Comparative Effectiveness Review Disposition of Comments Report

Research Review Title: *Tympanostomy Tubes in Children With Otitis Media*

Draft review available for public comment from July 20, 2016 to August 11, 2016.

Research Review Citation: Steel D, Adam GP, Di M, Halladay C, Pan I, Coppersmith N, Balk EM, Trikalinos TA. Tympanostomy Tubes in Children With Otitis Media. Comparative Effectiveness Review No. 185. (Prepared by the Brown Evidence-based Practice Center under Contract No. 290-2015-00002-I.) AHRQ Publication No. 17-EHC003-EF. Rockville, MD: Agency for Healthcare Research and Quality; May 2017.
www.effectivehealthcare.ahrq.gov/reports/final.cfm.
doi: https://doi.org/10.23970/AHRQEPCCER185.

Comments to Research Review

The Effective Health Care (EHC) Program encourages the public to participate in the development of its research projects. Each research review is posted to the EHC Program Web site or AHRQ Web site in draft form for public comment for a 3-4-week period. Comments can be submitted via the Web site, mail or email. At the conclusion of the public comment period, authors use the commentators' submissions and comments to revise the draft research review.

Comments on draft reviews and the authors' responses to the comments are posted for public viewing on the 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.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	General Comments	The report is clinically meaningful. The key questions are appropriate and clear. There are inconsistencies in the approaches used within and between the key questions which make it sometimes difficult to interpret the results; these inconsistencies, if they can't be resolved, could at least be better explained.	Thank you. We have addressed specific comments by this reviewer related to our use of different approaches tailored to the available evidence.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	General Comments	I expect that the report will be valuable to clinical practitioners. The key questions are stated clearly and concisely. Operational definitions are provided for key constructs, rationales for methodological decisions (e.g., inclusionary and exclusionary criteria) and explanations of corollary analyses (e.g., risk of bias and evidence strength) are likewise presented both clearly and completely. As noted below, there are a few issues that could be addressed to make the report even better, but overall I found it to be excellent.	Thank you.

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #1	General Comments	I want to congratulate the AHRQ for an excellent job on this systematic review in regards to "Tympanostomy Tubes in Children with otitis Media" I find the 5 key questions appropriate and well defined as well as the targeted population and audience	Thank you.

Commentator & Affiliation	Section	Comment	Response
Public commenter Iris Tam, Pharm.D. Otonom, Inc.	General Comments	Based on the PICOD criteria, we have identified additional studies that may be relevant for updating the draft report (see Table 1). Our comments pertain to KQ3. One Phase 1b and two Phase 3 studies[1,2,3] are relevant to KQ3 as they contained a TT treatment arm and measured safety outcomes in pediatric patients with middle ear effusion (MEE). Additionally, an ongoing Phase 3b single-arm study (NCT02600559)4 that has N>50 subjects, with TT + ciprofloxacin 6% as the treatment, may be potentially considered for KQ3.	For consistency with KQ 5, we have clarified in methods that we have also excluded studies reporting only postoperative otorrhea (first 30 days) in KQ3. The other adverse events reported in these studies are not in our pre-specified list.

[1] Mair EA, Moss JR, Dohar JE, et al. *Ann Otol Rhinol Laryngol*.2016a;125(2):105-114.
[2] Mair EA, Park AH, Don D, et al. *JAMA Otolaryngol Head Neck Surg.* 2016b;142(5):444-451.
[3] Park AH, White DR, Moss JR, et al. *Otolaryngol Head Neck Surg.* 2016;155(2):324-331.
[4] ClinicalTrials.gov Identifier: NCT02600559. Available at https://clinicaltrials.gov/show/NCT02600559. Accessed 8/2/16.

Source: https://effectivehealthcare.ahrq.gov/topics/tympanostomy-tubes/research-2017/
Published Online: May 5, 2017

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #2	General Comments	This extensive literature review and analysis of evidence about several key issues about tympanostomy tube placement addresses the main issues: indications (OME and rAOM)), outcomes (complications/adverse effects as well as benefits regarding hearing levels and speech and other neurocognitive outcomes), and care issues (use of water precautions with tubes and use of drops for otorrhea). The key questions are appropriate, target population is explicitly stated (children with and without special risk factors)	Thank you.
TEP/KI Reviewer #2	General Comments	please look through the manuscript as cholesteatoma and prophylaxis are occasionally spelled incorrectly	Misspellings of cholesteatoma have been corrected throughout. A search did not identify any misspellings of prophylaxis.

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #3	General Comments	This comparative effectiveness review is well-written, rigorously conducted and as far as I know all relevant studies have been included in this comprehensive report. Recognizing the completeness of the review and the particular importance of the Structured abstract and Executive summary, I have mainly focused on these sections but the comments below are also applicable to the full review.	Thank you.
TEP/KI Reviewer #4	General Comments	Throughout the document, the phrase "hearing test" should be replaced with the phrase "audiological evaluation". Assessment of hearing in children can only be reliably and validly completed by an audiologist. "Hearing test" implies that most anyone can test a child's hearing.	The phrase "hearing test" is used exclusively in KQ1 "Does obtaining a hearing test help identify which children are more likely to benefit from the intervention?" However, following AHRQ guidance, we have not changed the approved language for Key Questions. We agree that the reliability and validity of hearing evaluations in children is important, albeit difficult to assess in a majority of studies.

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #4	General Comments	Throughout the document, the phrase "may put at risk" should be replaced with the phrase "puts at risk". Use of the word "may" is superfluous, as "puts at risk" is always a true statement, and does not imply that the condition IS present, only that there is risk of the condition.	We agree, "may" deleted as suggested.
TEP/KI Reviewer #4	General Comments	Throughout the document, the use of the word "other" is inappropriate in this context: "...cleft palate, Down Syndrome or OTHER neurobehavioral condition". Cleft palate is not a neurobehavioral condition. The sentence should read (throughout the document) "...cleft palate, Down Syndrome, or any neurobehavioral condition".	Abstract, ES and Full report changed to "provides little guidance for the treatment of children who may be at increased risk for speech, language, or learning problems because of baseline sensory, physical, cognitive or behavioral factors."

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #5	General Comments	The authors are to be congratulated on an excellent systematic review, which clearly involved a huge amount of time and effort. The network meta-analyses are particularly novel and useful. My comments below relate primarily to the abstract, introduction, and summary, which are the parts likely to be most read (especially, the abstract). I would also comment that in reading this report I detect a general bias against tubes in the writing (e.g., selectively listing adverse events of tubes without stating adverse events of comparative strategies), which I would urge the authors to consider and be vigilant for in the revision.	Our descriptive survey of adverse events associated with TT directly reflects KQ 3 "what adverse events, surgical complications, and sequelae are associated with inserting TT"

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #6	General Comments	The key questions were explicitly stated and were appropriate. I believe the report was rigorously pursued and is clinically meaningful.	Thank you.
TEP/KI Reviewer #7	General Comments	Yes, I found the report to be clinically meaningful. Also, I thought the target population and audience were explicitly defined. Earlier, I had the opportunity to review the key questions, including their explicit content or "statement". I agree that these criteria were met in excellent fashion.	Thank you.

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #3	Structured Abstract	The Results section would benefit from adding the (magnitude of) effect estimates, the number of individual trials and participants included and the strength of evidence on which the general statements are based, e.g. “Overall, the evidence suggests that TT placed in children with persistent middle-ear effusion result in short-term (DEFINED AS ??) improvements in hearing (EFFECT SIZE, X TRIALS, XXX PARTICIPANTS) compared to watchful waiting (STRENGTH OF EVIDENCE: XXX), but there is no evidence of a sustained benefit (STRENGTH OF EVIDENCE)”. The same applies to statements regarding effectiveness of TT for recurrent acute otitis media (the authors need to be more concise than just “fewer”), treatment for otorrhea in children who have TT (magnitude of effect? NNTB?) and adverse events (“a variety of adverse events” is not very specific, the authors should specify the likelihood and type of (major vs minor?) adverse events).	Edited to provide specific time frames for KQ 1. Given space limitations, effect size, number of trials, strength of evidence and other details have not been added to abstract.

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #3	Structured Abstract	From the executive summary it is not clear which outcomes of interest are considered primary and secondary. In the abstract, the authors report hearing for KQ1 but the other outcomes (QoL and patient-centered outcomes) are also highly relevant to parents, physicians and policy makers involved in the care of children with otitis media.	From the systematic review perspective, it is not meaningful to define primary and secondary outcomes. We agree that QoL and patient centered outcomes are highly relevant to stakeholders. These are reported. However, given the multiplicity and heterogeneity of QoL and patient centered outcomes, meta-analysis of these outcomes was not performed. Hearing levels and duration of middle ear effusion were reported by a sufficient number of studies, allowing meta-analysis of these outcomes.
TEP/KI Reviewer #5	Structured Abstract	Page v, abstract/conclusions: Lines 37-38. The sentence "Overall, the evidence…a sustained benefit" is more accurately stated as "Overall, the evidence suggests that TT placed in children with persistent middle-ear effusion improve hearing at 1 to 3 months compared to watchful waiting, but there is no benefit at 12 to 24 months."	Edit made as suggested.

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #5	Structured Abstract	Page v, abstract/conclusions: Lines 38-39. The sentence "A period of watchful waiting does not worsen language, cognition, behavior, or quality of life" is NOT supported by your analysis. As noted, there is space evidence here, which creates significant imprecision and low statistical power that prevent concluding definitively there is no effect. Moreover, all of the 8 studies that reported these outcomes EXCLUDED children with baseline delays or disorders of language, cognition, or behavior. Obviously, the ability to show improvements in these areas on groups of children without any baseline problems is difficult to impossible. I suggest you amend this sentence to "TT did not consistently improve cognition, behavior, or quality of life, but low statistical power prevents any definitive conclusions and the results apply to otherwise healthy children without baseline disorders or delays in language, cognition, or behavior."	Abstract edited to: "TT did not consistently improve language, cognition, behavior, or quality of life, however, evidence is sparse, limiting definitive conclusions and is applicable only to otherwise healthy children."

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #5	Structured Abstract	Page v, abstract/conclusions: Lines 41-44. The sentence "Children with recurrent AOM may have fewer episodes after TT placement, but the evidence base is severely limited and it is unclear whether quality of life outcomes are improved" requires revision because (a) the evidence base is not "severely limited" (there are 5 RCTs cited in the text) and (b) only 1 small RCT looked at quality of life in very young children. It would be more accurate to substitute "limited" for "severely limited" and to conclude there is "…insufficient evidence to assess the impact on quality of life."	Edit made as suggested

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #5	Structured Abstract	Page v, abstract/conclusions: Lines 46-49. These lines support treating tube otorrhea with an "antibiotic-glucocorticoid drop. There are a few problems with this conclusion: (a) 3 of the RCTs that show efficacy (Van Dongen 2014, Strachan 2000, Granath 2008) used an off-label eardrop that is not FDA-approved for treating otorrhea, and in one case had an ototoxic antibiotic (neomycin), and (b) all 3 of the trials (Dohar 2006, Roland 2004, Roland 2006) that support using ciprofloxacin-dexamethasone drop were industry-funded by the company that developed the drops and conducted with investigators that had conflicts of interest. There is significant potential for bias. I would change your conclusion to "Should otorrhea develop, the evidence supports treating with a topical antibiotic drop, with or without dexamethasone, and not treating with oral antibiotic therapy. The key point here is the inferiority of oral antibiotics, not whether a topical antibiotic drop should or should not have a steroid.	Simplified discussion in abstract, edited to: "Should otorrhea develop, the evidence supports topical treatment rather than oral antibiotics or watchful waiting."
Peer Reviewer #1	Executive Summary	P12: Line 31-32 "The comparative effectiveness of TT for chronic OME and recurrent AOM is likely influenced by the many factors…" Comparative effectiveness compared to what? Other potential therapies?	"Comparative" deleted to clarify that we refer to potential modifiers of the effectiveness of TT. (same change in full report)

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Executive Summary	P14 Lines 15-26: The time frame defining intermediate outcomes is defined later on in the text, depending on the specific outcome, but the time frame of the QOL is sometimes vague. One might argue that even if QOL is not affected by TT in the long term, short-term QOL benefits may be worthwhile.	We did not define time frames of interest for QoL and other patient-centered outcomes. As noted under "Timing" header, we included studies with any duration of follow-up. Short term outcomes, if reported, were not excluded.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Executive Summary	P16 Line19. Age 1 month to 18 years old is a wide age range; a correspondingly wide range of developmental changes occur in middle ear and immunologic physiology during these years. Infants, for example, are at higher risk of AOM and OME than older children, and even in the same child, the risk of these conditions usually gradually decreases. Therefore the relative benefit that TT may provide will decrease accordingly. While KQ1a does address this, I feel that this issue should be explicitly stated in the paper, because it greatly affects generalizability; few of the studies will include high numbers of teenagers, for example. It also pertains to the significance of one of the listed "adverse events:" premature extrusion. As the child usually outgrows his/her need for the tubes, often by the time of tube extrusion, the child no longer needs the tubes, and the child avoids a second procedure for formal removal—I'm not sure how "premature" is defined here and if this is indeed an adverse event, overall.	We agree that age effects are of interest. The comment references the inclusion criteria for this review, designed to find studies relevant to the entire pediatric age group. As noted in the Limitations section, "individual studies did not often explore treatment effect heterogeneity across subgroups". Further, we were unable to conduct meaningful subgroup analyses across studies, because most trials used similar inclusion criteria, and thus were not highly variable in terms of proportions of age..." e.g. For KQ1, we note that "Among the 16 RCTs (Table 1), a majority enrolled children in the preschool and early school ages (mean age of enrolled children ranged from 1.6 to 5.4 years)." In table 18, we conclude that Strength of Evidence is Insufficient, regarding "TT efficacy ... by risk factors such as age, ..."

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Executive Summary	P16 Line 29: We excluded children with chronic suppurative otitis media." Should briefly say why.	Methods (ES and Full report) edited to read: "We excluded studies of children with chronic suppurative otitis media since it is associated with a persistently perforated tympanic membrane."
Peer Reviewer #1	Executive Summary	P17, line 21-24. The time course of these QOL, behavioral, and language outcomes is not specified. While it's reassuring that long-term language outcomes do not seem to be affected by watchful waiting, as above, there may be value in short term language, behavior and QOL outcomes.	(see also #8): We did not define time frames of interest for QoL and other patient-centered outcomes. As noted under "Timing" header, we included studies with any duration of follow-up. Short term outcomes, if reported, were not excluded.
Peer Reviewer #1	Executive Summary	P17 Line 49-50 "Studies with per ear assignment were excluded." Why?	This was an a priori decision during protocol refinement, made after consultation with the TEP, reflecting an emphasis on patient level outcomes.
Peer Reviewer #1	Executive Summary	P17 Line 51: Any reason that 50- and 1000-subject cut-offs were included here? Does this come from previous reviews?	This was an a priori decision made during protocol refinement, with consultation with the TEP. Ascertainment of more common adverse events was felt to be subject to less bias in prospective studies. For rare events, the TEP suggested including large registry based studies.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Executive Summary	Page 19 Lines 51-55: 481-306=175—the 178 included studies stated in the text does not follow from this explanation. In Figure D we learn that 2 additional publications were added from the hand search of reference lists, that brings us up to 177—where did the extra study appear from? Everything needs to add up.	Counts of included studies have been updated and reconciled.
Peer Reviewer #1	Executive Summary	Page 19 Lines 51-24: * risk of bias is discussed in the text for KQ#2 but not for KQ#1, why? * Apparently some of the papers reported data from the same study. Was this hierarchical nature of the data accounted for in the quantitative meta-analyses, or is this not necessary?	In the ES (for brevity) risk of bias is summarized by KQ in the "Overall summary and Strength of Evidence Section" for each KQ (deleted from Results in ES and included in Results (under "Risk of Bias" subheader) in Full Report. Instances elsewhere have been deleted in the ES. We counted patients only once when several papers reported on the same study.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Executive Summary	Line 3: For KQ#1, apparently the results from nonrandomized trials are presented separately. The NRCT section is set off and bolded (p 23, line 27), so it's clear where that presentation starts, but there is no previous bolded section heading indicating where we start finding results of only the RCT.	Added "Randomized Comparative Studies" header
Peer Reviewer #1	Executive Summary	P 20 Figure E. and the other network graphs: For those of us unfamiliar with network graphs, it would be helpful to have legends on these figures explaining what the lines' associated numbers and relative widths mean.	We have added additional explanatory text prior to figure E: "Such network plots are a visual representation of the evidence base. The network plot consists of nodes representing the interventions being compared and edges representing the available direct comparisons. The number of studies that include each comparison is indicated next to each edge (connecting lines with thickness proportional to this number)."
Peer Reviewer #1	Executive Summary	P21 If Figures E, F, G and/or H and/or Tables A and B refer to results of only RCT, they should say so in their titles or in legends.	Text of ES edited to clarify that RCTs only included in meta-analysis. "For the network meta-analysis of these RCTs"

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Executive Summary	P 23 Lines 45-48: Were results weighted to account for the fact that some papers reported the same study?	If multiple papers reported a single study, data were abstracted on a per study basis. Thus, no weighting is required.
Peer Reviewer #1	Executive Summary	P 23 Why does the section on nonrandomized studies for KQ#1 not include a forest plot, and why were randomized trials and nonrandomized trials presented separately? It's not necessarily wrong, it's just not explained and it's harder to compare results between them.	Throughout the report, forest plots are included when we have performed quantitative synthesis of multiple studies. For KQ1, we included RCTs only in the network meta-analyses.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Executive Summary	P23 Line 45 Why is “Quality of life and patient-centered outcomes” in a tinier font than “Non-Randomized Comparative Studies” above it? And it’s not capitalized, so it looks like this paragraph is a part of the Nonrandomized comparative studies paragraph, but it is not. The QOL paragraph includes results of RCTs and NRCTs and I believe should be set off more clearly here.	Thanks. Heading changed from Level 4 to Level 3 and capitalized
Peer Reviewer #1	Executive Summary	P23 Lines 45 vs. 51: Line 45 says 8 studies reported QOL but line 51 says 2 studies report QOL—this is confusing.	Edited to clarify distinction between patient-centered outcomes and QoL.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Executive Summary	Page 24 For KQ2, no meta-analysis is done here, and there is not forest plot; instead results of the individual studies were described. Why are these results presented differently than those of KQ1; again it makes it harder to compare them.	We have been consistent throughout in presenting forest plots when meta-analysis has been performed. For KQ2, we did not perform meta-analysis given the small number of studies and heterogeneity of reported outcomes. Hence, no forest plots.
Peer Reviewer #1	Executive Summary	P24 Also, the description of KQ 2 is much smaller and does not separate into RCT and NRCT, in contrast to KQ1. The relevant studies are presented individually here; they were not for KQ#1. Is this because there are fewer studies?	Yes. Given the small total number of studies (with a single NRCS) and substantial heterogeneity of reported outcomes, we chose to summarize these studies individually.
Peer Reviewer #1	Executive Summary	P24 Line 11: Risk of bias is presented here but not for KQ1—is there a reason for this?	In the ES (for brevity) risk of bias is summarized by KQ in the "Overall summary and Strength of Evidence Section" for each KQ (deleted from Results in ES and included in Results (under "Risk of Bias" subheader) in Full Report. Instances elsewhere have been deleted in the ES.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Executive Summary	P24 Lines 15-57 through page 25, line 1-43: are these descriptions of KQ2 results all referring to RCT data?	Yes. ES and Full Report edited to explicitly identify discussion of RCTs and single NRCS
Peer Reviewer #1	Executive Summary	P 24, Lines 23-45: Some results are presented with p values and some with 95% CI; it would be best to be consistent if possible.	We agree that confidence intervals are preferable, and report these when available. However, some papers report P values only, thus consistency is not possible.
Peer Reviewer #1	Executive Summary	P24 Line 53: amoxicillin is mis-spelled	Thank you. We have corrected misspellings (and replaced alternative spelling of amoxycillin) throughout the document and appendices.
Peer Reviewer #1	Executive Summary	Page 25, KQ3: here RCT and NRCT data are combined for this outcome. These choices and their rationales should be made explicit, if not in the Executive Summary, at least in the main text of the report.	Methods: Study Design of ES and Full Report edited to add "(including arms of RCTs or NRCSs with 50 more patients")
Peer Reviewer #1	Executive Summary	P25 Line 21: Can the timing of the hearing loss and speech impairment assessments be defined—are these long or medium term results?	Text of ES and Full Report edited to include timing of QoL assessments for this study

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Executive Summary	P26, Limes 42 – 51: This paragraph uses two different metrics-risk difference in episodes par month vs odds ratios, to describe the same type of results. Is there a way to put everything in the same metric so we can compare these results more easily?	No. Discussed separately as they represent different outcome metrics.
Peer Reviewer #1	Executive Summary	P 27, KQ4 gets a forest plot; again the rationale for presenting similar types of results (OR) in different ways should be make explicit somewhere.	A forest plot is shown as we have done meta-analysis of the NRCS. We preferred not to combine RCTs and NRCS in the meta-analysis. Hence, the two RCTs are described only and not shown in the forest plot.
Peer Reviewer #1	Executive Summary	P 28: Lines 36-56 The non randomized study accounts for a relatively large portion of the results; is there a way to compare results stratified by randomized vs nonrandomized study?	ES edited with additional detail (already in full report) which clarifies that the NRCS and one RCT were excluded in the network meta-analysis.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Executive Summary	Page 30 Line 23: The fact that hearing level was chosen as the primary outcome for KQ1 should have been presented in the Methods.	We pre-specified intermediate and quality of life and patient-centered outcomes. Meta-analyses were performed for specific outcomes when studies reported sufficient data. We do not distinguish primary and secondary outcomes. The ES and Full Report has been edited to delete the sentence, "Given the functional importance of hearing, we chose hearing threshold as our primary intermediate outcome for meta-analysis."
Peer Reviewer #1	Executive Summary	P30 Line 55: Again, why is risk of bias reported here only for KQ#2 but not for the others?	In the ES (for brevity) risk of bias is summarized by KQ in the "Overall summary and Strength of Evidence Section" for each KQ (deleted from Results in ES and included in Results (under "Risk of Bias" subheader) in Full Report. Instances elsewhere have been deleted in the ES.
Peer Reviewer #1	Executive Summary	P32, Table E: This table is helpful but if the results from KQ1 include only results from RCT, it should say so.	This table is highly summarized and does not rely exclusively on RCTs.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Executive Summary	P33 Lines 3-4: Either here or at least in the discussion at the end of the report, there should be an assessment of how the pneumococcal conjugate vaccine's success is likely to have affected our interpretation of the results shown here.	added: "It is unclear whether these or other factors affect the relative (current vs. historical) benefits of TT placement for recurrent AOM."
TEP/KI Reviewer #3	Executive Summary	General comment: I would like to encourage authors to add a general statement on the strength of evidence for each outcome across all relevant (Results) sections. This is very relevant when interpreting the data.	For brevity and ease of comparison, we have presented our assessments of Strength of Evidence this in Table E (ES) and Table 18 (Full Report)

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #4	Executive Summary	ALL PAGE NUMBERS REFERENCED BELOW ARE THE PAGE NUMBERS AT THE BOTTOM OF THE PAGES, NOT THE PAGE NUMBERS IN THE UPPER LEFT HAND EXECUTIVE SUMMARY p. 23: line 5 "late" hearing levels" should read "subsequent hearing thresholds"	Hearing "levels" changed to "thresholds" throughout. In this case "late" has been specifically defined as 12-24 months an has been retained in preference to subsequent.
TEP/KI Reviewer #4	Executive Summary	p. 23: line 36 strike word "measured" and replace with "expressed as", "described as" or "reported as".	Edited in ES and Full Report to "Hearing levels reported as pure tone averages"
TEP/KI Reviewer #4	Executive Summary	p. 30: what is "non-significant"? if it's non-significant then it is not an "improvement".	Changed in ES to agree with more nuanced language in full report: "A trend was noted for myringotomy with adenoidectomy, but credible intervals are wide and include the null effect"

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #4	Executive Summary	Throughout the document, the phrase "hearing level" should be replaced with "hearing threshold". THRESHOLDS are measured. "Levels" is a somewhat ambiguous term--"threshold" has a very specific definition.	"hearing level(s)" replaced by "hearing threshold(s)" throughout ES and full report
TEP/KI Reviewer #5	Executive Summary	Page ES-1, line 35. More accurate to state that "The AAO-HNS CPG recommends that clinicians offer TT to children with recurrent AOM who have middle ear effusion at the time of assessment for tube candidacy, and that clinicians do not perform TT insertion when middle ear effusion is not present." I strongly object to the current wording, and as the first author of the CPG can definitively state this was NOT the conclusion of the group (at least in the context it appears in the systematic review)..	ES edited as suggested.

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #5	Executive Summary	Page ES-1,line 39. You state "In addition, TT placement is associated with complications, such as acute otorrhea." If you want to state the downside of tubes then also state the downside of watchful waiting. I recommend "TT placement may result in acute otorrhea in some patients and watchful waiting may result in continued episodes of recurrent AOM, which may include tympanic membrane perforation and otorrhea."	ES edited along the lines suggested.
TEP/KI Reviewer #5	Executive Summary	Page ES-1, line 41. Again, you talk about potential downsides of tubes, without mentioning downsides of watchful waiting. More accurate to say "Otorrhea is rarely chronic and both tube otorrhea and recurrent AOM (in children without tubes) may negatively affect quality of life."	ES and Full Report edited to focus on treatment of otorrhea

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #5	Executive Summary	Page ES-9, ES-10, and ES-11: Figures E, F, and H, refer to “Tympanostomy,” which should more correctly be “Tympanostomy Tubes.”	"Tympanostomy" has been changed to "TT" in all instances
Peer Reviewer #1	Introduction	P 39: Consider subheadings for the background: Persistent OM, Recurrent OM, Water prophylaxis, Tube otorrhea.	Discussion of otorrhea moved to a new paragraph. Given the brevity of the discussion, subheadings were not added.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Introduction	P 39: Lines 49-53: "A risk-centered approach might involve differential management of children with otitis media by their risk of important outcomes, as obtained from risk prediction models, and may be preferable to algorithms that use a single threshold for duration or frequency of a diagnosis." I'm not sure that this is needed. All of our decision-making typically involves a risk-centered approach. For example, we all know that older children are less likely to have recurrent OM, that OM is more common over the winter months, that children with certain underlying conditions are more likely to have recurrent OM, persistent OM, or complications from OM.	We agree with the reviewer that clinical decision making should include an assessment of individual risks based on the specific clinical context. We have retained this text to emphasize the potential importance of risk prediction models which could aid clinicians.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Introduction	P41 Lines 9-15. These bullets are referred to as “a” and “b” on the Analytic Framework, they should be labeled consistently. Are they sub-questions?	To clarify that they represent subquestions, bullets are replaced by KQ 1, a) and b) and KQ 2 a). ES and Full report: Edited to add bullet (What factors ...)
Peer Reviewer #1	Introduction	P41 Lines 16-25: Why does KQ#1 have the bullets (a,b) but KQ#2 does not? Should be consistent unless there is a specific reason.	ES and Full report: Edited to add bullet (What factors ...)
Peer Reviewer #1	Introduction	P42: the bullets that are not labeled with specific KQs, does this mean that they apply to all of the KQs?	Yes, bullets applicable to specific KQ(s) are indicated by listing the KQ in parentheses.
Peer Reviewer #1	Introduction	P43: Somewhere it can be noted that some adverse events related to use or non-use of TT and antibiotic drops are not included here, such as expense, antibiotic resistance.	KQ 4, Results edited in ES and Full report to add "We did consider other adverse events such as antibiotic resistance, gastrointestinal side effects of antibiotics or pain related to ear drops.

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #4	Introduction	p. 59: line 40-41: "clinically important" should replaced with the phrase "clinically significant". AND ALL OTHER USES OF THE PHRASE "clinically important" should be replaced with "clinically significant".	The phrase "clinically significant" highlights the distinction between clinical and statistical significance. We prefer the phrase "clinically important", to emphasis that the difference is important in at least some clinical contexts.
TEP/KI Reviewer #4	Introduction	p. 25: line 39--"speech reception in noise"--is this correct? speech reception threshold is rarely performed in noise. speech recognition is often performed in noise.	Thank you. We have edited and expanded the summary of this study. "The MRC Multicentre Otitis Media Study Group 2004 report outcomes of a speech-in-noise automated toy test (SiN ATT). They hypothesized that a measure of understanding of speech in noisy situations may tap the disability experienced by children with OME."
TEP/KI Reviewer #4	Introduction	p. 26: line 17--"hearing levels measured as pure tone averages" is awkward. should read "average pure tone thresholds"	Edited as suggested

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #4	Introduction	p. 43: line 8: (and throughout paper)--"average hearing levels" should be specific as to what frequencies are averaged. typically the pure tone average is 500, 1000, 2000 Hz. for speech development, speech perception--2000 Hz is more important than 500 Hz. Some acknowledgement or description of frequencies being averaged, with some recognition of the idea that not all frequencies are equal in importance for perceiving and producing speech sounds, would be helpful.	Text added to results (Full Report) "Pure tone average (typically averaged over 500, 100, 2000 and 4000 Hz) hearing thresholds were extracted. Hearing thresholds were variably reported as: averaged over both ears, best and worst ear and right and left ear. When multiple averages were reported, we extracted for analysis the worst ear and the right ear." and Limitations (ES & Full Report) "Our meta-analysis of hearing levels used average pure tone hearing levels (typically reported as an average over frequencies of 500, 1000, 2000 and 4000 Hz). This simple measurement is likely insufficient to fully elucidate the complex relationships between hearing and speech perception and development in children."
TEP/KI Reviewer #5	Introduction	Page 2, line 15. “They note the overall favorable natural history of otitis media” should more correctly be “They note the overall favorable natural history of recurrent AOM without persistent middle ear effusion.”j	Edited in Full Report as suggested.

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #6	Introduction	The introduction clearly and succinctly gives the appropriate background and identified the key questions for analysis.	Thank you.
TEP/KI Reviewer #7	Introduction	The Introduction was informative and well written. I have no other comments or suggestions for the Introduction.	Thank you.
Peer Reviewer #1	Methods	Page 44, line 33-34: Why were children with chronic suppurative OM excluded?	Added in ES and Full Report: "since it is usually associated with a persistently perforated tympanic membrane"
Peer Reviewer #1	Methods	Outcomes P 45 Line 25-26: Aren't the outcomes used for KQ#1 relevant for KQ#2?	To clarify, we added "also" extracted, when referring to KQ2
Peer Reviewer #1	Methods	P46 Line 4: A brief justification for excluding studies with one ear assignment would be useful.	This was an an a priori decision during protocol refinement, made after consultation with the TEP, reflecting a need to limit the scope of the review and an emphasis on patient level outcomes.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Methods	P46 Line 39-40. "were independently screened by two researchers. At the start of citation screening, we implemented a training session, in which all researchers screened the same articles and conflicts were discussed." All researchers meaning all 2? Or does this instead mean that each citation was independently screened by at least 2 of the researchers, with discussion with the larger group until agreement was reached?	Yes. Edited for clarity.
Peer Reviewer #1	Methods	P 47 Line 30: How was sufficiently similar defined?	edited to specify the two categories of interventions: "(water restriction vs. ear protection)"

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Methods	P47 Lines 29-36: Why were the meta-analyses of KQ4 handled differently than the ones for KQ1, 2, and 5?	We did not perform meta-analysis for KQ2 due to heterogeneity of outcome reporting. For KQ4, we have added further explanation in Methods: Data Synthesis regarding our choice not to combine the two RCTs with the NRCSs. "The two randomized comparative trials were not combined in a meta-analysis on the basis of clinical heterogeneity (suggestion of higher baseline risk) as well as methodological heterogeneity. Rather, each was individually reported." In addition, we have specified that these meta-analyses were "direct pairwise" comparisons. As discussed in Methods, we performed network meta-analyses for KQs 1 and 5 to allow simultaneous comparisons of multiple interventions.
Peer Reviewer #1	Methods	P 48, line 45: “will” is a typo	edited, removing "will"

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Methods	P 49: Can consider making the distinction here between generalizability, which I believe in this case would apply to the populations of interest (chronic OME, recurrent AOM and children with TT) vs applicability, which I believe would apply to specific groups of children not addressed in the included studies, such as children with Down Syndrome, cleft lip and palate, etc. [Ramon et al. Addiction, February 2012; 107(9):1570-9]	Thanks for providing this reference. Text edited ES and Full Report to read "We assessed the direct applicability within and across studies with reference to children with specific comorbidities (Down syndrome, cleft palate, etc.), and whether interventions and comparators are used in current practice."
TEP/KI Reviewer #2	Methods	I don't feel qualified to comment on the statistical methods—I do think the search was thorough and had explicitly stated methods. Appears to have produced all of the studies I know about and a lot of cohort studies I did not know about.	Thank you.

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #3	Methods	From the executive summary it is not entirely clear which outcomes of interest are considered primary and secondary. This should be specified.	We pre-specified intermediate and quality of life and patient-centered outcomes. Meta-analyses were performed for specific outcomes when studies reported sufficient data. We do not distinguish primary and secondary outcomes.
TEP/KI Reviewer #6	Methods	The methods are clearly stated and are reasonable and logical.	Thank you.
TEP/KI Reviewer #7	Methods	Regarding Methods: I found the inclusion and exclusion criteria justifiable; the search strategies were explicitly stated and logical. I also thought the definitions and diagnostic criteria for the outcome measures were appropriate. In addition, I thought the statistical methods used were not only appropriate, but extremely well explained and explicitly stated.	Thank you.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Methods	My only comment concerns the network meta-analysis methods, which are alluded to in the ES and mentioned in the Methods section of the full report, and also are the basis for the network graphs in the ES and main Results sections. These methods are not explained in much detail until Appendix J (which isn't listed in the Table of Contents), which is long after readers will encounter (and if they are like me, be baffled by) the network graphs. I strongly suggest including more of an orientation to network meta-analysis here, so that readers will know what the graphs are intended to show. For example, no doubt the darkness and width of the lines, the numbers next to them, and whether the comparator nodes are connected or not are meaningful, but I couldn't interpret any of these even after locating Appendix J. A sample labeled graph explicitly interpreted with reference to more common displays such as forest plots might be helpful in the Methods section; I also recommend including a detailed note or key to aid in interpreting each of these figures. I was very interested in this approach, but despite being highly motivated to understand it I couldn't get there from the description that was provided, and I suspect that I might not be alone among readers who will use this report.	Additional explanatory text added in ES prior to Figure E.
Peer Reviewer #1	Results	Page 50: Lines 7-14: 481-306=175 but text says 178 Fig 4 says 2 more added from hand search of reference lists, that brings us up to 177; the remaining study is not accounted for. The # need to add up	Counts of included studies have been updated and reconciled.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Results	P50 Line 20: Table not referenced, should reference Appendix G. I suggest, instead of "NS" for nonsignificant p values, substitute the actual p value. The reader can then judge the meaning of that p value for his/her particular question.	We removed the paragraph describing reporting of quality of life outcomes (line 20-27). We reference Appendix G prior to Table 10. “Full details for all outcomes are in Appendix G” In some cases, studies report only 95% confidence intervals, not p-values. If the 95% confidence interval excludes the null value, we can only conclude that the p-value is less than 0.05.
Peer Reviewer #1	Results	P 52 Line 25: included children are 1.6-5.4 years. Somewhere there should be a comment about the generalizability of the results to other ages, especially to infants and toddlers (age 1 mo to 18 mo).	Text of Discussion in ES and Full report edited to add "The generalizability of results to infants and young toddlers and to school age children is also uncertain, given that children in these age groups are underrepresented in available trials."
Peer Reviewer #1	Results	Page 58: Line 15: Who are the second group? Kids with more significant hearing loss or is this a different group altogether?	Edited to clarify that these were the 23 patients with significant hearing loss

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Results	Page 59 “Table 3 …it appears that interventions that ventilate the middle ear (TT and TT and adenoidectomy) improved hearing levels by -9.1 dB and -10.5 dB respectively, with 95% credible intervals that exclude a null effect in the 1 to 3 month time frame.” Compared with watchful waiting? Shouldn’t this be “-10.3?”	Yes, corrected to "-10.3" and edited to include the comparison with watchful waiting
Peer Reviewer #1	Results	p 60 How was Table 4 produced? Does this take into account the confidence intervals or just the point estimates of the comparative effects?	The rank probabilities represent the cumulative probabilities obtained from the MCMC sample and reflect the Bayesian posterior distribution (i.e. they incorporate the full uncertainty in the model).
Peer Reviewer #1	Results	P60 Lines 45-51: This was presented earlier page 59 lines 41-44.	The first description on pg. 59 refers to the full matrix of treatment effects. The second description refers to Figure 7, describing the forest plot which compares multiple interventions with watchful waiting.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Results	Page 62 Lines 4-6: "…that interventions including adenoidectomy (TT with adenoidectomy) may be more effective than watchful waiting." And Myringotomy and Adenoidectomy.	edited to: "At 12-24 months, interventions that include adenoidectomy (TT + Adenoidectomy and Myringotomy + Adenoidectomy) "
Peer Reviewer #1	Results	Page 60, Tables 4&5, and Page 62, Tables 7&8: Do Tables 4 and 7 provide additional helpful info relative to Tables 5 and 8?	Tables 5 and 8 collapse information found in tables 4 and 7, and thus may be easier to interpret. We prefer to retain tables 4 and 7 in order to provide full information in the Full Report. In the ES, the collapsed categories only are presented as tables A and B.
Peer Reviewer #1	Results	Page 63, lines 8-9: The 95% credible interval of prophylaxis is the same as Myringotomy and Adenoidectomy so this explanation falls flat. I think what is instead driving the conclusion here that the interventions including adenoidectomy are more likely to be effective are the magnitude of the effect together with similar confidence intervals.	Edited to read "The point estimates suggest that interventions that include adenoidectomy (TT with adenoidectomy and myringtomy with adenoidectomy) are more likely to be effective at 12 to 24 months, although 95% credible intervals do not exclude a zero mean difference."

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Results	P63 Lines 36-56: Statistical significance level reported only for TARGET results in this section, but not the others. Why?	Chaudhuri, 2006 did not report any P-values. D'Eredita did not report any details beyond "Hearing levels were normal in all children, in both groups, at 1-year follow-up." We have edited the discussion of the 2003 MRC Multicentre Otitis Media Study group substudy of speech in noise and have included a P-value for the baseline by treatment interaction term.
Peer Reviewer #1	Results	P63 Lines 41-46: Can adenoidectomy be evaluated separately?	No. As noted, children in both arms underwent TT placement.
Peer Reviewer #1	Results	Page 64: Line 7 Would be helpful to define meaning of early vs. delayed TT	In both ES and Full report, we removed duplicated discussion and added "(mean age 3 months)" and "mean age 40.8 months or not at all in two subjects)" in summary of Hubbard, 1985.
Peer Reviewer #1	Results	P64 Line 23, CI or p value would be helpful here.	In both ES and Full Report, we added "(P=0.05 for ears with better hearing and P=0.10 for ear with worse hearing)"
Peer Reviewer #1	Results	Page 67: Line 24: What does it mean to favor delayed treatment? Does this mean that TT should be delayed or that it should not be done at all?	This sentence refers to specific outcomes on a test of nonword repetition at 4 and 6 years which favored the group randomized to delayed treatment with TT. We can draw no broader inferences regarding whether TT should be done at all.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Results	Page 67: Line 42-47: Whether we consider short vs long term QOL differences are quite important here. I could argue that even if there are no long term QOL differences, that short term QOL differences may be important and worthwhile pursuing.	We agree. However, as noted, information is very sparse.
Peer Reviewer #1	Results	Page 68 Table 10, What does statistically significant mean here? For an outcome showing TT effectiveness? For either direction, better or worse? In general: I don't think this part of KQ1 was addressed: Does obtaining a hearing test help identify which children are more likely to benefit from the intervention?	added as note to Table 10 "* No statistically significant effect of intervention on outcome (in either direction) reported."
Peer Reviewer #1	Results	Page 69 First paragraph: Was hearing addressed here as an outcome, and if not, why not?	Hearing was not an outcome specified by KQ 2. This has been clarified in Figure 1 (analytic framework) by adding "(KQ 1)" next to hearing and vestibular outcomes in the Intermediate outcome section.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Results	P69 Line 24: Were other NRCT studies found in addition to the additional 169 patients?	Edited in Full Report to: "The Mattila 2003 paper described two groups, an RCT which randomly allocated treatment in 137 patients, and a NRCS in which parental choice determined treatment in169 patients."
Peer Reviewer #1	Results	Page 70: Line 40-41: Number-needed-to-treat is introduced here as an outcome, which is helpful, but it does not seem to be included when discussing other studies or outcomes. If more NNT values could be calculated from risk differences they would be helpful to compare results.	We have calculated NNT based on results from meta-analyses for KQ 5, based on a baseline risk in a specific control group (identified with footnotes)
Peer Reviewer #1	Results	P 71 QOL outcomes: what is the time horizon of QOL here—can we make it explicit?	The time horizons for QoL vary by key question and by study. We did not attempt to define these a priori

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Results	P71 Line 35-37: “No study evaluated whether age, age of onset, number of recurrences, comorbidities, history of complications of acute otitis media, antibiotic allergy or intolerance or other”: Note that it says above (p 70 lines 33-35) that Casselbrant did address age effects.	A "Key Question 2a" header was added to ES and Full Report and clarified edits added.
Peer Reviewer #1	Results	P 73 Line 24-41: There is no summary OR from these two RCTs?	Summary odds ratios for the two RCTs are presented on pg. 73, lines 43-44.
Peer Reviewer #1	Results	P73 Line 36: does this include the children assigned to swimming who did not swim?	Text edited to reflect that this was an intention to treat analysis.
Peer Reviewer #1	Results	P73 Line 48 and p74 line 14: Does this mean that combined the children had 125 water exposure episodes?	Table edited to read "Children who each had"

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Results	P 75 figure 12 and: in general, why do only NRCT get forest plots and the RCT get described individually? Figure 12: the Key should say OR for what? For favoring plugs?	We have consistently used forest plots to display and summarize studies included in meta-analyses. To avoid confusion, individual studies such as the two RCTs are reported in text and/or tables. Table 12 has been added for clarity. A footer has been added to clarify the direction of effect.
Peer Reviewer #1	Results	P 79, Fig 14: Does this forest plot, unlike the previous, include the RCT? It must, because only one trial is nonrandomized. It's still not clear why some results get combined and some don't.	We excluded the single NRCS (Dohar 1999). The forest plots include data from RCTs only. Text edited to clarify this point. "Two studies were excluded from our meta-analysis. The first was a nonrandomized comparative study by Dohar 1999"

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Results	I'm not aware of any studies that were overlooked, nor of studies that should have been excluded. In general the Results are presented at the appropriate level of detail; the forest plots and tables are clear, but it might be worth adding a note to the tables presenting the probabilities that an intervention was in the top N for effectiveness that says that these were derived from results of the network meta-analyses (as noted explicitly in the body p. 10).	We have added "derived from the network meta analysis" when introducing these tables to the Full Report.

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #1	Results	Tube life span was mentioned but not taken into consideration in the assessment of benefits for TT in KQ1. "... evidence suggests that TT .... results in short term (1-3 mos) improvement in hearing compared to watchful waiting, ... no evidence of long term (12-24 mos) sustained benefit (except adenoidectomy) "The longer TT are functioning the greater is the chance that the child has grown out of the disease and normal hearing. Tubes with short life span (3-6 mos) are often replaced due to recurrence of COME/hearingloss as the tubes becomes nonfunctional within a short time. If children (ww vs TT) are properly randomized there should NOT be a major difference in COME/hearing between the children before TT and at >12-24 months as by that time most tubes have extruded and the children in both groups should have equally grown out of the disease and only few children in may still have COME The benefit of tubes on hearing is related to the life span of the tube.	We have edited the discussion of KQ 1 (ES and Full Report) to reflect these considerations.

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #2	Results	The network diagrams helped illustrate the comparisons between treatments in some of the RCTs, the Forest plots clear, and the tables that looked at probalities that an intervention was effective (top two, etc) all were clear and helpful to any reader. I don't think any studies were omitted. Key messages are clear. No surprises though.	Thank you.
TEP/KI Reviewer #3	Results	KQ1. No comments.	Thank you.

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #3	Results	KQ2. For further clarification and to ensure consistency with other sections (e.g. KQ5), I would suggest authors to rephrase "We identified six studies in seven publications" to "We identified seven papers, representing six studies, reporting five RCTs (six papers) and one NRCS with a total of xxx patients analyzed (xxx in RCTs and xxx in NRCS)." Would suggest to change the heading "Risk factors" to "Subgroup analysis".	Enumeration of publications/studies edited (and clarified) in ES and Full Report. "Risk Factors" has been replaced by "Key Question 2a" (text of this sub-question moved below header) following the discussion of NRCS.
TEP/KI Reviewer #3	Results	KQ3. I think this section would benefit from adding some general sentences on the frequency and type of (major vs minor?) adverse events rather than providing a Table only.	We have added a high level summary of Table 11. The table represents a descriptive summary of the observed median and range of estimates in studies.

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #3	Results	KQ4. The authors need to clarify why they decided to use "Odds Ratio (OR)" rather than "Risk Ratio (RR)" in their R-E meta-analyses. I prefer RR rather than OR, since OR is more difficult to interpret and tend to overestimate the RR, particularly when the outcome is relatively common (which is the case in the included studies) [Knol MJ et al. CMAJ 2012].	The scale in which the synthesis is done should be the scale where data appear to be more homogeneous. Typically, this is the odds ratio scale. When data are sparse, we also prefer the more parsimonious model. See Panagiotou OA, Trikalinos TA. Commentary: On Effect Measures, Heterogeneity, and the Laws of Nature. Epidemiology [Internet]. 2015 Sep;26(5):710–713. Available from: http://dx.doi.org/10.1097/EDE.0000000000000359 PMID: 26196685

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #3	Results	KQ5. 1. Network Graph - The heading should include “for the 10 RCTs” - The authors state that “Several studies had a watchful waiting or placebo arm” but as far as I know only Van Dongen 2014 (initial observation) and Ruohola 2003 (placebo) classify as such. According to the Network Graph, the authors identified three trials comparing watchful waiting/placebo with oral antibiotics, but as far as I am concerned only two studies classify as such?	"for the 10 RCTs" added to header for Figure 13. Details (with citations) added to identify the three treatment arms grouped in the Watchful waiting/placebo category.
TEP/KI Reviewer #3	Results	KQ5: The authors need to clarify why they used OR rather than RR as effect estimate (see also KQ4).	The scale in which the synthesis is done should be the scale where data appear to be more homogeneous. Typically, this is the odds ratio scale. When data are sparse, we also prefer the more parsimonious model. This can be easily translated to risk report a risk difference and NNT for a given baseline rate. We have additionally reported risk differences and NNT (presuming a specific baseline rate) from odds ratios. For KQ 5, we have calculated NNT for a given baseline rate.

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #3	Results	I would encourage authors to report the magnitude of the effect estimates including NNTB for all relevant outcomes.	The scale in which the synthesis is done should be the scale where data appear to be more homogeneous. Typically, this is the odds ratio scale. When data are sparse, we also prefer the more parsimonious model.
TEP/KI Reviewer #3	Results	KQ5: The statement regarding QoL is entirely correct as it stands. Generic QoL did not differ across treatment groups but for disease-specific QoL a small difference in favour of topical antibiotic-corticosteroid eardrops was found, citation van Dongen et al. NEJM 2014: "At baseline, the generic and disease-specific health-related quality-of-life scores indicated good quality of life and were similar across the groups. At 2 weeks of follow-up, the change in the generic health-related quality-of-life scores did not differ significantly among the study groups. The changes in the disease-specific health-related quality-of-life scores at 2 weeks were small but consistently favored eardrops (Tables S2 and S3 in the Supplementary Appendix)." I would therefore encourage authors to either add "generic" to current sentence or slightly amend the current statement.	Discussion expanded to include both generic and disease-specific QoL as suggested in both Full Report and ES.

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #4	Results	Concluding the results section with key bullet-points would be helpful. The detail is sufficient and appropriate, however, a succinct summary would facilitate reading.	We have provided a succinct summary of results in table E.
TEP/KI Reviewer #5	Results	Page 34, line 21. Listing outcomes for “otorrhea” as a general term is not very helpful, because otorrhea can be transient (which is of little to no concern), recurrent (which is of more concern, but usually readily managed), or chronic (which is of significant concern and is difficult to manage). Separating out the outcomes by type of otorrhea would provide more meaningful data to clinicians and patients.	We agree. In our discussion of KQ 3 we note that definitions used in individual studies are "highly variable" and conclude that otorrhea "is particularly complex to characterize". We have added the proposed classification of otorrhea as "transient, recurrent or chronic" as an example in Future Research Recommendations (ES & Full Report)

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #5	Results	Page 38. My earlier comments about the abstract (lines 46-9) pertain here as well.	We have redone the network meta-analysis, removing the studies which do not contribute indirect information (and compare dual treatments, e.g. oral antibiotics and topical drops). The posterior probabilities (shown in Table 15) from the NMA are consistent with the overall conclusion that topical treatments are more effective than oral antibiotics. The combination of direct and indirect information suggests that antibiotic-glucocorticoid drops are superior, and is suggestive, but not conclusive for antibiotic (alone) drops. Table 15 and following discussion and conclusions have been edited.
TEP/KI Reviewer # 6	Results	The results are clearly presented. Where detail is not available it is due to the level of detail in the reports analyzed for the analyses.	Thank you.

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #7	Results	I checked especially on some particular studies to learn whether they were included or overlooked. In fact, the ones I knew in advance and checked for inclusion were present and accounted for. I think the amount of detail is appropriate and that characteristics of the studies were very clearly described. Also, the key messages were explicit and applicable. I was especially impressed by the figures, tables and appendices. All were highly informative and appropriate.	Thank you.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Results	P 80: Line 40: Should say hearing level is primary intermediate outcome in the methods section and the results.	We pre-specified intermediate and quality of life and patient-centered outcomes. Meta-analyses were performed for specific outcomes when studies reported sufficient data. We do not distinguish primary and secondary outcomes. For KQ1 we report two intermediate outcomes: hearing thresholds and duration of middle ear effusion. Text edited to add a brief summary discussion of duration of middle ear effusion. "There was a trend toward shorter duration of middle ear effusion in children treated with TT with adenoidectomy and TT alone. However, credible intervals are wide and cannot exclude a null effect for any intervention compared with watchful waiting."
Peer Reviewer #1	Discussion/Conclusion	P 81: KQ2 Line 19: Describes that pooled results were not provided due to the small number of studies, multiple interventions, and outcome heterogeneity; shouldn't this be in the results section?	In KQ2:Outcomes, we have the text: " We did not quantitatively pool the results, primarily because of the small number of studies and substantial heterogeneity in reported outcomes."

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Discussion/Conclusion	Page 82, line 40: Re KQ#1: “Limited number of studies (less than 9, out of a total 68), each using different outcome definitions No quantitative synthesis done” This is in the summary table but not in the results.	Table 18 edited to read "Limited number of studies (8)" - now agrees with results
Peer Reviewer #1	Discussion/Conclusion	P83 Lines 27-28 “Many of these trials were performed prior to widespread use of conjugate pneumococcal vaccines and in an era where antibiotic resistance was less common.” Should say what this can mean in terms of interpreting these results and generalizing to today.	Full Report: added: "It is unclear whether these or other factors affect the relative (current vs. historical) benefits of TT placement for recurrent AOM."

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Discussion/Conclusion	I recognize that it's standard practice, but I think wording findings inconsistently with respect to their positive or negative valence, as in Table 18, interferes with comprehension (by contrast with the Detailed Strength of Evidence Assessment table in Appendix H, which is much clearer). For example, saying that strength of evidence is low to support the conclusion that "periods of watching waiting do not result in consistently worse.outcomes") seems to invite the interpretation that watchful waiting has negative consequences (rather than that the outcomes of watchful waiting are no worse than outcomes of TT). Similarly, some conclusions in this table are stated as unambiguous assertions ("... TT results in short term improvements...") but others are stated using the word "may", which implies that they are only possibilities (e.g., "Treatment with TT may not improve quality of life"). I would prefer to see all the conclusions stated as direct assertions so that implications about possible effects and relationships come only from the effect sizes and strength of evidence ratings.	We retained table 18 and moved it to into the full report, harmonizing as per the reviewers comments, to the extent possible.
TEP/KI Reviewer #1	Discussion/Conclusion	The findings and limitations are well described for each Key Question. The future research suggestions well described.	Thank you.

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #3	Discussion/Conclusion	Yes—clearly stated. Limitations of studies are clear, and to my knowledge no omissions occurred.	Thank you.
TEP/KI Reviewer #5	Discussion/Conclusion	The limitations of the available data and the implications of the findings are adequately and fairly presented.	Thank you.
TEP/KI Reviewer #7	Discussion/Conclusion	The implications of major findings were clearly stated in my view. Also, I found that the limitations of the review/studies were adequately described. As mentioned earlier, I thought the review was very thorough; I did not detect the omission of any important literature. Finally, I found the future research section to be clear and, also, that it could easily be translated into new research efforts.	Thank you.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Clarity and Usability	the tables in the various appendices are extremely valuable; they might be easier to use if they were all organized alphabetically by last name of PI (as is done for some but not all of them).	Appendix tables have been regenerated, sorted Alphabetically by author
TEP/KI Reviewer # 1	Clarity and Usability	My concern is if this systematic review will further improve the quality of recommendations in the healthcare decision making process in regards to tympanostomy tube insertion in children compared to the present "Clinical practice guideline: Tympanostomy tubes in children" as approximately 80% of the articles are published 2010 or earlier	We agree and encourage new pragmatic trials, particularly for KQs 1 and 2. For KQ1, we note that "The majority completed enrollment more than a decade ago." No trials were found subsequent to 2006 regarding KQ4. The Kujala 2012 (KQ2) and the von Dongen 2014 trials provide at least some incremental evidence.

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #2	Clarity and Usability	Tympanostomy tube placement is common and over use has been suggested by some. This CER mirrors SRs done in the past, as well as evidence assessments done for CPGs. The strength of evidence for benefits of tubes is low for most measures. This report is clear and the evidence review and analysis comprehensive. Hopefully the knowledge gaps and lacking evidence will encourage additional research.	We agree.

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #3	Clarity and Usability	The authors correctly state that "The majority of trials utilized similar inclusion criteria and subgroup analysis of higher or lower risk groups is sparse." However, some IPDMAs are available investigating which subgroups of children do benefit more than others from TT or adenoidectomy and TT for otitis media (Boonacker et al. Health Technol Assess. 2014 & Rovers et al. Arch Dis Child 2005). It may be worthwhile to acknowledge these IPDMAs and add some of their results to the review? In particular, the IPDMA of Rovers et al. 2005 has been influential by informing current clinical practice guideline recommendations on this topic (NICE Surgical management of otitis media with effusion in children).	We have briefly summarized conclusions from Rovers 2005 and Boonaker 2104 in discussion section (ES and Full Report).

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #3	Clarity and Usability	Although recognising the importance of surveillance of otopathogens in future TT trials, I don't think that bacterial cultures are helpful for outcome assessment in future trials. Otorrhea in itself is considered the result of AOM irrespective of otopathogen involved, so I don't agree with the authors that bacterial cultures will substantially benefit outcome assessment in future trials.	We agree that the benefit of culture is unclear. However, there does appear to be some potential to identify superinfections associated with chronic otorrhea. Text edited for clarification. "Bacteriologic evaluations performed in the research setting may assist in differentiating otorrhea resulting from infection with organisms associated with AOM (e.g. Streptoccus pneumoniae, nontypable Haemophilus influenza)from superinfections with organisms associated with chronic otorrhea (e.g. Staphylococcus aureus and Pseudomonas aeruginosa).{cited: Idicula WK, Jurcisek JA, Cass ND, et al. Identification of biofilms in post-tympanostomy tube otorrhea. Laryngoscope. 2016 Aug;126(8):1946-51. doi: 10.1002/lary.25826. PMID: 27426942.} "

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #3	Clarity and Usability	The authors state that "In studies with complete followup, the intervention itself is subject to natural attrition due to extrusion of the TT over time, which complicates the interpretation of intention-to-treat comparisons." I however, respectfully disagree with this statement. The extrusion of TT is inherent to this specific intervention and should be taken into account when balancing the pros and cons of this surgical procedure. As such, I don't think that extrusion of the TT over time complicates interpretation of ITT comparisons.	We agree. Text (ES and Full report) edited to remove "In studies with complete follow-up, the intervention itself is subject to natural attrition due to extrusion of the TT over time, which complicates the interpretation of intention-to-treat comparisons."
TEP/KI Reviewer #4	Clarity and Usability	Yes, well-structured and well-organized. Reading the document in this (on-line) format makes it somewhat difficult to judge.	Thank you.

Commentator & Affiliation	Section	Comment	Response
TEP/KI Reviewer #6	Clarity and Usability	The report is well structured and organized. The conclusions are relevant to clinical decision making.	Thank you.
TEP/KI Reviewer #7	Clarity and Usability	In summary, I found the report to be exceeding well structured and organized. The main points were quite clearly presented. The conclusions are highly relevant to policy and practice decisions. I certainly believe this compendium of information, so nicely organized and presented, will contribute to the acquisition of new information and understanding.	Thank you.
Peer Reviewer #1	Appendix	p 169 Appendix G: would be helpful to have actual p values here instead of just NS	We have chosen to emphasize effect sizes of results reported to be statistically significant. P values are often not fully reported in original reports.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Appendix	P204: I'm not sure premature extrusions is a valid adverse outcome. How is premature defined?	We note that study specific definitions of adverse events (including premature extrusion) are poorly reported and/or highly variable between studies. Study specific definitions are given in Appendix I. We have further noted that definitions are often different, poorly defined. Future Research Recommendations (ES and Full Report) "In some cases, e.g. premature extrusion, one author's premature extrusion may be another's time extrusion, depending on the duration of anticipated need.{Kay, 2001 #654}"
Peer Reviewer #1	Appendix	Page 235: are these the references for the appendix? They could be labeled this way	Appendix I edited to specify "KQ3 References"