Author’s response to reviews

Title: Safety, efficiency and health-related quality of telephone triage conducted by general practitioners, nurses, or physicians in out-of-hours primary care: a quasi-experimental study using the Assessment of Quality in Telephone Triage (AQTT) to assess audio-recorded telephone calls

Authors:

Dennis Graversen (d.graversen@ph.au.dk)
Morten Christensen (mbc@ph.au.dk)
Anette Pedersen (afp@ph.au.dk)
Anders Carlsen (andershelles@gmail.com)
Flemming Bro (fbro@ph.au.dk)
Helle Christensen (helle.collatz.christensen.02@regionh.dk)
Claus Høstrup Vestergaard (cv@ph.au.dk)
Linda Huibers (huibers@ph.au.dk)

Version: 2 Date: 11 Dec 2019

Author’s response to reviews:

Dear Editor,

We submit our revised manuscript entitled "Safety, efficiency and health-related quality of telephone triage conducted by general practitioners, nurses, or physicians in out-of-hours primary care: a quasi-experimental study using the Assessment of Quality in Telephone Triage (AQTT) to assess audio-recorded telephone calls" (FAMP-D-19-00246). Thank you for considering our manuscript for publication and the reviewer for the constructive comments. The valuable remarks helped us to improve the manuscript.

We carefully revised our manuscript according to the reviewer’s comments. We include a point-by-point response detailing our considerations and reactions upon the comments and suggestions raised. We revised the manuscript accordingly, adding a clean copy and an annotated copy using Track Changes to highlight our changes.
We hope that you will find that the suggestions and criticism have been met in a satisfying way and that the manuscript has improved accordingly.

Yours faithfully, on behalf of the other authors,
Dennis Schou Graversen

Point-by-point comments
Technical Comments:

1. Please rename Objectives to Background.

2. Please rename Introduction to Background.

3. Please move the position of Abbreviation after conclusions.

4. Please include Figure legends headings.

• The above comments have been addressed and all Figure and Table headlines and legends have been gathered in a section.

Editor Comments:

Reviewer report reviewer 1:
Comment 1). Authors have addressed some comments appropriately; however, statistical analysis is still incorrect at several places. I would encourage authors to get the manuscript reviewed by a statistician.
• We appreciate the reviewer’s thorough review and relevant comments. We mistakenly did not implement some of the described changes in the submitted version. We apologise for the confusing and time-consuming process. Some of the comments seem to be caused by this confusion. We have had a comprehensive discussion of all the included comments together with a statistician, who is co-author.

Comment 2) As previously mentioned, the Table 1 has incorrect statistical tests; authors mentioned in the revised manuscript that pair-wise comparison is done using chi-squared test (categorical variables) and student's t-test (continuous data). Chi-square test is incorrect for the type of data they have mentioned Mean minutes and seconds. Also, if authors preferred not to use ANOVA for three way comparison and preferred to use pairwise comparison, the p value needs to be adjusted with Bonferroni corrections in this case.
• Table 1 includes both categorical (sex, age group, time of call) and continuous (length of call) variables. Thus, as length of call is considered a continuous variable, we used non-parametric tests. Moreover, we have re-evaluated the comments proposed by the reviewer in the prior review by making an overall comparison of the baseline characteristics in table 1 of all three
triage professionals using three groups comparison statistics. Thus, we have compared sex, age group, time of call using chi-square test and length of call using Kruskal-Wallis.

• In case of a significant difference on the three group comparison, we have conducted pair-wise comparison with GP as reference using chi-square tests for categorical variables and Mann-Whitney U-test for the continuous variable length of call, adjusting level of significance with Bonferroni (0.025). This made it possible to see the direction of the differences found in the three groups comparison. We made adjustments accordingly, in the statistical analyses paragraph and table legends. As Bonferroni correction in general is known to be rather conservative (1), therewith increases the risk of type-2 errors (i.e. underestimating the statistical power of the study). We added this to the limitations. P-values have been added enabling the reader to interpreted appropriately.

Table 1 Legends: “£ Indicating a significant difference (p<0.05) between all three groups of triage professionals, using chi-square test for categorical variables and Kruskal-Wallis for length of call. *Significant difference between nurses or physicians in pairwise comparison with GPs as reference group (Bonferonni adjusted p<0.025), using chi-squared test (all categorical variables) and Mann-Whitney U-test (length of call). ”

o Statistical analyses section: “We conducted an overall comparison of patient and call characteristics using chi-square test for categorical variables and Kruskal-Wallis test for continuous variables (significance level <0.05). In case of a significant difference, we conducted a post-hoc pair-wise comparison using chi-squared test for categorical variables and Mann-Whitney U-test for continuous variables with Bonferroni adjusted significance level (<0.025).”

o Limitation: “In the analyses we have performed many tests so significance by chance cannot be excluded. A solution could be adjusting significance levels by Bonferroni consistently in all analyses, but this has been suggested to be too conservative and associated with increased risk of type-2 errors(1).”

Comment 3) Authors mentioned that they have changed confidence intervals to standard deviations. However, it has not been changed at several places. Also I am wondering if authors just mentioned confidence intervals previously and now changing to standards deviations, the calculations are different and the numbers should change as well.

• Thank you for highlighting this again. As mentioned, we have mistakenly overlooked to save these changes in the submitted manuscript. Mean length of call is now consistently supplemented with standard deviations.

Comment 4) In Table 2, authors previously used the categories 4 - 5 versus 1 - 2 and clearly mentioned that responses to 3 were excluded. In the revised manuscript, they changed the categories to 4-5 versus 1-3; however, none of the readings changed in the table which is not possible if the categories were revised.

• This is an essential part in order to interpret appropriate, and apparently, this needs further clarification. It is a misunderstanding that “3” was excluded or that “3” was included in the poor quality “1-3”. We stated that “we categorized the outcomes into poor quality (rated “1” or ”2”) and sufficient quality (rated “3”, “4”, or “5”). “Not applicable” (“N/a”) was recoded into “missing”. We have made an effort to clarify our categorization, to limit misunderstanding, by emphasizing that “3” was included in what we define as “sufficient quality”
We calculated the relative risk (RR) of having poor quality (i.e. rated “1” or “2”) versus sufficient quality (i.e. rated “3”, “4” or “5”).

Legend table 2: “The RR for “poor quality” (i.e. “1” or “2”) was analysed using binomial regression model (GP as reference group). *Significant differences: p<0.05.

Comment 5) Authors mentioned that it was a typo to write General Linear Regression Model which was actually "binomial regression model". However, they have still not addressed my comment regarding the assumptions for the model to determine if it was a good fit.

- We apologize for missing this point in our earlier correspondence. We thought that the original comment stemmed from confusion over the term "General Linear Regression", by which we meant to write "Generalized linear MODEL". As this seemed to be a source of confusion, we decided to specify which type of GLM was used, i.e. binomial regression (binomial family, logit link). Binomial regression does not make many of the key assumptions of linear regression and as the covariates are categorical, tests for linearity are not applicable. Therefore, we are not sure what the reviewer means by "if it was a good fit". Describing the model fit statistics in binomial regression is rather uncommon and we feel it would distract from the findings of the study. In addition, this seems to go somewhat beyond the usual standards of the field, especially when we have not used predictive modeling and have not used statistics to predict outcomes. We will gladly provide fit statistics if the editor finds this necessary, but we struggle to see the additional value of reporting such statistics for our study.

Comment 6) In the revised manuscript, the title of the Table 2 is missing
- Thank you for highlighting this. The title has been added again.

Comment 7) Authors responded mentioning that the limitations have been revised so did not change the conclusion as previously suggestion. However, despite addressing the limitations, the conclusion should be balanced to provide realistic information to the readers
- We acknowledge that our study has limitations and we have addressed limitations further as described above. We have changed the conclusion concerning comparison of GPs and nurses as follows:

- Adding “keeping limitations in mind”
- Adding “explorative study”
- Adding “indicated”
- Deleting “significantly”
- Adjusted our study aim to “explore and compare”
- We have provided p-values enabling the reader to interpret the results with this in mind.

- As described in the prior revision, due to the similar-to-me bias, we consider our comparison to be a conservative estimate of the differences seen between nurses and GPs. Thus, we think that
the reader would be able to get a realistic and balanced picture. We have changed the conclusion as follows;

“Keeping limitations in mind, our explorative study indicated that nurses using CDSS performed better than GPs in telephone triage, especially in four out of ten specific health-related items. Those items concern identification and uncovering of the problem.”