing threshold, type, and degree of HI. In

order to appropriately stress the fact that successful identification of cases with HI may nonetheless not be very successful, we modified the last sentence of the Abstract.

P3, index 29 to 31: Similarly I don't see how the results obtained justify this very strongly worded dot point.
As explained above, the conclusion that parents do recognise the decrease in auditory and language skills in children with moderate to severe HI was based on the significant correlations between the degree of HI and the results from the questionnaire on language and auditory skills, and on the significant differences between the group of children with and without HI.
However, based on the recommendation from the Editor, the Strengths and Limitations section was changed to reflect the methodological issues and strengths.

P5, index 8 to 12: This statement does not seem accurate given the availability of PEACH.
The PEACH questionnaire was validated both for NH and HI children in various languages, and we thank again to the reviewer for the mention of these studies. Since it was not validated for Italian, and since one of the scopes of the study was to directly compare auditory skills to other language-communicative skills in various age groups, we believe that the results from this new questionnaire represent novelty in the literature. A more detailed description of the differences between the existing questionnaires and QUAC is now inserted in the Introduction section.

P5, index 45: In what respect were the responses to the other scales used as control? I can't see that they have been used in that way, other than being "included" in the discriminant analysis, though it is not stated how they were included (P10, index 10 to 12).
The responses to scales A, P and C correlated with the Degree and Type of HI, so all scales were included in the discriminant analysis. They were indeed not used explicitly as the control variables in any of the final analyses. We have changed the section in the Introduction accordingly.

P5, index 52 t0 56: Is this section saying that the questionnaire itself sets out the expected behaviours for older children, but not for younger children? If so, that's not how the questionnaire reads.
Thank you for pointing out this inconsistency. Indeed the questionnaire itself was not created to set out the expected behavior for older children only. We have therefore omitted the last sentence of the Introduction.

P7, index 38: Why constrain the Cronbach alpha analysis to the A scale? Was it because of the smaller number of questions in the other scales?
The Cronbach alpha analysis was constrained to the A scale because the A scale contained several questions, whereas other scales were only represented by one or two questions per each age-group. This is now explained in the text.

P8, index 15: What does it mean "Missing data were only accounted for in the analysis of the data"?
Missing data were not interpolated in any way, and this was the intended meaning of the sentence. It reads as somewhat redundant, so we omitted it.

P9, index 47 and 56: Values of R-squared cannot be negative. Also, t-values are given in this section, but it is not clear from the description just what t-tests are being carried out.
Here we reported the Pearson's product-moment correlation with the t-value computed for the hypothesis that true correlation is equal to 0. To maintain clarity, and given that Pearson's r rank correlation coefficients in a multiple correlation table represent the same result, we now modified the Results section by taking away the t-values and inserting a more comprehensive table of multiple correlations.

P10, index 10 to 15: "Positive predictive value" is the probability that subjects with a positive screening test truly have the disease. This sounds different from "children … successfully classified as with or without HI"?

Thank you for noting the inconsistency, we have now removed the words "or without" from the sentence. Indeed, Positive predictive value is the proportion of positive results in statistics and diagnostic tests that are true positive.*

*Fletcher, Robert H. Fletcher ; Suzanne W. (2005). Clinical epidemiology : the essentials (4th ed.). Baltimore, Md.: Lippincott Williams & Wilkins. pp. 45. ISBN 0-7817-5215-9.

P10, index 17; What does "run over the respective train subgroups" mean?

We thank you for pointing out that this description was not clear enough. Originally, the positive predictive value was computed by randomly dividing our dataset into the subgroups in order to verify the quality of the classification method. This procedure was repeated 500 times, and mean values were reported. In order to maintain transparency, and because this gives more stable estimates, all signal detection values (sensitivity, specificity, PPV, odds ratio) and the non-parametrical AUC-ROC values are now computed by averaging the values computed from 500 repetitions of the linear discriminant analysis. We have corrected all results regarding the discriminant analysis accordingly.

P10, index 22: What makes 0.5 the optimum cut-off?

The value 0.5 is the cutoff value for the results of the discriminant analysis, and it does not refer in any way to the ray scores of the questionnaire. Since all the parameters of the discriminant analysis are needed to produce estimates for each new dataset, whereas the cutoff values only serves for the final prediction, we have now omitted the explicit mention of the cutoff values. We thank you again for pointing out the need to clarify this point.

P10, index 31: Does "the existing results" mean "completed answers"?

This was the intended meaning; thank you for the correction.

P11, index 31: Again, in what sense were scores on the other scales used as a control?

Table 2: It's not appropriate to express hearing threshold to a precision of three decimal places (i.e. 0.001 dB).

As mentioned above, we agree with the note that none of the analyses used these variables as controls. We have therefore omitted the part of the sentence that refers to scale I as control variable. The dB values are now corrected.

The supplementary material on discriminant analysis is not expressed in a publishable form. Some more explanation is needed.

We have explained in more detail the applied analysis using the LDA in the main tex now.

Given that there are no standardised means of reporting the results of the LDA, we have assumed that the visual representation of the results and predictive value of the classification method will be more informative and have therefore replaced the raw output of the models with the visualisation of the grouping results of the LDA and the ROC curves. The replication of the results is possible based on the statistical code and the raw dataset available at the OSF repository.

VERSION 2 – REVIEW

REVIEWER	Dillon, Harvey The University of Manchester, Audiology
REVIEW RETURNED	29-Nov-2020

GENERAL COMMENTS	General comments

The authors have dealt with my concerns except for my comment that "The supplementary material on discriminant analysis is not expressed in a publishable form. Some more explanation is needed."

However, on a fresh read on a version with fewer problems, I have noticed some additional concerns. They are all minor, and I expect the authors will have no trouble attending to them.

Specific comments
Page numbers refer to the numbers at the top of the page (53 in total) not the numbers at the bottom of the page.
P3, index 42: Please acknowledge the weakness of the relationship by inserting "weakly but" before negatively. Also, how can anything correlate with "type of HI", as type is not a numerical value?
P6, index 29: Based on what is written, the authors have misinterpreted the Bartlett's K-squared test. A significant p value (in this case 0.008), indicates that the Bartlett's statistic (which has a Chi-squared distribution) is sufficiently large that the two distributions are very unlikely to have the same variance. I have no idea what the significance of this is for conducting the linear discriminant analysis that follows, but it still may be possible to use it. Statistical advice should be sought. If the variation in variance between the two groups invalidates ding the discriminant analysis, then the discriminant analysis should be dropped from the paper.
P12, index 3: With an area under the curve of 0.662, it is obviously not possible to have a high sensitivity and a high specificity at the same time. By changing the cut-off point, many combinations of sensitivity and specificity are possible. While I think I can infer from the distributions shown in the Supplementary material why the authors chose such a low sensitivity and high specificity, it would be useful to add a sentence saying why this particular choice was made.
P11, index 56 to P12, index 13: I find the odds ratios confusing in this context, and actually hard to believe. What are they indicating? For example, an odds ratio of 82.67 (P12, index 10) is associated with a probability of 0.988, calculated from R/(1+R), where R is the odds ratio. Given the relatively weak association between the questionnaire scores and the measured hearing status, I am struggling to see where such near perfect probabilities and such high odds ratios come from, nor how they are interpreted. The figures for area under curve, sensitivity, specificity and positive predictive value all look to be consistent with the data, however.
P12, index 52: I don't understand the phrase "in specific screening evaluation".
P13, index 17 to 19: This is unclearly expressed. I think the intended meaning is that the HI children did not differ from the normally hearing children in their interactive skills.
P14, index 19 to 22: This is an overly strong statement as the children's auditory and language skills have not actually been measured. If the parents have not reported a deficit in these skills, then it might be that there is a deficit they have not noticed, or it might be that the hearing loss has not (yet) caused a deficit in these children.
Typos and grammar
P3, index 48: Add space after "with".
P3, index 42: This is one of many place in the manuscript where there are double full stops separated by spaces.

	P3, index 52: Add "have some capacity to" before "recognise", or else this gives an overly optimistic impression of how well the parents recognised hearing loss.
	P5, index 45: Delete "a" before "specific".
	P6, Index 19: typo "mentioned".
	P6, index 31: typo "frist".
	P6, index 56: typo "extend" should be "extent".
	P8, index 17: "hearing" should be capitalized and "Degrees" should not be.
	P9, index 22: "Estimate" should not be capitalized.
	P11, index 17: Typo "post-hock"
	P13, index 8: Insert "of" after "severity".
	P13, index 40: Delete "age" before "larger".

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2

Prof. Harvey Dillon, The University of Manchester, Macquarie University

Comments to the Author:

General comments

The authors have dealt with my concerns except for my comment that "The supplementary material on discriminant analysis is not expressed in a publishable form. Some more explanation is needed."

However, on a fresh read on a version with fewer problems, I have noticed some additional concerns. They are all minor, and I expect the authors will have no trouble attending to them.

Specific comments

Page numbers refer to the numbers at the top of the page (53 in total) not the numbers at the bottom of the page.

P3, index 42: Please acknowledge the weakness of the relationship by inserting "weakly but" before negatively. Also, how can anything correlate with "type of HI", as type is not a numerical value?

Thank you for pointing out this minor inconsistency. In fact, we have transformed the type of HI into numerical values as well for the purpose of the analysis, and our decision to label conductive HI with 1 and sensorineural HI with 2 was based on the knowledge that sensorineural deficits more often cause long lasting language deficits than conductive deficits. We now report this numeric transformation in the methods.

However, upon the second read and the re-run of the correlation analysis, we noticed that the method reported was that of Pearson instead of Spearman (which is indicated in the methods section and is more appropriate since all our score variables were non-normally distributed). Hence some differences in the results section that reports the multiple correltation analysis.

7

P6, index 29: Based on what is written, the authors have misinterpreted the Bartlett's K-squared test. A significant p value (in this case 0.008), indicates that the Bartlett's statistic (which has a Chi-squared distribution) is sufficiently large that the two distributions are very unlikely to have the same variance. I have no idea what the significance of this is for conducting the linear discriminant analysis that follows, but it still may be possible to use it. Statistical advice should be sought. If the variation in variance between the two groups invalidates ding the discriminant analysis, then the discriminant analysis should be dropped from the paper.

Thank you for noticing this! During the first revision of the manuscript a phrase went missing by mistake. We have added it now: "The Bartlett's test of homogeneity of covariance matrices showed that the variances of the questionnaire scales of subsets with and without HI were equal for all but the comprehension scale (Bartlett's K-squared(1) = 6.997, p = 0.008)."

Bartlett's K-squared test was used here to see if the covariance matrices of the dependent variables are homogenous for the populations used. If they are not homogenous, the Quadratic Discrimination analysis should be used. As there are 4 independent variables in our questionnaire (A,I,C,P), Bartlet's test was done for the distribution of the independant variable in each of the dependent variables. Only the comprehension scores (C) had a non-homogeneous covariance matrices (p=0.008), whereas for others p>0.05. Therefore, we have continued with the linear discriminant analysis.

P12, index 3: With an area under the curve of 0.662, it is obviously not possible to have a high sensitivity and a high specificity at the same time. By changing the cut-off point, many combinations of sensitivity and specificity are possible. While I think I can infer from the distributions shown in the Supplementary material why the authors chose such a low sensitivity and high specificity, it would be useful to add a sentence saying why this particular choice was made.

Thank you for this thought; the reference to the effects of the cutoffs to the different success measures indeed helps to further clarify the results. We have now looked in more detail into the effects of various cutoff, and the modified paragraph now reads:

"The best performance of the model, with 77.7% children correctly classified as having HI or not (PPV = 0.777), was achieved at the sensitivity (true positive rate) and specificity (true negative rate) scores of 0.388 and 0.933. By changing the cutoff of the model to increase correctly recognised cases with HI, the number of falsely positive cases would increase significantly. Given that the percentage of children with HI in the population is relatively low, this would, however, mean a great number of children falsely identified at risk for HI. The area under the ROC curve (AUC) for the model at different cutoff values was 0.661, which confirms the above observation. The output of the discriminant analysis, along with the predictions of the model at various cutoff values, is available in the Supplementary material A4."

P11, index 56 to P12, index 13: I find the odds ratios confusing in this context, and actually hard to believe. What are they indicating? For example, an odds ratio of 82.67 (P12, index 10) is associated with a probability of 0.988, calculated from R/(1+R), where R is the odds ratio. Given the relatively weak association between the questionnaire scores and the measured hearing status, I am struggling to see where such near perfect probabilities and such high odds ratios come from, nor how they are interpreted. The figures for area under curve, sensitivity, specificity and positive predictive value all look to be consistent with the data, however.

Regarding the odds ratios, we have revisited the analysis, and realised that the reported ratios did not correctly measure the probability of the given score for the diagnostic outcome. Given that sensitivity, specificity, and PPV give a rather accurate prediction about the quality of the questionnaire to predict the diagnostic outcome, we have thereof omitted the odds ratios.

P12, index 52: I don't understand the phrase "in specific screening evaluation".

The word specific referred to the fact that various developmental problems were correlated with the corresponding parental concerns, in the case of the reference 22 the autism spectrum. We have now replaced the word with an indefinite article to avoid misinterpretations.

P13, index 17 to 19: This is unclearly expressed. I think the intended meaning is that the HI children did not differ from the normally hearing children in their interactive skills.

Thank you for the note; we have reformulated the sentence now: "The increased degree of HI weakly correlates also with lower language (production and comprehension scores), but not interaction skills."

P14, index 19 to 22: This is an overly strong statement as the children's auditory and language skills have not actually been measured. If the parents have not reported a deficit in these skills, then it might be that there is a deficit they have not noticed, or it might be that the hearing loss has not (yet) caused a deficit in these children.

Given that the indexation and the pages do not correspond to the proof pdf that we received upon resubmitting, we had some difficulties in establishing the exact paragraph that the comment refers to.

If it refers, however to the sentence: "Nonetheless, the majority of parents did not correctly recognise the decrease in auditory and language skills in children with HI," then we would like to confirm your thought - it might be that there is a deficit that the parents have not (yet) noticed.

To avoid misunderstanding, we have added the clarification few lines below:

"Again, such results can be attributed to the young age of the participants: hearing deficit may be difficult to notice in very young children, especially if it had not (yet) caused a significant change in their behaviour."

Typos and grammar

P3, index 48: Add space after "with".

P3, index 42: This is one of many place in the manuscript where there are double full stops separated by spaces.

P3, index 52: Add "have some capacity to" before "recognise", or else this gives an overly optimistic impression of how well the parents recognised hearing loss.

P5, index 45: Delete "a" before "specific".

P6, Index 19: typo "mentioned".

P6, index 31: typo "frist".

P6, index 56: typo "extend" should be "extent".

P8, index 17: "hearing" should be capitalized and "Degrees" should not be.

P9, index 22: "Estimate" should not be capitalized.

P11, index 17: Typo "post-hock"

P13, index 8: Insert "of" after "severity".

P13, index 40: Delete "age" before "larger".

Thank you very much for your time in noting the typos and grammar; all the points noted above have been addressed in the text accordingly.

VERSION 3 – REVIEW

REVIEWER	Dillon, Harvey The University of Manchester, Audiology
REVIEW RETURNED	03-May-2021

GENERAL COMMENTS	The revised version has largely met my concerns, and I have only the following minor specific comments. In the following, the page numbers refer to those at the foot of the page, not those at the top. Specific comments

Abstract, line 40: Here and throughout the paper, figures should be reported with a number of decimal places commensurate with the precision of the figures. For example, the Chi-squared value would be 14.4, and percentage of children would be 59%.

P3, line 17: It's not clear how "comparing" age-appropriate auditory skills to language and communication ability results in a more accurate recognition of hearing impairment. Also, more accurate than what? Do you mean that the three types of abilities were compared to determine the best way to recognise hearing impairment?

P3, line 22: In more detail than what?

P10, lines 38 to 49: It is unusual to perform correlations between other variables and both Hearing threshold and Degree of loss, because Degree of loss is a categorical variable determined almost solely by Hearing threshold. The only difference between the variables (other than being continuous versus categorical) is that the authors have decided that unilateral losses have a severity intermediate to that of normal hearing and mild bilateral. My preference would be to delete the correlations with Degree of loss, although deletion is not essential. The main problem with having both of these variables is that the additional correlations creates more opportunities to get "significant" correlations by chance alone.

P11, line 31: The area under the ROC curve does not depend on the cut-off value used to measure sensitivity and specificity. Therefore, delete the words "at different cutoff values".

P13, lines 3 to 6: It is misleading to say "We therefore conclude that parental developmental concerns regarding their children's auditory abilities could be assessed during paediatric visits and taken into consideration in overall risk assessment." This currently sounds like a conclusion from this study, and if so, it is certainly not supported by the data, given the poor performance of the predictions of hearing status from the questionnaire. I suspect the authors meant to say "We therefore speculated whether parental developmental concerns regarding their children's auditory abilities could usefully be taken into consideration in overall risk assessment."

P13, line 22: Replace "age larger groups" with "larger age groups".

Figure 1: Please indicate in the caption what the error bars correspond to.

P33, line 9: Spelling error in "Predicition".

VERSION 3 – AUTHOR RESPONSE

Dear prof Dillon,

Thank you once more for the thorough read of the manuscript. We are deeply honoured by the time and the precision invested in the reviews of this manuscript. We have now addressed all the issues brought up in the third review, as follows.

11

Abstract, line 40: Here and throughout the paper, figures should be reported with a number of decimal places commensurate with the precision of the figures. For example, the Chi-squared value would be 14.4, and percentage of children would be 59%.

Corrected.

P3, line 17: It's not clear how "comparing" age-appropriate auditory skills to language and communication ability results in a more accurate recognition of hearing impairment. Also, more accurate than what? Do you mean that the three types of abilities were compared to determine the best way to recognise hearing impairment?

The corrected sentence is now as follows:

Parents' recognition of age-appropriate auditory, language and communication abilities was evaluated as a possible tool for a more accurate recognition of hearing impairments that progress or have a later onset.

P3, line 22: In more detail than what?

The word more is removed now.

P10, lines 38 to 49: It is unusual to perform correlations between other variables and both Hearing threshold and Degree of loss, because Degree of loss is a categorical variable determined almost solely by Hearing threshold. The only difference between the variables (other than being continuous versus categorical) is that the authors have decided that unilateral losses have a severity intermediate to that of normal hearing and mild bilateral. My preference would be to delete the correlations with Degree of loss, although deletion is not essential. The main problem with having both of these variables is that the additional correlations creates more opportunities to get "significant" correlations by chance alone.

Thank you for this consideration; in fact, the correlations with the Degree of HI yield very similar results to the ones with the Hearing threshold. As the story doesn't change with the removal of the results for the degree of HI, we have now removed them.

P11, line 31: The area under the ROC curve does not depend on the cut-off value used to measure sensitivity and specificity. Therefore, delete the words "at different cutoff values".

The words "at different cutoff values" are deleted.

P13, lines 3 to 6: It is misleading to say "We therefore conclude that parental developmental concerns regarding their children's auditory abilities could be assessed during paediatric visits and taken into consideration in overall risk assessment." This currently sounds like a conclusion from this study, and if so, it is certainly not supported by the data, given the poor performance of the predictions of hearing status from the questionnaire. I suspect the authors meant to say "We therefore speculated whether parental developmental concerns regarding their children's auditory abilities could usefully be taken into consideration in overall risk assessment."

Thank you for your suggestion. We have corrected the sentence accordingly.

P13, line 22: Replace "age larger groups" with "larger age groups".

Thank you; corrected.

Figure 1: Please indicate in the caption what the error bars correspond to.

Thanks for noticing; we added the indication about the error bars now.

P33, line 9: Spelling error in "Predicition".

Corrected.