Peer Review File

Article information: https://dx.doi.org/10.21037/ajo-23-38

Reviewer A

This is a very useful systematic review and meta-analysis of balloon dilatation of the Eustachian tube in children (22 months age to 18 years of age; n= 677 procedures). Despite the fact that this review is limited to the 11 studies available up to January 2023, it is well structured, follows protocols and is excellently written. It provides valuable information albeit all from retrospective or cohort studies. The manuscript gives valuable insight into further warranted research in this area.

Reviewer B

The authors have produced an excellent paper. It is acceptable in its current form.

Editorial Comments

This systematic review and meta-analysis quantitatively and qualitatively analyzed objective and subjective outcomes for balloon dilation of the eustachian tube in the pediatric population. The innovation of this article compared with two systematic reviews that also focused on the efficacy and safety of balloon dilatation of the eustachian tube in children is the more comprehensive synthesis of a large amount of objective and subjective outcome data. However, there are a lot of errors in the data. Moreover, there are also confusing information in the manuscript. Some major and minor issues should be addressed.

Major Comments

Comment 1. The rationale for this paper is unclear. A systematic review focusing on the outcome of BDET in children has been published in 2020 (Saniasiaya, J., Kulasegarah, J. and Narayanan, P., 2021. Outcome of Eustachian Tube Balloon Dilation in Children: A Systematic Review. Annals of Otology, Rhinology & Laryngology, p.00034894211041340.). The authors do not explain why this topic was reopened at such a short time interval when it was known that there were too few current RCTs and a systematic review was available. Please describe in the introduction the new contribution of this study compared to previous systematic reviews.

Reply 1: Saniasiaya, Kulasegarah and Narayanan (2021) performed a systematic review of seven studies published up to December 2020, focussing on the indications for BDET and associated complications, while only touching briefly on outcomes. Comparatively, Aboueisha et al. (2022) performed a meta-analysis of seven studies, collating data on complications, tympanometry and PTA. This study pooled data across 11 studies, and add parameters of otomicroscopy, tubomanometry, Valsalva manoeuvre, satisfaction and quality of life.

Changes in the text: The following has been added to the last paragraph of the introduction: “This review aims to build on the works of Saniasiaya, Kulasegarah and Narayanan (2021) and Aboueisha et al. (2022). Saniasiaya, Kulasegarah and Narayanan (2021) performed a systematic review of seven studies published up to December 2020, focussing on the indications for BDET and associated complications, while touching briefly on outcomes. Comparatively, Aboueisha et al. (2022) performed a meta-analysis of seven studies, collating data on complications, tympanometry and PTA. This is the first study to comprehensively synthesise data across a large number of objective

and subjective outcomes."

Comment 2. "however, two randomised control trials (RCT) have been identified that demonstrated a statistically significant improvement in symptoms and tympanometry up to 12 months post-dilation (8,10)". Why were there no RCTs among the 11 studies included in this systematic review? In addition, please cite the original data source for these two RCTs instead of references 8 and 10.
Reply 2: The two RCTs were only for adult patients, therefore not included in this systematic review. On performing a comprehensive search, three RCTs were found.
Changes in the text: "… in adults" has been added to clarify the above, and the original data sources for the three RCTs is included.

Comment 3. Methods: We suggest the authors add an independent supplement table to present the full search strategies for all databases, registers and websites, including any filters and limits used. Example: https://jhmhp.amegroups.org/article/view/6853/html (Table 1).
Reply 3: A supplementary table has been created to reflect the full search strategy.
Changes in text: Supplementary Table 1 reflects the above recommendation.

Comment 4. Results-Table 1: Please clarify the meaning of "Patients": does it mean "the number of pediatric patients who received BDET"? If so, the number of patients in the study by Oehlandt et al. was not 23 but 7 because "In the subgroup of children (n = 23), BET as the only intervention was performed to 7 patients". Also, should the number of patients in the study by Tisch et al. be 126 or 60 instead of 94? Please check carefully to make sure the data are correct.
Reply 4: Patients refer to individuals who have consented to being part of the study to evaluate BDET. In RCSs, it refers to number of patients who have undergone BDET. But in HCSs, it refers to the total number of patients from the BDET treatment group and control group. In Oehlandt et al.'s (2022) paper, 7 patients underwent only BDET, but there was an additional 16 patients who underwent BDET in addition to another procedure, making a total of 23 patients who underwent BDET. With regards to Tisch et al.'s (2017) paper, the first group included 60 children, while the second group included 66 children. However, in the second group, "of the 66 parents who were contacted during the second part of the study, 34 (51.5%) agreed to take part." Hence, the total number of patients who consented to be included in the study was 94.
Changes in text: No changes were made.

Comment 5. Results-Table 1: Accurate reporting of age ranges is recommended. For example, the age range in the study by Howard et al. was "6.6-17.7 yo" instead of "6-18 yo", "The youngest patient was 22 months (16)" should be 28 months.
Reply 5: This has been updated.
Changes in text: The age range for Tisch et al. (2017) has been changed to 28mths – 13 yo, and Howard et al. (2021) changed to 6.6 – 17.7 yo.

Comment 6. Results-Table 4: Please explain the meaning of "n" (e.g., "the number of pediatric patients who received BDET"). Then, please check the data, which should have corresponded to Table 1. For example, should "Chen et al. (2020) (n=25)" be 49 and "Howard et al. (2021) (n=42)" be 43?

Reply 6: In Chen et al.'s (2020) paper, 25 patients underwent BDET and the remaining 24 patients underwent myringotomy and tube insertion. Hence, the complication statistic for BDET was only based around the 25 patients. The 'n' for Demir & Barman (2020) was incorrect as only 30 of the 62 patients underwent BDET. Hence, this was changed. For Howard et al. (2021), the n=42 was a typo, and this should be 43.
Changes in text: Table 4, n=30 for Demir & Barman (2020), n=43 for Howard et al. (2021). Figure 2 has been updated based on the changes to the sample size. The relevant pooled estimate was also changed in the main text and discussion. Clarification has been added to the caption for Table 4 to reflect that the n relates to BDET procedures: "No serious complications were reported after balloon dilatation of the eustachian tube in the included studies, but a summary of self-limiting adverse events is provided."

Comment 7. If the above data issues would affect the results of the meta-analysis, please redo it.
Reply 7: This has been updated.
Changes in text: Figure 2 was re-performed based on changes to the sample sizes for Demir & Barman (2020 and Howard et al. (2021). The remaining meta-analyses did not need to be re-completed.

Comment 8. "Substantial improvements were noted in otomicroscopy, tympanometry, PTA and Valsalva manoeuvres when comparing pre-operative and post-operative results". Please discuss the heterogeneity results. Figures 4 and 6 show heterogeneity >50%, have the authors considered Meta-regression analysis or subgroup analysis to confirm the source of heterogeneity, as well as sensitivity analyses to determine the stability of the meta-merged results?
Reply 8: For most of outcomes, we found low to moderate heterogeneity except retraction and PTA. Although it would have been ideal to do moderator analyses, we could not perform meta-regression or sub-group analyses due to the small number of pooled studies for these outcomes.
Changes in text: The limitations section includes the following sentence: "Heterogeneity among objective outcome measures also poses a challenge. Due to small number of studies, moderator analyses could not be performed for otomicroscopic retraction and air-bone gap, both showing heterogeneity >50%."

Comment 9. We recommend including a separate section on the STRENGTHS and LIMITATIONS of this review to promote a more intellectual interpretation.
Reply 9: A 'Strengths & Limitations' section has been added to the end of the discussion.
Changes in text: The following text has been added to the end of the Discussion section: "As mentioned previously, this systematic review and meta-analysis is more comprehensive than previous publications, encompassing a larger number of objective and subjective parameters across a greater number of studies to strongly support the safety of BDET in children. This review provides a strong platform to setup prospective studies and randomised Controlled trial of BDET in children.

In spite of this, several key limitations hinder its ability to draw robust conclusions. Firstly, the reliance on a small sample size of only 11 studies raises concerns about the generalizability of findings, limiting the strength of any conclusions. Moreover, most of the included studies are retrospective case series without control groups or historic cohort studies, which are inherently

susceptible to bias and confounding variables. Additionally, the variations in follow-up periods across these studies create challenges when trying to assess the effect size at different time points, making it difficult to draw a clear temporal relationship between interventions and outcomes. Heterogeneity among objective outcome measures also poses a challenge. Due to small number of studies, moderator analyses could not be performed for otomicroscopic retraction and air-bone gap, both showing heterogeneity >50%. Lastly, the lack of a validated questionnaire to assess ETD-related symptoms and quality of life in children underscores the reliability of the subjective outcomes. These limitations highlight the need for more rigorous and standardized research in this area to draw more robust conclusions regarding the effectiveness of BDET in children."

Minor Comments

Comment 10. Title: One key piece of information is missing from the title: "Outcomes".

Reply 10: This has been updated.

Changes in text: The title now reads "Outcomes after Balloon Dilation of the Eustachian Tube in Children: A Systematic Review and Meta-Analysis".

Comment 11. Abstract: We recommend the authors simplify the abstract (200-350 words max). Authors may consider rewriting the Abstract-Conclusion rather than copying the Conclusion.

Reply 11: The suggestion to simply the conclusion has been taken.

Changes in text: The conclusion has been reduced to ensure the abstract's word count is less than 350 words.

Comment 12. Abstract-Methods: Please specify the methods used to assess risk of bias in the included studies and to present and synthesize results.

Reply 12: This has been updated.

Changes in text: The following text has been added to the Abstract-Methods: "Risk of bias was evaluated using the Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I) tool."

Comment 13. Abstract: Please specify the register name and registration number at the end of the Abstract, e.g., "PROSPERO registration: CRD42023430248".

Reply 13: This has been updated.

Changes in text: The abstract has "PROSPERO registration: CRD42023430248" at the end.

Comment 14. Introduction: "Tubomanometry, however, is a less common technique that measures the latency () of a pressure change transmitted to the TM when pressure is applied to the nasopharynx: < 1 indicates immediate opening, > 1 indicates delayed opening, and = 0 indicates no opening (Leichtle et al. 2017)". What about "=1"?

Reply 14: R = 1 indicates accurate opening.

Changes in text: The text now reads "R = 0 indicates no opening, 0 < R < 1 indicates immediate opening, R = 1 indicates accurate opening, and R > 1 indicates delayed opening (15)".

Comment 15. 1) Methods: "and the PRISMA guidelines were followed". We suggest the authors fill out and submit the "PRISMA 2020 Checklist" (https://cdn.amegroups.cn/static/public/12-PRISMA-2020-Checklist.pdf). The relevant page/line and section/paragraph number in the

manuscript should be stated for each item in the checklist. Here is an example for your reference: https://www.theajo.com/article/view/4638/rc. 2) A statement "We present this article in accordance with the PRISMA 2020 reporting checklist (availble at https://www.theajo.com/article/view/10.21037/ajo-23-38/rc)" should be included at the end of the "Introduction".

Reply 15: Apologies, that this was not uploaded earlier. We have attached the PRISMA checklist.

Changes in text: There have not been any changes to the text. The checklist has been uploaded separately.

Comment 16. Methods-Literature Search: Please report the date when each source was last searched or consulted (specified to date, month, and year).

Reply 16: This has been updated.

Changes in text: Within the Literature Search section of the Methods, "All searches were completed on 24 October 2023" has been added.

Comment 17. Methods-Literature Search: "Articles published prior to January 2023". This is not consistent with the timeframe "be published prior to 1 January 2022" in the registration, please confirm it. In addition, as far as the search is concerned, the 2023.1.1 time is too old. You need to update the search time to 2023.10

Reply 17: The search has been recompleted up to and including 24 October 2023. There have been no new papers that meet the inclusion and exclusion criteria to be included in the analysis.

Changes in text: The methods section and Figure 1 (PRISMA flow diagram) have been recompleted to reflect the updated search.

Comment 18. Methods-Selection: "involved patients under 18 years old". Suggest changing to "involved patients aged 18 years old and under" because Table 1 includes patients 18 years old.

Reply 18: This has been updated.

Changes in text: The phrase now reads "involved patients aged 18 years old and under" within the Selection section of the Methods.

Comment 19. Methods-Selection: "and no restriction was placed on language". Did the authors search in English and then check all other languages for studies that met the criteria? Please kindly give the databases and their complete search strategies as well.

Reply: The search was performed in English, but where the search yielded non-English papers, they were translated to English for analysis.

Changes in text: The Selection section of Methods now includes "non-English papers were translated to English through Google Translate."

Comment 20. Methods-Validity Assessment: The included studies were appraised by two reviewers using criteria such as risk of bias (randomisation, allocation concealment, blinding), loss to follow-up and selective reporting. The risk was valuated as high, low, unclear or not applicable (n/a) using the Cochrane Handbook. Please select the appropriate evaluation tool based on the type of study included. For example, cohort studies should use the NOS scale. In addition, the evaluation tools mentioned in Table 7 should be modified accordingly.

Reply 20: A general appraisal tool was used, but this has now been changed to the Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I) tool, which is the recommended tool for non-randomised studies, including cohort studies and case series, as per the Cochrane Handbook.
Changes in text: The Methods section now reads, "The included studies were appraised by two reviewers using the Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I) tool as described in the Cochrane Handbook. This tool evaluates the risk of bias related to (1) confounding, (2) selection of participants, (3) classification of interventions, (4) deviations from intended interventions, (5) missing data, (6) measurement of outcomes and (7) selective reporting. Each domain was valuated as low, moderate, serious, critical or no information. Discrepancies were resolved with discussion."
Changes in text: The Results section now reads: "An overview of the validity assessment is provided in Table 7. The overall risk of bias was evaluated to be low across all the included studies. However, several studies had moderate risk in one or more domains. Firstly, attrition over the follow-up period was moderate in three studies (15,16,21). This could not be assessed for Jenckel et al. (2015) and Oehlandt et al. (2012) as sample sizes were not reported over the follow-up period. Additionally, the risk of selective reporting was moderate for two studies. Oehlandt et al.'s (2022) study included adults and children, but only some of the outcomes were segregated for children. Toivonen et al. (2020), on the other hand, had a control group where they assessed the number of failed procedures and the 2-year failure-free probability, but no other objective or subjective outcomes were reported from this control group. No explanations were provided for the omission in either paper."
Changes in text: Table 7 has been changed to reflect the outcomes of the ROBINS-I tool.

Comment 21. Methods-Statistical Analysis: Please provide information on the software, packages, and version numbers used to implement synthesis methods, e.g., "metan in Stata 16, metafor (version 2.1-0) in R". Also, please provide the rationale for choosing the linear mixed-effects model.
Reply: The metafor package (Version 4.0-0) in R was used to conduct the meta-analyses. We used generalized linear mixed-effects model due to the repeated nature of the data (pre / post paired cohorts). The GLMM also incorporates an additional random component that has a variance component associated to it. By fitting GLMM, these variance components are estimated accurately.
Changes in text: The statistical analysis section has a sentence stating "the metafor package (Version 4.0-0) in R was used to conduct the meta-analyses." The rationalization of the linear mixed-effects model is also listed in this section as "For all outcomes, separate random-effects models were fitted using a conditional generalized linear mixed-effects model (GLMM). This framework accounts for the repeated nature of the data (pre/post paired cohorts), while also incorporating an additional random component that accurately estimates the associated variance components."

Comment 22. "values of 25%, 50% and 75% indicated low, moderate, and high heterogeneity, respectively". It should be a range. For example, "75%<I2≤100%: high heterogeneity".
Reply: This has been updated.
Changes in text: The statistical analysis section has the following "heterogeneity was examined using the Higgins I2 statistic, where low, moderate, and high heterogeneity were indicated by I2 values in the ranges of 0-25%, 25-75% and 75-100%, respectively."

Comment 23. Results-Figure 1: "Records identified from: Databases (n=6):". Suggest changing to "Records identified from six databases (n=303):". Also, please specify the reason for exclusions. E.g., "Records excluded (n=102): - no relevance to BDET in children (n = 78); - review-type studies (n = 24)".

Reply: This has been updated.

Changes in text: The 'Results of Search' paragraph in the Results section has been updated with the new search results. Figure 1 has been updated to reflect the recommendations above.

Comment 24. Results-Table 1: "If studies reported outcomes across multiple follow-up periods, the measure at 12 months or less was taken to ensure a sufficient sample size". Please report the follow-up periods in Table 1. Also, please report in Table 1 the information in the pre-formulated data extraction form that is relevant to the purpose of this review (e.g., department, setting). In addition, please explain the meaning of "Procedures": does it mean "the number of eustachian tubes"?

Reply 24: A column has been updated to include follow-up periods. The country is important, but the department and setting is not relevant. Procedures relates to the number of BDET procedures performed.

Changes in text: A column has been added to Table 1 for follow-up periods. The "Patients" and "Procedures" columns in Table 1 have been amended in Table 1, and the following clarification has been added to the caption: ""Patients" refers to the number of patients who underwent either unilateral or bilateral balloon dilatation of the eustachian tube (BDET), while "procedures" refers to the number of BDET procedures. For historic cohort studies, the total number of patients and procedures performed among the BDET and control groups is given in brackets".

Comment 25. Results-Table 3: There is no "satisfaction" in Table 3, did the authors combine satisfaction and quality of life?

Reply 25: The satisfaction data was combined with QOL for Demir & Barman (2020a) but not for the other studies.

Changes in text: Table 3 and 6 has been adjusted to add "satisfaction" as a new parameter.

Comment 26. Results-Table 6: The p-value for Demir & Batman (2020a) is blank.

Reply 26: This has been updated.

Changes in text: The p value of 0.018 has been added. Clarification as to what this refers to is also placed in the Result section: "OM-6 questionnaire scores were significantly reduced from pre-op (BDET 31 ± 5 vs VT 29 ± 4) at 6 mths (BDET 15 ± 6 vs VT 18 ± 7) and 12 mths (BDET 8 ± 2 vs VT 8 ± 2) ($p<0.001$). The reduction was greater in the BDET group at 6 mths (($p=0.018$) but not 12 mths ($p=0.510$)."

Comment 27. Missing the funnel plot results.

Reply: The funnel plots are included as supplementary figures.

Changes in text: See supplementary figures 1-6 for funnel plots corresponding to each meta-analysis.

Comment 28. Data should be consistent throughout the text. For example, "Higher incidence of normal TM in BDET (93%) vs VT (28%) group" (Table 5) and "(93.3%, n=30) than VT insertion (28.1%, n=32)" (in the text); "The estimated rate was 0.033 (95% CI: 0.020 to 0.057, $p<0.001$)"

(Figure 2) and “with the estimated rate of minor complications being 3.3% (95% CI: 2.0 to 5.7, p<0.001)” (in the text). Please check the entire manuscript to address similar concerns.

Reply 28: This has been updated.

Changes in text: Data has been amended across the text, figure legends and tables to ensure consistency.

Comment 29. 1) For ease of reading and double-checking, it would be better to add the reference number of the included studies in the tables. 2) Line 24 (…"equalise pressure (Schilder **et al**. 2015)."): Please add the reference number in this sentence. Also, the same goes for sentences that contain “**[name(s)] et al…**” 3) Ref No.28 is missing in the main text.

Reply 29: (1) and (2) have been updated. Ref 28 was included as Aboueisha et al. (2022) within the main text, rather than a reference number. However, this has now been updated.

Changes in text: References have been added to the Tables, and all citation has been changed to reflect the reference number. The missing reference in the main text has also been fixed.

Comment 30. "Only Jenckel et al. (2015) customised the adult applicator system": Should that be "pediatric applicator system"?

Reply 30: Jenckel et al. (2015) customized the original applicator system, which was built for adults. Hence, it was an adult applicator system.

Changes in text: The phrase has been changed to read “Only Jenckel et al. (2015) customised the applicator system for their pediatric patients,” for clarification.

Comment 31. "The timeframe for post-operative care varied among studies and is summarised in Table 1". It should be "Table 2".

Reply 31: This has been updated.

Changes in text: The sentence now reads “The timeframe for post-operative care varied among studies and is summarised in Table 2.”

Comment 32. "The pooled estimate of the rate of complication was 3.3% (95% CI: 2.0 to 5.7, p<0.001)": Should the p-value be 0.06? The p-values in Figures 3 and 4 also seem to be inconsistent with the descriptions in the text. Please confirm and refine it.

Reply 32: The p-value in the forest plots is for tests of heterogeneity and P > 0.05 means no significant evidence of heterogeneity. The estimated pooled effect is the one mentioned with the corresponding 95% CI in the figure description and main text.

Changes in text: No changes were made to the text.

Comment 33. Reference: It is recommended to cite references rather than just mention the author and year of the study. For example, "R > 1 indicates delayed opening, and = 0 indicates no opening (Leichtle et al. 2017)" should be "R > 1 indicates delayed opening, and = 0 indicates no opening (15)".

Reply 33: This has been updated.

Changes in text: The citation has been changed to the reference number.

Comment 34. Reference: "Maier, Tisch and Maier (2015) used Bielefeld catheters": It is

recommended to refer to the study in the text by mentioning only the first author and citing the reference, i.e. "Maier et al. (13) used Bielefeld catheters".

Reply 34: This has been updated.

Changes in text: Citations have been adjusted within the main text to reflect the above suggestions.

Comment 35. Reference: "and one study measured both (22)": It should be Ref. 21.

Reply 35: This has been updated.

Changes in text: The citation has been updated to reflect Toivonen et al.'s (2021) reference.

Comment 36. "The procedure was first reported in 2010". Please add the reference.

Reply 36. This has been updated.

Changes in text. There is now a citation for Ockermann et al. (2009), and the text has been changed to "first reported in 2009."