Peer Review File

Article information: https://dx.doi.org/10.21037/ajo-23-33

Reviewer A

Comment 1: A novel paper comparing 2 common-sized chip in the tip endoscopic nasal endoscopes. Excellent evidence-based research and necessary new knowledge documented.
Reply 1: Reviewer A's feedback is well received
Changes in the text: Nil changes to the text

Reviewer B

Comment 2: The paper presents a comparison between two endoscopes in a single prospective study. It demonstrates that the smaller endoscope may have advantages. There are some methodology concerns that need to be addressed by the authors.
Reply 2: Reviewer B makes some valid criticisms of the study that we address below.
Changes in the text: Nil changes to the text

Comment 3: The study was not randomised. A repeated measures study could have still been performed with randomisation resulting in different sized scopes for the first scope.
Reply 3: First use or order of endoscopie size was not randomized for each patient. It was decided to make the examination procedure as controlled as possible with a standardized procedure order and only a single operator used. The right nasal cavity was examined first by the 2.9mm endoscope followed by the 3.6mm endoscope. The left nasal cavity was examined in the reverse order. There is a possibility of an undetected Order by Side confounding effect but this was considered/hoped to be trivial. The assessment of the primary outcome was not made at the time but by post-examination video with two independent and blinded expert reviewers. The presentation of the video footage to the reviewers was randomized.
Changes in the text: The current draft has been updated to include the reversing of the order of the size with side. This was inadvertently left out in the former draft. Please see line 127-129.

Comment 4: The operator was not blinded to the scope.
Reply 4: The operator could not be blinded to the treatment arm, given the palpable difference in scope diameter. The two independent expert reviewers were blinded to the treatment arm.
Changes in the text: The current draft has been updated to include a statement specifically reflecting that the clinician was not blinded to the scope diameter. Please see lines 131-132.

Comment 5: Could the authors explain the rationale for using the McNemar paired test?
Reply 5: Differences in proportions between the site visualisations were tested using the McNemar paired test. The 2.9mm and 3.7mm scopes were being taken repeatedly from the same 'sample' (patient-side-site). Standard tests of proportions (Pearson's

X^2 or Fisher's Exact test) assume impendence between samples that would have been violated here. None of the above three methods take into account the hierarchical nature of the study design and this was a limitation of the NcNemar test. A more complex generalized linear mixed model (GLMM) was developed as part of the analysis, to account for the hierarchical study design, but this was excluded from the submission as the results were considered to be too difficult to interpret for the wider audience.
Changes in the text: Nil changes of the text

Comment 6: Line 173 ("Reviewer 1 gains only a 173 1% improved visualisation…") - could the authors clarify what does "gains only" mean?
Reply 6: The gain of 1% visualisation reported with the 2.9mm scope for reviewer 1 was in comparison with the 3.6mm scope visualisation rate also for reviewer 1. The visualisation percentage difference reflects the improved visualisation across all 9600 anatomical subsites reviewed, when comparing the 2 differing diameter scopes. This is also the case for the 5% gains noted for reviewer 2. The word "only" was removed from the line describing the improvement noted with the 2.9mm endoscope for reviewer 1.
Changes in the text: The current draft has been updated to include clarification regarding what the visualisation percentage indicates. Additionally, the word "only" has been removed. Please see line180-182.

Comment 7: Line 176(…increased visualisation rates of 175 3.1% over the traditional diameter…) - is the increased visualisation rate of 3.1% statistically significant
Reply 7:
The p-values of ($p < 0.0001$) based on a McNemar test for the 3.1% increase in visulasation is statistically significant. The hierarchical nature of the study, the large number of replicates, produces overly optimistic/sensitive p-values and hence only the percentages were reported.
Because of the large sample size of 9,600 visualisations, an effect or interaction might be statistically significant but not necessarily clinically significant. Clinical significance is to be determined by the reader based on the accuracy they required from their endonasal examinations.
Changes in the text: The current draft has been updated to include the McNemar's test p value data for the differences in visualisation based on scope size. Please see line180-184.
Further discussion was introduced to reflect the tendency to overly optimistic p values. Please see line 286-291.

Editorial Comments

1. Please specify the date and settings where the data were collected in the abstract and the main text. Please briefly report the statistical methods in the abstract.
Reply 1: All of these recommendations have been attended to.
Changes in the text: The current draft has been updated, please refer to line 39-40, 42-44, 125-126.

2. All significant findings should be reported with specific data and their respective precisions in the abstract instead of using generalized terms like "significantly outperformed" or "significantly lower."

Reply 1: Specific data has been added to the abstract result section with the removal of generalised terms in isolation.
Changes in the text: The current draft has been updated, please refer to line 51-55, 57-58.

3. “The sample size of 50 patients was established by a power analysis on a pilot study of 15 patients”, it’s suggested to report the parameters involved in the power analysis, such as α and β values if applicable.
Reply 1: The α:0.05 and β:0.2 values from the pilot study have been included as recommended.
Changes in the text: The current draft has been updated, please refer to line 141.

4. Are there any differences in appearance between the two flexible nasal endoscope of different diameters? Alternatively, how can one discern between the two following visualization scoring?
Reply 1: The appearance of each scope was similar but not identical and as such the endoscopist could identify each scope from appearances alone, aside from shaft diameter. Although the patients were not informed of which scope was being used, close observation on their part would have appreciated a difference in appearance. Determination of the endoscope size from the recorded data after visualisation scoring would require correlation with the master datasheet where the scope size is recorded in correlation with each endoscopic recording file number.
Changes in the text: The current draft has been updated, please refer to line 155-161.

5. , “There were no observable differences in image quality or lighting between the two endoscopes”, please provide an explanation in the Methods section regarding how they were measured or evaluated.
Reply 1: This statement has been removed from the manuscript as the image quality and lighting were not outcome measures tested in the study. The statement on image quality and lighting reflected the authors opinions.
Changes in the text: The current draft has been updated, by the removal of the line “There were no observable differences in image quality or lighting between the two endoscopes”.

6. To provide a more comprehensive context, consider adding additional relevant baseline information, such as age, sex, and patients' diagnoses, in the Results section.

Reply 1: The inclusion of demographic (age and sex) have been included as recommended. Diagnosis was not recorded as part of the study designed and as such was not part of the data collection.
Changes in the text: The current draft has been updated, please refer to line 214-215.

7. Ensure accuracy in the reported percentage values. For instance, if reviewer 1 gained an 1.1% improved visualization and reviewer 2 gained a 5.1% benefit, align these values consistently with the results in Table 1.
Reply 1: The reported percentage values have been adjusted in the manuscript to include the 1st decimal place so as to align with Table 1.
Changes in the text: The current draft has been updated, please refer to line 219-220, 222.

8. “The 2.9mm diameter endoscope was on average 2 discomfort ratings lower than the 3.7mm endoscope on a scale of 1-10”. When discussing discomfort ratings between different endoscope diameters, provide not only the median value but also the mean (SD) value for a more comprehensive understanding.
Reply 1: The median, mean and standard deviation data present in Table 3 has been added to the text to provide a more comprehensive understanding as recommended
Changes in the text: The current draft has been updated, please refer to line 235-236, 249-250).

9. Consider merging Table 3 and Table 4 to present the data more cohesively and facilitate easier access for readers.
Reply 1: Table 3 and Table 4 have been merged to facilitate easier access to the data for readers.
Changes in the text: The current draft has been updated, please refer to line 263, 557.

10. If P value <0.001, report "P<0.001" to avoid reporting unnecessarily excessive precision.
If 0.001≤ P value <0.01, report the specific P value to 3 decimal places.
If P value≥0.01, report the specific P value to 2 decimal places.
If the P value is >0.99, report "P>0.99".
Reply 1: All P values have been adjusted to reflect the precision as per the above framework
Changes in the text: The current draft has been updated, please refer to lines 51-53, 55, 57, 203, 220-223, 227-230, 255, 262, 266, 309-310, 371, 532.