Comparative Effectiveness Research Review Disposition of Comments Report

Research Review Title: *Otitis Media With Effusion: Comparative Effectiveness of Treatments*

Draft review available for public comment from September 27, 2012 to October 18, 2012.

Research Review Citation: Berkman ND, Wallace IF, Steiner MJ, Harrison M, Greenblatt AM, Lohr KN, Kimple A, Yuen A. Otitis Media With Effusion: Comparative Effectiveness of Treatments. Comparative Effectiveness Review No. 101. (Prepared by the RTI-UNC Evidence-based Practice Center under Contract No. 290-2007-10056-I.) AHRQ Publication No. 13-EHC091-EF. Rockville, MD: Agency for Healthcare Research and Quality. May 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Comments to Research Review

The Effective Health Care (EHC) Program encourages the public to participate in the development of its research projects. Each comparative effectiveness research review is posted to the EHC Program Web site in draft form for public comment for a 4-week period. Comments can be submitted via the EHC Program Web site, mail or email. At the conclusion of the public comment period, authors use the commentators' submissions and comments to revise the draft comparative effectiveness research review.

Comments on draft reviews and the authors' responses to the comments are posted for public viewing on the EHC Program Web site approximately 3 months after the final research review is published. Comments are not edited for spelling, grammar, or other content errors. Each comment is listed with the name and affiliation of the commentator, if this information is provided. Commentators are not required to provide their names or affiliations in order to submit suggestions or comments.

The tables below include the responses by the authors of the review to each comment that was submitted for this draft review. The responses to comments in this disposition report are those of the authors, who are responsible for its contents, and do not necessarily represent the views of the Agency for Healthcare Research and Quality.

Source: http://effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1484
Published Online: May 4, 2013

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #9	General	In overall assessment it would not be greatly helpful to publish this work in its current form. It is likely to create, overall, more confusion than help for clinicians. It is also quite likely that a great deal of work would be needed to reassess the data and methodology prior to consideration of publication.	We hope that we have provided greater clarity through our revisions. In the revision, we have included more data from the original source studies.
Peer Reviewer #1	General	Page: Main report All pages The main evidence report is voluminous, and does a marvelous job of summarizing the current evidence. This will be very useful to researchers and policy makers Suggested change: Despite the wealth of information in the full report, there is no doubt that most readers will not get beyond the Structured Abstract or the Executive Summary. I have raised several points above concerning these that will not be repeated here. What I do wish to add, however, is the need to keep this as an "evidence report" and not make the mistake of giving management recommendations (which are for guideline developers, not systematic reviewers). The overriding importance of understanding the limited generalizability of the evidence and our lack of confidence in much of it is important to emphasize.	Thank you. We appreciate your cautious note. We have changed some of our language in the discussion chapter to be responsive to your concerns. We did not intend to give management recommendations. However, based on our review of the literature, we do offer our views on potential areas of future research and clinical concerns that we believe have not been settled through earlier studies.
Peer Reviewer #2	General	The report is extremely exhaustive and well written. The reviewers have done an excellent job of completely reviewing the available literature.	Thank you
Peer Reviewer #2	General	yes it is clear. Usable is a little harder because the document is very long. It is always frustrating when a well done study has conclusions that the presently available modalities of treatment are not very helpful but that does not mean the information is not useful	Thank you
Peer Reviewer #3	General	The report is very clinically meaningful and key questions are approriate. Overall, a well written document.	Thank you
Peer Reviewer #3	General	Well written, and easy to follow the flow.	Thank you
Peer Reviewer #4	General	This Comparative Effectiveness Review (CER) is generally quite comprehensive, providing much detail on a range of OME-related issues. However, regarding the issue that is the most fundamental and of greatest clinical and societal concern--namely, whether OME causes important harm to children and whether it requires any treatment at all--a series of reports providing much evidence appears to fallen through the cracks.	This was addressed in both the Introduction (page 1) and the Executive summary (page 1), where we acknowledged that the impact of OME on developmental outcomes may be doubtful but indicated that the universality of the condition and the high costs of treatment were justification for this review.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #4	General	The reports, from a long-term study and supported by NICHD and AHCPR, provided evidence that in children less than 3 years of age who have persistent middle-ear effusion, prompt insertion of tympanostomy tubes does not improve developmental outcomes as compared with watchful waiting and delayed insertion in the subgroup of children in whom effusion fails to clear spontaneously and continues unremittingly. The significance of the study's findings was discussed in an editorial accompanying the study's final report (Berman S. The end of an era in otitis research. New Engl J Med 2007;356:300-302): "These new findings . . . provide reassurance that developmental impairments that are not identified at an earlier age do not come into play later in the setting of greater academic challenges." Further details are discussed below in the second comment concerning page ES-10 of the CER.	We have added additional information in the report concerning this study, both a description of the study and specific outcomes. We have increased the transparency of all of the research studies that we included from the earlier systematic reviews.
Peer Reviewer #4	General	It would appear that omission of reference to the Paradise study findings in the CER might have been attributable to an assumption by the CER's authors that a previous Cochrane report by Browning et al. had described the findings, since the reports from the study were listed in the Browning et al bibliography. However, inexplicably, none of the actual findings were described in the Browning et al text.	We have added many additional findings in the revised report directly from the series of articles describing the findings of Paradise and colleagues about their study that compared tympanostomy tubes to delayed treatment. We have also revised the referencing approach that we use in summary tables in the results chapter so that readers can more easily identify and find this information.
Peer Reviewer #4	General	The CER conveys to the reader the impression that persistent OME in the first few years of life may have dire long-term consequences. Thus, "Although children 3 years or older may be able to tolerate a mild-moderate hearing loss for a period of 3-6 months or longer without risk to language outcomes, the effect of the same hearing loss on children 24 months or younger is unknown" (page 81), and " . . . a key clinical decision concerns the length of time that mild to moderate hearing loss needs to be present to negatively impact important outcomes . . ." (page 78). It would be regrettable for that expression of uncertainty and anxiety to be disseminated as part of a federally-sponsored report, since it would likely serve as justification for many unnecessary tympanostomy tube operations, and since there is now solid evidence that for most children with persistent OME there need be no fear of adverse long-term developmental effects.	We have modified the language as follows: Additional research needs to determine the appropriate criteria and waiting period before surgical intervention with children. Analyses by Paradise et al. suggest that mild hearing loss in preschool children for periods of up to 9 to 12 months does not affect subsequent speech or language outcomes. Whether toddlers are able to tolerate the same degree of hearing loss without risk to their language development is not known.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	General	The key questions are specific, appropriate, and clinically relevant. In particular there is a great need for evidence concerning effectiveness of treatments for OME in populations of children other than the "otherwise healthy" subjects who have been studied to date. As the authors suggest, evidence (e.g., Browning et al., 2010) that OME poses little or no threat to the development of healthy children cannot as yet be generalized to children for whom OME could conceivably be among a set of risk factors that collectively have adverse effects.	We agree with the reviewer and have included that in the introduction and discussion of the review.
Peer Reviewer #6	General	The report seems likely to be clinically meaningful with respect to clinical outcomes and utilization patterns for the interventions considered. However, as described below I believe that the analyses of some of the functional outcomes would benefit from additional clarification and that certain aspects of the Discussion are in need of revision.	Thank you. We address individual comments.
Peer Reviewer #6	General	The structure and organization are clear, but the report would be stronger if decisions, conclusions and interpretive statements concerning KQ 2 were more consistent with the available evidence.	It is not clear what available evidence the reviewer believes we missed. However, to answer KQ2 more thoroughly, we have added more detailed findings from the Paradise and other studies.
Peer Reviewer #7	General	Thank you for the opportunity to review this paper. This is a much needed review with the changes in health care. I would suggest that the authors include the paper why it is important to treat OME - why should there be a paper talking about the efficacy of treatment?	We have addressed this in the 2nd paragraph of the introduction. We have cited the high prevalence of OME among children, and indicated despite the uncertainty surrounding the impact of OME on developmental outcomes the condition has resulted in large expenditures for treatment. These factors suggest that it is important to examine the effectiveness of different treatments.
Peer Reviewer #7	Abstract	Based on the questions asked, the abstract is hard to follow, specifically the results. Headers need to be adjusted on the Executive Summary. See attachment.	We have reviewed the header structure in the Executive Summary. It is consistent with AHRQ requirements.
Peer Reviewer #7	General	Never addresses why it is important to treat OME – why is it important to study this disease if left untreated?	Please see comment above
Peer Reviewer #7	General	Future tense should be removed throughout the document examples, page 6, line 33 therefore , will not be included should be therefore, were not included or line 51, We will not duplicate their…should be We did not duplicate	The entire final report was reviewed by a copy editor prior to submission.
Peer Reviewer #7	General	Font changes in tables and they are inconsistent – check document for consistent size font as well as type of font example Table 3. Inclusion and exclusion criteria for studies of OME, page 12 and Table 29. Functional and health related quality of life outcomes	We addressed these specific problems and the entire final report was reviewed by a copy editor prior to submission.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #7	General	Maybe this topic did not make study inclusion but Otoacoustic Emissions can also be a tool to identify OME. If OME are absent that can be a sign of middle ear and/or cochlear dysfunction…were any of these papers reviewed? Serbetcioglu B, Ugurtay O, Kirkim G, Mutlu B. (2008). No association between hearing loss due to bilateral otitis media with effusion and Denver-II test results in preschool children. Int J Pediatr Otorhinolaryngol. Feb;72(2):215-22. Epub 2007 Nov 28. Hunter LL, Davey CS, Kohtz A, Daly KA. (2007). Hearing screening and middle ear measures in American Indian infants and toddlers. Int J Pediatr Otorhinolaryngol. Sep;71(9):1429-38. Epub 2007 Jun 27. Gravel JS, Roberts JE, Roush J, Grose J, Besing J, Burchinal M, Neebe E, Wallace IF, Zeisel S. (2006). Early otitis media with effusion, hearing loss, and auditory processes at school age. Ear Hear. Aug;27(4):353-68. Boone RT, Bower CM, Martin PF. (2005). Failed newborn hearing screens as presentation for otitis media with effusion in the newborn population. Int J Pediatr Otorhinolaryngol. Mar;69(3):393-7. Epub 2004 Dec 30. Yeo SW, Park SN, Park YS, Suh BD. (2002). Effect of middle-ear effusion on otoacoustic emissions. J Laryngol Otol. Oct;116(10):794-9.	This is not a study of diagnostic procedures but rather of treatments for OME; as such we did not review any papers concerning different methods of diagnosing OME.
Peer Reviewer #8	General	The Report is carefully done but nonetheless incorporates important errors and draws incorrect conclusions. There is a danger in building off the work of others and it includes the potential to reify the less than sufficient thought processes of others, and that is evident here.	We responded to specific concerns individually.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #8	General	While there are a number of errors or concerns that I will mention, the major one is the completely unsupported and false assertion of evidence to support adenoidectomy over non-surgical intervention. So far as I can tell from both the original source article and this EPC report, the cited Cochrane review that originally drew this conclusion (van den Aardweg) and that provides the major evidence in support of it included studies that compared adenoidectomy plus unilateral tympanostomy to unilateral tympanostomy. There was no non-surgical or watchful waiting arm. So the reasonable conclusion is that there is evidence of a small advantage with unknown clinical significance (at the level of effect size seen) of adenoidectomy over tympanostomy tubes but no direct evidence of advantage over watchful waiting.	The studies that the reviewer mentions are randomized by child and by ear. Therefore, they would allow for measures of adenoidectomy with TT (or without) and no adenoidectomy with tube (or without). The last group would be considered the no intervention group. We found limited evidence from one well-designed study - TARGET - where watchful waiting was a comparison group. The outcomes of this study were published during the period of peer review and we've included it in the report (see Table 20). Please note that we re-reviewed the original studies to ensure the accuracy of our summary based on the earlier systematic reviews.
Peer Reviewer #8	General	As I have not reviewed all the original material, I will simply cite the above as a real danger of incorporating others' reviews: even if the methods are exquisite, there is a reliance upon the judgment of others and the framing that occurs in its wake. This is not to say that the efficiencies gained are not worth the risk in aggregate, but to note that for any specific example, this is a risky methodology, albeit a well accepted one. I think this risk should be explicitly stated and it should be recognized as a weakness of all EPC reports that incorporate it. We are far less critical or examining of the risk for bias or other errors in those who synthesize the literature than for those who create it.	Because the reviews were conducted recently by well-regarded sources, we proposed to not replicate their work. We have also reviewed the original studies to abstract additional data so that we are not reliant on the original reviewers' determination of what data is important. We have not included the conclusions from the earlier reviews but their summary data.
Peer Reviewer #8	General	The exclusion of non-comparative studies of harms from surgeries is a serious mistake given that observational studies dominate the literature. Evidence of tympanosclerosis and associated (small) hearing loss with tympanosclerosis is at least moderate with consistent findings dating back at least 3 to 4 decades.	The focus of the review was on the comparative effectiveness of treatment options and because of this, we limited our literature to head-to-head comparisons. We were able to observe harms within this literature base. However, we included a caveat that acknowledges that we may have eliminated studies that examined harms (p. 105): "At the outset of the review, we established that we would only include head-to-head trials, including active monitoring. We recognize that by excluding single arm studies, we may have eliminated studies that examined important outcomes, particularly harms."
Peer Reviewer #8	General	The description of the methods and he genral quality of the writing are very good.	Thank you

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #8	General	I would have liked the subgroup mentions in the ES and explored in the literature to have acknowledged the specific ethnic predilection for OME mentioned in the body of the report.	We have included a sentence under KQ 4 subgroups that states that we wished to include studies that were focused on subpopulations at particular risk: "One of the explicit goals of this review was to examine treatment options for subgroups of patients including individuals defined by age groups and subpopulations at greater risk for OME such as individuals of American Indian, Alaskan, and Asian backgrounds, individuals with cleft palate, Down syndrome, and other craniofacial anomalies. Our search found very few studies of any subgroups that met our inclusion criteria nor did we find any studies that analyzed their data by the subgroups that we had targeted."
Peer Reviewer #9	General	There are some other conclusions - as pointed out in the attached document that are likely not to be clinically useful and may actually be detrimental to clinical care.	Each of the reviewer's specific comments are addressed separately.
Peer Reviewer #9	General	Creating an Appendix for excluded studies is helpful – there should be an Appendix for Included studies.	In compliance with the format required by AHRQ, all references included in the document are provided in the reference list but not in a separate appendix. However, we intend to make our database available for AHRQ's systematic review data repository (SRDR).
Peer Reviewer #9	General	I am very concerned about the overall studies that were used for the comparison of myringotomy with tube insertion and myringotomy alone and adenoidectomy. A number of these studies have significant overall flaws. I would be happy to delineate these if desired. However, as an example of one of the "better" studies. The study by Papova et al: 1. The patients average age is 5 years of age – this is an example of a need by the authors to determine which age group they are examining. Adenoidectomy for OME is likely to be more effective in older patients. Especially in this group a VERY high percentage of patients had nasal airway obstruction (80+%) and likely may have had some of degree of their difficulty related to adenoid obstruction. 2. The authors do not statistically handle the patients in this study correctly for those with recurrent OME 3. The study used a very short-acting tube and the article did not comment on the length of tube insertion for patients. 4. Overall there were a very small number of patients studied to make the conclusions that were made.	We have noted in the text the age groups of the children included in the studies. We agree that many of the studies are small, lack important information about the samples, are under-powered and do not state whether tubes are still in place at the time of the follow-up assessment.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #9	General	Very well constructed.	Thank you
Peer Reviewer #9	General	Some of the conclusions will be able to be used for policy and practice decisions.	Thank you.
Reviewer # 5	General	I appreciate the opportunity to review these guidelines which are obviously the product of much hard work. I've listed some comments below.	Thank you
Peer Reviewer #8	General	The Report is generally well structured.	Thank you
Peer Reviewer #9	General	General Comments: There are some clinically meaningful pieces to the report such as the ability of interventions to improve OME, the overall duration of those interventions, the delineation of some degree or likelihood of risks with certain interventions.	Thank you.
Peer Reviewer #4	Abstract	P. vi, line 30: Study by Paradise et al involved infants and young children aged 2 months to 3 years. For details, see second comment below regarding Page ES-10.	We have added additional detail concerning the Paradise studies.
Peer Reviewer #4	Abstract	P. vi, lines 37-38: tubes vs. watchful waiting during first 3 years of life resulted in no significant between-group difference in hearing levels at age 6 years (see Johnston et al, cited on page B-10 in the CER among the studies excluded.)	We have added the results of this report from the Paradise study to the review.
Peer Reviewer #4	Abstract	P. vi, lines 47-48: No mention of differences between tubes and watchful waiting, or of differences in other developmental outcomes or at other ages.	We have added the comparison group to the reporting of results.
Peer Reviewer #7	Abstract	P. iii, Preface: – a period is needed in line 30	Thank you. We have added a period.
Peer Reviewer #7	Abstract	P. vi, Structured Abstract: Objectives line 16 – and other treatment strategies (watchful waiting) – they only mention one treatment strategy – state what it is instead of using other.	We removed "other treatment strategies" from the sentence.
Peer Reviewer #7	Abstract	P. vi, Structured Abstract: When using KQ# - there are times there is a space between KQ and the # whereas other times where there is no space Line 32 (KQ 1) vs. line 49 (KQ3) or line 53 (KQ4) or line 55 (KQ5)	We fixed the spacing that there is a space between KQ and the number.
Peer Reviewer #7	Abstract	P. vi, Structured Abstract: Results section was difficult to follow	Thank you. We have rewritten the results section with an eye to clearly presenting our key findings.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #9	Abstract	I am quite concerned about the conclusion statement from the abstract," Tubes do not add any benefit to adenoidectomy in comparison to adenoidectomy plus myringotomy". Although the studies examined in this review may support this statement to a very modest degree there is a significant difficulty in the methodology of many of these studies as pointed out above including the types of tubes utilized, length of insertion time, length of time and intervals during which patients were followed and the types and ages of patients included in the studies. It would appear, based on some of the set-up and analysis, that not enough individuals with content expertise (a pediatric otolaryngologist for example) were involved in some of the design and crafting of this initial draft.	We have deleted that sentence from the conclusions in the abstract.
Peer Reviewer #5	Structured Abstract	Structured Abstract Abstract Objectives: For the first sentence, consider: "To compare benefits and harms of strategies currently in use for managing otitis media with effusion (OME)."	We have added the words "of strategies" to and replaced the word "treatment" with "managing" in the first sentence.
Peer Reviewer #5	Structured Abstract	Abstract Review Methods The results section confused me somewhat, because it refers to the key questions but the KQs are not directly described in the abstract. I think listing them briefly here would help to structure the rest of the results and make them easier to follow. Some of the results could be eliminated, or presented more succinctly if word count is an issue.	We have added the specific KQ's in the abstract results.
Peer Reviewer #5	Structured Abstract	Abstract Conclusions "However, there is evidence that adenoidectomy and myringotomy improve effusion and hearing more rapidly than myringotomy alone through 2 years post surgery." This is not presented in the results section of the abstract; I think it should be if it's in the conclusion	The result section had findings through 12 months: OME was more likely to resolve in children after adenoidectomy than in those with no treatment at 6 and 12 month followup (high SOE)." We added the following statement "Adenoidectomy and myringotomy were superior to myringotomy alone for reducing time with effusion and improving hearing at 24 months (low SOE)."
Peer Reviewer #1	Structured Abstract	P. vi, line 10: Sentence lacks a subject "To compare benefits and harms currently in use…" Suggested change: "To compare benefits and harms of interventions currently in use…"	We added a subject. " To compare benefits and harms of strategies currently in use for managing otitis media with effusion (OME)."

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Structured Abstract	P. vi : Structured Abstract. Nowhere in the structured abstract is the target population identified to which these results can be generalized; this should be stated clearly and early on. Suggested change: It needs to be made crystal clear that the studies in this CER almost exclusively relate to "otherwise healthy children" with otitis media, and CANNOT be generalized to children with otitis media an comorbid conditions that include developmental delays, speech/language delays or disorders, craniofacial anomalies, or syndromes that involve the head and neck region (e.g., Down)	We added a qualifier, "Generally, studies examined interventions in otherwise healthy non-infant pediatric populations."
Peer Reviewer #1	Structured Abstract	P. vi, line 52: This sentence is very confusing: "We were unable to disentangle results related to watchful waiting from myringotomy." Suggested change: I assume you mean that in assessing myringotomy results it was impossible to tell if what occurred was simply natural history or there was a real effect of myringotomy (which is unlikely). Reword as "Myringotomy did not offer any significant advantages over watchful waiting."	This phrase was written in relation to meta-analyses that were described by the authors as combining studies that compared tympanostomy tube arms to either watchful waiting or myringotomy. We have clarified that in some of these meta-analyses, the studies that were quantitatively synthesized were limited to tympanostomy tubes versus watchful waiting. We have eliminated the confusing sentence.
Peer Reviewer #1	Structured Abstract	P. vii, lines 14-15: You make a blanket statement that is not true: "Tubes do not add any benefit to adenoidectomy in comparison to myringotomy, and placement of tubes increases side effects. Suggested change: This statement, taken in isolation (as it stands) implies that clinicians should do adenoidectomy and myringotomy instead of placing tympanostomy tubes. There is a single randomized controlled trial (RCT) by Gates in 1987 that showed in a very select group of children (age 4y or older with chronic, bilateral OME) similar outcomes for tubes alone vs. adenoidectomy with myringotomy vs. adenoidectomy plus tubes. His conclusion was that tubes alone were the best first line intervention, but that adenoidectomy plus myringotomy was an alternative. This was not simply a trial comparing adenoidectomy plus myringotomy vs. adenoidectomy plus tubes. Moreover this is a single RCT with limited generalizability. Unless you can provide more context in the abstract, the statement as stands is misleading.	We have deleted this sentence from the abstract.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Abstract	P. ES-2, lines 53-54: KQ5, regarding pneumococcal vaccine is listed as a "key question," but is not even mentioned in the structured abstract. Suggested change: Add a statement to the structured abstract about your findings regarding KQ5.	We added a statement at the end of the first paragraph of the abstract, " We also asked two additional questions: what the comparative benefits and harms of these treatments are in subgroups of patients with OME, and whether the comparative effectiveness of treatment options is related to factors affecting health care delivery or the receipt of pneumococcal vaccine inoculation."
Peer Reviewer #1	Executive Summary	P. ES-1, lines 10-19: There are additional causes of OME Suggested change: You do not mention in this paragraph the 3 leading causes of OME in children: (a) viral upper respiratory infection, (b) sequelae of acute otitis media, and (c) chronic Eustachian tube dysfunction in childhood.	Added this change to the Introduction (page 1): " In addition to chronic dysfunction of the Eustachian tube, the leading causes for OME viral upper respiratory infection and acute otitis media (AOM)" and to the ES in a slightly different form (page 1, 1st paragraph).
Peer Reviewer #1	Executive Summary	P. ES-1, lines 42-44: The consistency of fluid in the middle ear is irrelevant in diagnosing OME. Suggested change: You state the MEE is "sticky or thick fluid behind the eardrum." MEE can also be thin, watery, serous, mucoid, purulent, mucopurulent, or just about any consistency possible. There is no need to mention anything here about fluid consistency	We removed the modifier of the type of fluid. The sentence now reads: Diagnostically, the core feature of OME is middle ear effusion (MEE)—i.e., fluid behind the eardrum in the middle ear.
Peer Reviewer #1	Executive Summary	P. ES-2, lines 5-6: Incorrect statement: "Given the natural history of OME including spontaneous resolution in most patients over time, clinical decisions are complicated." Suggested change: If it were simply a matter of waiting for "spontaneous resolution in most patients" then decisions would be remarkably easy, not complicated. The reason they are complicated relates to (a) difficulties in accurately predicting the natural history in a given patient, (b) a tendency for OME to recur even if it does initially resolve spontaneously, (c) the varying impact of OME in a given child on hearing, speech, language, cognition, school performance, and development, and (d) limited information from RCTs on managing OME in children with common comorbid conditions (developmental delays, Down and other syndromes, cleft palate and craniofacial disorders, concurrent sensorineural hearing loss, etc).	We have modified this section so it now reads: Despite recent practice guidelines and systematic reviews,[8, 13-20] the comparative benefits and harms of treatments and treatment strategies for OME are uncertain. The uncertainty stems from the lack of consensus regarding the clinical and functional outcomes of OME. The difficulty predicting the course of recurrence for individual patients, especially those with co-morbid conditions, is one factor that makes clinical decisions difficult. Secondly, the authors of the most recent systematic review of the natural history of OME[8] had mixed findings regarding the impact of early OME on later developmental outcomes. Although they concluded that children with early OME were at greater risk for subsequent conductive hearing loss, they were unable to draw strong conclusions about the effect of early OME on later speech and language development.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Executive Summary	P. ES-2, lines 28-31: I disagree with the statement that the "intent of the review was to cover the entire range of individuals with OME… Suggested change: This implies that the intent of the review was to draw conclusions generalizable to these populations, which presupposes a sufficient evidence base to do so. The intent of the review, as I recall from early discussions, was to *assess the generalizability, or lack thereof of existing evidence, particularly from RCTs* to these populations. The lack of generalizability is an important message, one that (as stated above) is not even mentioned in the structured abstract.	We have changed the sentence as follows: The intent of our review was to cover the entire range of individuals who have OME; in particular, we sought evidence specific to populations who have not been examined in past reviews, such as adults and children with special conditions such as Down syndrome, cleft palate, or existing hearing loss. A goal of the study was to look for evidence that was specific to subpopulations that had not been included in reviews in the past. Secondarily, we examined whether subpopulations were included in sufficient numbers in more global studies to reach any conclusions about these groups. However, we did not find evidence through either approach.
Peer Reviewer #1	Executive Summary	P.ES-5, lines 37-44: The inability to perform ANY pooled analyses in this review is very interesting. Suggested change: This contrasts with other published systematic reviews, and even contrasts with the Cochrane Collaboration, which is probably the most "conservative" group when it comes to pooling data. The implication here is that perhaps the criteria for pooling were overly restrictive, requiring a degree of precision in the data that does not exist and may never be achievable. Having only "qualitative" conclusions limits the utility of this report	We did not conduct any new pooled analyses but we include a number of pooled analyses that were conducted by the earlier Cochrane studies. We did not find many additional RCT studies and as a result, did not have new data to add to the existing quantitative syntheses.
Peer Reviewer #1	Executive Summary	P. ES-8, line 20-27: The lack of any quantitative information here really limits the utility of results. Suggested change: The quantitative outcomes that belong in this section are (a) change percentage of time with MEE and (b) change in hearing levels in decibels. Even if you could not pool the data, at least give a range of results and possibly some median values. Simply using "increased" or "decreased" is of little meaning.	We added quantitative information: "Tympanostomy tube placement decreased time with middle ear effusion by 32 percent at 1 year post surgery and by 13 percent through 2 years post surgery." In contrast, tympanostomy tubes only improved hearing through 9 months post surgery; hearing improved by 10 db at 4-6 months post surgery and by 4 db at 6-9 months post surgery The same information was added to the results section.
Peer Reviewer #1	Executive Summary	P. ES-9, lines 22-29: Thank you for providing quantitative results. The results would be more meaningful if 95% confidence intervals were added.	We added CIs to ES text where it was available. We provide CIs in the results section of the report where it was available as well.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Executive Summary	P. ES-11, lines 9-15: These broad, descriptive results offer no new information. Suggested change: The issue of harms/adverse events of tubes vs. watchful waiting are of major concern in clinical decision-making. Some quantitative information would be useful to included.	Generally, we believe that more broadly stated information is appropriate for the ES. We have included some of the key quantitative findings in the ES. In relation to these specific findings, based on limited evidence we were able to determine the direction of effect. For example, more otorrhea in ears with tubes versus in ears without. However, data was limited and therefore we could not determine the magnitude of the effect.
Peer Reviewer #1	Executive Summary	P. ES-11, lines 41-46: Just as important as what you DID find regarding subgroups (which is not much) is to mention what you did NOT find. Suggested change: This is the place to comment on the disturbing lack of information on subgroups most likely to have problems with OME, such as children with developmental delays, speech/language problems, syndromes, etc.	We have indicated that we could not find studies focused on subgroups (page ES-20, Research Gaps): The first area is to expand research in subgroups that were targeted in this review but for whom there was no evidence. This includes infants and toddlers who are developmentally vulnerable for language acquisition and for whom a mild conductive loss over a shorter period of time can be more detrimental than for older children. Children with craniofacial anomalies such as cleft palate and other developmental disorders including Down syndrome and sensorineural hearing loss have not been a part of most treatment studies. When we did find studies, they were excluded for reasons such as having no valid comparison group (e.g., case series with no comparator), including children with acute AOM, or the study was not available in English.. Additionally, there is only limited research on treatment efficacy in adults as we were only able to identify one study that concerned treatments for adults.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Executive Summary	P. ES-12, line 23-27: There is no mention here of how the duration of tube function (e.g., how long it remains in the tympanic membrane) relates to outcomes. Suggested change: This is an important point. You use the word "tubes" here as if to imply there is only one kind of tube and that it will yield the outcomes stated. The proper terminology would be "tubes used in included studies," since in most of the studies (in the Cochrane review, for example), these were short-acting Sheppard type tubes, that tend to extrude after only 6-8 months. Obviously, one would not expect a benefit of persist AFTER the tube was gone. In contrast, most tubes used in the US are Armstrong-type types, which have a median intubation period of about 14-16 months. There needs to be some acknowledgment of how the generalizability of the results stated here depends on using tubes similar to the ones in the included studies.	We have examined short v. long term tubes in terms of benefit and harms. We have included what we found in the Executive Summary and in the Results section. What we wrote in the Executive Summary follows: ES 10: Length of tube retention was higher in the longer-term TT. Other TT comparisons and endpoints differed across studies. Because of sparse data, the diversity of comparisons, and inconsistent findings, the evidence is insufficient for comparisons of other design features or for hearing outcomes. OME recurrence was mixed in the shorter term versus longer term TT(insufficient strength of evidence ES-13: Otorrhea rates differed by tube type, with placement of longer-term TT related to a higher probability of otorrhea (low SOE). As suggested, we have now made the change that adds the following qualifier: "TT used in the included studies" We have also indicated in the discussion that information regarding tube type was often not included in the report.
Peer Reviewer #1	Executive Summary	P.ES-13: Table A This is a very helpful table, with good quantitative information. Suggested change: Add 95% confidence intervals to the outcomes for autoinflation (lines 47-49, column 3)	We have inserted confidence intervals for autoinflation into the table
Peer Reviewer #1	Executive Summary	P. ES-14, line:14-27: Generalizability not addressed. Suggested change: Again, the generalizability of these "negative" findings needs to be qualified by the type of children studied (e.g., otherwise healthy without any comorbid conditions that would make it "unethical" to randomized them).	We have added the following language to be clear about all of the evidence across the KQs. "Except where otherwise noted, across KQs, studies were generally limited to otherwise healthy children".
Peer Reviewer #1	Executive Summary	P.ES-15: Table C Utility of this information is severely limited by the absence of any quantitative data Suggested change: Try to add some actual numbers showing prevalence or incidence to put this in perspective	The findings concerning harms were sparse. They are included in the results section of the report. We do not believe that data on prevalence of specific harms can be determined from our findings.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Executive Summary	P. ES-15: Table C All of the evidence strength in column 4 is "low" Suggested change: I realize you have followed strict criteria in how this was determined, but your outcome defies common sense and conventional wisdom. I am not aware of any clinician who would disagree with the statement that tympanostomy tubes result in structural changes to the tympanic membrane (sclerosis, atrophy, retraction, performation), which, although non-serious, are clearly much more common than for ears treated without tubes. Much of the best data for this comes from long-term follow-up of cohorts from RCTs. When you state there is "low" strength of evidence for harms from tubes vs. watchful waiting it is hard to understand how. You need to clarify this in the text.	Due to the reviewer's concern about our presentation of harms, we re-reviewed all of the studies to ensure that we had included all of the harms that had been reported by the study authors. We include that information in our KQ3 results section. Because many of the studies were small, harms data is sparse. We conducted SOE grading based on the information that we found. We believe that the reviewer is correct in relation to several harms of tubes versus no surgery and we increased the strength of evidence as "moderate" in relation to otorrhea and tympanosclerosis. We did not find sufficient evidence to increase the SOE grades beyond "low" for any other harms.
Peer Reviewer #1	Abstract	P. ES-16, line 9-1: FINALLY you comment on the limited generalizability of this evidence. This belongs up front and center in the report, certainly in the Abstract. It should not take until page 16 to make this clear.	We have added this to the abstract.
Peer Reviewer #1	Executive Summary	P. ES-17, line 3-6: The wording of this statement is deceptive, and implies equipoise of harm vs. benefit, despite the differing SOE. Suggested change: The SOE for the various meta-analyses and RCTs for the benefits of tubes mostly have and SOE that is "moderate" or "high." In contrast, the SOE for all of your adverse event comparisons is "low." This is not equipoise and does not allow, or support a simple statement (as you have done here and in the abstract) to the effect that "There are benefits…however, they are offset by harms." If our confidence in benefits is higher, this needs to be reflected in the wording.	In the concluding paragraph of the ES, we have clarified that we have strong evidence of benefit from tubes but weaker evidence relating to harms.
Peer Reviewer #4	Executive Summary	P. ES-1, line 12: Not clear which listing of references is being used. Would help to alert readers that references in this section are those listed at the end of the Executive Summary, not the set of references listed subsequently beginning on p. 85.	We added a footnote at bottom of the first page of the ES indicating the location of the reference list for the ES.
Peer Reviewer #4	Executive Summary	P. ES-3, line 43: The 5 reviews should be referenced.	We added reference numbers.
Peer Reviewer #4	Executive Summary	P. ES-4, line 22: PICOTS not previously spelled out	We spelled out the definition.
Peer Reviewer #4	Executive Summary	P. ES-4, line 25: Again, the 5 reviews should be referenced.	We added reference numbers.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #4	Executive Summary	P. ES-4, lines 25-27: Language not clear	We modified the language so that it makes better sense. It now reads: "We included the five systematic reviews identified during topic refinement as well as eligible studies not included in those reviews, which included randomized controlled trials (RCTs), nonrandomized controlled trials and cohort studies."
Peer Reviewer #4	Executive Summary	P. ES-5, line 50: References #18 and #19 are identical. Numbering of references in text skips from #18 to #20 DELETE #19 IN REFERENCE LIST AND ALL OTHERS WILL NEED TO BE MOVED UP ONE IN LIST AND IN TEXT	Thank you. We have revised the reference list in the ES.
Peer Reviewer #4	Executive Summary	P. ES-7, lines 14-34: Numbers in flow diagram don't add up; 4798 - 4037 = 761 (not 750); 714 + 49 = 763 (not 750); 25 + 23 = 48 (not 49). Clarity could be improved by repeating the word "wrong" before each item in the "Full-text articles excluded" box.	We have revised the PRISMA flow diagram based on our updated search.
Peer Reviewer #4	Executive Summary	P. ES-10, lines 14-15: States that functional outcomes are described under KQ1--where? In executive summary (I can't find) or full text?	We added a phrase to further clarify the meaning that the description of the studies themselves (not the outcomes) is presented in KQ1: "The studies that are included to address KQ2 are described under KQ1."
Peer Reviewer #4	Executive Summary	P. ES-10, lines 17-24: Major omission here of relevant evidence in reports referenced, but not detailed or discussed, in the Browning review. The reports, by Paradise et al, detail long-term developmental outcomes of tube insertion vs. watchful waiting for persistent OME developing during the first 3 years of life. Followup covered 8 or more years post-treatment, with evaluations of children at ages 3, 4, 6, and 9 to 11 years. The outcomes, involving a total of 119 outcome measures, variously concerned speech, language, cognition, auditory processing, attention and impulsivity, academic achievement, behavior, psychosocial function, literacy, phonologic awareness. The study was supported by NICHD and AHRQ. The reports are mentioned briefly in the Browning meta-analysis and are listed in the Browning bibliography, but most of their findings are nowhere actually described. The reports are listed on page B-11 of the CER as having been excluded from the CER because of prior inclusion in a review, presumably the Browning review.	We have expanded our description of the Paradise study and our discussion of the findings in the revised report. Many of these findings are discussed in relation to KQ2 (functional outcomes) in the ES.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #4	Executive Summary	P. ES-11, lines 11-14: I believe that SOE regarding tympanosclerosis in tubes vs. watchful waiting was strong, as summarized in Table 3 in the report by Johnston et al that is cited on page B-10 among the studies excluded. There is no reason to anticipate that otorrhea would occur in children managed with watchful waiting, whereas there is strong evidence that otorrhea is a frequent complication of tubes in young children (see Ah-Tye et al, Pediatrics 2001;107:1251 and also references in that report).	Thank you. We appreciate that otorrhea would not occur in children who do not have tubes. It may occur in children in the watchful waiting group because some of these children eventually get tubes. We have added the Johnston study to our review.
Peer Reviewer #4	Executive Summary	P. ES-14, lines 15-19: See second comment above re page ES-10	We have reviewed the Paradise study for additional outcomes including harms. If we found harms in just one study we generally consider that evidence to be insufficient until it has been replicated in at least one additional study.
Peer Reviewer #4	Executive Summary	P. ES-14, lines 48-53: See comment above re page ES-11	I repeat the response here that we provided above. Thank you. We appreciate that otorrhea would not occur in children who do not have tubes. It may occur in children in the watchful waiting group because some of these children eventually get tubes. We have added the Johnston study to our review.
Peer Reviewer #4	Executive Summary	P. ES-18: Following two pages are confusingly numbered ES-2 and ES-3	Thank you for noting that problem. We have checked and fixed all of the page numbering
Peer Reviewer #7	Executive Summary	The headers are difficult to follow/not formatted correctly.	We have had a copy editor review the entire report prior to final submission.
Peer Reviewer #7	Executive Summary	Never addresses why it is important to treat OME – why is it important to study this disease if left untreated?	We have addressed this by citing the universality of the condition and the high expenditures on treatment.
Peer Reviewer #7	Executive Summary	Page ES-4: line 44 had two periods	We removed the second period.
Peer Reviewer #7	Executive Summary	P. ES-4: line 26, abbreviates NRCTS but then never references it again but spells it out everywhere else in the entire article (p ES-6, line 34, line 41;page 21, line 48, page 22, line 16, page 71, line 33, etc.)	We have removed the abbreviation from the ES

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #7	Executive Summary	P. ES-5, line 22: what is the name of the "instrument"?	We have changed the language as follows to be more specific: The risk of bias assessment was conducted using two tools, one appropriate for trials based on the Cochrane risk of bias tool and modified by our EPC to be used to evaluate observational studies (including instructions to reviewers that some questions concerning trial study design would be considered not applicable) and AMSTAR, appropriate for systematic reviews.
Peer Reviewer #7	Executive Summary	During the Executive Summary, they mention two reviewers reviewed for various items (risk of bias, study selection, strength of body of evidence, etc and when there was conflict a third person what brought it. Were the same two people used for each section of the review? Was the third person the same person each time who resolved the conflict? Also, the body of the paper they never mentioned how many they agreed upon and how many times the third person had to be brought in to resolve the conflict.	We have elaborated the review and abstraction process to indicate that there was a group of 6 trained reviewers and 6 abstractors who were involved with the task and that different pairs of reviewers were responsible for this task. The entire team took part in this process.
Peer Reviewer #7	Executive Summary	P. ES-10, line 48: Results page, Key Question 3, Tube versus tube studies – please clarify	We have clarified that this was comparisons between different types of tubes.
Peer Reviewer #7	Executive Summary	All pronouns referring to the authors should be removed – there were hundreds of "we" – they should all be removed.	Our style guide encourages us to use active voice which necessitates the use of "we."
Peer Reviewer #7	Executive Summary	P. ES-17, line 3 – followup should be follow-up Paper	The AHRQ style guide requires us to use "followup" not "follow-up."
Peer Reviewer #8	Executive Summary	I was less happy with the executive summary than I was with the document as a whole. It included important errors and omissions, as well as more modest imperfections. In the executive summary I had the following concerns:	We addressed each specific concern.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #8	Executive Summary	P. ES- 1: Prevalence. last sentence, the word but seems highly inappropriate suggesting opposition to an earlier construct that is artificial. OME is characterized by recurrences and spontaneous remissions, is for many children a dynamic and not a constant state and the notion of cases lasting more than a year implies a constancy that is probably only true for a moderate fraction of those who are cited as having it for a year or longer. There are characteristics of thickened fluid, the so called glue ear, that is less likely to remit and recur, but that is not the majority of children with OME. I found missing from the literature cited for this area Teele (1983 JAMA if I recall) and more so Paradise (Pediatrics, 1997) that are critical studies of the prevalence and natural history of this disorder. I don't think the condition was accurately or sufficiently portrayed, nor was it raised the extent to which middle ear effusion is a normal rather than an abnormal condition (i.e. an incidental finding rather than a disease). This conceptualization becomes central to considering what are meaningful or valuable clinical outcomes.	In the prevalence section we state: "Many episodes of OME resolve spontaneously within 3 months, but 30 to 40 percent of children have recurrent episodes and 5 to 10 percent of cases last more than 1 year".
Peer Reviewer #8	Executive Summary	P. ES-1: Diagnosis. OME is diagnosed with evidence of any fluid, not only sticky or thick. This is a small percentage of effusions and a larger proportion of the pathology, but to suggest that all OME or middle ear effusions are characterized by thick or sticky fluid is incorrect.	We have removed this qualifier. The sentence now reads: "Diagnostically, the core feature of OME is middle ear effusion (MEE)—i.e., fluid behind the eardrum in the middle ear."
Peer Reviewer #8	Executive Summary	P. ES-2: Scope and key questions. There is a statement that is either wrong and contradicted later in the report, or just badly worded" "which have been extensively reviewed previously..." appears to modify antibiotics as well as antihistamines and decongestants, which it should not. As noted later the reason for excluding antibiotics was an upcoming Cochrane review and an efficiency decision, not settled science. Indeed the 1994 3 academy guideline concludes that there is a number needed to treat of about 7 to eliminate one first effusion which suggests a real if limited benefit	This is correct. The Cochrane review of antibiotics as a treatment for OME was published in September 2012 and this is acknowledged in the text. It currently says: Antibiotics are the subject of a Cochrane review that was published in September 2012 after the deadline for including new reports.
Peer Reviewer #8	Executive Summary	P. ES-6: Literature Searches. the exclusion criteria of unknown time with OME at presentation seems wrong to me, since that is often encountered in the real world. The previous state of the ear may well have been unexamined or unreported or misreported and there is a need to make clinical decisions in the context of that as an unknown. Hence such studies may add value without introducing serious bias. They represent a model for effectiveness rather that efficacy studies.	We adopted a 3 month criteria for time with OME prior to randomization. This was based on clinical guidelines (AAP, 2004) that recommend a period of more than 3 months with OME before treatment. However, we did not exclude studies solely based on unknown time with OME at presentation. Rather, we excluded studies that did not adequately define their population with respect to a number of criteria since lack of information made it impossible to examine comparability between treatment groups.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #8	Executive Summary	P. ES-9: As noted above so far as I can tell from the van den Aadweg study there was no group that did not get treatment. All were subjected to one surgery or another and hence the conclusions are disturbingly misleading and in error. General anesthesia and whatever may be the aftermaths of dealing with the other ear surgery are not at all the same as no intervention or watchful waiting. A serious error is perpetuated in the current draft.	It is true that the preponderance of the evidence is from "by ear" studies. That may limit differences in the harms presented in the original studies. However, we do not believe that it would limit the review of benefits.
Peer Reviewer #9	Executive Summary	The sentence, "Diagnostically, the core feature of OME is middle ear effusion (MEE)—i.e., sticky or thick fluid behind the eardrum in the middle ear" is imprecise. It is well known, now, that the primary cause of OME is mucoglycoproteins which cause the hearing loss and much of the fluid. (Int J Otolaryngol. 2012;2012:745325. Epub 2012 May 22. Mucin production and mucous cell metaplasia in otitis media. Lin J, Caye-Thomasen P, Tono T, Zhang QA, Nakamura Y, Feng L, Huang J, Ye S, Hu X, **Kerschner JE**. And Laryngoscope. 2007 Sep;117(9):1666-76. Mucin gene expression in human middle ear epithelium. Kerschner JE.)	This has been fixed in the ES. The core feature of OME is middle ear effusion (MEE)—i.e., fluid behind the eardrum in the middle ear space over a period of time, commonly 3 or more months.
Peer Reviewer #9	Executive Summary	This point made on Page 5 of the Executive Summary – "Because we determined that quantitative analyses were not appropriate, we did all analyses qualitatively. Evidence used in the synthesis included the results from the earlier meta-analyses, additional data from individual studies contained in those systematic reviews, and data from the articles included from our own searches. "is very important and this should be highlighted in some fashion even more so than it currently is.	We have highlighted our approach by stating it in the abstract, executive summary and methods section of the main report.
Peer Reviewer #9	Executive Summary	P. ES-8: This sentence gave me a great deal of pause as I read it and suggested a possibility of bias in the authors? "In contrast, tympanostomy tubes only improved hearing through 9 months post surgery (high [4 to 6 months] to moderate [6 to 9 months] SOE". The use of the words "In contrast" and "only" carry substantial editorial connotation and really should be avoided. This sentence would read much better with just the facts. "Tympanostomy tubes improved hearing through 9 months post surgery…"	We modified the sentence as suggested.
Peer Reviewer #9	Executive Summary	The summary is lacking in a discussion or comment regarding the average length of tube insertion for the tympanostomy tube studies. The overall length of tube insertion is the defining piece of data, that if not assessed or taken into consideration, impacts all of these studies the greatest.	We have included greater detail concerning short-term versus long-term tubes in the results section of the main report. We did not have sufficient information concerning length of time of tube insertion to make this the defining piece of the data. Studies varied in the types of tubes were used and children varied in the length of time tubes were retained.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #9	Executive Summary	In the conclusions the word, “multiple” should be excluded from the sentence, “However, tube placement also increases the rate of multiple side effects ….” The word multiple is imprecise and does not add to the sentence where examples are used.	We removed the term “multiple”
Peer Reviewer #9	Executive Summary	ES-8: KQ1 Tympanostomy Tubes Plus Adenoidectomy Versus Myringotomy Plus Adenoidectomy or Adenoidectomy Alone We identified eight studies examining outcomes in relation to tympanostomy tubes plus adenoidectomy as compared to myringotomy plus adenoidectomy or adenoidectomy alone.29-36 Three of the studies compared tympanostomy tubes in one ear to a control ear in children who all had adenoidectomies. One of the studies was an RCT,29 and the other two were nonrandomized controlled studies.30, 31 The other five studies compared tympanostomy tubes to myringotomy, among children who all had adenoidectomies; four were RCTs,32-35 and one was a nonrandomized control study.36 The evidence was insufficient for examining OME signs and symptoms, with only two single studies examining middle ear effusion and recurrence of OME. We found no differences in hearing at any endpoint in five studies between tympanostomy tubes and myringotomy among children who also received adenoidectomies (low SOE). We found mixed results for tubes compared with watchful waiting in children who also received adenoidectomies (insufficient SOE). - The data and methodology for this paragraph as well as all of the studies included in making the conclusions in this paragraph need to be much more closely examined before publication. o Many of the studies utilized in the analysis have some significant flaws o Again, it is critical to comment, when making a conclusion as this, many other factors _ Age _ Patient selection for the study and other associated conditions _ Types of tubes used _ Average duration of TT insertion _ Type of follow-up _ Potential causation of tube failure • Plugging of tubes and how that impacted data o The authors should understand that this conclusion is contrary to the majority of clinical practice currently employed today. The relatively poor evidence and studies used to “refute” common practice will be readily transparent to those who care for these children and will likely invalidate – in the minds of many, some of the very good evidence in this report.	A greater level of detail is provided in the full report. We appreciate the reviewer’s detailed knowledge of the studies. We have included many of these details (to the extent they were available) in the main body of the report. . We have also noted in the limitations section the deficiencies of many of the studies that were included.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #9	Executive Summary	ES-10: KQ2 - It would appear that the report misses a significant opportunity on QOL studies to report on the overwhelming evidence that in children who undergo myringotomy with tube insertion, for a variety of otitis media conditions, that there is a significant improvement in QOL. Although many of these studies do not have a comparison group – the purpose of this report is to evaluate interventions for OME. Many of these are very good studies, using excellent QOL tools and have consistently demonstrated an enhancement if QOL for TT insertion.	Quality of life was an outcome of interest. We provided QOL data from the 2 studies where it was reported (see Table 29 pages 73 in Results and Table 30 pages 74 in Results) – Rovers and colleagues, (2000) that was in the Browning systematic review[17] and the Vlastos and colleagues study.[52] The Rovers and colleagues study failed to find quality of life improvements as measured on the TAIQOL at 6 and 12 months post intervention between the group receiving tubes and those in watchful waiting. The Vlastos and colleagues study did not find any differences between groups receiving tubes plus adenoidectomy and those receiving myringotomy plus adenoidectomy on quality of life as measured by the OM-6. We thank the reviewer for drawing to our attention the non-comparative studies that have included quality of life measures. As noted, we did not include such studies. At the outset of the review, we established that we would only include head-to-head trials, including active monitoring. This decision was made with input from our technical experts. We recognize that by excluding single arm studies, we may have not included studies that examined important outcomes, including quality of life.
Peer Reviewer #9	Executive Summary	ES-14: I have concerns that the authors stated that, “we found no evidence concerning harms from adenoidectomy”. A number of the studies utilized in this review did not evaluate for harm or effectively look for this. It is well documented that in general populations, although rare, adenoidectomy is associated with an incidence of velopharyngeal insufficiency. The report might suggest that in other broader studies (and those with better methodology) looking at this procedure there exists evidence of harm. Perhaps outside of the scope of the report – but in a report of this type I think it is important to acknowledge what other issues are likely considered by the practitioner.	We have added additional data that we found in our studies on harms from adenoidectomy. In particular, hemorrhage. Please see our response above which also eliminated one-arm studies that may have examined harms as a result of adenoidectomy.
Peer Reviewer #5	Executive Summary	Diagnosis of OME Does the effusion have to be ‘sticky’ or ‘thick’, doesn’t just plain old fluid count as OME?	We removed the words “sticky” or “thick” in the ES (page ES-1).

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	Executive Summary	The Introduction is concise and clear with respect to the nature, hypothesized pathophysiology, diagnostic criteria, and course of OME. Minor suggestions for the Executive Summary (p. ES-1): clarify that effusion viscosity varies and specify that the average conductive hearing loss during episodes of OME is approximately 10 dB (as noted on p. 2).	We removed the types of fluid as a qualifier in the first sentence under Diagnosis. We added information regarding the degree of hearing loss "that measures at 25 decibels (dB) or an elevation of 10 dB relative to that of children with normal hearing" to the sentence that states that children with OME have a conductive loss (last sentence in the Diagnosis paragraph).
Peer Reviewer #2	Introduction	well written. The overall outline of the work into the reviewed categories were well explained	Thank you
Peer Reviewer #4	Introduction	P. 2, lines 15-16: Statement is dubious in light of recent work. See comments above. Should at least qualify "hearing loss" with "severe, protracted."	As recommended, we qualified hearing loss as severe and protracted.
Peer Reviewer #4	Introduction	P. 2, line 26: Presence of bubbles indicates an air-fluid mix and constitutes the mildest form of OME, usually without hearing loss. Far more important as diagnostic indicators (although often less obvious and more difficult to ascertain) are opacification and impaired mobility of the tympanic membrane	Thank you for noting this. We have changed the content and wording appropriately.
Peer Reviewer #4	Introduction	P. 2, line 55: I suggest that reference to CT be deleted. MEE may be visible incidentally on CT, but it would be unthinkable to use CT specifically for diagnosing MEE.	We have eliminated this mention of CT scans.
Peer Reviewer #4	Introduction	P. 2, lines 25-39: See comments re page ES-7.	The comment related to ES-7 concerns the article inclusion flow diagram. The information on page 2, lines 25-39 concerns diagnosis of OME. Therefore, we are unable to determine the reviewer's concerns on this page.
Peer Reviewer #5	Introduction	Background, line 12 Definition of OME I think it should read: "Otitis media with effusion (OME) is defined as a collection of fluid in the middle ear without signs or symptoms of acute ear infection." I think that's how it's defined in the AAP OME guidelines.	We added the word "acute" to the definition in the intro as well as ES.
Peer Reviewer #5	Introduction	The end of the paragraph should say something about functional outcomes, such as "…and can sometimes lead to a 'fullness' sensation in the ear, pain from pressure changes, and/or decreased hearing, with the potential for related developmental language delay."	We have edited the introduction to include a consistent approach to sequalae of OME.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #5	Introduction	P. 7-9: Scope and Key Questions Is hearing a clinical or functional outcome? It's listed in Figure A, and treated in the discussion, as both. Maybe this should be addressed specifically.	We have revised the key questions to more clearly state that hearing is a clinical outcome (objectively measured) and a functional outcome when measured as perceived hearing level based on patient and/or parent report.
Peer Reviewer #7	Introduction	P. 9, line 52: All pronouns referring to the authors should be removed – there were hundreds of "we" – they should all be removed along with "our"	Our style guide strongly discourages the use of passive voice and encourages us to use active voice which necessitates the use of "we."
Peer Reviewer #8	Introduction	The same criticism as above regarding the natural history of OME, although the Paradise study is cited. Ratio of secondary to primary sources too high in Introduction for my judgment.	We adjusted the discussion of natural history as suggested.
Peer Reviewer #8	Introduction	P. 1: "will be affected more dramatically" is probably not correct and certainly off in tone. The marginal consequences of small hearing loss may be greater in these children.	We have modified this to read: "In addition, children with existing hearing loss will experience poorer hearing thresholds as a result of the secondary conductive hearing loss that occurs with OME."
Peer Reviewer #8	Introduction	P. 2: I was unnerved by the statement attributed to a Clinical society Guideline (reference 12). It is directly contradicted by the Paradise clinical trial reported in the NJM and elsewhere and it requires a whole lot more support than an assertion in a review or guideline. I don't believe it and it is irresponsible to state it as fact in the background/introduction section of this document.	We have rephrased the statement to clarify that the concern is with the role of hearing loss on these outcomes. We do not state that hearing loss is caused by OME and in the results section of the report, we discuss the Paradise and other finding. Because protracted hearing loss in young children may delay or permanently change their communication skills and may lead to behavioral and educational difficulties,[13] clinicians and others are concerned about the possible role of OME on these outcomes.
Peer Reviewer #8	Introduction	P. 2: Diagnosis. More jarring writing: eliciting a history of x,y, and z is critical is absurd. The presence or absence of any or all of these factors does not make OME unlikely. These may all be provocative of consideration of OME, but the phrasing is way overstated and clinically nonsensical. And grouping URI with Down syndrome or cleft palate is also wrong. A pediatrician should help rewrite this section.	A pediatrician changed the content and writing of this section. We limited our discussion to tympanoscentesis, pneumatic otoscopy, and hearing loss; we also indicated that it must be distinct from AOM. We have deleted the language concerning taking a careful history and that it is critical.
Peer Reviewer #8	Introduction	P. 6: Antimicrobials first statement is false according to NAMCS data analyzed and referenced in the AAO-HNS guideline. It's overstated and not in consideration of the mixed evidence that is later described.	We have emphasized that there is conflicting evidence on efficacy of antimicrobials in treating OME.
Peer Reviewer #2	Methods	The definitions and various categories reviewed are clear and logical. The inclusion and exclusion of studies were logical and well explained	Thank you

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #3	Introduction	Overall, good inclusion and exclusion criteria and definitions except the definition on page 34, describing hearing aids as this definition sounds like a specific style of hearing aid (an in-the-ear hearing aid definition) instead of a generic definition of hearing aids.	We have modified this definition to be: A small electronic device that amplifies sound, worn behind the ear (children and adults) or placed into the external ear canal (adults).
Peer Reviewer #6	Methods	The search methods, inclusion and exclusion criteria were clearly stated although as noted below I questioned whether they were applied consistently. The outcome measures appear to be appropriate. No statistical methods were employed; a more explicit description of the methods used in synthesizing results qualitatively would be welcome.	We have revised the sentence to provide additional explanation of the qualitative synthesis process. Thus, we did all analyses qualitatively, based on our reasoned judgment of similarities in measurement of interventions and outcomes and homogeneity of patient populations.
Peer Reviewer #8	Methods	Methods: I have commented some concerns above. I think observational studies are often the best for estimating prevalence either of illness or of harms from a given treatment. I think the acceptance of the methods and conclusions of previous literature syntheses introduces risks that in the present case have bitten the authors badly creating serious errors in conclusions.	We included observational studies in our inclusion criteria, but most were of too poor quality to be informative for either benefits or harms.
Peer Reviewer #8	Methods	Search strategies are generally logical and well described, even where I don't always agree with the judgment's they manifest.	Thank you
Peer Reviewer #9	Methods	I believe the overall criteria are justifiable. However, the criteria did not necessarily always lead to the selection of good evidence. For instance, just because a study is a RCT, if the trial is done with poor controls, poor patient selection, questionable follow-up or definitions of the patients enrolled the quality of evidence is likely to be flawed and skew conclusions.	We agree and because of this, we have rated the risk of bias of each included study and systematic review. That information is available in the report.
Peer Reviewer #9	Methods	Statistical methods are appropriate but the portions above probably needed some additional help with selection of data to be analyzed by additional content experts.	We used strict inclusion and exclusion criteria that erred on the side of inclusion. A pediatrician was our scientific director.
Peer Reviewer #3	ES and Results	Figure B on page ES-7 and later on 21 as it relates to the number of studies included--the last of bottom-most box adds up to only 25, but 26 were included for 25 studies. While the previous page does state the inclusion of articles vs. studies, the table appears to mix the two.	We have revised the counts in the PRISMA figure based on our update search. We have also separately counted the number of studies and the number of articles. Our goal in presenting both is to clearly state the included evidence; some studies are reported in more than one article.
Peer Reviewer #2	Results	There are no studies that applied to this issue who it seems were not evaluated. The extensive nature of the review I am afraid is a problem- the lengthy nature will make it hard to use this document even though the reading is very logical.	We appreciate that this is a large detailed document. We have tried to enhance the readability by extensively using tables and a clear table of contents. We have also focused on the presentation in the Executive Summary, which we acknowledge is the primary resource for most readers.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #3	Results	The detail is appropriate and tables and figures are well utilized and key messages are explicit and applicable.	Thank you
Peer Reviewer #4	Results	P. 46, line 21: Presumably “uninterpretable” was meant, rather than “uninterruptable.”	Because the Williams study was included in the new Simpson systematic review, this section has been removed.
Peer Reviewer #4	Results	P. 52, lines 47-56: See second comment above re page ES-10.	We have expanded our description of the Paradise study and our discussion of the findings in the revised report. Many of these findings are discussed in relation to KQ2 (functional outcomes).
Peer Reviewer #5	Results	KQ1: The document doesn’t address the part of the key question regarding recurrent AOM (a very important outcome to consider when deciding on treatment) or health care utilization.	If treatment outcomes were included in studies that we reviewed, they were included in our tables and text. This was not a common outcome included in studies. The only treatments to examine recurrence of AOM were those comparing tubes v. watchful waiting or myringotomy and adenoidectomy added to tubes and myringotomy. Key points in the results section comparing tubes to watchful waiting or myringotomy (page 28) said “We found one small RCT measuring AOM outcomes at 3 years that found no difference between groups. Strength of evidence is insufficient.” The strength of evidence table in the results section indicated that there were mixed findings for AOM as an outcome in which adenoidectomy was added to tubes (page 41). We’ve added this to the Key points in the results section. We have added the following in the research gaps section: “We recommend that future research include recurrence of AOM as an outcome. It is important to know whether an OME treatment shows reductions in AOM, even if hearing and functional outcomes do not show an effect.”

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #5	Results	KQ3: Harms Associated with Interventions to Treat Otitis Media with Effusion It concerns me that there's so little mention of potential harms from adenoidectomy. The only mention is:"We found no evidence concerning harms from adenoidectomy, including no data on any risks from having a surgical procedure." The way it's presented, there's more evidence from potential harms from myringotomy, or even oral steroids than adenoidectomy. But I consider adenoidectomy to be a more invasive procedure than myringotomy and oral steroids, and certainly the bleeding and anesthesia risks bear more consideration when comparing effectiveness and harms. I'm interested in the pediatric ENT surgeons' perspective on this.	We have revised the results concerning harms from adenoidectomy. We found limited additional evidence of harms from surgery and have added those to the report.
Peer Reviewer #5	Results	KQ4: Subgroups I think it could explicitly say up front in the results that while you intended to address adults and children, there were insufficient data available for adults so the rest of the review focuses on children. Adults aren't really a 'subgroup' of children and the rest of the document discusses children.	Added the following statement to the introduction of KQ4 (page 68). "One of the explicit goals of this review was to examine treatment options for subgroups of patients including individuals defined by age groups; adults were of particular interest. Our search found very few studies of any subgroups that met our inclusion criteria. We did find one study of adults examining autoinflation and one study of children with sleep apnea who received tubes or myringotomy."
Peer Reviewer #6	Results	It would be very helpful to list study citations for specific findings in all of the tables. I found it very difficult to separate findings that had been reported in previous systematic reviews from findings that were newly analyzed for this report. When I focused on the results in my area of expertise concerning KQ2 I was left with a number of questions about the basis for including and reporting findings as well as their sources. For example, Table 28 (p. 52) presents certain findings on language outcomes for the TT - watchful waiting (WW) comparison, citing Browning et al. (2010). However, the Table only shows the results reported by Browning et al. from the three studies that employed the same language measure – not included are language (and other) findings from the Paradise et al. studies (2001, 2003, 2005, 2007) that were also addressed by Browning et al. (e.g., p. 13).	We have included the quantitative syntheses included in the systematic reviews. We have added outcomes of interest (including functional outcomes) by reviewing the individual study findings directly and including additional information in the summary tables and text, including those of Paradise et al..

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	Results	Table 29 (p. 53), by contrast, does present findings for a measure of cognition and a measure of child behavior from one of the Paradise et al. studies (2001). I don't understand the reason for this inconsistency, and it made me question the exclusion from the Report of a substantial amount of evidence concerning language as well as other outcomes that could have altered the SOE rating of "low for no difference" to be more consistent with statements by Browning et al., e.g., "This raises the further question as to whether in the long-term there are any detrimental effects in the vast majority of children with otitis media, even at the more severe end of a spectrum of persistence and hearing loss. The hypothesis here being that the hearing-deprived period is rapidly compensated for by the flexibility of development in children. No study that randomised children to grommets versus 'watchful waiting/active monitoring' demonstrated a significant effect on any developmental outcome in either group compared with 'normal' non-otitis media with effusion controls" (p. 15). The Report seems to take issue with this conclusion without providing evidence that appears sufficient in quality or quantity to counter it.	Table 28 contains findings from all of the studies that were included in the Browning review. The low SOE is due to few studies examining either cognition or behavior in which the investigators used different measures at different outcome points. We do not take issue with the conclusion that it is unclear whether there is an impact of OME on developmental outcomes and have noted this in the introduction
Peer Reviewer #7	Results	P. 20, line 48: identified 4798 should be identified 4,798 – insert comma	Thank you. We've added a comma and adjusted the counts based on our updated search.
Peer Reviewer #7	Results	Page 37, line 34 – hearing levels were similar. – to what? What were the hearing levels similar to?	Added the following: A second study[46] found hearing improvement during the first 3 months in the tympanostomy tube group, but by 5 years hearing levels were similar **in the two groups (i.e., tubes plus adenoidectomy and adenoidectomy alone).[46]**
Peer Reviewer #7	Results	Key Points – in some places there are several "Key Points" but there are also several places where there is only one key point yet the header is plural. Page 38, line 17 – there is only one key point so the header should be "Key Point" – remove the s from the headers on pages 38, 55, 56, 57, 66, 67, 68, 69, & 70.	We have corrected the text as you have recommended.
Peer Reviewer #7	Results	P. 46: Key point bullets are not aligned	We have reviewed the entire report for formatting concerns
Peer Reviewer #7	Results	P. 47, line 39: tere should be a space between the footnote superscript an the next sentence… (N=243)20The Williamson et al…insert space. (N=243)20 The Williamson et al…	We have reviewed the entire report for formatting errors.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #8	Results	I am concerned about the use of "worst ear" hearing outcomes as having any meaning for development in healthy kids, since what matters for speech development is how well the child hears, which is likely to come from "best ear" hearing. This has implication regarding the management of unilateral disease in particular, but also regarding the assessment and treatment for bilateral OME. Of course worst ear may be important when considering quality of life, although I am not aware of this having been well assessed.	We believe that you are referring to the Gates study. The author reports both best and worst ear and we have included both in our summary table.
Peer Reviewer #8	Results	I found the tables challenging and the presentation of results in particular hard to glean form the tables requiring extra work. I would prefer more columns for each group and results or effect sizes or differences as laid out better than they were. It was hard to view multiple comparisons from multiple studies at once with the current tables. They layout I suggest would have made explicit the flaw in the adenoidectomy comparatives. here would have been a column for watchful waiting or active observation and it would have been empty, making clear that no conclusions could be drawn compared to that group, as the report currently does. It is not systematically presented in a clear, organized, and consistent fashion.	We are sorry that you do not like the presentation of our tables. We appreciate your suggestion but are unable to accommodate this request for the report.
Peer Reviewer #8	Results	I have commented above that I think insufficient use of observational studies and too much use of secondary reports shaped the framing of this document.	We searched for observational studies and included all that met our inclusion criteria. We started with secondary reports to not repeat recent work of other reviewers but independently sought additional evidence that they may have missed or not included because of the focus of their review.
Peer Reviewer #9	Results	The data and results are very well described. It was quite easy to follow how the authors had arrived at conclusions and to follow the trail of evidence. Although in some cases the trail of evidence was flawed (as mentioned above) it was not difficult to understand how the methodology led to results, how those results we assessed, categorized and constructed.	Thank you

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #9	Results	KQ1 - Comparison between types of tympanostomy tubes (TT) (ES-7) – The writers missed an excellent opportunity to really call attention to the lack of research in this area. TT insertion is the most common surgical procedure in children and the fact that there are so few good investigations comparing types of TT is appalling for something that is this expensive and common. - TT vs. WW – The authors need to comment on the importance of the results in studies with shorter-acting compared with moderate compared with long-acting tubes. The evidence is quite clear that shorter acting tubes – the subject of several studies including those included in this data analysis contain short-acting tubes which provide little longer term benefit – as expected.	We agree with these points. We believe that these summary conclusions are most appropriate for the discussion section and we have added them there. We have changed the wording of that section in the Results to affirm that the number of studies is low- "There are relatively few studies that compare different types of tubes, approach to insertion, and topical prophylaxis therapies for hearing and other outcomes. The studies that are available are generally small and present insufficient evidence." This is an important point and we have adjusted summary results to include the statement that " The included studies used tympanostomy tubes with varying design and retention times. This may have an impact on the length of time with persistent middle ear effusion and hearing outcomes."
Peer Reviewer #9	Results	P. 41: The most likely reason for the finding that, "We found that results were mixed across studies concerning whether the addition of tympanostomy tubes to adenoidectomy improved OME-related outcomes" , is the heterogeneity of the study methodology and quality of the studies that are included in this analysis. The real conclusion for this report – almost throughout- is that there needs to be additional, high-quality randomized clinical trials to assess these issues, done through sponsored and supported programs with excellent quality control. If the purpose of this work is to guide clinicians, reduced variability based on evidence and provide the greatest value for healthcare interventions (Quality/Cost) then we MUST resist the temptation to generate a paper based on the "best data we have". It is as harmful to use what are clearly relatively poorly controlled studies with a rigor that almost no one would accept who does clinical trials at a major center. If we want to answer these questions we need to sponsor studies that are designed to answer them, pay for them and truly get data we can trust. ESPECIALLY for such a common problem as otitis media where there are literally millions of patients just in the US that could be entered into a clinical trial it is a significant disappointment that we are relying on data published in some of the journals utilized.	We share the reviewers concern about overall study quality and the confidence with which we can draw conclusions. We have amended the wording of that section to include a statement that "The studies have varying methodology and quality and…". We agree that this likely contributes to an inability to combine results and produce confident conclusions

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Discussion	It is clear that the present wealth of studies do not reach a clear conclusion. That is probably because the modalities available are not very effective. Future research ideas are listed but may not be needed. The key is to perhaps look at other issues rather than restudy the prsent modalities	We agree that the research does not clearly point to one intervention as superior to all others However, we have added a paragraph in the discussion concerning the potential role of vaccinations as modifying the impact of OME.
Peer Reviewer #4	Discussion	P. 72, lines 52-53: All watchful waiting studies were combined with myringotomy" seems unclear. Do you mean compared with myringotomy? Also, not consistent with studies of watchful waiting vs. tympanostomy tubes.	We expanded the results to compare TT with either watchful waiting or myringotomy separately when possible. Table 9, provides strength of evidence for TT v. watchful waiting, TT v. myringotomy, and TT v. TT or watchful waiting.
Peer Reviewer #3	Discussion	Yes, although on page 78 line 24-26, there is summary of findings that is not stated earlier and not clear where the conclusion came from. That is, while there was no evidence of functional or quality of life difference, it was not previously stated with respect to mild to moderate hearing loss.	We have deleted the phrase "with respect to mild or moderate hearing loss."
Peer Reviewer #4	Discussion	P. 72, line 8 The study by Hubbard et al, cited on page B-27 of the CER as among studies of "the wrong population," compared developmental outcomes in children with cleft palate undergoing early tube insertion vs. those managed with watchful waiting.	We had strict criteria for inclusion that required explicit diagnosis of OME prior to treatment. The Hubbard et al. study did not meet the inclusion criteria because of the assumption that middle ear disease had been present in all children. Some more recent work suggests that OME may not be universal in individuals with cleft palate (Chen et al. Is otitis media with effusion almost always accompanying cleft palate in children? The experience of 319 Asian patients, Laryngoscope,122:220-224), 2012). We have added a sentence to the introduction to KQ 4 of subpopulations: Although we did find OME treatment studies for individuals with cleft palate, the studies did not provide data on pre-treatment diagnosis of OME using validated procedures. We have added a sentence to our previous explanation: We were unable to find studies on individuals with cleft palate or sensorineural hearing loss that met our inclusion criteria, and we found only one study that targeted individuals 16 to 75 years of age. In particular, although we did identify studies that included individuals with cleft palate, in no cases were there any studies in which OME was unambiguously diagnosed prior to treatment.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #4	Discussion	P. 72, lines 13-14: See second comment above re page ES-10	We believe that the reviewer is concerned about missing information concerning functional outcomes. We have revised the report to ensure that we have clearly included the research by Paradise and colleagues that includes functional outcomes. We do not think that the sentence quoted below needs to be changed because we believe that Paradise was generally alone in reporting these outcomes. "We tried to examine a broad range of clinical, functional, and quality-of-life outcomes and harms of treatment. Although most of the studies examined middle ear status (e.g., presence of effusion or recurrence of OME), and many examined hearing and harms of treatment, only a handful included measures of speech, language, behavior, or quality of life."
Peer Reviewer #4	Discussion	P. 74, lines 18-19: See second comment above re page vi	This phrase was written in relation to meta-analyses that were described by the authors as combining studies that compared tympanostomy tube arms to either watchful waiting or myringotomy. We have clarified that in some of these meta-analyses, the studies that were quantitatively synthesized were limited to tympanostomy tubes versus watchful waiting. We have eliminated the confusing sentence.
Peer Reviewer #4	Discussion	P. 74, lines 8-9: See second comment above re page vi Tubes vs. watchful waiting during first 3 years of life resulted in no significant between-group difference in hearing levels at age 6 years (see Johnston et al, cited on page B-10 in the CER among the studies excluded.)	We have added the Johnston article on the Paradise study to our review results.. We did not include this report of longer term hearing outcomes in our discussion because the findings were limited to one study, and as such, we had insufficient evidence. .

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #4	Discussion	P. 75, lines 14-22: See second comment above re page ES-10 P. ES-10, lines 17-24: Major omission here of relevant evidence in reports referenced, but not detailed or discussed, in the Browning review. The reports, by Paradise et al, detail long-term developmental outcomes of tube insertion vs. watchful waiting for persistent OME developing during the first 3 years of life. Followup covered 8 or more years post-treatment, with evaluations of children at ages 3, 4, 6, and 9 to 11 years. The outcomes, involving a total of 119 outcome measures, variously concerned speech, language, cognition, auditory processing, attention and impulsivity, academic achievement, behavior, psychosocial function, literacy, phonologic awareness. The study was supported by NICHD and AHRQ. The reports are mentioned briefly in the Browning meta-analysis and are listed in the Browning bibliography, but most of their findings are nowhere actually described. The reports are listed on page B-11 of the CER as having been excluded from the CER because of prior inclusion in a review, presumably the Browning review.	We have expanded our findings and have added data directly from the Paradise and Maw studies which provide additional functional outcomes for longer periods of time. We report findings for up to 8 years of age.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #4	Discussion	P. 78, lines 17-20: However, we now do know that in otherwise healthy children less than 3 years of age, bilateral OME for as long as 9 months and unilateral OME for as long as 12 months do not impact any of a large range of developmental outcomes adversely. See second comment above re page ES-10.	We have expanded the paragraph as follows: Overall, children with TT placement for OME lasting greater than 3 months are more likely to have resolution of middle ear effusion for up to 2 years after the procedure. We noted a similar difference for hearing loss up to 6 months after tube placement. This difference and the physiological and developmental plausibility that the hearing loss could worsen speech and language outcomes in either the short or the long term has driven clinicians to intervene on prolonged OME. Because, in the longer term, effusions resolve in the vast majority of patients without any intervention, a key clinical decision concerns the length of time that mild to moderate hearing loss needs to be present to have an important negative impact; similarly, how these outcomes may differ for individuals at different developmental stages and ages remains a crucial unanswered question. The series of studies by Paradise et al. suggests that delaying TT insertion for 9 to 12 months after OME develops with mild hearing loss does not worsen long-term functional outcomes compared with providing earlier insertion.
Peer Reviewer #4	Discussion	P. 80, lines 44-46: Power was provided in the study referred to in the second comment above re page ES-10.	We added a qualifier to the sentence that included among the few studies that provided power was Paradise and colleagues and Black and colleagues:
Peer Reviewer #4	Discussion	P. 81, line 9: Statement is not correct. See comment re page 78.	We have changed the statement as follows: Additional research needs to determine the appropriate criteria and waiting period before surgical intervention with children. Although children 3 years or older may be able to tolerate a mild-to-moderate hearing loss for a period of 3 to 6 months or longer without risk to language outcomes, analyses by Paradise et al. suggest that mild hearing loss in preschool children even for longer periods (up to 9 to 12 months) does not affect subsequent speech or language outcomes.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #4	Discussion	P. 81, lines 47-50: One can assume that all children had otitis media at the time of initial assessment, since otitis media is virtually universal in infants with cleft palate. Episodes of suppurative otitis media commonly supervene in children with OME; should not be a reason for exclusion in my judgment.	Our study inclusion criteria specified that we had explicit evidence that the sample had OME at the time of treatment (i.e., diagnosis using a validated measure such as otoscopy, tympanometry, or myringotomy). Thus, we did not assume that all children had OME at the time of initial assessment. Moreover, a recent study of Asian children with cleft palate indicates that about ¼ did not have OME during their initial palatoplasty (Chen et al. Is otitis media with effusion almost always accompanying cleft palate in children? The experience of 319 Asian patients, Laryngoscope,122:220-224).
Peer Reviewer #4	Discussion	P. 82, lines 25-30: In the Paradise et al study referred to above, auditory processing was among the developmental outcomes studied at ages 6 years and 9 to 11 years (Reports #73 and #77 on page B-11 of the CER). No difference between early tubes and watchful waiting groups.	We have added reference to the Paradise study outcomes that related to auditory processing.
Peer Reviewer #4	Discussion	P. 83, line 16: Reference #73 appears incorrect. Is a study of proton pump inhibitors rather than of diet.	Thank you. The references have been fixed.
Peer Reviewer #4	Discussion	P. 83, lines 27: I believe that the reference shown (# 72) is the one listed as #73.	Thank you. The references have been fixed.
Peer Reviewer #4	Discussion	P. 86, line 3: References #30 and #34 are identical.	Thank you. The references have been fixed.
Peer Reviewer #4	Discussion	P. 303, line 16: Decibels is misspelled.	Thank you. We have edited the text for spelling errors.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #5	Discussion	I'm concerned that a conclusion that certain treatments vs. nontreatments make no difference in the long run doesn't adequately address the important issue. The issue of the developmental importance of the time-to-improvement is not directly discussed. If we do nothing at all, in the vast majority, or all, children, OME will eventually resolve. The problem is that poor middle ear drainage, associated with OME, can lead to recurrent AOM, causing short-term potential morbidity such as pain, sickness, extra antibiotic use and allergy, and $ cost and missed school and parent work days. Perhaps more importantly even short to medium duration OME can interfere with hearing, and in the age group experiencing most of the OME, this is a critical time period to develop language skills; language skills foster all kinds of important intellectual development in this age group, and the concern is that even short to medium language delay can have lasting consequences for the child. Jack Paradise and his group published work showing no evidence of long term intellectual adverse consequences from use vs. non use of tubes for OME; I'm sure this study was included in the reviews you use, and language outcomes are addressed briefly, but I think this is the critical issue and you should consider discussing it more directly.	We have expanded the detail we provide on the Paradise study. We have also pointed out in our discussion that recurrent AOM was a very uncommon outcome in the studies. Pain was not an outcome in any study. We found no studies concerning the relative costs of various treatment options, either in direct medical costs or lost income due to sickness or caring for a child.
Peer Reviewer #6	Discussion	I found the implications concerning most of the key questions to be clear. However, given that the Report is designed to synthesize objective evidence I would suggest reducing both the number of speculative comments in this section and the emphasis accorded to them. I've listed some examples (in quotation marks) and my comments below. "Shorter time periods are likely to be more important for the youngest children (less than 3 years of age) who are still developing their speech and language skills" (p. 74; see also p. 81). This is speculative and arguable, given the robustness of speech and language skills to enormous variation in environmental conditions as well as the high prevalence of OME in young children, virtually all of whom end up with normal speech and language skills.	It is true that gross measures of language support the notion that development of speech and language skills are robust in children who have no sensory or cognitive challenges. This has not been demonstrated to be the case for those who are more vulnerable (have hearing loss, Down syndrome etc). Modifying the text to include those children should adequately address the issue of vulnerability to longer periods during which access to speech is compromised. "Shorter time periods *may* be more important for the youngest children (less than three years old) who are still developing speech and language skills and who have sensory or cognitive challenges."

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	Discussion	I found the implications concerning most of the key questions to be clear. However, given that the Report is designed to synthesize objective evidence I would suggest reducing both the number of speculative comments in this section and the emphasis accorded to them. I've listed some examples (in quotation marks) and my comments below. "Few studies provided a power analysis so making it difficult to interpret failure to find differences. However, we suspect that power was low given the relatively small, heterogeneous samples and extensive attrition in some trials" (p. 80). I agree that statistical power is addressed far too infrequently. However, some studies in do report it, so the suspicion that power was generally low seems to be an exaggeration.	The statement now reads: Aside from several exceptions—notably, studies by Paradise and colleagues[8] and Black and colleagues[97]—most investigators did not present a power analysis. Without such information, we could not determine with confidence whether a failure to find differences in individual studies was because the study was underpowered. We suspect that power was low for many of these studies, given the relatively small and heterogeneous samples and the extensive attrition in some trials.
Peer Reviewer #6	Discussion	I found the implications concerning most of the key questions to be clear. However, given that the Report is designed to synthesize objective evidence I would suggest reducing both the number of speculative comments in this section and the emphasis accorded to them. I've listed some examples (in quotation marks) and my comments below. In the Research Gaps section: "Research in the area of infant speech perception and later outcomes has demonstrated that babies who were able to distinguish between two simple vowels /i/ (tea) and /u/ (two) at 6 months of age had larger vocabularies when they were 18 an 24 months old than those who could not. Early vocabulary development is important because it is one of the strongest predictors of academic achievement" (p. 81). Vocabulary is associated with many other cognitive, linguistic and academic tasks. However, there is a considerable literature questioning the validity of such vocabulary-dependent measures for children whose exposure to vocabulary input may differ by virtue of factors such as cultural norms and values and parental educational level. I was unable to access or locate the single citation [ref 61] that is listed in support of this statement, but in several large, carefully conducted, longitudinal studies efforts to predict children's outcomes based on vocabulary size at age 2 (or at age 3, for that matter) have yielded "disappointing" results (e.g., Dale, Price, Bishop & Plomin, 2003, JSLHR 46, 544-560). If the issues of prediction and vocabulary are to be raised in the Report, a more balanced presentation of findings based on empirical evidence is recommended.	Instead of vocabulary, this section should more specifically state " babies who were able to distinguish between two simple vowels /i/ (tea) and /u/ (two) at 6 months of age demonstrated better word understanding, word production and phrase understanding at 13, 16 and 24 months of age on the MacArthur Bates Developmental Communicative Inventory than those who could not. (Kuhl citation as is in text) Phonological and morphological proficiency is associated with reading ability at school age. First grades performance on morphological and phonological awareness tasks accounted for a significant portion of the variance on second and third graders word analysis and reading comprehension tests. Carlisle, J. (1995). Morphological awareness and early reading achievement . In L. B Feldman (Ed). Morphological aspects of language processing. Hillsdale, NJ: Lawrence Erlbaum Associates.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	Discussion	I found the implications concerning most of the key questions to be clear. However, given that the Report is designed to synthesize objective evidence I would suggest reducing both the number of speculative comments in this section and the emphasis accorded to them. I've listed some examples (in quotation marks) and my comments below. "In many instances children younger than 2 years of age are underrepresented in studies and when they are included the results are not portioned by age. We recommend that RCT's that include children at the most vulnerable ages examine effects of OME on morphosyntactical development (0-36 months) and report results partitioned by age groups reflecting developmental vulnerability" (p. 81). Clinically useful and/or predictive measures of morphosyntactic skills for children as young as 3 years are difficult to find; many expressive morphological and syntactic forms are not mastered until after 36 months. In addition, it might be worth noting that more than 85% of the children in the in the Paradise et al. RCTs met the criteria for severe, persistent OME before 24 months of age.	It is the case that tests of morphological and syntactical development at age 3 are scarce. M.P. Moeller at Boys Town and colleagues at U of Iowa have recently reported that the morpho-syntactical performance of children with mild to severe hearing loss (in ranges similar to those of children with OME) are statistically different from control subjects. Koehlinger, K., Owen Van Horne, A. J., & Moeller, M. P. (in press). Grammatical outcomes of 3 & 6 year old children with mild to severe hearing loss. *Journal of Speech-Language-Hearing Research*
Peer Reviewer #6	Discussion	I found the implications concerning most of the key questions to be clear. However, given that the Report is designed to synthesize objective evidence I would suggest reducing both the number of speculative comments in this section and the emphasis accorded to them. I've listed some examples (in quotation marks) and my comments below. "For instance, we had targeted auditory processing as an outcome of interest because research has demonstrated that OME can affect skills such as binaural auditory perception,69 and speech recognition in noise" (p. 82). The two studies cited have small Ns, minimal description of participants, and are uncontrolled for a number of factors other than (retrospectively ascertained) OME history that could contribute to group differences on the tasks reported in these studies, which have not to my knowledge been causally linked to performance variations in more naturalistic contexts. The Paradise et al. studies of TT and WW reported measures of auditory processing in 2005 (SCAN test) and 2007 (Hearing in Noise Test; HINT); again, the Report would be strengthened by a more balanced presentation of the findings concerning OME and auditory processing.	We added the following: "In the current review, the only study that reported auditory processing was that of Paradise and colleagues. Using the SCAN test with 6 year old children [Paradise 148] and the Hearing in Noise test with children 7 – 9 years old [Paradise 96], they found no differences between their early and delayed tube groups. Replication with additional samples would be extremely useful."

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #8	Discussion	Discussion/ Conclusion: The conclusions are incorrect in two respects. As I noted above, the reported studies do not compare adenoidectomy to no therapy and no conclusions can be drawn to that. Full stop Period. None. So please stop it and correct what you have. One surgery is compared to another surgery. An error in the Cochrane report is perpetuated. The only other possibility is the reports of the studies are flawed that he studies themselves are not comprehensible from either the van den Aardweg work or the current product.	We have reviewed the original studies. A number of studies do compare adenoidectomy to no therapy in children with unilateral tubes. In these studies, the non-tubed ear in both groups is compared. Some of the same studies also compare tubed ears.
Peer Reviewer #8	Discussion	I also think the evidence for harms from tubes, specifically tympanosclerosis and associated hearing loss (3 dB is the best estimate I have seen, form work in Boston in the 1970's and 80's I believe) is understated. Consistency is absolute in those who have looked it (32% tympanosclerosis post tubes in May's metaanalysis).	Thank you for the comment. Hearing was not consistently measured in the included studies, and rarely after 2 years post intervention. There appears to be debate about the long term impact of TT insertion in regard to hearing. We have attempted to better frame this part of the Discussion. " We found consistent evidence that tympanosclerosis was more common in children who had TT than in those who were actively monitored or who had myringotomy; these results pertained whether or not the children had an adenoidectomy (strength of evidence low). Otorrhea was also more common among ears with tympanostomy tubes (strength of evidence low). While tympanosclerosis and otorrhea could be debated as important harms, there is concern that hearing outcomes could be worsened long term in children who received trauma to the tympanic membrane (Pichichero, Pediatr Infect Dis J. 1989 Nov;8(11):780-7). However, this has not been definitively demonstrated in controlled studies (Valtonen, Arch Otolaryngol Head Neck Surg. 2005 Apr;131(4):299-303.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #8	Discussion	The future section omits several important considerations, such as the need to develop better validated measures of functional status related to OM. Also there is a need for basic research to report the extent of placebo by proxy that exists when parents who chose to subject their child to a surgery are then asked to report on outcomes --cognitive dissonance makes such a phenomenon almost unavoidable so it should be corrected for in outcomes studies. Yet no work has been funded in this area. We have sought it in our group without success.	Thank you for the thoughtful comments. Discussion was amended as suggested. *There were no studies that examined either health care utilization or parent satisfaction with care. Better quality of life and functional measures for children (and adults) with OME are needed. Whether any of these treatments under question reduce time spent at the physician's office, along with costs associated with loss of productivity are not known. Anecdotally, we know that parents often request tympanostomy tubes because they hope that it will reduce time that their children are ill. The unexamined issue is whether receiving tubes or another treatment options affects these secondary outcomes. Functional and parent-reported outcomes are often measured using parental report. Beyond the usual issues of high placebo response rates in surgical studies, proxy reporting of outcomes may have other important effects. We do not have adequate understanding of proxy reporting particularly for placebo assignment when parents have consented for children's assignment to groups. Basic research in this area will help OME and other similar comparative effectiveness reviews.*
Peer Reviewer #8	Discussion	There is also a need to enhance the use of conceptual models about the natural history and pathology of this condition, how the normal and abnormal interrelate and the consequences for that on future research design, analysis and interpretation. Clearly much more comparative research is needed. The failure to recognize that there were no comparisons between adenoidectomy and no intervention means the report short sells the importance of doing some initial work in that area.	Because of your comments, we reviewed the included studies that you are referring to and believe that there is a no intervention ear that is compared in children where one group received an adenoidectomy and one did not.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #9	Discussion	I think the limitations are not clearly described. I would lead with the fact that despite aggregating data a major limitation is that the quality control of the studies and design obviously have some holes. I think the study falls short in this regard. The major implication of this study should be apparent to all who have worked in this area for years - we need well-controlled RCT, which need to be sponsored by institutions or governments (costing millions of dollars by the way) which will monitor quailty control and are specifically designed to answer some of the questions raised. Short of doing this we will always be left with many questions about data and design. The real value of this work is that it makes it ever more abundantly clear that there is not short-cut to this kind of work and no aggregation of data based an hetergenous quality, outcome and purpose will EVER achieve what this report is striving to help with.	We have included a greater amount of information on the original studies. We have also conducted a qualitative summary of evidence that could not be quantitatively combined in a meta-analysis. We did identify that many of the studies are small, conduct analyses by ear, do not include very young children or other subgroups of interest. We have indicated in the research gaps the methodological limitations of many studies. We have further indicated that use of a common outcome protocol would facilitate pooling results.
Peer Reviewer #9	Discussion	There is a lot here that would guide future research. I think the authors miss a chance to carry a torch for the comments above - which in doing this report they must know is true.	Thank you. We have expanded the section on research gaps dramatically. Given the severe limitations of the evidence base—both gaps in study topics (interventions, appropriate outcomes, relevant populations) and in methods, we have included recommendations for future directions.