Appendix I. Peer Review Comments and Responses [Editorial comments were excluded]

Area Addressed	Comment	EPC Response
Title	• Although entitled “Diagnosis and Treatment of OME” the report *does not* deal with treatment. Some other word would be more accurate, perhaps “management.”	• Title has been changed to “Diagnosis, natural history, and late effects of OME”.
Abstract	• page v – the objective statement makes no mention of treatment for OME; further the search strategy does not include a “treatment” module. If treatment is not addressed then the title of the report should be modified. • On page "v", I suggest that the first sentence of Objectives would read more smoothly as: "the impact of otitis media on hearing and on long-term speech and language development, and the operating characteristics...." • page "v", the first sentence of Search Strategy should read consistent with page 2; for page "v", I suggest "including otitis media, otitis media with effusion, non-suppurative otitis media, fluid ear...." • The refinement of the purpose of the guideline on page V in the objectives section states the purpose of the guideline more clearly in comparison from report I reviewed in August of 1999. • Move **(Results: line 1, page vi)** of the results to the conclusions. The remainder of the results are clear within the structured abstract. • Page vi - the ages of the patients are unclear. Outcomes were followed in children up to 22 years of age. A general statement regarding the ages of patients in the studies would be helpful. • Page "vi", the last sentence of Data Collection and Analysis would, I think, be better by omitting "in" at "non-English language", and omitting the comma after "craniofacial deficiencies" [I prefer the word anomaly rather than deficiency]. • In last sentence, add “reports *in non-English language”* • page "vi", in the first sentence of Main Results, the phrase "of ears" disturbs me. I'd prefer "patients" or "children" -- as pages 9 (lines 4-7 from the top of the page) and 154 (lines 1-5 from the top) state. However, I do realize that the data depicted in Tables 21 and 22 do not allow such a change in words on page "vi". • Page vi - what is meant by early life otitis media - does that include	• The title was changed to read “Diagnosis, Natural History and Late Effects of OME”. • Sentence revised. • Revised. • Noted. • Revised. • Revised • Revised • Added • Data was reported as ears for these particular estimates • Revised • Added: ‘defined as greater or equal to 20

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response

	AOM and OME.
	• Page vi - hearing loss should be qualified - mild, moderate, severe, profound.
	• page vi – bottom paragraph, how is "hearing loss" defined – qualitatively or quantitatively? Is this any detectable hearing loss or significant hearing loss; are the percentages for resolution of OME cumulative or simply point estimates at the stated time periods?
	• Page vi, 3rd paragraph and elsewhere. It is not clear whether the resolution rates are cumulative or represent the total resolved to that point in time. This is particularly true because the numbers are so similar. Some comment is needed about why it might seem that no one recovers between 6 weeks and 3 months. Obviously, it is because the results come from different studies, but this is not mentioned anywhere in the document. It might be helpful to select the studies that measure at multiple time points and show the progression over time (or use the formulas that some studies have to show what the progression appears to be).
	• Page vi, 3rd paragraph. Your statement about hearing loss in children in later life needs to be clarified as to whether this is with or without treatment of the effusion (or some hint that we don't know about treatment effect on this outcome).
	• I read the report in sequence, from the abstract to the summary to the text. In the abstract and summary, it would have been helpful to include very early in these sections, more context about:
	  • What is OME?
	  • Why is there concern about OME?
	  • What are the possible interventions for OME?
	Without knowing this information, the summary of evidence seemed a
		dB threshold at any frequency with or without treatment."
		• Added definition of hearing as above. We deleted all meta-analyses in Table 23 and only presented the third and fourth meta-analyses of point estimates in Table 24 and the third and fourth meta-analyses of cumulative estimates in Table 25 and made appropriate changes in the summary, results, and conclusions. We commented on this issue in the results and conclusions.
		• This issue has been discussed in Conclusion and Future Research sections. Also, see immediately above.
		• Definition of hearing loss included a phrase on treatment. A discussion of treatment was added in Results.
		• We added definition of OME early in Abstract and Summary.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	little abstract and dry. If you need to reduce the length of other parts of the abstract or summary, I think most readers will be less interested in the detail given about the methods than in the contextual questions above. • p. vii and other places: You refer to the need for a "coordinated uniform approach using a rational conceptual framework." This is a little abstract. It would be good to show or talk a little about the examples, or to try to put what this means in plainer English for clinician-readers. • page vi-vii – no mention of treatment in "Main Results" section. • page vii – conclusions – what percentage of OME persists after 3 months?; it would be helpful to define what is meant by "early life otitis media." • **(Conclusions: page Vii, line 1).** I recommend simplifying line 1 of the conclusions in the structured abstract so that is presents more generalization from the results and not just a restatement of the results. • Bias: It may be helpful to add a section on the Biases associated with the types of studies selected, i.e. case control, cohort, and randomized controlled trial.	• Sentence revised and concept clarified. • Title changed. Treatment will not be addressed in this Report. • 59-78% persist after 3 months. Results revised. Definition of "early life otitis media" added in appropriate places in Abstract, Summary and Results chapters. • Revised. • A discussion of study design has been included in the Limitation section.
Summary	• The review process and scope is clearly defined **(page 1)** and in more detail **(page 34)** of the report. Some bias exists in the composition of the panel as there is heavy representation toward specialty composition which may influenced selection of diagnosis and long term outcomes as key questions developed for this evidence report. The authors did review previously developed guidelines from 1994 but chose a different area of focus based on the rankings from the panel. The information presented is most useful for beginning to structure uniformity in randomized clinical trials and prospective studies. Primary care practitioners and consumers may be more interested in treatments as the immediate results are observed in primary care. This may be a reasonable focus for building on future evidence based studies if the interest was to focus on the consumer perspective.	• Comments noted. • Added definition of OME at the end of

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	• Page 2 - the definition of AOM should be provided. • I miss the four research questions. They are only mentioned in the section reporting the evidence. • page 2: The search strategy is not very clear described. The reader can not know yet what the different concepts or modules are. You use the words *concept* (the otitis media with effusion concept*) , component* (the speech and language component…*)* and *module* (the natural history module) not every time in the same way. I was confused by reading this section the first time. • In discussing pneumatic otoscopy here and later in the report, it would be helpful to note whether the report conclusions should be limited to examiners trained in the technique, or can be generalized to untrained examiners. (see Comment, pg 88.) • Page 3 - the degree of OM was graded in some ways - is this AOM or OME? • page 3 – was age limit "up to 22 years" (abstract) or ""under 23 years"? Is age limit 22 years or 21 years? • Page 3 last paragraph: Effect of OME in first three years of life on later language limits the question. • Page 3 and 4: Diagnostic method. Algorithms are generally accepted in the diagnosis of OME because the accuracy of most diagnostic methods is disappointing. Excluding the literature using algorithms to diagnose OME limits this evidence assessment. • Page 4 - generally, researchers refer to adherence rather than compliance. • Page 6 - hearing loss should be qualified - mild, moderate, severe, profound. • Page 6 last sentence. This sentence is ambiguous. Better might be to	Reporting of Evidence • Sentence revised mentioning the 4 key questions. • The search strategy was re-written. • Included the following statement in the conclusions: "The important question may be what degree of training will be needed for the clinician to be as effective with pneumatic otoscopy as in the studies reviewed in this report." • OM is considered as a general term including all types. • Age limits were clarified both in Abstract and in Summary. • Comment noted. • This point was included as a limitation to this assessment. • Changed. • Added: 'defined as greater or equal to 20 dB threshold at any frequency." • Changed • Corrected. • Only rate difference was reported.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	say something like "Neither the studies pooled for the rate difference nor the studies pooled for the risk ratio were statistically heterogeneous." • Pg. 6, line 3: "synthesize" was misspelled as "synthesis" • p. 6, last para: Various measures of risk and rates and ratios are given. Can you please help the reader decipher which of these is most relevant to interpret? • P. 6, Line 6:Unsure what you mean by "underlying concepts measured in each group were of questionable similarity." • P.6, Line 13: Hearing loss, couldyou be more specific such as degree degree of loss? • Page 7 - the skill of the performer of pneumatic otoscopy is critical. This is a recurring theme in the report. Either clarification is needed, or some statement that who performed otoscopy was not examined in the analyses. • page 7 – is there data on variation in diagnosis of OME using pneumatic otoscopy by specialist (e.g., pediatrician, family physician, otolaryngologist)? • The report defines the positive and negative predictive values for the pooled at the pooled prevalence rate of the studies involved (63%). It might be useful to also calculate what the +PPV and –PPV would be at an estimate of outcome (e.g., 3 months) prevalence rate to give the reader an appreciation of the impact of prevalence on test performance. • page 8, "Diagnostic Method fo OME", I would like for there to be elaboration as to why more comparisons could not be made. This 9-page summary will likely be the most-read portion of the report, and must be most communicative. • Page 9, last sentence. This isn't clear. I would doubt that most algorithms are so complex that computer programs are used in practice. The actual instructions for applying the algorithm are probably what is needed. If a program is used, then, of course, it should be supplied.	• Paragraph rewritten. • Definition of hearing loss had been added in various places in the document. • Addressed in Results and commented in Summary. • Addressed in Results and commented in Summary. • We revised our Results according to the recommendation. Figure 7 which plots PPV and NPV by prevalence rate was added. • We limited the minimum number of studies to be analyzed by meta-analysis to 3 because lower than this it would not be statistically sound. • Sentenced changed.
Introduction	• p. 11: Did you have an operating definition of OME for this review? I	• Yes, the project definition of OME has

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	realize you didn't get consensus on a single definition, but even the simple definition given in the conclusions section would be helpful. • Pages 12-16. The discussion contains no reference to the data in our epidemiological report (Paradise et al 1997) although the report is included in the reference list. In particular, no mention is made in the discussion of low socioeconomic status as a major risk factor, and the issue of daycare attendance is dealt with in isolation rather than as one type of exposure to large numbers of other children. Specifically, being a member of a large family of other children is also a risk factor. That report also provided detailed data on the prevalence of OME. • Page 12 - the final two sentences seem to be at odds with one another. • Page 13 - some discussion of diagnostic coding of otitis media would be helpful. It is not clear how these data sets distinguish AOM from OME. • Page 13, first paragraph. This sentence is a bit unclear, because it isn't clear what the denominator is (it can be dug out, but it isn't clear). I think it would be better to use a small inset table that shows the breakdown by age since reading requires one to build one to see what's going on anyway. • Page 13 3rd parag. 2nd sentence. This is so obvious as to not need stating since pediatricians only see kids while the others have larger patient bases. Sentence deleted. The next sentences are also not clear because the denominator isn't well stated. Is it per unit of population or per member of the physicians group? • Page 14, second full parag. It would be interesting to also have earlier numbers for myringotomy with tubes to see if there was an impact from the earlier guideline. • p. 15 last sentence. Since there are many articles indicating increased risk of acute otitis media and number or procedures for placement of tympanostomy tubes in children who attend day care contrasted with children who are in home or family care, there must be more OME in children in day care. I don't know why the guideline panel was limited in their statement but you should not repeat the mistake. • Page 16 table (and many other tables). Something is wrong with the table settings because the tops of the characters intersect the table	been added to the introduction and methods. • The purpose of this section on prevalence is to establish the importance of OME. The section which previously dealt with risk factors has been deleted, and risk factors are mentioned • The two sentences are consistent, one addressing aom and the other addressing ome. • Defined at appropriate places. • The Introduction Chapter has been revised. • Paragraph revised. • Comment noted. • We did not change this statement because there was not sufficient evidence to support such a change.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	lines. This is annoying, particularly in the evidence tables and can probably be fixed with a global change in your style or somewhere.	• Tables deleted. Introduction greatly revised.
	• The scope of work reviews previous literature that analyzes the natural history (page 16) and common outcomes (page 24). The information is comprehensive, however there may be biases[2] in this information which is not described in detail in this report such as how some of the cohort studies are different from the general population, lost to follow up and whether there was any confounding. Given the heterogeneity noted in studies done previously, it is not clear if all outcomes are included or are the subjects representative of the sample. There is no mention of potential gaps in practice or newer outcomes, such as health status and satisfaction with treatment which would be a subject of future research.[3] (Page 10-29). 2 Calognge, N. *Examining the evidence.* Evidence-Based Medicine Briefing. US. Capitol Building Washington, D.C. Kaiser Permanente. January 28, 2000. 3 Stuart, M. *The evidence-based medicine process.* Evidence-Based Medicine Briefing. US. Capitol Building Washington, D.C. Kaiser Permanente. January 28, 2000.	• We mentioned that "potential gaps in practice or new outcomes, such as health status and satisfaction with treatment which would be a subject of future research" in Future Research chapter.
	• Page 17 last parag. I would avoid the use of phrases like "Interestingly". This implies a judgment by the writers that is probably inappropriate in this type of document.	• Paragraph deleted.
	• Page 18, line 8 - is should be are.	• Corrected.
	• The remark about assessing middle ear function, at the bottom of p. 18 reads oddly. Is that not what the entire topic is about? Or is the point being made that clinical assessment relies too much on otoscopy (structure) and insufficiently on tympanometry (function)? If so, say so.	• Paragraph deleted.
	• pag 19 Rosenfeld is cited that OME should be managed by a multidisciplinary team. In the Netherlands we wait and see, and most children did not even see their GP. Why should you manage a disease that isn't a disease at all and even when it is, it will be self limiting in most cases.	• Paragraph deleted.
	• Page 20, section e. The word "as" is omitted.	• Paragraph deleted.
	• Page 21 Middle of page. The term subacute OME is not defined anywhere in this document that I've noticed. It should probably be	• Paragraph deleted.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	omitted or else defined. • Page 21, point 1c: Should this be broken into two points, one for "general hygiene maintenance" and a second for smoke avoidance? (At least restate the stem so these two do not run together. • p.21: Point 2(c): It is undesirable to perpetuate in the literature, without questioning or reservations, the unfortunate wording about giving adenoidectomy for extruded tubes, for 3 reasons. Firstly re-insertion of tubes should depend on recurrence of fluid and hearing loss confirmed over time, with regard for time elapsed, rather than the mere fact of extrusion, which depends on the individual, on the tube, and possibly on infection, and extrusion as a dominant criterion may lead to over-treatment. Likewise where adenoidectomy is going to help in a child that meets a clear overall criterion for initial surgery plus specific indicators including age, adenoidectomy need not await the 2nd set of tubes. The Report should avoid lending its authority by uncritical repetition of this simplistic rule. • Page 22 - actors should be factors, unilater - should be unilateral. • Page 22, para 1: Is "Certain actors" meant to be "Certain factors"? • page 22, I suspect a misprint of "actors" instead of "factors"; "actors" gets the message across, but may be considered slang. In that same paragraph, "dysarticulation problems" seems redundant; is the term not "articulation problems". • Page 22, first full paragraph. The word "actors" should probably be "factors". However the entire paragraph smacks of being a recommendation and is not really appropriate for this document, particularly since treatment is not covered in this review. • Page 22, last paragraph: Item 1, Hearing loss" suggests that the average hearing loss is 27db. This is potentially an important point, given that decisions regarding intervention, as discussed by the early 90s OME Panel, might depend on the hearing level. I looked at the literature some years ago, and found a number of studies that provided enlightening data. These references, and any other studies known to the team, should be discussed in an additional paragraph at some point in the Introduction. There is a suggestion in the literature I reviewed that the average hearing loss might trend upwards as one moves from	• Paragraph deleted. • Paragraph deleted. • Corrected • Corrected • Paragraph deleted. • Paragraph deleted. • Introduction greatly revised.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	community, to primary care, and to specialty based studies. (Culpepper L, Froom J: Otitis media with effusion in young children: treatment in search of a problem? J Amer Board Fam Pract 1995;8:305-16.)	• Paragraph deleted.
	• p. 23 nos. 3,4, and 6. Since OME is not suppurative it could not be responsible for the supparuative complications of AOM such as mastoidits, petrositis or labyrinthitis.	• Paragraph deleted.
	• Page 13, page: Mastoiditis, petrositis, suppurative labyrinthitis, and facial paralysis all should be moved to the list of potential complications more relevant to AOM, with this latter list expanded by dropping the word “intracranial”. The concept of infectious extension of AOM beyond the middle ear seems to be what is most important, not the intra or extracranial site of extension.	• Paragraph deleted. • Paragraph deleted.
	• p. 24: This section doesn’t seem to differentiate between antibiotics for prophylaxis, vs. antibiotics for treatment.	
	• Page 24. The statement, “Paradise (1995) also listed susceptibility to middle ear infection and impairment of psychosocial development as additional OME complications” requires qualification. The text in that report made it clear that any possible developmental effects were uncertain.	• Paragraph deleted. • Corrected.
	• pages 22-24, Outcomes – what is incidence of complications that are described? Some figure should be included since these are outcomes that generally we would want to avoid.	• Paragraph deleted.
	• Page 25 - why not put the OME guideline findings (numbered 1-4) on the graph above, it would make for easier comparison.	.
	• Page 26, bottom of the page. You might want to recheck the studies you cited on adenoidectomy. I seem to recall that they dealt primarily with older children. You might mention this in regard to the recommendations quoted.	• Paragraph deleted.
	• p.26: Whatever other authors may have said, it is irresponsible to give consideration to systemic steroids as a main line of treatment without similar cautions, and reviewing them is totally unnecessary when servicing the 4 key non-treatment questions addressed.	• Paragraph deleted.
	• page 27, in the paragraph in the middle of the page, "inconsequential hearing affected less than 0.5 dB" should be re-written. Depending on the size and mass of the tympanostomy-tube, and the location in which	• Paragraph deleted.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	the tube is placed in the tympanic membrane, the ear may have a low frequency conductive hearing loss of 20 dB. • Page 27 middle of page. "Interestingly" pops up again. I believe that when looked at swimming before we also found that it had no impact on otorrhea. You might be surprised, but I think this has been known for a while. • Page 27, bottom. You mention three meta-analyses but then list two outcomes. This may be confusing. One of our meta-analyses was tubes vs. myringotomy so I'm not surprised you omitted it, but mentioning three and then listing two is confusing. • p.27: Swimming. There are many studies of this, mostly showing no difference. The point is that they are mostly underpowered. • Page 28 middle. "Billinon" should be "billion." • Page 29, 1st paragraph: Just as an expansion on the note above on hearing loss levels, as I indicated in my JABFP critique of the earlier Panel report, the rate of intervention is highly dependent on the threshold hearing level adopted, and that panel adopted the 20 dB level with almost no discussion in its rush to finish discussion on the last day of its last meeting. A different level might cut costs dramatically. • Two generally important issues are not clear from the introduction: (a) why this major effort was undertaken now, and (b) who is considered to be the main audience? On point (a), reasons might be emergence of new results, pressures of economics e.g. via HMOs, public opinion, a federal review of medical training etc. On many of the questions there have been recent (attempts at) meta-analyses. Although this does not pre-empt the issue, it does limit the scope for radically new conclusions. The funding agency or SCEBPC must have considered this as a global issue at some stage. • Many of the summary statements in the introduction are not from primary sources but are convenience citations. These can be so summary that they could be misleading if quoted out of the context of the original studies which they summarize. An example occurs on p 17: "Rosenfeld (1994)…. duration of 6 years." The point about both the cited studies lies in the particular selectivity of entry that leads to the estimate. The citation of these estimates is of little value out of that	• Paragraph deleted. • Paragraph deleted. • Corrected. • A statement was added in the Summary and Conclusion chapters to alert caution. • Introduction has been greatly revised. • The Introduction has been greatly revised. • The Introduction has been greatly revised. • The Introduction has been greatly revised.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	context. • If the Report is to be widely distributed as such (rather than being a source for 4 review articles on the KQs) then the introductory chapter needs further work, preferably to cluster it better around the four questions actually addressed. • I found the introduction a little disconcerting in that selected reviews were cited with tables of conclusions by the review authors. It was unclear to me (again because I did not analyze the bibliography in detail), if these were the only reviews available or how these articles were chosen for detailed citation in the introduction. In the evidence report itself, as opposed to summary articles, I would suggest adding sentences such as "There were XX reviews of the subject by...." Then summarize each published review. If not all available reviews are summarized, state what criteria are used for choosing what to cite. This would be more consistent with the exhaustive nature of the evidence report and what follows after the introduction. • Introduction – the inclusion of an introductory section describing the results of several review articles seems contradictory for an evidence-based report. Regardless of the intended justification, strong consideration for removing this section is recommended. • It was surprising to note the footnote explaining the introduction section (page 10) as an “overview on otitis media.” The inclusion of a traditional literature review within the body of an evidence report he introduction section seems contradictory, especially since selected members of the expert panel produced a significant amount of the cited literature. These persons are cleared well qualified but it might suggest a possible bias in terms of key questions and findings to others who are less familiar with their work. How did other reviewers reaction to the inclusion of this section?	• Recommendation taken. New Introduction was written. • Noted. New Introduction was written.
Methods	• p.30: "Variation in practice". Good to introduce this issue, as a main objective of doing reviews and disseminating them is to reduce such variation . However more emphasis is needed on the findings about variation in OME, and why it is important. Clinicians' understanding of variation and its implications is poor. • Page 31 middle. Format the list of criteria in an outline form with	• Noted. • Done

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	indents. As it stands the criteria with subparts are very hard to decipher. • p.32 and 35: There is some inconsistency of terminology over "interaction". Popular use of the term is incorrect, as it is used to mean co-action including additivity or some other unspecified form of co-action. I think that the meaning intended here is the same and therefore also incorrect. Correct statistical usage means non-additivity when there is a significant interaction, i.e. sub-additivity , supra-additivity, or more exceptionally a cross-over. p.32 needs to be reworded, because it is clear from p.35 (KQ2) that separate main effects (co-action not interactions) for the dependent variable is what is intended there. Synergistic (supra-additive) interactions are indeed likely in OM, e.g. language effects and behaviour effects "more worse" in low socioeconomic groups, but unfortunately direct evidence for them is slender so far, due to conceptual failure of investigators and underpowered studies. A report having a statistician as author should take the opportunity for some terminological hygiene here and correct the incorrect popular use, not perpetuate it. • Page 33 middle (and elsewhere). The listing of gold standards uses the terms “vs” which implies that there is some comparison of the two items. Propose using the term “or” rather than “vs”. Further, it seems that looking at the meta-analyses done that only myringotomy was actually used as a gold standard. If this is the case, it should probably be mentioned here and elsewhere in the text. If not, then the meta-analysis results should be modified to indicate what was used. • Page 34 - don=t refer to the definition of AOM in a reference, just put it in this report. • Page 36-7. The description of the nature of Cochrane, Medline, and EMBASE is unnecessary in this type of document. Anyone who could possibly understand this document should be familiar with those sources or be able to find out. Also, EndNote is no longer from Niles Software, but from ISI Researchsoft. • Medline search (pag 37) Is “controlled vocubulary” the same as MESH terms? Could you describe the different concepts and modules a little bit clearer. I’m not a native speaker but I think that the word ‘cluster’ is more appropriate for all the different terms.	• P.32 lists the wording of the exact wording of the questions and was left as such. However, the correct wordings are noted and used wherever appropriate. • Agreed and changed. • Paragraph revised. • Left in for completeness. The EndNote version we used was from Niles Software. • The search strategy has been rewritten. • Even though the search used the word “placebo” , we did not use randomized controlled trials in the natural history

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	• Is it justified to search for 'placebo' when looking at a natural course of a disease? Placebos can influence the natural course in the same way interventions do. How many studies are describing placebo cohorts and does this influence the outcome of the evidence? (I have not looked for this my self due to lack of time) • Page 39 You mention interrater reliability statistics. Do you supply those anywhere in this document? • Page 41, question 1, #2. I don't understand this as a criterion. Does the other ear have to be the control or not be the control and why is this relevant for the natural history study? I would think you would be most interested in bilateral disease, but can't tell from this. • Page 41, questions 2&3. I don't understand why you excluded studies that were reanalyses of prospective studies. This review is essentially a reanalysis of prospective studies so I see no reason to exclude such studies, but don't really understand what types of studies you are referring to here. • Page 41, second paragraph: Criterium 3: degree of OME graded in some way. Specification of what is meant by degree -lenght of time, persistence, recurrence- would be helpful. • Page 41, question 4. I don't understand why algorithms were excluded. • Page 41. Question 4. You excluded algorithms as diagnostic test. But in real life physicians use algorithms all the time. Diagnostic research is extremely difficult because of the correlation between all small steps in the diagnostic procedure. A better way to investigate these problems is looking at different diagnostic common pathways. (But I guess that non of the retrieved studies has done this kind of analysis). • Page 42 2nd to last paragraph, #5. What were they blinded to since there was no treatment involved in the natural history? How was this assessed and why is it relevant in this case? • Page 42, last paragraph #5. What hospital stay? Everything here is outpatient, even tube insertion. I suspect you modified these criteria for your use and would suggest you include the modified criteria.	assessment. • This was done but not reported. • This criterion applies to RCT only as indicated. • The evidence report focuses on original data, which we then analyzed. Therefore, we did not need to include re-analyses of original studies, as the data were already included, and to do so may have led to double counting. • Examples added. • Technical experts are more are interested in the effects of individual tests and thus algorithms were not included. Also, algorithms would take more time to evaluate. • Same comment as above. • We are referring to blinding of the previous condition, not treatment, as stated in the Methods. • Changed to 'illness'. • "hospital stay" was corrected to illness.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	• Page 42: The 8 components of cohort studies under 2). What did you do with studies in general practice, where there is no uniform point in hospital stay ? The problem of making a consistent cohort is the fact that parents don't come to the surgery with exact data or complaints. Often a teacher or a school physician sends the children to a GP, who looks in the ears and "Ah, I see a little bit of fluid, …will you participate in a nice study" . Even in studies were a whole age cohort is followed, the exact time of duration can not be determined. • Pg. 42, 2nd line from the bottom: "were" was misspelled as "was" • Page 43, 1st full parag., #4. I'm not sure I would exclude or down grade a study that dealt with only one level of severity. I might analyze it separately, but wouldn't consider it low quality. • Page 43, 2nd full par. This paragraph is misleading in that implies that articles were excluded for quality reasons. I am under the impression from later that was not so. If there were exclusions, then the criteria for exclusion should be given. • Page 43 last par., Does partial resolution mean resolution in one ear or does it mean reduction in fluid levels (improvement in tympanogram, etc.)? • At page 43 you describe six components of a diagnostic research. May be you can discuss the problem of correlation between observations in the discussion section. • Page 43 - more clarity is needed around the quality reviews. Was each quality indicator simply assigned a value of 1 and the numbers added up. This implies that all quality indicators are created equal (not true). This is a potential limitation and should be stated as such (see below). As a consistent theme, I could not find analyses based upon high quality reports - this is critical. • Page 44 - funnel plots check for publication bias and other biases (BMJ, Egger). However, these funnel plots have very few points and are probably not reliable • Page 44, middle of page, last sent. in par. I'd be interested in knowing	Concern was dealt with in Conclusions. • Corrected • We did not exclude any studies based on study quality. We agreed that sensitivity analysis should be done, if adequate number of articles is available. • Paragraph revised. • Revised the phrase. • The six components referred to quality of diagnostic studies. Discussions on study quality and outcomes were added in the Results chapter and Limitations section. • The following was added in the Methods section: 'Each component of the quality was assigned a score 1 if present and 0 if absent. The total score was the sum of the components.' The study quality issue was addressed in Results and Limitations sections. • Agree. A caution of interpreting the funnel plot results with small number of articles was provided in the Results section. • Comment noted. • We checked 100% and corrected all

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	whether your manual scan found anything and how many. • Page 45, top. How big a random sample did you cross check and what was the error rate detected? I'd be interested in your results using highly trained reviewers. • Page 46 top, barotrauma shouldn't be capitalized. • Page 46 middle. Why didn't you go ahead and combine when you only had two studies? At a minimum, reporting a weighted mean would be helpful. Remember guidelines still have to be built. Also this paragraph suggests that you are pooling effect sizes, but later it looks like you use rates (a good choice) rather than effect size. • States on p. 46, para 2: "Furthermore, the type of study is an important consideration for the assessment of natural history. The stratified random sample of a broad well-defined population forms the best evidence whereas a single arm of a clinical trial represents worst evidence. For this evidence assessment we used only prospective cohort studies." There are several problems with the above statement. First, most prospective cohort studies use unselected or population-based samples with OME detected by screening. This group of children often has transient and asymptomatic OME that would never have reached the healthcare system in the absence of a systematic detection program. The result is to have "rosy" estimates of natural history (up to 44% at 1 month!) compared with the more meaningful group of children with OME sufficient to warrant seeking medical attention. The control groups in clinical trials of medical or surgical therapy better represent this latter group. Second, I don't see how you can condemn clinical trial control groups as the "worst evidence." Usually these groups have much better methods of detecting OME and documenting duration and resolution compared with simple cohort studies. They also tend to use pneumatic otoscopy (often as part of an algorithm), instead of tympanometry alone which is the typical measure in nearly all cohort studies. Your own analysis shows that pneumatic otoscopy is superior later in the report. Further, the huge variability in resolution rates based on choice of	discrepancies. However, we did not keep track of the error rates. • Changed • In the random effects model, we need to estimate the between-study variation. Estimating variance is a difficult problem under any circumstances and a sample size of at least three is required. Also, changed effect size to rate in paragraph. • We revised the statements and added a section to the Conclusion which includes some of these comments as appropriate.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	tympanometric criteria (B->A vs. B/C2->A) creates additional problems in interpretation. There is certainly some bias introduced by the restrictive selection criteria in most RCTs, but I do not believe this makes them useless raw material for determination of natural history. By restricting to the natural history analysis to cohorts only, you immediately eliminated the ability to assess most of the OME "types" deemed important by the expert panel: a) OME after discrete AOM episode (very important to clinicians!), b) OME for weeks or months (typically represented by control groups in RCTs of medical therapy), and c) OME 3 months or longer (typically represented by control groups in RCTs of surgical therapy). Basically, all studies in which duration of OME was prospectively documented (eg, RCTs) were excluded! • pag 46: 'The first step of the analysis was to obtain a distribution of studies by the 5….' Which 5?	• The 5 diagnostic groups of the natural history question: (a) OME persisting after a discrete episode of acute otitis media, (b) newly diagnosed OME of unknown duration (unilateral or bilateral), (c) OME persisting for weeks or months (unilateral or bilateral), (d) unilateral OME lasting 3 months or longer, (e) bilateral OME lasting 3 months or longer. This has been added for clarification. • Statement taken out.
	• I missed the ranking of the distribution of non-treatment factors (the next sentence) in the description. Please make it a little bit clearer here. • The whole section on supplemental analysis need subheadings to make it more clear to read. • p47 (bottom): These seem to be the risk factors for OM and not necessarily co-factors for the ultimate outcome. Only a few of them e.g. socioeconomic status would be expected also to be strong co-factors for the dependent variables of interest, (here speech and language), in the way that the wording implies, working other than through OM. It surely does not matter what combination of all the RFs the child has, although it would be interesting and methodologically sophisticated to quantify composite risk as well as superficial severity of disease as independent variables. Confounding only comes into the issue when badly controlled group designs fail to equate groups for co-factors. An	• Subheadings added. • Noted. The use of words was corrected and a statement has been added to clarify the concept.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	OM risk factor can also be a confounder, but only if there is reason (e.g. 1 published article of high quality) suggesting a separate causal path from the risk factor to the dependent variable that does not act chiefly through otitis media, or an article in another area where there may be a causal path (e.g. breast feeding and language or intelligence) but where the authors of the article were ignorant of a likely role for OM as the mediator, so it was not controlled for. All this needs to be made clearer. • Page 48 2nd full paragraph. As I read this I wondered how you dealt with the different test involved. Later I found out you didn't do meta-analysis for that reason. I'd mention it here. • Page 48 3rd full par. It isn't clear what is meant by the second part of the question. If one looks at the question on page 46-7, it appears you are referring to the third question (or the second part of the specific formulation.) I'd restate exactly what you mean here without assuming the reader remembers the exact question structure from two pages away. Further when the questions appear in the text, I'd consider using boldface font to make them stand out. They get lost in the current formatting and are hard to find. • Page 48: Why do you describe plans to do a meta-regression, while you don't do it. Leave this information out, it only confuse readers. (Put it in your letter to your funding organisation) • Page 49 first full parag after question 4. As worded, "gold standard" should be "gold standards". In the current wording one could assume that all tests in combination were used. • Page 49 last full par. There is a strange tense shift in this sentence. • Page 49, bottom. You should probably indicate what the outcomes are that you are analyzing. • p. 51 – should say nurse practitioners instead of nursing. I think that there were only NPs serving as the technical panel expert and the reviewer. • Page 53 it is Family Voices, not Family Voice. • p.55 : The column head should be Rank Total, not total rank (which could only have 20 ranks**)** • Table 5. This table refers to key questions. However, in the text the term applies to the four questions selected, not the larger number in	• Meta-analysis was performed and results presented in Results section. • Second part of question repeated in paragraph. • Plan left in but stated reason why not done. Two other reviewers found this informative. • Changed. "Gold" standards has been changed to "Referenced" standards. • Corrected. • Added. • Changed. • Corrected. • Changed • Kq was changed to pq in Table 5. • Noted.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	table. This should be cleared up (maybe called potential key questions or something.) • Tables 6-8. The items at the bottom such as examiner type, monitoring time, monitoring personnel don't influence outcomes although they may influence the reports of outcomes. • table 7 and 8: The box with all non-treatment factors is well filled, but not very clear. May be you can refer to table 10 for a full list of potential factors and give only a few examples here? • Table 9, page 63: In the Non-Condition Factors box, what is the rationale for including "Age at first OM"? This does not seem to pertain to this point in the pathway. • p.68: I do not think that the 170 publications not in English can be ignored without some survey of the possibility that some of them might be valuable. When I did a comprehensive review 10 years ago with one part-time assistant, I accessed English translations of the major available abstracts of work in the Finno-Ugric, Oriental and Slavonic languages (which I cannot read) as evidence, before concluding that I had probably not missed much of importance, but I included those in the Germanic and Romance languages in the evaluation and summarised them according to quality. A large explicitly funded project should attempt to do similarly, especially as much good work on OM is done in Netherlands and Scandinavia . • page 68, table 11 – does "age<12" need to be changed to "age <22", or does this reflect the change in criteria described on page 41? • Table 15: Does this list includes all of the measures reviewed or just ones accepted? A. Numbers 5, 6, 8, and 9 would not consider language tests but assess other several developmental domains. Could you check the test manuals or test measurement book to see how they describe the test? (Let me know if you want me to do it.) Safer to say "developmental test." B. #45: I believe this is a subscale of an IQ test (DROP) since at other times you did not mention the subscale, only the test.	• Noted. • The technical experts decided this was an important non-condition factor. • We added the rationale of why we restricted to the English literature in Methodology and discussed the issue of non-inclusion of non-English language in a new section entitled "Limitations of Evidence Report" at the end of the Conclusions chapter. • The initial scope for age was age<12. It was later relaxed to 22 for Q2 and Q3 only. Q1 and Q4 had age<12 as the limit. This was clarified in text. • How the list was compiled was added in the Methods. Table 15 was revised based on input from technical expert. • How the list was compiled was added in the Methods. Table 15 was revised based

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	C. #46 & 47: Are these actual tests or names of an informal method? D. #48 & 49: Not language tests -- perceptual motor DROP. E. #50 & 51: Not tests, but I think a variable studied. (Consider dropping because other times you did not use this.) • Table 15 (page 72), the definiton and method of assigning to a "Grouping Category" is not apparent to me. • Instructions (Table 17): I believe that the final paragraph should refer to otitis media with effusion rather than acute otitis media.	on input from technical expert. • Corrected.
Results	• Page 77 - again, are all quality scores given equal weight? • P. 77, Line 15: What are the quantifiers for quality scores of 1 to 6? • P. 78, Table 19: Zeisel (1999) was not included, perhaps it was too recent. Zeisel, S. A., Roberts, J. E., Neebe, E. C., Riggins, R., & Henderson. F. W. (1999). A longitudinal study of otitis media with effusion among 2- to 5-year-old African American children in child care. Pediatrics, 103(1), 15-19. • Table 19: Should Zeisel (1995) be excluded (see last paragraph of study results) Zeisel, S. A., Roberts, J. E., Gunn, E. B., Riggins, R., Evans, G. A., Roush, J., & Henderson, F., W. (1995). Prospective surveillance for otitis media with effusion among African-American infants in group child care. Journal of Pediatrics, 127, 875-880. • Page 79 middle. What about children less than three? Were there any results? • Why are all meta-analyses restricted to children *greater than* age 3 years? This is NOT the primary population at risk for OME, and not the population likely to have morbidity from prolonged disease. Unless *something* is said about younger children you are providing a very incomplete picture.	• Yes, as mentioned in the Methods section. • The "quantifiers" for the quality score components were described in the Methods. • This article was excluded because it included cases of AOM, and the findings were not stratified by OME and AOM. • The reasons for exclusion were given in Table 19. We re-examined the studies in light of this reviewer's concern, but reached the same conclusion that the studies should be excluded for the reason listed. • There were only two articles for the <6month and for 6month-<3years each. A statement has been added to indicate this. • We appreciate this reviewer's desire for additional information about children younger than 3 years of age. In our literature search, however, we did not find much evidence-only 2 studies. This information has been included in this report, but we could not say anything more due to a lack of evidence.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	• Page 79 middle. There is something strange going one here. The text mentions statistically significant heterogeneity but the table lists p=.19 and p=.14 or not statistically significant. Also I wonder about the heterogeneity numbers they are similar but the first set of studies seems much more homogeneous. I'd recommend rechecking this. • Pg. 79: Doesn't restricting the samples to children not receiving any intervention bias the natural history samples to children with less OM (or milder OM) to start with? • page 79 "Heterogeneity, however was evident statistically in the first synthesis and clinically might not be unexpected?" There something wrong in the formulation of the second part of this sentence. • Could you refer more to the tables in this part of the results? You only do this at the start of the analysis. The reader has to look for the specific table himself. There is sometimes a difference in the figures in the running text and the corresponding table (43,1% in the text on page 80 , 3th line versus 41.3% in table 24; and 24,3% in tympanogram B to A transition in the text (mid page) and 22.4 % in table 25) • Page 80 middle. I'd reference the specific table rather than just saying the "next set of meta-analyses". The numbers in this paragraph don't match the numbers in table 25) so I'm lost. The paragraph mentions that Holm-Jenson et al. was older, but the table looks like the citation should be Holmquist 1987. Same comment about the end of the paragraph perhaps. Also you need to use article ids to keep the Fiellau-Nikolajsen 1979 articles straight. • Page 81. In the final paragraph, the meaning of the statement "speech or language outcome was measured for under 22 years of age" is unclear. This must be a typo. • Page 82, 1st paragraph: The statement that the Rach study (and follow-up) was excluded because it was not a prospective cohort study and OM was not measured before age 3 years puzzles me. In this study OM was measured serially from ages 2 to 4 (See page 228). Language was	• Statements were corrected accordingly. • By definition, natural history is the course of OME without intervention. We do include groups of children who may have had OME for weeks-months or three months or greater, so we do not necessarily exclude children with "less OME". • Statement reworded. • Numbers corrected and more reference to Tables made. • Numbers were revised and corrected accordingly. Holm-Jenson et al. was corrected to Holmquist. Table numbers were added in text. • The sentence has been revised to read "speech or language outcome between 4 to 22 years of age was measured" • Although the Rach study measured OM severity at 2-4 years of age, it still violated the criterion that OM severity is measured under 3 years of age. We could not separate the children whose severity were

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	evaluated prospectively with an interval of 6 months at the age of 3 years and again at age 7 years (Peters et al., J Learn Disabil 1994; Grievink et al. J Speech Hear Res 1993). On what basis have this and perhaps other studies been excluded? Inclusion of these studies, however, would probably not have affected the results and conclusions.	measured before 3 and after 3 years of age.
	• Pg. 82, line 14: "reject" is misspelled as "rejected".	• Corrected
	• Page 82. In the final paragraph, "Dollaghan" is misspelled. With respect to that report, I do not understand the rationale for excluding the report "because it did not report on speech and language and development outcomes beyond three years of age." The report dealt with outcomes at three years of age, which issue seems to me to be within the parameters listed for key questions 2 on page 32 of the Draft Evidence Report. Similarly, our report (Paradise et al 1999--included in the bibliography but not in the reference list) dealt with parent-child stress and behavior at age three years and, it seems to me, might also have been appropriately referred to in the Draft Evidence Report.	• "Collaghan" was corrected to "Dollaghan". The assessment was on long-term effects and it was decided that beyond three years of age would not include outcomes measured up through 3 years of age.
	• P. 82, Line 1 & Table 29: I question using studies where OM data were collected retrospectively and outcomes were prospective. There is a problem with OME data collected by parent's report, which has many methodological problems. I would not incude the Freeark (1992) and Paul (1993) study.	• We share the concern of the reviewers. However, we decided that as long as we identified these studies as retrospective-prospective, readers would be aware if and could do sensitivity analysis, when possible.
	• P. 82 & Table 29: I would delete the Klein (1988) study. First, it was published as an article in 1990 (same data, I believe), and there are other references from the Recent Advances in OME that I believe were not included.	• It appeared that the Klein (1988 article) published results on the same cohort as article by Teele (1990). The results in Klein was not contained in Teele. Thus did not exclude.
	• P. 82 & questions: I would have considered 3 years of age also as an outcome and not only studies beyond 3 years.	• Comment noted.
	• Roberts is not included in review, maybe too current: Roberts, J. E., Burchinal, M. R., Jackson, S. C., Hooper, S. R., Roush, J., Mundy, M., Neebe, E. C., & Zeisel, S. A. (2000). Otitis media in early childhood in relation to preschool language and school readiness skills among African American children. Pediatrics 106:4, pp. 1-11.	• Yes, it was too current.
	• (Table 27) I think that Teele (1984) was done at 3 years, and this was	• Teele (1990) was included (Table 26) because it had outcomes beyond 3 years

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	included.	of age. Teele (1984) was NOT included (Table 27) because it did not report outcomes beyond 3 years of age.
	• P. 83 & Table 27: Should be considered in review and then dropped: Roberts, J. E., Burchinal, M. R., Zeisel, S. A., Neebe, E. C., Hooper, S. R., Roush, J., Bryant, D., Mundy, M., & Henderson, F. W. (1998). Otitis media, the caregiving environment, and language and cognitive outcomes at two years. Pediatrics, 102(2), 346-352.	• This is article #2264. It was on the Table 27 list because it did not report outcomes beyond 3 years of age.
	• Page 84 - particularly for the relationship between OME and speech and language outcome, the quality of report is critical - but I can find no analyses based upon quality.	• The issue on study quality was discussed in the Findings of each question. There is not enough studies in these two questions to do sensitivity analysis.
	• P. 84, Line 7: McCarthey & Binet do not measure expressive language.	• We meant 'cognitive verbal intelligence' in the sentence. Error corrected. • Both sentences revised.
	• Page 85 last paragraph, 2nd sentence. I had to read this three times to interpret it. I'd reword it. The sentence on heterogeneity reads somewhat awkwardly as well.	• Additional analysis and findings had been added to section on speech and language.
	• Page 85 - contains specific findings, while in the previous section you allude to findings and refer readers to the tables - more consistency from section to section would be helpful.	• Definition of hearing loss had been added throughout the Report.
	• P. 85 & Tables 35&36: I did nto go back to review these studies, but hope that an audiologist has reviewed them for their methodology and quality. What is percent hearing loss? What is considered a hearing loss? Were all assessments only finding conductive losses? How was the hearing loss measured?	• We clarified the 20-25 dB was air-conduction threshold in the Results.
	• Page 85: I miss discussion of the degree and nature of hearing loss. Is a loss of 20-25 dB due to persistence of OME, is it conductive hearing loss due to ossicular chain dysfunction or tympanosclerosis, is there a sensorineural component? How should we interpret a RD of hearing loss of 11% without at least some of that information?	• We could not contact the authors due to time limitations. Thus the Tables were left as is. We shared the concern of the reviewer
	• Page 87 bottom. This is not clear. You should indicate that table 49 has the complete set of articles while table 50 deletes the duplicate articles for the same study. I would, however, recommend that you contact the	

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	authors of the studies to find out if the articles really do have duplicate patients/cases. If they do, then delete as appropriate. If not, then include both. It seems to me reasonable to reduce this to one correct table rather than two of unknown validity. It is particularly puzzling that the including the possibly duplicated data would increase heterogeneity. This would be the opposite of what would be expected and should be looked into. • p. 88 – shouldn't this say professional tympanometry instead of tympanometer? • Page 88 1st full par., 1st sent. It should be made clear that we are now talking about the results including all studies as opposed to the results with possible duplicates deleted. All other statements in this paragraph should be similarly tagged so the reader knows which group of studies is being discussed. Further an in-text table would be more readable then the strings of numbers. • Page 88 - the diagnostic methods section - I could find no analyses based upon who performed the test. Its accuracy will vary depending upon its use by a general practitioner or a trained researcher. • Page 88, 1st paragraph: In discussing pneumatic otoscopy, it would be useful to clarify how many of the studies involved trained and untrained examiners, and the qualifications of the examiners. This would help the user understand the generalizability of the pneumatic otoscopy data. • Page 88, 2nd paragraph: Adequate performance of professional tympanometry does not receive much credit. In many countries outside the USA doctors are not trained to use pneumatic otoscopy, so it might be worthwhile to include performance of the second best diagnostic method –professional tympanometry- in the abstract, summary and conclusions. • Your statement about hearing loss in children in later life needs to be clarified as to whether this is with or without treatment of the effusion (or some hint that we don't know about treatment effect on this outcome). • Table 22 Probably shouldn't underline the superscripts in the footnotes. • Table 26. There is a missing space between villages and in under ID 1623	• Term changed. • Tables numbers attached to statements. • An analysis of the examiners who performed the diagnostic tests has been added Table 51 and a discussion of study quality and quality of documentation of test performer was added. • See above. • Already in Conclusions. Added comment on tympanometry to Abstract and Summary. • Definition including with or without treatment was added throughout the document. • Corrected. • Corrected. • We included publications from the Proceedings but we were aware of the duplicated findings.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	• Table 28: Why is Roberts' speech study (1988) listed under two numbers--3118 and 4806? Use only article. • Table 28: Ruben (1997) is only a 2-page extended abstract of data reported elsewhere. Would not include it. • Table 28: WRAML assesses narratives, not overall expressive language. Not sure that Verbal Scale Index is expressive language. Several of the measures (e.g., MLU) focus on grammar, one aspect of expressive language and are not overall expressive language measures, while others such as the SICD are more overall measures. • Table 28: Harsten (1993)--Not sure phonology or receptive language is correct here, need to check article. • Table 28: Roberts 88 use Goldman-Fristoe as a test; not phonology. • page 105-6, Table 28 has at least a couple of duplicate row entries: Fischler and Gravel • Table 34: Would not have considered Roberts don't think there was audiology data (1988). Would have considered Roberts (1995) and (1998) and, if you went through 2000, Roberts (2000). • Table 36. There is something wrong here. First the confidence interval for the risk ratio for Fischler seems way to big. I recomputed it using other software and got a much smaller interval. I'd recheck it. Second, it is odd that the heterogeneity for rate difference is much greater than for risk ratio. I'd recheck the numbers but suspect much of the problem comes from the wide variance in OME- percent hearing loss. The 20% number for the Sorri study is very hard to believe. I'd try to find out if that was an error or if there was something special about that population or way of measuring loss. I might exclude it if it seems to have some special properties that make it non-comparable. • Tables 40-50 and figures 5-6. You title these tables (or rows within tables 49-50) as tool "and" myringotomy. Sometimes you use "with". Either of these terms seems to me to imply use of two tests rather than comparison between the two. Here I would use the term "vs." rather than "and" or "with" to make clear what is going on. In figures 5 and 6, I think I would put the fact that myringotomy is the gold standard in the	• We included publications from the Proceedings. • Comments added to the explanation of the Table. • It was linguistic analysis. • It was phonologic analysis. • Table 28 listed studies by outcomes. Thus, duplicates are expected because of multiple tests in one study. • These articles were not included because they did not report outcomes after 3 years of age. • The 95% CI of risk ratio for Fischler was corrected. The EpiInfo program was used to calculate it based on 9/96 and 1/70. We added another meta-analysis in Table 36 taking out the Sorri articles and the heterogeneity was greatly reduced. The results were discussed. • Changed made as suggested.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	title or someplace general rather than repeating it for each line in the legend. • Table 44 Heading has a typo (+A24) stuck in the middle of myringotomy. • Tabl4 44, page 123: Typo in last word "M+A24yngotomy" • Table 49 The "Number of Articles" is not wide enough, cutting off the column heading. • Figures 5-6. The figure uses asterisks to note the points with duplicates while the legend and footnotes use a and b. This leaves the asterisks undefined. You should be consistent here.	• Corrected • Corrected • Corrected • Corrected
Conclusion	• Page 144 1st sent. This could be worded better. I suggest something like "We were able to conduct sets of meta-analyses of OME resolution at 3 follow-up intervals. These sets were stratified, when possible, by unit of analysis, age group, OME type, and diagnostic method. • Page 144 2nd sentence. You again need to make clear whether these numbers are cumulative or not and indicate what the problems are that lead to these strange results. • Page 144 2nd par. I'd put the word "section" after "Results" in the first line. Also the first sentence implies there is more about these studies within this review than actually exists. I'd just say something like "we looked at the isolated studies of …" • Page 145 1st full par. I'd indicate the number of studies or percent of studies rather than just saying the "majority". Also might change "control of" to "control for" in the 1st sentence. • p. 144 – is this correct? Is this the same 41? That may resolve by 1 month or an additional amount? This seems fairly low; especially the previous report states that 13-44% resolve at 1-month follow-up. How could the 3-month resolution rate be less? • P. 147: Other NIH supported prospective cohort studies that have published OME and speech/language data with similar methodological rigor (but no randomization) are also ongoing. Studies in North Carolina (Roberts et al., 1995; 1998; 2000 should be cited. There are other ongoing prospective cohort studies, but they have not yet published their data.	• Done • Done • Done • Done • Done as above • Roberts, Burchinal, Zeisel et al., 1998; Roberts, Burchinal, Jackson et al., 2000 were added as ongoing prospective cohort studies.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	• Page 148, 2nd paragraph and page 149, 1st paragraph: See comment on page 85. Is the long term effect on hearing due to persistence of OME and expected to resolve? Is it a conductive hearing loss with an aerated middle ear, is it a sensorineural hearing loss? I believe some information on this subject can be found in the studies and should be included in the abstract and summary as well.	• It is conductive hearing loss. This term was added to appropriate places in the abstract and summary.
	• Page 149 2nd full par. Another point is that we don't know if treatment is effective in changing the long-term hearing outcome. This is perhaps the most important point.	• Sentence added.
	• Page 149 bottom. It is a little difficult to make a judgment that pneumatic otoscopy is "best." Best depends on the relative values of missing true positives or true negatives. A more sensitive, but less specific test may be better if the treatment alleviates much suffering. The reverse is true if there are significant harms to false positives. I'd stay away from such value judgments in this evidence report. I'd also change the words "The pooled" to "Its pooled" at the very bottom of the page for clarity.	• Revisions made. • Done
	• Page 150 1st full par. Last sentence. Actually we did, but as noted earlier elected not to publish them. It might be better to say that the OME Guidelines did not include quantitative syntheses of the evidence.	• Reason added.
	• Page 151, top. I'd mention again why you didn't look at combination methods.	• Done
	• Page 151 7th line from the bottom. Either put commas around "over time" or move it to after "improved".	• Corrected
	• Page 151 3rd line from the bottom. "thdiagnosis" should be "the diagnosis"	• Already in the Conclusions and added to the Abstract and Summary
	• Page 150-151: Adequate performance of professional tympanometry does not receive much credit. In many countries outside the USA doctors are not trained to use pneumatic otoscopy, so it might be worthwhile to include performance of the second best diagnostic method –professional tympanometry- in the abstract, summary and conclusions.	• We sympathize with the reviewer's desire for recommendations, but developing recommendations are the function of a guideline panel and outside the scope of Evidence Based Practice Center Evidence Reports.
	• There are several places in the conclusion section, however, I would make greater efforts to summarize as recommendations. Pages 146, 148, 149, and 152 should each have as last sentences. Therefore…..	

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	I know to some this might make it too simplistic but if we want to change behaviors BEFORE we have the definitive study, we need to provide some real guidance to the reader and practitioners. You have nicely done that for the researcher; the practitioner or provider needs some of the same structure.	
Future Research	• Page 9, last sentence. This isn't clear. I would doubt that most algorithms are so complex that computer programs are used in practice. The actual instructions for applying the algorithm are probably what is needed. If a program is used, then, of course, it should be supplied. • Page 153: The Future Research chapter would benefit from adding a general introductory section, and moving comments/recommendation that pertain to the broad scope of OM research to this new section. These include the discussion (now in the diagnostic methods section) of the need for clear definition of OME, the need for agreement on standard research follow-up intervals, the use of the child or episode rather than or in addition to ear as the unit of analysis, clarity on treatment received by cohorts, inclusion of univariate as well as multivariate analyses, cost-effectiveness, etc. • p.153: Moller and Tos. This relates to the point about persistence versus diagnosis in my pre-amble. I do not see the Moller-Tos findings as undermining the position I state in my 4th introductory paragraph above. If I really have missed something, then the facts and the authors' interpretation need to be made more clear. • In the sections on future research, you emphasize the need for consistency in definitions and diagnostic procedures. Beyond this, however, it was not clear to me whether you were advocating any randomized controlled trials of interventions, or whether you advocated only meta-analyses of multiple cohorts. One concern is that even these meta-analyses will not be able to rule out the uncontrollable confounding that plagues any observational cohort study. Are you certain that you believe that RCTs are not achievable? I thought I heard that Jack Paradise conducted one of tympanostomy tube placement whose results are imminent. • Page 154: Section title – would add "Effects of*Early Life OM . . ."* • Page 154 top. Perhaps as important, we need to know if resolution of	• Statement changed. • Added a general issues section. • We further expanded the comment stating that "The issue of assessment of OME duration or recurrence is as important as the issue of diagnosis of OME at a single point in time." • Agreed with comment on randomized controlled trials is confusing. Comment deleted. The important issue was that of assessing the role of influencing factors and interventions. Reference to ongoing studies added as well as the uncertainty with regard to areas for further prospective studies which will be dependent on the results of these ongoing studies. • Added to all subtitles in the Report. • Comment added

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	the effusion has an effect on the outcomes. • p.154: "Research on influencing factors" recommended… What precisely is being said here? More studies of simple risk factors? Possibly there are too many, especially small bad ones! What specific gaps are there? Or are synergistic (comorbidity) conditioning factors what is intended here? Surely the problem is to get clinicians to use a risk-based approach with information already existing! When we've shown that they are prepared to heed evidence of this type that is neglected in ORL although common in cardiology, we can then worry about improving the evidence. Politicians sometimes fund research rather than facing up to lobbies, a process which degrades us all. Unless something more specific is said, this recommendation also will appear to be ducking the issues. • P. 154: Other OME developmental conceptual frameworks are also cited in the literature including support for a transactional model (Roberts & Wallace, 1997). Roberts JE, Wallace IF. Language and otitis media with effusion. In: Roberts JE, Wallace IF, Henderson F, eds. *Otitis Media in Young Children.* Baltimore, MD: Paul H. Brookes Publishing Co, 1997. • P. 155, Line 3: Paradise is a randomized controlled study. • P. 155: NIH is currently supporting a study that does include children followed prospectively from both NC and NY. • p.155: "Randomised trial of the effect of early OM". This phrase is nonsense, as well as the sentence being far too long. Perhaps there has been a word-processing error. The only meaningful type of RCT has treatment as independent main effect. If that is what is meant here, say it more clearly. You can't give children OM experimentally. A trial could have OM disease markers, or speech and language, or both as outcome (dependent variable) and many do. An argument often put by Jack Paradise is that the best-controlled answer on developmental sequelae questions is obtained by doing a treatment RCT. I do have some sympathy with the logic of his argument but would call the application of it only "one useful source of evidence". The view is overstated, and does not lead to a sufficiently powerful design within the US system, where only 6 months withholding is permitted by parental	• Revised. • Done. • Reference to RCT deleted. • Noted • Reference to RCT deleted. • Reference to RCT deleted.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	and clinical pressures. It leads to underpowered and expensive research. If this is what is intended, simply mentioning the idea without listing the difficulties is a little irresponsible. • Page 155, paragraph 2. I take exception to the statement "it is likely that a randomized controlled trial of the effect of early otitis media on speech and language development is not ethically possible at the present time. . . ." We have been conducting just such a trial over the course of the past ten years, as described in my 1998 report referenced on page 167. (Incidentally, initial results of that clinical trial will be published in *The New England Journal of Medicine* later this month) It is likely that no associational study such as discussed on pages 155-156 can definitively answer the question of causality because of the multiplicity of known, and particularly, unknown developmental risk factors. • Page 155, 2nd paragraph: In this very important, methodologically correct evidence report it is concluded that no conclusive evidence can be provided for the effect of early life otitis media and long term speech and language development. In the ongoing Pittsburgh study only weak to moderate correlations between early life OME and later language were found, and OME explained only 1.2-2.9% of the variance in the language scores. Also, you have shown that the natural course of OME is favourable. Then why do you 'close the road' for randomized controlled trials on this subject by suggesting that this is not ethically possible? RCT's assessing the efficacy of ventilation tubes in children with OME have been performed in Europe in recent years (Maw in Bristol, Rovers in Nijmegen, TARGET study in Nottingham) and their results have proven that these RCT's were ethically correct. It is quite unlikely that results would have been different if these studies were carried out in the USA. • Page 155, 3rd paragraph, Page 156, 3rd paragraph: 'Individual-level-data-meta-analyses' on several aspects of OME will indeed give valuable information. Such analyses are currently being carried out by Maroeska Rovers at the Nijmegen University, The Netherlands (e-mail m.rovers@mie.kun.nl) in collaboration with Mark Haggard (Nottingham, UK) and Richard Maw (Bristol, UK). So far, the investigators only had	• Reference to RCT deleted. • Noted. • Comment added • Comment added

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	access to the raw data of the European studies, but Dr Paradise and Dr Gates have been approached by Maroeska Rovers and Mark Haggard with a request to co-operate. It would be very valuable for European and American researchers to collaborate in projects such as these. • Page 156, bottom. Similarly, treatment effect on long-term outcomes is still the main issue. • p. 156: In the sections on early life OM on long term hearing, I found myself wondering what the potential gains from improved diagnosis and effective treatment really would be. A decision analysis/cost-effectiveness analysis would be useful here. You suggest a CEA on p. 158, for diagnostic methods, but I think such a model would ideally be expanded to include treatment and long-term outcomes as well as short-term diagnostic outcomes. • p.157: Agreement on borderline between AOM and OME. Likewise this statement makes the report seem insufficiently joined-up. Earlier in the Report, many authorities were rightly quoted on the difficulty of this distinction. It is pointless to subscribe to the view that improving diagnosis of OME is a main important question, and that a main route is tidying up this diagnostic boundary, if the authorities considered that to be impossible in the first place! The answer is to educate the professions into acknowledging the more important questions. Deciding whether 2, 4 or 8 weeks of effusion after AOM should now be called OME, one version of the compulsion for a single diagnosis, will certainly not be helped by a new gizmo, and possibly not even by an algorithm, because the question of a single categorical diagnosis itself is fundamentally misposed and is of limited clinical usefulness. • Page 157 1st par. Last sentence. The first guideline panel spent hours on this with no resolution either. Very frustrating. • Page 157. The statement "whether diagnosis of middle-ear effusion in the context of OME was different than in the context of acute otitis media" is not clear to me. • Page 157, second paragraph: I agree that the diagnosis of middle-ear effusion is different in the context of OME than in the context of AOM. • Page 157, second paragraph: What is meant by "different"? The allusion to a difference in diagnosis of effusion in the context of OME	• Comment noted • Comment noted • Comment noted • Comment noted • Comment noted • Comment noted • Comment noted • Comment incorporated.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	and AOM without clarification of the way in which it is different is confusing. • page 157 first paragraph. Here is the opportunity to bury the OME concept. Why more agreement about what OME really is? Is it really important to diagnose OME in an exact way, when it is self limiting? Ok, I understand that this is not the place and moment to discuss these problems. But they are interesting and should be discussed by guideline developers who will use your evidence report. • Page 158. The discussion of cost effectiveness analysis seems to me impractical. It is clear that, of readily available methods, pneumatic otoscopy is potentially the most accurate, but it is also clear that its accuracy depends on the skill of the examiner.	
Evidence Tables	**Evidence Table 1**. • I am concerned that the table doesn't record losses to follow-up. This could be a very significant issue, particularly for a guideline panel trying to make sense of the data. • Page 178. How do we have an N of 103 with #at risk of 137? I suspect that N is children and # at risk is episodes, but the section under gender confuses this so I'm not sure what's going on. • Page 190. The extra underscores are decrease readability. Also I cannot figure out what the column headings mean for the table on the left (e.g. Mean duration (1-r)/r means nothing to me. I think I know what mean duration is and 1-r/r looks kind of like an odds number, but I'm not sure how odds work with duration.) • Page 201 Formatting badly messed up. • Page 225 N=433, but number at risk is 443. Is there a typo somewhere? • Page 227. Why is this N1 rather than N and why is the number at risk so much smaller? • Evidence Tables 1-3 addressed questions relevant to understanding the natural history and short and long term outcomes. The evidence tables contain very comprehensive information but there is a great deal of variability in the populations studied regarding geography and ethnic background which may account for the variation seen in the outcomes.	• Both total number of cases and number of cases used are reported. • There are 103 children and 137 episodes. Units were added. • Evidence Table 1 has been reformatted and indicated changes made. • Redone. • Corrected • The use of N's and N1's had been changed throughout the Evidence Table. • One study which we abstracted addressed the issue of season and that information is noted in the evidence table and the Results.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	There is no mention here of a seasonal prevalence of variation seen during the winter months which also may be more useful to answering questions of persistence or recurrence seen in the natural history.	
	Evidence Table 2	
	• Page 230 GCI and PPVT-R are not McCarthy scales. The table should be reworked.	• Clarified.
	• Page 232 Here "Grp1" is in a different font. This recurs intermittently throughout these tables. Sometimes the group definition is also in this roman font. This also happens in Evidence table 3	• The fonts have been made consistent throughout the Table.
	• Page 239 I have no idea what TOJxxx means. Perhaps someone in the field would, but it's Greek to me.	• Explanation of abbreviation added.
	Evidence Table 3	
	• No comments other than the one for page 232 above.	• The fonts have been made consistent throughout the Table.
	Evidence Table 4	
	• I don't think it is necessary to list N and N1 if there is only one group and the numbers are the same.	• The use of N's and N1's had been changed throughout the Evidence Table.
	• Page 261. I don't understand this group 1(a) vs. group 1. What does the "(a)" mean?	• Clarified and reformatted.
	• Page 262 and later. Here again we have a 1(a), 1, and 1(b) and I don't know what they mean.	• Clarified and reformatted.
	• Page 278-9. Here we have groups 1,2,3 and 9? What about 4-8? Why is there no N for group 9? Why are groups 1,2,3 mentioned if there are no data for them?	• Clarified and redone.
	• Page 282 and later. Here again we have group 9 with no N and group 1 with no data. This pattern occurs later in the evidence table (sometimes with a size for group 9, sometimes not). These need to be found and fixed (or at least explained.) Also the font size is wrong for "outcome" under "findings" for comparison 1. This also occurs throughout the remaining tables sporadically for different comparisons.	• Clarified and redone.
	• Evidence Table 4 contains information of the accuracy of the diagnostic test in a quantifiable form which is useful information for translation of evidence into practice and research.	• Noted

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
Bibliography	• page 623 Rovers MM, Zielhuis …. Head & Neck Surgery *1999* (and not 1203)	• Corrected.
Appendices	**Appendix D** • Page 729 Again, I'd change "vs." to "or". Also all the non-treatment factors listed here seem to be basically ignored in the main document. I realize it is probably impossible to do, but that fact should probably appear in the summaries and conclusions. (Maybe it does and I missed it.) **Appendix F** • Some type of total column would be nice.	• Changed. • Totals are presented in Table 10 of the text.
Thematic Issues	• A limitation section should be added • Clarification around hearing loss is necessary • Presenting information based upon quality of data would be helpful • Analyzing the diagnostic accuracy by performer of pneumatic otoscopy would be helpful • Some experts believe that the relationship between OME and speech and language outcome is impacted upon by social class - did I miss any discussion or analysis of this issue?	• A limitation section has been created in the Conclusions Chapter • Definition of hearing loss has been added in Abstract, Summary, and Results sections. • The issue of study quality has been addressed in the Results and Conclusions sections. • Analysis by performer has been added, Table 51. • We were not able to synthesize the findings because of inadequate number of articles addressing social class. Added as a limitation. Added to the Abstract, Summary, Results, Conclusions, Limitations, and Future Research sections.
Overall Evaluation	• Overall it is a superb report. Obviously, it represents an exhaustive review and analysis of the literature. It is somewhat dismaying that we cannot answer the basic question - does otitis media impact on	• Noted

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	cognitive development? If it does not, then all of the other issues in the report are far less important. • Overall evaluation- an extraordinary compendium of information. It is clear what was done and how it was done what was done and how it was done. • Although the criteria for inclusion were clearly stated it would appear that rules were not followed as noted on pages 147 - 148 in the conclusion section that mentions an ongoing study which, to some, has serious methodological problems. This is also inconsistent with your statement on p 82 where it was stated that this same study was not included because it did not include results > 3 year. The inclusion of this work in progress by a member of the panel suggests that your process is less than objective. There are other completed works that could have been used such as those from the United Kingdom under the direction of M. Haggard PhD which were not mentioned. The inclusion of the preliminary reports by a panel member seriously compromises the objectivity and validity of this report. • Additionally, the second part of the language study was not performed (pp48) but it should have been done. Thus the conclusions concerning language have to be modified, and it should be noted in the final report that there are data that have not been examined. The meta regression analysis of this data may have provided some of the most useful information as to who is susceptible and who is not susceptible. • This is a topic for which I have both affection and aversion…I was disappointed that despite the report's volume (several times the AHRQ report), the scope is narrower and less useful. The title is entirely inaccurate and misleading, because the review does not touch treatment. I find this personally amazing and disappointing, because it seems to me that the major issues out there on OME are precisely the ones that the review omits: the efficacy and effectiveness of treatment. The panel clearly came up with the "right" questions (appendix A). In my view the fact that the scope of work ended up excluding most of them represents an astonishing failure of the process. • There are certainly no problems with clarity. • The generally negative findings of this literature review reflect the well	• Noted • All criteria were established a priori and applied equally to all studies. Several studies by Haggard were reviewed but did not fulfill the inclusion criteria. • Meta-regression analysis would be desirable but could not be done within this time frame. • The title of the Report has been changed. The issue on selecting key questions for Evidence Reports has been brought to the attention of appropriate parties. The importance of treatment and other aspects of OME was included in the Limitations of the Evidence Assessment. • Noted • Noted.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	known variability in both the disease (OME) and the study of the disease. Given the fact that few investigators have the time, talent, funding, facilities, and study cohorts to answer the questions posed by this review, it is not surprising that the reviewers discovered a variety of weaknesses in these studies.	• Noted
	• Perhaps some experiential observations might provide some perspective. First, the natural history of otitis media is well known: it gets better with age. Second, children with severe childhood otitis media are at risk for damage to the middle ear structures, developmental delays, and language impairment. The degree of abnormality, logically, varies with the severity and chronicity of the disorder, the impact of treatment, and the effects of remedial education. Trying to quantify these diverse elements has frustrated researchers for decades. Third, the value of pneumatic otoscopy is well established but its value depends on the training and expertise of the examiner. Tympanometry is a very useful objective measure that assures both the patient and physician when it is normal but has less value when it is not. Binocular otomicroscopy with pneumomassage is the gold standard for the diagnosis of middle ear effusion. Unfortunately, it is not available in the primary care physician setting. Fourth, residual scar tissue formation in a chronically infected ear may limit sound transmission and result in a permanent conductive loss. Fortunately, these losses tend to be mild, some are surgically correctible, and all are remediable with amplification.	• Noted and revised.
	• The recommendation in the Future Research section that a cost-effectiveness analysis will lead to a “truly informed decision on the best diagnostic method” is difficult to understand, given the unambiguous conclusion that pneumatic otoscopy is the superior diagnostic method. Pneumatic otoscopy is widely used, highly regarded, and is the teaching method of choice in nearly all educational programs. It seems to me that this is sufficient justification for any decision-maker and that the money proposed to be spent on such a study might be better used for educating decision-makers.	• Noted
	• While the stated goal of determining the natural history of otitis media seems reasonable on the surface, the judgment of the present review,	

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	namely " ... we... view these estimates of OME resolution with great caution..." is appropriate and unlikely to change with further study. Simply put, the natural history is variable and has been so since medical records have been kept. It is highly unlikely that funding for natural history studies can be obtained and even more unlikely that subjects could be found to participate in such a study. Indeed, a current funded randomized clinical trial of surgical treatment is seriously undersubscribed because of the changing referral patterns of contemporary medical practice.	• Noted
	• Given the variable natural history, it is also a fact of life that the impact of otitis media on child development is also variable. That severely affected children have moderate delays is well known. Documenting such delays in a controlled study is and has been extremely difficult. I agree with the conclusions of the review that the evidence is sketchy. I do not agree that the ongoing Paradise study (anonymously cited in the conclusions) will shed much light on this issue because surgery (tube insertion) is being used much earlier in that study than in normal practice and the interval between early or late surgery is only 6 months. Thus, it is unlikely that much useful information will accrue from that report.	• Noted • Noted • Noted
	• My overall evaluation of the report is that it was a very clearly written document. I found the findings and conclusions to substantiate the previous guideline, and I was interested in the fact that the literature to support hearing loss and speech delay was once again not clearly documented. Thus it supports the guideline results in '94. • What an impressive piece of work. • The description of the approach to the Evidence Report is clear. The questions (scope of the report) and the methods used to determine the questions were clearly presented. I had no difficulty in following the chronology of the development of the Evidence Report. • It is entirely clear what was done, and much of it is obviously of value. • Although the document is well written and easy to understand, the problem is that it could be so much more if the authors had some additional time and resources. It is inexcusable that an agency whose last name is Quality provides inadequate funding and time to create the	• Noted • Noted • Noted. • Noted

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	quality document needed. Specifically, the report does an excellent job of obtaining and evaluating the evidence, but the time constraints do not leave sufficient time for the proper synthesis of this evidence. This same problem has occurred with several other evidence-based reports. • The methods used to determine the primary questions to be answered, determine causal pathways, search for articles from multiple databases, and abstract articles was clear and understandable. The report is comprehensive and provides sufficient details to be able to understand the process and methods used. • The descriptions of the process, search criteria, article selection and review criteria were exceptional. It will take a careful reader significant time reviewing the evidence tables to understand each of the analyses and conclusions. Since I did not spend that time, I found the text somewhat dense, especially as the conclusion often was that the literature quality was poor and few definitive conclusions could be reached. As I am sure that there will be summary articles written, I trust that they will distill the information into more readable form. The research is complete and exhaustive. • The report is well-organized and its methods are clear. • The methods used to create the evidence report are clearly described in the report. • The methodology used in this report is clearly described and can be followed without difficulty. • My over-riding comment is that it seems sad that the report is limited to the natural history of otitis media with effusion, the impact of otitis media on long-term speech and language development, the impact of otitis media on long-term hearing and the operating characteristics of the various diagnostic methods. This seems to negate the title, which refers to treatment. Indeed, treatment is a very important part of management. Whilst the text does review medical management in relation to the natural history of otitis media, much of this work has previously been carefully reported by Rosenfeld, to whom you variously refer. However, you do not seem to take precise evidence from surgical trials in the manner that you have	• Noted • Noted • Noted • Using a standard consensus method, the technical expert panel chose the four key questions to be addressed by this study. Questions on treatment were not included among those four questions. We agreed that treatment is not addressed in this evidence report, and the title has been changed to “Diagnosis, natural history and late effects of OME”. For the natural history question, we decided that cohort studies would give the best estimate of OME resolution rates. We agreed that the placebo or no intervention arm of a clinical trial may provide information on OME resolution in groups of children with OME who, for whatever reason, receive close follow-up and refer to this in the revised Conclusions. We appreciated the references would be useful when treatment is studied.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	evaluated medical trials. There is of course information about the natural history of OME from surgical trials, though the entry criteria of these trials are often quite different. If the review is to cover treatment, then I think you should evaluate all aspects of treatment. Much of the cost and morbidity and possibly mortality is related to surgical treatment of this condition and I would recommend that the report be extended to cover this aspect. May I refer you to three recent papers from our department, which look at the prevalence of otitis media and its effect and the effect of surgery on behavioural problems. <u>References:</u> *1. The frequency of otitis media with effusion in British pre-school children: a guide for treatment. Midgley EJ, Dewey C, Pryce K, Maw AR, and ALSPAC Study Team. Clin Otol 2000, 25:485-491* *2. The relationship between otitis media with effusion and contact with other children in a British cohort studied from 8 months to 3½ years. Dewey C, Midgeley E, Maw R, The ALSPAC Study Team. International Journal of Pediatric Otorhinolaryngology 55 (2000) 33-45* *3. Randomised controlled trial of early surgery versus watchful waiting for glue ear: the effect on behavioural problems in pre-school children. Wilks J, Maw R, Peters TJ, Harvey I, Golding J. Clin Otolaryngol 2000, 25.209-214*	• Noted • Comment noted.
	• what was done to produce the report seems clear and the methodology seemed generally appropriate. As noted above, I thought that certain information was omitted or overlooked. The information contained in the report would clearly be useful to anyone developing clinical practice guidelines or medical review criteria for diagnosis and treatment of OME.	• Noted
	• I found the report very clear and easy to follow. I have focused my comments mainly on the speech/language question as listed below. I did not check to make sure that every article in the table matched the data in the article. I have some concerns about the classification of the	• Noted • Noted.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	measures mentioned and why some articles were excluded and others included. I also think that 3 years should have been included as an outcome for speech and language and not just articles above 3 years. I did not evaluate the studies included in the audiology question with the same scrutiny as the results in the area of speech and language. • As usual with your group, the clarity and thoroughness are exceptional. The meticulously prepared evidence tables do a terrific job of summarizing the existing raw material in the field, as well as highlighting the inconsistencies and deficiencies. On a purely technical level, I think you did a marvelous job. • It is very clear what the reviewers have done. • The report is the most systematic and comprehensive approach to these important questions that exists to date. The expert panel contain superb and respected individuals who have long been investigators in this field and whose opinion is valued by the greater medical community. It is quite clear what was done and the outcomes of this process are also well presented. I do believe that the answers to the questions are substantiated by the systematic review of the literature and by the experience and opinions of the expert panel. It is also important to frame the conclusions from examining the four questions in similar terms that the guidance for future research has been done quite carefully. Obviously, the greater public will need to have greater synthesis of the tabular/figure supporting data but that will be the challenge of the team responsible for crafting a final report that will be effectively disseminated. • You and your colleagues have produced a very comprehensive document,…. I will not comment on the methodology or review of the literature which is very extensive. I do feel that this is a valuable document for organizations who wish to write guidelines as it saves them the effort of reviewing the literature again. Unfortunately there has been very little new in good studies since the 1994 guideline and this may emphasize the importance of providing more funding for research in this frequent occurring disease which still has many variations in treatment and which constitutes an expense both direct and indirect. • Very good.	• Noted. • Noted • Noted. Discussion included in Future Research. • Comment noted.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	• The evidence report is again a thorough and useful product of evidence based medicine. The methods are very explicit and almost every step in the process can be reproduced. The answer on your question "is it clear what we did" is a loud and admiring "yes, you do". Searching, judging and compiling evidence of such a large amount of literature is almost Sisyphus labour. Too much labour for a disease that doesn't exist, as one of my colleagues told me, when we spoke about your 3,5 kg weighting report. Your panel could only described OME according the OME guidelines as "OME is fluid in the middle ear without signs or symptoms of ear infection." Why should we bother? Has anyone studied the natural course of a little bit of fluid in the knee in patients without complaints? A little discussion about the relevance of this whole concept of disease (other than costs and financial interests of ear throat nose surgeons) could make the report more attractive to readers and guideline developers in other countries. • Overall the revised draft and tables included in the appendix is more concise and clear in the writing style. The key questions and results present outcomes that are more important for further research questions particularly in defining studies more clearly in order to study risk factors, interventions and outcome measures in a uniform fashion. Consumers and primary care practitioners may be more interested in some additional information regarding immediate outcome in addition to duration of natural history, persistence, and recurrence.[1] 1 Geyman JP. Evidence-based medicine in primary care: an overview. Journal of the American Board of Family Practice. URL. Http:// www.medscape.com/ABFP/JABFP 1-18; January 31, 2000 • Yes, I do find the description of what was done clear and understandable. This report could be used to reproduce a similar investigation. • It is clear what you did. I think the detail will please the academics. The average doctor out there will not be interested in all the details and would prefer to read the executive summary. I think the findings and conclusions are what I would conclude them to be. I was a bit pleased and surprised that pneumatic otoscopy is still the preferred and simplest way to diagnose OME. I learned more about the methodology from the	• Noted. • Noted

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	draft than from the teleconferences.	
Methodology	• Methodology was appropriate • Your methods were well thought out and well applied to the available literature. I agree the areas narrowed down for study (page 10) were appropriate. • The methodology was appropriate in identifying the key questions of interest to the panel of technical experts. The literature review and the methods used to obtain and extrapilate the literature were very clearly deliniated. The body of literature out there on OME is enormous, but the value of much of the literature is still of little value. It never ceases to amaze me that when a study is written that many professionals use such small numbers to extrapilate their information. • In my opinion, the methods used in deriving the four key questions of interest from the panel of technical experts were appropriate as was the searching and reviewing of the identified literature. The synthesis of the literature was appropriate. Evidence tables were supportive; inclusion and exclusion criteria for studies were specified. • I have no criticisms of the reviewing methodology or coverage. • However, as with many of the Cochrane reviews that I have to scrutinise, I have some doubts about the depth of analysis of issues and the degree of familiarity with the clinical and biological interpretations of the literature that lie behind the interpretation of the review findings and particularly the introduction, and I am not sure that the procedure for identifying key questions has produced sensible answers. • An outstanding job. I doubt that anyone could find fault. (See #4 below, however, for some comments on how key questions are chosen). • The methods are appropriate. I liked your conservative approach to using meta-analysis – ie, you seemed to avoid it when studies were very heterogeneous. • The methods used to derive the key questions from the panel are described in detail and accurately reflect the process the panel used to limit the scope of the guideline. The process of the literature review and the search terms are outlined. Synthesizing the literature is obviously	• Noted • Noted • Noted • Noted. • Noted • Introduction greatly revised. • Noted. • Noted • Noted

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	extremely difficult and will require a detailed interpretation for the process to be excepted by clinicians. The meta-analyses are detailed and the process used to exclude specific studies are specifically described.	• Noted
	• I was pleased to note that definitions used in the recent AOM guideline and 1994 definition for OME were both endorsed without changes or additions.	• The age issue has been clarified.
	• There is some question in my mind about the age cut-offs. One section state up to 22 years, other sections note up to 12 years. Was there a typo or was age restriction changed?	• Agree. As long as in one single meta-analysis we did not include multiple studies, that meta-analysis would be fine. However, when outcomes are aggregated, then bias would exist. Discussion of this issue included.
	• The methodologic limitations of published studies on this topic seem to be adequately described. The issue of multiple publications resulting from a finite number of study cohorts is complex and likely results in some degree of bias (albeit unmeasureable). There might be some consideration for adding further emphasis on this point.	• Noted
	• The methodology seemed appropriate for identifying key questions of interest, reviewing the literature, and synthesizing the literature.	• Comment noted and addressed in Limitations
	• The methodology for selecting the questions was a combined Delphi approach with a nominal group process. Both of these are designed to maximize input and develop priorities. This was done quite successfully. My only regret was the subject of comparing intervention, especially surgical vs medical, was not selected well. The literature review addressed this issue frequently and it seems to me that the team could have developed answers to that question as well. The literature search and synthesis thereof was complete and scientifically performed.	
	• seems appropriate to me.	• Noted
	• The chosen quality scales are adequate and well described, even so the data abstraction and procedures to reduce bias.	• Noted
	• The methods for identifying the key questions were certainly democratic. The method seems counter intuitive and awkward in going through the process. The most awkward part of the process was the personal evaluation and shooting from the hip on speech and language and other aspects of the questions. In looking at the tables generated, we seemed pretty close in agreement on most of it. I think the literature part of the process was well done. There was a lot out there to consider	• Noted

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	and the meta analysis was used appropriately. You were explicit about why articles were excluded and included in the evaluation.	
	<u>Selecting key questions</u> • Personally, I am disappointed that none of the selected topics deal with updating the treatment portion of the guideline. The results of this evidence report will make updating the guideline difficult because of that lack. • I suspect that more topics could have been covered had this been treated as an update process without spending time going back over the same articles that the original guideline panel had dealt with. I believe the EPCs and AHRQ need to think about what are appropriate activities in an update evidence report. • It was not possible to review the details of the evidence tables due to time constraints and the fact that I don't have the articles to double check the data in the tables. However, I did scan the tables and noted a few difficulties, mostly in formatting. I would recommend having an editor go through the tables to improve readability. In particular there are some strange uses of fonts (e.g. group names and definitions), inconsistent table formats (e.g. Evidence Table 1), inconsistent definitions counts of groups and uses (e.g. Evidence Table 4). • At the level of choice of the 4 key questions (KQs), it seems that the nominated experts were asked to rate importance of questions. But the equally important matters of the degree of current uncertainty about them, the prospects of a review reducing that uncertainty, or the prospects for influencing practice as a consequence, all of which should be criteria, do not seem to have come through. The impact of this omission is seen in the audience issue (b). For example, I would not have thought that a review was needed to endorse the evidence of general nullity (except in clearly defined extreme or co-morbid cases) of effects on formal speech and language performance (when the question is, rightly, put for outcomes beyond age 3). The issue of subtler cognitive effects after the age when language effects are measurable is subject to more uncertainty and could merit review, but was overlooked.	• Treatment was not one of the top four questions according to our methodology • Noted. • changed • We sympathize with the desires of the reviewer to have different questions assessed, however we had an explicit process for prioritizing the questions to be assessed given the time and resources available, and this was how the aggregate results turned out. No doubt some individual expert panelists, as some individual reviewers, would have preferred a different rank order or a different set of questions, but its a group process with representation from many of the interested parties, and the rank order that we obtained was how we proceeded.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	The problem is that practitioners ignore the nullity of the evidence except in clearly defined extreme cases, particularly because of vested interests in the speech/language pathology profession, and hence they continue to focus on the wrong aspect of the problem. • Is the report chiefly a means of directly accessing audiences without access to the specialised literature, e.g. parents, family practitioners, or speech language pathologists? Is the authority of the panel peer process, combined with recency of this review, a necessity to preceed an anticipated surge of practice-changing guidelines? Perhaps this is so, but it would have been useful to have the strategic appraisal that it was so. This prioritisation of questions looks more like a lament that the evidence is still not getting through to belief and practice, rather than an informed judgement that a review will be productive now. Thus the methodology is appropriate, but has not been appropriately deployed, and the implementation issue has been addressed only indirectly. It's like building a bigger bomb rather than aiming a small one accurately. I do not consider the issue of diagnosis as important as the panel did, as it misses out the main point about OM (recurrence or persistence over time), which single diagnostic measurements do not adequately address. Nevertheless, I do not think that there is elsewhere as comprehensive a synthesis of the data sources of this question nor so clear a conclusion elsewhere. Just to show how unbiased I am (!) I am prepared to acknowledge that the value actually added by the reviews in this section is, paradoxically, rather high. Perhaps it is easier to make progress on what is a conventional elementary issue than on what is truly important; because the data are not so difficult to get, they happen to exist. However the published article must still emphasise that this question only addresses single frames in the "movie" that we need of OM histories. It is very reasonable to select four questions agreed as important for the concentration of evidence reviewed. However the Kendall's W for the prioritisation is only modest, and a treatment issue comes in a close 5th at rank total. The stopping at question 4 makes the Report's title, which refers to 'treatment' inappropriate. None of the top four questions is about treatment, which thus becomes only strictly relevant as the side-issue of treatment-free	• Again, we sympathize with the desires of the reviewer to have different questions assessed, however we had an explicit process for prioritizing the questions to be assessed given the time and resources available, and this was how the aggregate results turned out. No doubt some individual expert panelists, as some individual reviewers, would have preferred a different rank order or a different set of questions, but its a group process with representation from many of the interested parties, and the rank order we saw was how we proceeded. The title of the Report has been changed. Introduction has been revised.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	groups in natural history. This selection also makes some aspects of the general background introduction stick out as of less relevance. It is clearly stated that the introductory chapter is intended to be general background and not an overview of evidence. Yet certain aspects of it may be counter-productive. There is little appeal to understanding of process or therapeutic hypotheses, that would inform the critical interpretation of results. • for identifying the key questions of interest from the panel of technical experts: Yes, the ranking system used was a good method to get the technical panel to consider the rank-ordered importance of the questions, and the final ranking reflects questions that were not only important to answer, but also quantifiable. • the approach using technical experts to focus in on key questions seems reasonable (and is consistent with that used for the AOM report. • There is no better way to do this, only different ways. How did you get to the originally 20 questions? You only described this topic a little bit on page 31. But why did the original task order and letters from the nominating agencies propose only these questions? Based on difficulties in formulating guidelines? On questions of patients or problems of practising physicians? Experts (and other people) find it often difficult to arrange topics in order of importance. In my comment on the OMA report two years ago I made the same comment. However, in this former case experts had only to arrange 5 questions instead of the 20 now. The framing of the questions and the original order on the form can influence the final outcome. Probably other experts will have asked other questions. The technical experts are all Americans (seems logical), but the four key questions are therefore also culturally determined. The question of tympanostomy tubes is not investigated while this topic scored almost as high as the question on diagnostic methods. (51 *vs* 57). This problem is in my opinion more important in daily care of family physicians than the debate whether I should use a tympanometer or pneumatic otoscopy. • The key questions (page 31) were developed in a systematic way and the evidence report documents this process clearly. • The causal pathway for each key question is more clear in this draft	• Noted. • Noted • Noted. • Noted. • Noted.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	than from August 1999. The causal pathway (beginning on page 59, Table 5) should be developed more fully with each diagnosis having specific outcomes as outlined by Sakett's clinical decision making for each key question.[4] For example, Table 6 lists 5 diagnosis for otitis media with 4 outcomes and it is unlikely that all five diagnosis would yield all five outcomes? If there is a causal link, probability or prevalence data that support the likelihood of the diagnosis with the outcome, this information would be useful for standardizing future research and dissemination of the standard. 4 Sackett, DL, RB Hayes, GH Guyatt, P Tugwell. *Clinical epidemiology a basic science for clinical medicine.* 1991. Little, Brown and Company: Boston/Toronto/London	• Although non-English articles were not included, studies from other countries were included. Large studies from UK and The Netherlands are included. • True but prospective is a better design than case-control or retrospective studies.
	Searching, Reviewing, and Identifying the Literature • The literature in other languages should have been considered.	• Noted. • Comment noted.
	• Consideration should have been given to those contributions that were not prospective as there is valuable information. This is especially important in the large population studies in the UK and Finland where small differences become significant because of the power of the sample size.	
	• All that is not prospective is not bad and all that is prospective is not good!	• We shared the concern. However as can observed from Table 28 that there was insufficient data for this age group.
	• The exclusion of language data from prospective studies which were for children < 3 years of age eliminates consideration of the problems which these children may or may not have, In the consideration of the development of language, deviations and/or deficiencies at < 3 years can and do have effects later in life that manifest themselves in other areas and also may be found in more sophisticated measures of receptive language. The exclusion of this data from the report is a very serious flaw.	• The analysis was conducted by cohort and the evidence tables indicated when multiple articles are from the same cohort.
	• A number of the studies for question 2 – language- on the surface appear to have poor quantification of the duration of Otitis – most particular are #'s and these are 3/20 or 15% of the data	• Comments added to the Abstract, Summary, Results, and Conclusions.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	1. 1277 – Freeark **2.** 1623 – Kaplan **3.** 2135 – Paul • Of the 20 studies for question 2, 6/20 (30%) are from one of the panelists' research efforts and 4/20(20%) from another research group. Thus at least 50% of the reports really represent two populations of patients reported at different times. This should be noted in the discussion and in the conclusion for it does limit the generalizations that can be made. • In table 26 and evidence table 2 there are of the 20 citations really at best 11 different populations. Of these, a possible 6 are from special groups, Native American, lower SES etc. Any conclusions from these data need to modify to indicate the lack of generalization to typical populations. • satisfactory for what was reviewed • Where studies were excluded because of flaws was any attempt made to get the needed data from the authors so the study could be used? • Yes, the various databases used and the search terms used were appropriate, with the exception that the term "mastoid" seemed an odd term, as it is a specific portion of the middle ear. The general term "middle ear" would have seemed more appropriate. • The categorization of Speech-Language Tests was missing several tests, such as: Test of Language Competence, Woodcock-Johnson tests of Cognition, Kaufman Assessment Battery for Children, Wechsler Intelligence Scale-Revised, Stanford-Binet. Clinical Evaluation of	• Noted. • Requesting additional data from original authors is, like searching for non-English language literature, one of those decisions where a balance must be struck between time and resources. Our EPC has not been particularly successful in past attempts at obtaining additional data from original authors within the time frame needed to complete the evidence report, and therefore we did not elect to spend the resources for this report in that way. • Noted. • Revised and additional write-up inserted. • Experts assistance sought. Additional materials provided in Methods and Results chapters.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	Language Fundamentals –Revised (standard version). On page 84, the Stanford Binet is mentioned as a test of "expressive language". It is actually a test of verbal and nonverbal intelligence. It should be included in the table since it is referred to.	
	• The categorization of Audiologic tests was not described at all. There was a request from this technical expert for assistance in categorizing tests, and extensive information was returned, but this input, and the final categorization used is not mentioned or described in the methods. This is an important issue, because throughout the diagnostic tests section, terms are used that are not defined anywhere in the methods, and the cutpoints used are not defined for the various tests. For example, Type A and Type B tympanograms are introduced on pg. 78 with no definition or categorization provided. Another example is on pg. 81 – the terms "impedance tympanoscope" and Impedance audiometer" are used without definition. These are non-standard terms, so it is impossible to determine what they are.	• A discussion of this issue is included in the Limitations of the Evidence Report section of the Conclusions chapter.
	• The method is comprehensive with good described steps, but by selecting only English literature there could be some (?) language bias. For example, looking at the included studies for question 2 all studies were conducted in the USA and I can't imagine that good quality studies were only conducted on that site of the Atlantic. The Dutch OME groups (Maastricht Otitis Media with Effusion and KNOOP, Nijmegen) are publishing most of their important work in English, but I'm not so sure that all Scandinavian groups are also expanding their markets to English language journals. Whether OME is a problem in French or German speaking countries I don't know, but I can't imagine it isn't. Searching for foreign languages in Medline isn't very useful, while most of these journals aren't indexed. Only contacts with experts abroad could solve this problem. A discussion on this topic should be included in the discussion section of the report.	• Noted. • Case control studies were not considered to be sufficiently strong evidence by the technical experts to be included in this report. We synthesized all the evidence that we did find that met the a priori criteria specified by our technical experts. • Noted.
	• An important part of translating evidence into implementable guidelines for many users is quantifying the outcomes. Within the scope of this work, the sections titled natural history (page 16) and common outcomes (page 24), there is an attempt to quantify outcomes by a number needed to treat analysis.	• We revised our analysis to just pooling sensitivity and specificity and deriving PPV and NPV for various prevalence levels. A new figure (Figure 7) was generated for the plot for pneumatic otoscopy. We also

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	Synthesizing • This is incomplete because of the omission of the children < 3 years, the case controlled studies and other mentioned deficiencies. • The criteria for the language of the children with < 3 compared to > 3 months duration should have had some form of correlation function to determine a worse *case best-case scenario.* (p47) . • You have performed meta-analyses on all variables even though you note that there is significant heterogeneity in the prevalence rates. Of course the other option is to do an ROC meta-analysis a la Littenberg, Moses, et al., but that really doesn't yield a result that is very useful for guidelines. My recommendation would be to do meta-analyses of sensitivity and specificity as you have done, but not do them for PPV, NPV, accuracy or prevalence. I would then take the resultants of the meta-analyses for sensitivity and specificity and use them to compute PPV, NPV and accuracy over a range of prevalence values. These could be plotted • For many meta-analyses, there was significant heterogeneity for the estimate of the pooled effect. In these analyses, it would seem appropriate not to present the pooled estimate due to heterogeneous effects among the individual studies -- if there is significant heterogeneity than it is appropriate to combine results across the studies. Eliminating the pooled effect estimates in these situations would avoid the potential for readers to mis-state results. • seems appropriate, however I cannot oversee the details of including and excluding studies. Comment on page 82 worries me, however, I don't believe that inclusion/exclusion of a several studies would have altered the conclusion. Qualitative reviews, such as by Mark Haggard, have reached the same conclusions.	pooled prevalence rates to evaluate the heterogeneity of the studies. • We put in a caution in places where heterogeneity was encountered. • The inclusion and exclusion criteria were established a priori and described in Methods. For example, the Rach study which measured OM severity at 2-4 years of age, violated the criterion that OM severity is measured under 3 years of age. Since we could not separate the children whose severity were measured before 3 and after 3 years of age, it was excluded. • We used what were reported in the literature. It contained both ear and child as units of analysis. We reported both. However, there was not enough data to synthesize findings by child. This issue was addressed in Conclusion and Future Research chapters.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	• Is it appropriate to analyse 'ears' instead of children in question 1? I don't think so. In bilateral OME there is, I suppose, a correlation between the 'cure rate' in both ears of the same child. Reumatologists did analyses of joints instead of patients and the whole world was laughing at them. At least you should discuss the reason to do an analysis on this unit of analysis.	
Evidence	• This is all quite straightforward, and in my spot-checking of sources and references I did not turn up any significant literature missed, nor inappropriate judgments of quality, nor inappropriate syntheses. • Your conclusions appear to be that no definitive conclusions can be made. It appears from your conclusions that although there is much written there is a need for better research. You reiterated much that was reported in the 1994 Guidelines (Stool, et al). • With respect to missing any crucial pieces of information in the literature search, I feel strongly that the literature search was very complete. There will always be professionals who will feel that certain articles should have to be included, but I think that you have done an excellent job in reviewing the present literature. The report does support your conclusions as written. • The point about the many natural history and sequelae studies not sufficiently stratifying by treatment status is correct, but is overstated for three reasons: (1) Rightly or wrongly, in the US system the urge to treat makes it very difficult to obtain untreated controls, and any wholesale culture change on this is likely to be slow. (2) Most treatments are known to be of limited duration and effectiveness, so the disease effect (developmental sequelae) is thereby only slightly underestimated: (3) In the context of removing the financial interest bias which usually leads to overestimation of the impact of OM by professions with a vested interest, this failing is as least a conservative one. • A) I do not see any crucial pieces of information missing in the literature search. B) Yes, the evidence does appear to support the conclusions. • I did not read the evidence tables in detail. Probably, very few readers actually will. I will leave it to others, who know the literature better, to	• Noted • Noted • Noted • Noted • Noted. • Noted. • Noted.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	comment if you missed anything (which I doubt). • Since I'm not an otitis expert, I can't comment on whether any pieces of evidence were missed. The evidence reviewed seems to support the conclusions drawn. It looks like a really frustrating body of evidence to review. • I did not identify key studies missing from the search. Given that the conclusions are broad, the literature does support these conclusions. The conclusions are very conservative in their scope and do not extend beyond the literature review. With regard to the Natural History question, it may be worthwhile to use descriptive terms to present a range of time expected for effusions to clear after an episode of acute otitis media. Obviously, the exclusion of the current Pittsburgh study eliminates a potentially large body of important information; although, the length of follow-up disqualifies the interim reports from this study. Given the importance of the speech/language data, it may be worth mentioning the importance this study will have on understanding the impact on speech and language development. • it appears that a thorough literature search, including supplemental sources of data, was completed. The conclusion section makes appropriate mention of methodologic limitations in many of the published studies. As noted above, some further mention of the potential implications of multiple publications resulting from the same cohort of children (e.g., multiple publications by same research group along with a brief comment about how this was managed in the analyses) should be included in the conclusion section (and also in the summary). • Early Life OM and Long-Term Speech and Language: Your analysis supports that no conclusions can be drawn about the early life impact of OM on speech and language development in otherwise healthy children without preexisting developmental delays. I am concerned that this will be broadly interpreted as stating no relationship between OM and S&L development in all children. You need to state prominently and repeatedly the following points: 1) this analysis does not preclude an impact of *prolonged OME, especially bilateral,* on S&L development, 2) this analysis *does not apply at all* to children with OME who already	• The issue of middle-ear effusion persistence after AOM was addressed in the AOM evidence report, pages 80-82 and Table 24 page 119, and is essentially OME after AOM. In the Conclusion chapter, we pointed out that the results of several ongoing cohort studies would be useful for the speech and language question when they come out. • Agree. As long as in one single meta-analysis we did not include multiple studies, that meta-analysis should not be subject to bias from double counting data. However, when outcomes are aggregated, then bias would exist. Discussion of this issue included • Disclaimer added in the Absract, Summary, Results, and Conclusions.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	have speech or other developmental delays, and may have progress impaired by hearing loss or auditory input degradation from persistent middle-ear effusion. The concept here is to provide enough disclaimers so that children who suffer from delays secondary to OME *on an individualized basis* are not denied care based on lack of association at a group level. • Early Life OM and Long-Term Hearing: In Table 36 you report an RD of 11% (95% CI, 3-19%) for sensorineural hearing loss (SNHL) in children with early life OM vs. those without early life OM. This suggests that for every 9 children with early life OM we get one additional hearing-impaired child. Given that OM is a nearly universal experience in childhood, we would expect a correspondingly huge population of hearing-impaired kids, which is clearly not the case. This inconsistency most likely stems from defining SNHL as a threshold of >20-25db HL at any frequency in either ear. Using this rather liberal criterion, about 1:16 (6.4%) of OM negative kids also had hearing loss (a whopping 20% in the Sorri 1995 study!). Bottom line: hearing loss is NOT a dichotomous outcome; need to delve deeper here to describe the laterality, frequency or frequencies involved (eg, 8kHz is much less relevant than 1kHz or 2kHz to daily functioning), and magnitude of impairment in dB HL. There is also a huge issue of external validity: how representative are the 346 OM+ kids in these 4 studies of the larger population of OM+ in general? Unless the magnitude, clinical relevance, and generalizability of this hearing loss vs. OM relationship is put in better context, you are likely to instill unnecessary fear in the minds of parents of children who suffer from early life OM. • Diagnostic Methods for OME: A real meta-analysis picnic here! Very interesting: confirms value of low-tech clinical skill (pneumatic otoscopy) as preferred measure. Some other interesting data in Table 50: portable tympanometer has only fair specificity (68%), which leads to over diagnosis of OME; ditto for professional tympanometer with B or C2 curve (57%). The acoustic reflectometry analysis does not take into consideration the	• We did a sensitivity analysis, excluding the Sorri study and added comments in the Results section about the heterogeneity of the four studies with regard to exclusion criteria, type of OM history and how information on it was obtained. • Comment added to Conclusions. • Conclusions and Summary edited accordingly.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	impact of spectral gradient analysis. As noted in Table 8-10 of the Evidence-Based OM book, this technique gives better results that the dichotomous cut-point you used in the evidence report. The use of spectral gradient at least deserves some comment and mention. • I believe the evidence supports the conclusions. My understanding of the literature is quite in agreement with the conclusions. The major frustration for the professional readership is the continued biases from their experiences and their own interpretation of the literature. The conclusions from this report that basically says there is little certainty about some of the previous cause and effect of OME and the sequelae will be hard to swallow. Therefore, clear and concise summaries of the full impact of supporting data will need to be pre-digested and presented carefully. • I am not aware of any crucial prieces of info that were missed. The evidence does seem to support the conclusions. • Useful information for development of practice guidelines is the development of a balance sheet or quantification of the benefits, harms and costs for each study.[5] Adding a quantification analysis within the evidence table would allow for translation of an evidence based guideline into a clinical practice to assist with grading the evidence. The quality scores for each evidence table help evaluate the overall usefulness of the article for inclusion in the results for the evidence report. [5] Eddy DM. Chapter 7: Comparing Benefits and Harms The Balance Sheet. Clinical Decision Making From Theory to Practice. *A collection of essays from the Journal of the American Medical Association.* Jones and Bartlett Publishers: Sudbury Massachusetts. 1990: 48-56. • I don't think you missed any crucial evidence that I am aware of. The evidence does support the conclusions • **Omissions:** A. <u>Child behavior and quality of life outcomes</u>: Disturbances in children's behavior associated with otitis media have been reported to include restlessness and fidgetiness, frequent, disobedience, impaired task orientation in the classroom, inattention, short attention span and/or distractibility, attention	• Noted. • Quality scores are contained in the evidence tables and discussions of study quality for each key questions were added. • Noted • Our technical experts did not specify behavior as an outcome of our evidence analysis. We commented on the socioeconomic status. Quality of life is important but is also not a focus of this evidence analysis. Parenting is also important but not a focus of this evidence analysis. We agree that cost analysis is important, but it is outside the scope of our study. The pneumococcal vaccine is important but is not the focus of our evidence analysis. In addition, a recent

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	deficits and restricted social interaction. Paradise, Feldman, Colborn 1999 found that parent-child stress and behavior problems were consistently highest among children from the most socioeconomically disadvantaged homes. Haggard and Smith (1) reviewed the various studies of the impact of otitis media on the quality of life of the child. Since the Paradise article is listed in the bibliography I may have missed the discussion of this important issue. The Haggard article is not specifically referenced but others by the same author are in your bibliography. B. Parents also suffer: Chase (2) noted that parents of 1 year old children who had experienced otitis media were less effective teachers in structured interactions. They were less effective in gaining the child's attention, less able to respond effectively when the child was distracted from the task, and less able to help the child understand and perform the task. C. Cost analyses should be considered more completely. D. Pneumococcal conjugate vaccine should be described as having decreased the incidence of number of acute episodes and decreased the number of surgeries for placement of tympanostomy tubes for severe and recurrent episodes of AOM and OME. **References** 1. Haggard MP and Smith SC. Impact of otitis media on child quality of life. In: Rosenfeld RM and Bluestone CD, editors. Evidence-based otitis media. Hamilton, Ontario: Decker, 1999 pp. 375-378. 2. Chase C. Hearing loss and development: a neuropsychologic perspective. In Eavey RD, Klein JO, editors. Hearing loss in childhood: a primer. Report of the 102nd Ross Conference on Pediatric Research. Columbus, OH: Ross Laboratories 1992:88-94. **Any missed crucial pieces of information** • The large population studies which were not included because they were not "prospective"	article does not show a decrease in overall AOM, only a decrease in AOM due to penumococcus (Eskola, Kilpi, Palmu et al., NEJM 2001;344(6):403-409) • We agree. We are aware of several prospective studies being conducted. This was commented in Conclusions. • Case control studies are generally perceived to be more prone to bias than either cohort studies or randomized controlled trials. Empiric evidence exists to support this conclusion (Lijmer, JAMA, 1999:282:1061-1066). Therefore, while there may be 'substantial information' in such studies, the validity of conclusions based on this information is questionable, and therefore we did not include case

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	• As all case controlled and cross sectional studies were excluded, there is the probability that much information was missed. I could not determine how many of the articles were excluded on these grounds. There is substantial information available from these studies in the language domain that should have been included. • The second part of the language study was not performed (p48) which had to be done. Thus the conclusions concerning language have to be modified and it should be noted that there is data that has not been examined. The meta regression analysis may have provided some of the most useful information – who is susceptible and who is not. • by limiting the evidence included in the report, several important papers are 'missed'. This is however described clearly in the method section. In 2000 several important papers addressing question 1 (natural history) and question 2 (language) have been published: Maw et al. Rovers et al. • You missed a study of *Rovers MM et al* about the large Dutch KNOOP study in Pediatrics *2000* in answering question 2. (But that was published after your searches, so be excused). RCT with a non intervention arm. • In the conclusion (page 146 and 147) you mention large studies coming up the coming years. Please look also to Europe were MOMES and KNOOP are following children for several years. • The literature is focused and well documented regarding methods. It would be helpful to explain the screening tool in appendix H. Is this a strategy for keeping the evidence up to date and at what interval for review? A more clear explanation in the text and an independent explanation in the appendix would be helpful on its purpose. **<u>Does the evidence support the conclusions?</u>**	control or cross sectional studies as evidence in this report. • Due to time constraints, the second part of the question was not done. We hope that this will be done in the future. This issue was addressed in the Methods section. • Noted. They are after our cutoff date. • Yes the cutoff point was January, 2000. • Noted and added to Conclusions. • The initial screening for this study was described in the Methods section. • Unfortunately, properly designed and executed prospective studies are needed in this area. Case-control studies are not the answer to the problem. • Noted • Details on the type of hearing tests and how children were classified were added to the evidence table, Table 35 and Table 36.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	• The answer, for the language section, is unfortunately **NO** for critical portions of the evidence have not been included. This is a serious shortcoming as the expectation of such a government-initiated report is both accuracy and precision: as well as, to be free of bias. The language section fails in all three areas. • To my knowledge, no crucial pieces of information were overlooked in the literature search. • Question 3 - Early OME and Hearing outcome: it would be useful to specify the audiometric range considered in the studies. It is unclear whether all of the studies included in the evidence tables used the same criterion for hearing loss (e.g., 3-frequency or 4-frequency pure tone average). Further, the type of hearing loss experienced by children in the studies considered is not specified: in the four studies considered, was later hearing loss conductive or sensorineural in type? • Question 4- Diagnostic Methods: the classification scheme for types of tympanometry/acoustic immittance technology is not clear: 'professional' versus 'portable' may or may not mean the same thing as qualitative (tympanogram classification by pattern: A, B, etc) versus quantitative (specified in specific units) tympanometry. More explanation would be useful. The question of whether or not in every clinician's hands pneumatic otoscopy achieves better operating characteristics than tympanometry was not addressed. • The evidence supports the conclusions. • The evidence supports the conclusion well in this report. The report also indicates appropriately the need to be cautious with the interpretation of the studies on natural history given the poor quality of the data. Previous guidelines (1994) have looked at the effect on short term outcomes such as speech and language and the results and report similar conclusions. Given that the type and number of research studies hasn't changed the conclusions for key questions 1, 2 and 3, more standardization and direction is research is needed.	• Definitions, a table, and clarifications had been added to both the Methods and Results sections. • Noted. • Noted and addressed in Future Research.
Utility	• As a state of the art review, the evidence report will be of importance. The conclusions are of necessity limited because there are still more data to come to appropriate conclusions about the importance of OME.	• Noted

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	• There is utility for questions 1,3,&4 but not for 2. • Personally, I am disappointed that none of the selected topics deal with updating the treatment portion of the guideline. The results of this evidence report will make updating the guideline difficult because of that lack. • I am very disappointed. The only part of this that has potential practical utility is the section on diagnostic methods; and, even there, very little more is provided than what the AHRQ panel did seven years ago. The rest of the material provides intrinsically interesting summaries on natural history and the effects of OME on hearing and learning, but has no direct applicability to practice and, again, hardly advances where we were seven years ago. The conclusions on the relationship between OME and outcomes are expressed in many thousands more words and tables than we did on the AHRQ panel, but are not materially different. • I wish the EPC had spent a great deal less time in constructing the elaborate tables, meta-analyses, funnel plots, shrinkage plots, and ROC curves; instead expanding scope on the treatment questions that would be useful in developing a better clinical guideline. This strikes me as an example of methods run amok, losing perspective on why we should be investing in these reports in the first place. The report might be useful to someone planing more epidemiologic or intervention studies; but I cannot see that it will be useful to clinicians caring for children at risk for this condition. I believe it fails to provide useful materials "to develop clinical practice guidelines or medical review criteria." • The biggest roadblock to better research to write guidelines for treatment of OME is there is no definitive diagnosis. I am not alone in this thought as I read of disagreement among your technical panel on page 157. I think until all can agree on a definite diagnosis, treatment options will remain varied. Without a doubt, being in clinical practice, I concur pneumatic otoscopy is the best diagnostic tool with tympanometry as a good backup too. The problem is with the rising	• Noted • Addressed in section: “Limitations of the Evidence Report.” • Noted • We disagree with this reviewer’s assertion that this would be a better report if we had spent less time on analyzing the data for the key questions we did study and instead used this time to analyze more key questions, specifically those on management. We sympathize with the reviewer’s desire for more questions to be assessed, but we cannot do this by cutting corners on the analysis of those key questions we did include. There are, unfortunately, no short cuts in this process, and the assessment of key questions on management will need to await another day. • Noted.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	incidence if resistant bacteria (i.e. strep pneumonia) and the recommendation to be more judicious in our use of antibiotics we need to first clarify the definition of OME versus AOM. The problem lies with no consensus on a definitive diagnosis among the experts in the field. Until this is done definite conclusions and recommendations can not be made.	• Noted
	• With respect to finding this information useful in developing a clinical practice guideline, I do not feel that enough new evidence is there to go on further. However, that will be up to the guideline committee. You do have plenty of evidence to support the writing of a guideline or at the very least to updating the present guideline. The medical criteria for the diagnosis and treatment of AOM always gets resurfaced; but when the reports are written, it is always OME that is researched. This once again makes me realize the overtreatment of OME.	• Noted
	• I would find this information useful for the development of a clinical practice guideline in the speech-language-hearing domain, even though the conclusions of this Evidence Report on certain aspects (natural history, speech-language, diagnostic methods) of OME do not differ substantially from the recommendations of the previous Guideline (Stool et al., 1994). The evidence suggesting long term effects of early OME on hearing would be particularly beneficial in the development of a clinical practice guideline focused on speech-language-hearing and is consistent with recent reports (Bess et al., 1996) of the prevalence of minimal forms of conductive and sensorineural hearing loss at school age. Recommendations for future research are useful.	• Noted. Findings will be published in peer-reviewed journals.
	• This information would be useful in developing guidelines if synthesized into a more readable and accessible format (i.e., with a more brief executive summary, and inclusion of the final tables and conclusions.	• We have related to AHRQ this reviewer's comment that perhaps the composition of the technical expert panels for EPC reports might be changed to better reflect input from guideline developers.
	• The actual conclusions that are supported by the evidence report are limited, due to the stated problematic quality of the literature. A guidelines committee will have to wrestle with this reality, but at least they will know what evidence is available (and not available) for the four key questions. Since one of the purposes of an evidence report is to support guideline development, I think this should be taken into account prior to choosing the key questions. Factors to consider are what	

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	practical questions must be answered in a guideline, and the likelihood that there is adequate evidence available to draw conclusions. In the case of the OME report, each of the questions chosen is important and would probably have been selected by a guideline committee. In addition, however, questions regarding efficacy of each potential intervention, would also probably have been chosen, since this is what the guideline committee must eventually make recommendations on. Recognizing the limits of resources available for this evidence report, it would take a like effort to answer the next set of questions. The more academic members of the expert panel probably knew that the literature would be inconclusive on some of the key questions. Would a more guideline oriented panel have preferred a more limited preliminary review followed by an analysis of likelihood of useful conclusions? If the likelihood of definitive answers was low, the team could then move on to other questions. This would potentially give a guideline committee more to work with when it met. I would suggest that the sponsoring organizations empanel the guideline committees as part of the process of developing future evidence reports. Perhaps, the guideline committee, including consultants and liaisons should be one and the same as the technical expert panel.	• Noted
	• Since I am on the US Preventive Services Task Force, I'm quite accustomed to using reports like this one as a foundation for recommendations. It is a useful report. It would be hard to write recommendations for otitis media with effusion -- I can imagine the USPSTF would give diagnosis and treatment for this condition a rating of "insufficient evidence". However, I think this is due to intrinsic limitations of the evidence, more so than the methods you used or the report you wrote.	• We have added comments in the Abstract and Summary.
	• The evidence presented in this report supports a limited number of conclusions that can be directly applied to clinical practice. Diagnostic techniques are presented clearly and offer support for pneumatic otoscopy and the value of tympanometry. I would emphasize the importance of "well-trained" otoscopists and the possibility that the predictive values presented for pneumatic otoscopy may not realistically reflect the situation in clinical practice and therefore increase the need	• Noted.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	for other diagnostic tools and methods. The ambiguity associated with speech and language problem and risk for permanent hearing loss draw attention to important gaps in our knowledge and provide background information about otitis media without supporting strict time limits for surgical or medial interventions. • within the context of limitations resulting the available body of published studies, this document provides a comprehensive summary of published evidence relating to the key questions examined. Many of the published studies are of limited quality. It is unclear whether there is sufficient evidence upon which to develop a guideline. However, the report does identify areas requiring further research. • I would find it useful if I were developing a practice guideline. • I am a bit concerned here. The sound bites are: up to 40+% of OME resolves in 1-3 months, relationship of OME to speech & language development is anyone's guess, OM increases risk of SNHL by over 2-fold (the panic sets in…), and pneumatic otoscopy is best. Some polish and perspective are needed here, as suggested above. There is certainly some utility here for developing a clinical practice guideline. The info on natural history is useful, when viewed in conjunction with other data from RCT control groups as collated in the Evidence-Based OM book. The speech and language analysis saves a lot of effort in reviewing the literature, but doesn't necessarily provide any therapeutic guidance. The hearing loss analysis, in my opinion, doesn't accomplish much except generate unnecessary worry on the part of parents and providers (at least as presented in the draft report). Probably the most benefit comes from the diagnostic analysis. Great to stress pneumatic otoscopy. Also great to put in perspective the various tympanometric results (by static admittance, peak shape, and handheld vs. professional). • I do not see how the report helps much in formulating *treatment* guidelines. Clearly, much of this stems from the choice of key issues decided upon by the panel at the start. It also highlights the need for a follow-up to this report, perhaps focusing exclusively on issues of treatment. • I understand that there is a minimum 13-month publication delay for	• Noted. • Disclaimer added in the Abstract, Summary, Results, and Conclusions. For hearing, we did a sensitivity analysis, excluding the Sorri study and added comments in the Results section about the heterogeneity of the four studies with regard to exclusion criteria, type of OM history and how information on it was obtained. • The title of this Report has been changed. • The time from delivery of the final document to publication by AHRQ is out of our control. • Noted

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	evidence reports at the AHRQ. Given that an evidence report is, by definition, retrospective research, publication delays of that magnitude add insult to injury. Considering that the AOM report is (unbelievably) still not published, I suspect that 13 months is a rather optimistic estimate. Bottom line: timely publication is essential for utility. • Yes, this report is useful for professionals attempting to develop clinical practice guidelines for diagnosis and treatment of OME (not AOM). The conclusion that on the basis of the current published evidence these guidelines cannot be defined is most valid. • The information is crucial to the development of clinical guidelines for the approach to and management of OME. Incorporating the actual recommendations of the 1994 Guidelines and modifying them from the conclusions of this report would serve several functions: 1) Practitioners might cease some of the clearly ineffective current practices. 2) Educators would incorporate this new knowledge when they approach their students or residents. 3) Researchers in the field or considering entering the field could start from a more knowledgable foundation and learn from past errors in definition, study design and methodology, and validity of outcome measures tried before and failed. 4) Potential funders, both private and governmental, would have better guidelines for evaluation future research and promote investigators who would be less likely to repeat past errors. The key, as in everything else, is the executive summary. • Good, though discouraging in that even the definition and diagnostic criteria are not so well established. • The evidence is very useful for making guidelines. But it answers only 4 questions of the many you have to address in a clinical practice guideline for GPs. And more important, the whole reports says completely nothing about treatment, so may be its is wise to change the title in 'Diagnosis, natural course and late effects of otitis media with effusion'. • Since the report is intended to have the production of clinical guidelines by sponsoring organizations as one of its uses, it might be useful for the	• Comment noted. • Noted. • Title was changed accordingly. • We tried to incorporate all the comments in the Conclusions and Future Research for reference by those who develop guidelines. • Comment noted.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	members of the evidence based review to team to add a few paragraphs summarizing and recording any additional insights – if any given the useful comments in the various chapters of the report – they might have that would be of benefit to guideline development panels. • (Appendix D page 727) is a helpful tool to key persons involved in constructing a clinical practice guideline as it provides a summary of the different components regarding how the information is useful and what focus the guideline would apply to for population, intervention and the potential biases. A balance sheet (Eddy DM. Chapter 7: Comparing Benefits and Harms The Balance Sheet. Clinical Decision Making From Theory to Practice. *A collection of essays from the Journal of the American Medical Association.* Jones and Bartlett Publishers: Sudbury Massachusetts. 1990: 48-56) may be helpful in making recommendations (this may be difficult to do if there is poor quality of the evidence). The difficulty of a developing strong evidence for a guideline is the lack of Randomized Controlled Trial (RCT) and heterogeneity of the data. • (page 153). It would be useful in this report to have a template or a model which is more precise which could serve as a starting point for a standard development for these three key questions. Tables 5-9 could serve this purpose and it would be helpful to make reference to this in the conclusions and implications for future research. (Sackett, DL, RB Hayes, GH Guyatt, P Tugwell. *Clinical epidemiology a basic science for clinical medicine.* 1991. Little, Brown and Company: Boston/Toronto/London) • Currently there is a great deal of practice variation in the management and treatment of otitis media which is going to influence the results for natural history and short term outcomes which were two key questions of this evidence report. Future research should use uniform standards for monitoring, management, and treatment and then potentially more useful information will be available to improve the research for natural history and short term outcomes. (Berg Alferd O. Dimensions of Evidence. Journal of the American Board of Family Practice. 1998. 11(3): 216-223.) • Clinical judgement plays a large role in the assessment and	• Tables 5-9 were intended for this purpose. • In Future Research. • This is not a focus of this evidence analysis. • Comment noted.

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	management of otitis media in primary care practice. Standardized clinical charting forms may help reach more uniformity in documentation understanding the disease process. However, strategies regarding changing clinical practice styles needs to be incorporated. (Isues in Permanente Medicine. Focus: Evidence-Based Medicine. July 1999. 1-8. **)**	• Comment noted.
	• Patient preferences regarding management and patient relevant outcomes are becoming more important to clinicians and clinical practice regarding reaching shared decision making. Currently in this evidence report there is a great deal of back ground on the expert preferences and consensus. Focus groups and patient rankings would be helpful in developing patient education materials for understanding the report. Also, if additional key questions were to be addressed it would be useful to obtain the patients point of view for Table 3 and Appendix A, page 709 to understand their priorities. (Djublbegovic, B., Hozo I; Lyman GH. Linking evidence-based medicine therapeutic summary measures to clinical decision analysis. Department of Mathematics Indiana University Northwest. January 31, 2000. Available from URL: http:// www.medscape.com/medscape/)	• Comment noted.
	• Clinical guidelines can be developed by several methods including global-subjective judgement, evidence based, outcomes based, and patient preference based. Outcomes base is an evidence based approach with more quantifiable estimates of the benefits, harms, and costs. Additional analyses that could support this quantification would be helpful in making the guideline more useful to clinical practice and a broader audience such as your report is intended. Also, a grading scheme that can be developed to rate the evidence would be useful. (Geyman JP. Evidence-based medicine in primary care: an overview. Journal of the American Board of Family Practice. URL. Http:// www.medscape .com/ABFP/JABFP 1-18; January 31, 2000	• Noted. • Noted.
	• A recommendation regarding how frequently this should be updated should be included in the future directions. It would be helpful to describe a starting process for automatic data collection cycle for the next review if the intent of this guideline is to have a database for new article submission available online for ongoing update by researchers.	

Appendix I. Peer Review Comments and Responses (Continued)
[Editorial comments were excluded]

Area Addressed	Comment	EPC Response
	• I think that those persons developing clinical guidelines would have an adequate of information. I think that a team approach consisting of a statistician, researchers and clinicians would be most appropriate for developing these guidelines. • This information is useful for a large segment of our pediatric practices. It also will be a good source of information on areas in need of future research projects. With e-mail and e-commerce taking off, office based research can quickly yield results on large populations seen in day to day practices across the country.	
Others	• Way too much information to thoroughly digest in three weeks. • Upon receiving the report, I was a bit overwhelmed at the size of the material. I am happy to report that it really was not that difficult to review. • It would be useful to describe the pathophysiology of OME.	• Noted • Noted • We greatly revised and reduced the Introduction.