Author's response to reviews

Title: Efficacy of ambroxol lozenges for pharyngitis: a meta-analysis

Authors:
Jean-Francois Chenot (jchenot@uni-greifswald.de)
Peter Weber (weber.peter@gmx.net)
Tim Friede (tim.friede@med.uni-goettingen.de)

Version: 3  Date: 5 March 2013

Author's response to reviews: see over
Point by point response

Reviewer’s report
Title: Efficacy of ambroxol lozenges for pharyngitis: a meta-analysis
Version: 2 Date: 21 December 2012
Reviewer: Christian de Mey

Reviewer’s report:

A reply and a revised manuscript were received - most of my comments were not addressed. Several issues were left unresolved.

Due to an oversight we previously did not submit the complete point-by-point response to your comments and we apologize for any inconvenience. However, we did address most of your comments, in particular the major ones, in the manuscript.

GENERAL: by adding new references, the list of references has changed. Cross-references to the listed references have not been correctly updated: for instance, previous ref.12 continues to be referred to as ref.12 in the table or as ref.13 in the text; previous ref.12 now is ref.14.

We have corrected that. We apologize for the technical problem with the automatic reference system.

ABSTRACT – results: ... "Main outcome was a ratio of pain reduction measured repeatedly over 3 h compared to baseline on 5-item verbal rating scale" …
– Comment: this does not describe the effect criterion correctly – see below.

You have chosen a very complicated outcome which cannot be explained in all details in the abstract. We have however responded to your concerns in the method section (see below).

ABSTRACT – conclusions: ... "However, the additional benefits of ambroxol beyond three hours, though clinically important, remain unclear given that more than 50% of patients using mint flavoured lozenges for pain relief reported satisfactory results after 1 day" …
– Comment: as extensively explained in present ref.14, two approaches were used to investigate the effects of ambroxol relative to placebo: a so-called "pharmacodynamic" evaluation focused on the pain scores over the first three hours after a first lozenge (observations at the practice of the investigator) and a subsequent ambulatory so-called "pharmacotherapeutic" evaluation under conditions of relatively free use of the investigational medication (up to 6 lozenges per day with a minimum interval of 0:30 h between consecutive lozenges). The latter was based on the daily diary records of overall pain reducing efficacy using a 4-point VRS (1=very good", 2="good", 3="not so good", 4="poor"). As reported in present ref. 14, 69% and 53% of the patients reported an overall efficacy that was good or very good for the treatment with ambroxol (20 mg) and placebo respectively, after the first day. At no time, the pain relieving efficacy was evaluated in terms of being "satisfactory".

We have summarized “good” and “very good” as “effective pain reduction”. We do not believe that this is a misrepresentation of your results. To respond to your concern,
however, we have added “(“very good” and “good”)” and clarified that we refer to patients' evaluation of overall efficacy.

MANUSCRIPT – METHOD SECTION - Data extraction and analysis: … "The primary outcome was a time and baseline adjusted value for pain reduction on a verbal rating scale ranging from 0 to 5. Measurements were done at baseline after 30, 60, 120 and 180 minutes. The values were subtracted from the baseline and weighted to adjust for time" ...

– Comment: this approximation does not describe the effect criterion correctly; as detailed extensively in ref.14: pain intensities (PI) subsequent to dosing were expressed as arithmetic changes (PID) from baseline intensity … The cumulative sum of the post-dosing PID was weighted for time elapsed since previous assessment: SPIDAUC = 0.5*PID30 + 0.5*PID60 + PID120 + PID180. Additionally, the relief over the 3 hours after the first lozenge was expressed as SPIDnorm = SPIDAUC/(3*Plbaseline). This SPIDnorm equals the ratio of the achieved SPIDAUC relative to the maximum achievable effect; hence, a SPIDnorm = -1.0, means that full pain relief had been achieved already after 30 minutes and was maintained up to 180 minutes after dosing.

We believe that our wording is in essence but less complicated and in less detail what you have elaborated here. For the sake of clarity we have expanded this explanation in the method section.

MANUSCRIPT – METHOD SECTION - Data extraction and analysis: … "The secondary outcome was patients' satisfaction regarding efficacy was measured in all studies at the end of each treatment day using a 4-point verbal rating score (“very good”, “good”, “not so good”, “poor”)" ...

– Comment: the sentence is grammatically incorrect; moreover, as detailed above (and explained in ref.14 extensively), the patients scored their assessment of the treatments' efficacy as either “very good”, “good”, “not so good”, or “poor”; they did not score "satisfaction regarding efficacy".

We are now using the exact term in the whole manuscript.

MANUSCRIPT – METHOD SECTION - Data extraction and analysis: … "To facilitate interpretation of the clinical relevance of the treatment effect on the primary outcome we expressed the treatment effect also in terms of the probability that a patient treated with ambroxol achieves a greater or faster pain reduction within three hours than when treated with the control. This effect measure is known as the probabilistic index or relative effect and was suggested for the assessment of clinical relevance [11]."

– Comment: In the article by Kieser et al. [11] the "probabilistic index" was used to derive appropriate sample size formulae for the design of studies aiming at demonstrating both a statistically significant and clinically relevant effect. Referring to recent studies in multiple sclerosis, they discuss potential issues in the application of this approach. Whether and how this approach is applicable in the present context is uncertain – see also comments below.

The statistical arguments are independent of the indication. Furthermore, your argument works both ways. You are unsure if the conclusions of this paper allow to
draw conclusions for a different medical condition. This also holds true for the studies done by Farrar et al. Please see also comments below.

RESULTS - Study characteristics and assessment of reporting: "There are inconsistencies in the reporting participants’ age as inclusion criterion".
– Comment: in present ref.14, the eligibility range is reported in section 2.2 (Patient & methods – Study population) as 18 to 80 years of age; in section 3.1 (Results - 3.1 - Demography, patient disposition and baseline features) the actual age is reported as "Within each study, the subjects were on average about 36 years old; overall, ages ranged from 16 to 80 years". This difference between the eligibility range and the observed range reflects that a small number of patients were younger than the eligibility range, but that this deviation was not considered relevant (as assigned at the blinded report planning meeting).

In the method section you state 18 to 80 (Ref 14). We do not believe that including patients as young as 16 is a scientific problem but it is inconsistent and a violation of the inclusion criteria. In Section 2.2 the overall age of all five trials is reported, which does not allow to report age of two individual trials. We have added a footnote clarifying the inclusion criteria to table 1.

RESULTS - Study characteristics and assessment of reporting: "although it is stated that all patient were outpatients it remains unclear if patients were recruited in ambulatory care or emergency departments"
– Comment: as already pointed out in our previous review and as explicitly detailed in present in present ref.14, no doubt was left in this regard: "The studies were conducted in ambulatory fashion by qualified primary care physicians" … "All subjects were investigated in a real-life primary care practice setting" … "these studies were conducted as clinical phase-III trials in a real-life primary care setting. Others have conducted such trials as phase-II investigation in professional human pharmacology research clinics [ref]; such approach might yield higher sensitivity due to more homogenous investigator performance; however, it remains difficult to extrapolate from such stringently standardized experimental conditions to real-life practice"

We agree that it remains difficult to extrapolate results from one setting to another. The setting does not necessarily depend on the qualification of the health care providers and primary care is umbrella term including emergency rooms. We only make the reader aware that patients might not reflect what is encountered in general practice.

RESULTS - Study characteristics and assessment of reporting: "Risk of bias is presented in Table 2. The publications do not meet the CONSORT statement standards of reporting [16]"
– Comment: as already pointed out in our previous review, present ref.14 (previous ref.12) is a compact summary analysis of all five trials that were addressed by the present meta-analysis. Because of its specific objective and format such summary publication was neither meant nor is it able to comply with the criteria of the CONSORT-initiative (reporting of “parallel group randomized trials”). Characterising this publication as deficient, while not complying with the CONSORT-criteria is inappropriate. The authors had been made aware of this, but have not replied to this concern.
You basically confirm that the studies were not reported according to the CONSORT statement. This is a factual statement. We agree that not complying with CONSORT is suboptimal.

RESULTS - Study characteristics and assessment of reporting: "For instance, for only two trials patient flow charts were reported [15]"

– Comment: an extensive disposition table is presented by study in Table 1 of ref.14 (i.e. of the paper that presents a compact summary of all studies). Additionally, this paper summarises the disposition in explicit detail as "The data relate to 5 trials. In these five trials 1,777 patients were enrolled; 1,772 patients were randomised and treated. 1,713 were evaluable with regard to efficacy; 95 were discontinued from the trial prematurely; 1,609 were evaluable in terms of efficacy without confounding protocol deviations (see Table 1)" … "Nineteen discontinuations were due to AE (1% of the treated patients)" … "The safety database relates to five studies involving 1,772 adult patients who were treated and randomised: 23, 24, 620, and 240 patients assigned to treatment with 5, 10, 20, and 30 mg ambroxol, respectively, 613 patients assigned to placebo, and 252 patients assigned to treatment with benzocaine. This dataset is larger than that for the efficacy evaluations since not all patients treated were evaluable for efficacy" … Although not presented in single graphs (mainly due to editorial constraints), the patient disposition both in total and for each study separately was well documented. We agree with the authors that it might have been preferable to have been permitted to describe the disposition in (even) more detail.

You confirm our factual statement that for only two trials patient flow charts were reported as required by CONSORT. We show the data you reiterate.

RESULTS - Study characteristics and assessment of reporting: "From these two charts it appears that all screened patients were randomized, which is rather unusual for a clinical trial setting and is therefore casting doubt on the accuracy of reporting. The number of eligible patients screened is not shown for any of the other trials included in this review.

– Comment: as explained in ref.14, patients with acute uncomplicated sore throat of recent onset without signs suggestive of bacterial infection were to be considered as possible candidates for enrolment. Such patients were then screened after providing informed consent. Patients who provided informed consent and who participated in the screening evaluation (irrespective of the outcome thereof) were considered "enrolled". Enrolled patients who were confirmed to be eligible were then scheduled for an in-house investigation at the earliest possible opportunity. Patients presenting at that time were randomized and treated. As detailed in ref.14, 1,777 patients were enrolled, 1,772 patients were randomised and treated. In my experience it is not unusual that with such rapid progression between screening and investigation, the difference between the number of patients screened and the number of patients randomised is indeed so small.

We beg to disagree, unless you had a very good incentive, which is not reported (and you denied), the reported number of participants which were included in the studies after screening is unusually high. This is also one reason why we are concerned that the trials might not reflect general practice or primary care setting, where refusal to
participate in trials after screening for eligibility is much higher. In our own study, the participation rate ranged from 50 to 80%, see for example Gágyor I, Bleidorn J, Wegscheider K, Hummers-Pradier E, Kochen MM. Practices, patients and (im)perfect data--feasibility of a randomised controlled clinical drug trial in German general practices. Trials. 2011 Apr 1;12:91. doi: 10.1186/1745-6215-12-91.

RESULTS - Study characteristics and assessment of reporting: "Age of the included patients cannot be assessed for the two newly reported trials in the summary report [14]."
– Comment: in present ref.14, the actual age is reported as "Within each study, the subjects were on average about 36 years old; overall, ages ranged from 16 to 80 years".

This is a factual statement. In Section 2.2 the overall age of all five trials is reported, which does not allow to report the age of two individual trials newly added, as pointed out previously.

RESULTS – Primary outcome: "The summary of the observed difference of pain reduction is -0.11 (CI95 [-0.15; -0.07]) for 20 mg and -0.17 (CI95 [-0.24; -0.10]) for 30 mg of ambroxol. Although it seems that Ambroxol 30 mg is more effective than 20 mg the difference is statistically not significant (p=0.15). The standardized mean differences are -0.34 (CI95 [-0.46; -0.23]) and -0.43 (CI95 [-0.62; -0.24]) for 20 mg and 30 mg of ambroxol, respectively. These translate into probabilities of 59% and 62% that the pain reduction is greater or faster with ambroxol compared to placebo".
– Comment: These probabilities are of interest, but they do not solve the issue whether and how such treatment effect can be accepted to be clinically relevant or not. I certainly agree with the author that it is difficult to interpret this outcome in clinical terms. However, I disagree (as I already explained in the first review of the manuscript) that the primary outcome had been evidenced to be small. Firstly, the mean pain reduction (relative to the maximum theoretically achievable effect) with ambroxol ranged from 37 to 42% for ambroxol lozenges 20 mg (and 40 to 49% for ambroxol lozenges 30 mg) vs. 27 to 35% for mint flavoured placebo lozenges. The estimated mean overall difference between ambroxol lozenges 20 mg and placebo was about 11%. On average the effect of ambroxol lozenges 20 mg might (also) be understood as about 33.3% higher as that of sucking a mint-flavoured lozenge. Secondly, as already pointed out in our previous review: reasonably sound approaches have been proposed to qualify the relevance of a pain relieving medication effect: in-depth research by Farrar et al identified cut-off points of 30% (chronic pain) and 33% (acute pain) in pain intensity difference/pain relief to be clinically relevant/meaningful (Farrar JT, Portenoy RK, Berlin JA, Kinman JL, Strom BL. Defining the clinically important difference in pain outcome measures. Pain. 2000 Dec 1;88(3):287-94; 3. And Farrar JT, Berlin JA, Strom BL. Clinically important changes in acute pain outcome measures: a validation study. J Pain Symptom Manage. 2003 May;25(5):406-11). In all five trials presently analysed, treatment with ambroxol lozenges met this criterion, whereas treatment with placebo failed this criterion except in one small pilot study (N:21 treated with placebo).
The study (Farar et al. 2000) investigates the effectiveness of buccal fentanyl for cancer pain. We do not believe that this is a relevant publication to resolve the issue what is a meaningful difference in sore throat, which is most likely not the same. First of all it is acknowledged that it is disputed what is considered a meaningful difference in pain reduction. We state the same in the discussion. Secondly the gold standard used to identify meaningful pain reduction was the need for rescue doses. You did not collect data or control for use of other pain medication for pain relief as stated in reference 12. Thirdly the publication states that an absolute 33% of PID is meaningful with the regard to the use of rescue pain medication. However the mint-flavored lozenges also reduced pain in absolute values pretty close to 33% ranging from 27 to 29%. This is strengthening our conclusion that mint-flavored lozenges provide a meaningful pain relief. A conclusion also drawn by many others (e.g. Prescrire Int. 2009; 18(100):52. arznei-telegramm)
The publication additionally suggest to publish the % of patients with total pain relief (TOTPAR), data which is not available, but would be useful to draw a better conclusion. The second study (Farar et al. 2003) confirms the previous finding and does not change our assessment.

RESULTS – Secondary outcome: "As secondary outcome patients’ satisfaction regarding efficacy was measured in all studies".
– Comment: as detailed above: "patients’ satisfaction regarding efficacy" was neither measured nor reported.

We have changed that to the exact wording you use, i.e. "evaluation of overall efficacy".

RESULTS – Secondary outcome: "In two studies treatment only lasted 1 and 2 days respectively [14, 13]. For these two studies it remains unclear whether intention to treat or per protocol analyses were reported and if there was missing" – Comment: as detailed in ref.14: "treatments lasted 1 (BI 18.174), 2 (BI 18.173) and 3 days (BI 18.446, BI 18.468, and BI 18.489)". Accordingly, as detailed in Table 1 of ref.14, data were available for 1713, 1623, and 1415 patients treated for at least one, two and three days respectively. As detailed in ref. 14, all evaluations were based on the ITT and considering the data as available. Note: something is wrong with the last sentence of this statement from the revised manuscript.

We have corrected that.

RESULTS – Safety: "There was inconsistency in the report of the number of adverse events in the summarizing publication".
– Comment: I don’t really understand what the authors refer to. I checked, re-checked and cross-checked the data, but could not find inconsistency.

Well, look at table1 in your summary report and compare the numbers with table 4 reporting adverse events in the same publication, for e.g. 105 subject received treatment with 20 mg ambroxol in trial 18.173 but in Table 4 you have suddenly 107, in the placebo arm the number is increasing from 103 to 108. Half of all other numbers surprisingly do not match. We do not belief that this matters in terms of assessing safety of ambroxol but make our statement a factual one.

RESULTS – Safety: "In four trials five (1.36%) out of 368 patients treated with 20mg Ambroxol discontinued the study prematurely because of adverse events"
"In one trial the number of patients who discontinued treatment is unclear [14]."

**It is unclear if the reported dropouts were in the placebo arm, the benzocaine arm or the ambroxol arm. Selective dropout cannot be excluded.**

– Comment: I don’t understand how the authors constructed these numbers (neither the number of discontinuations nor the number of patients exposed to 20 mg ambroxol is correct) and why the authors conclude that the number of patients who discontinued treatment is unclear in one out of the five studies. The paper describes AE-related discontinuations as follows: "95/1,772 (5.4%) patients were discontinued prematurely. Nineteen discontinuations were due to AE (1% of the treated patients): Trial BI 18.173: one patient each with moderate nausea, moderate otitis media and severe pharyngitis treated with ambroxol (20 mg); BI 18.174: moderate abdominal cramps in one patient treated with ambroxol (20 mg); BI 18.466: one patient each with face oedema, upper respiratory tract infection, sinusitis (placebo group), one patient with upper respiratory tract infection (ambroxol 20 mg), one patient each with bronchitis, gastroenteritis and nausea (ambroxol 30 mg). BI 18.468: one patient with rash (placebo), one patient with worsening pharyngitis (ambroxol 20 mg) and one patient with dry mouth (ambroxol 30 mg); BI 18.489: one patient with fever (subsequently developing tonsillitis, nasal oedema, nasal congestion) and one patient with mild myalgia and pyrexia (placebo); one patient with moderate oral hypoaesthesia, one patient with gastritis, and one patient with eye allergy, increased lacrimation, and allergic rhinitis (benzocaine)". This accounts for 6 patients treated with placebo, 6 patients treated with 20 mg ambroxol, 4 patients treated with 30 mg ambroxol, and 3 patients treated with benzocaine who were discontinued prematurely from the five studies as listed.

*We agree the sentence above is a misrepresentation and we unfortunately not able to reconstruct how that happened. We have erased that sentence.*

**DISCUSSION:** "It is noteworthy that patients in the placebo arms with mint flavoured lozenges also had good reduction of pain in more than 50% compared to baseline".

– Comment: a "reduction of pain in more than 50% compared to baseline" was not investigated in any of the trials; this was neither observed nor reported.

*We have clarified that we refer to evaluation of overall efficacy.*

**DISCUSSION:** "The primary outcome as defined in the studies cannot be interpreted easily in clinical terms. Therefore, it is unclear what should be considered a minimal important difference (MID) on that scale for sore throat. The summarized observed pain reduction of -0.11 (CI95 [-0.15; -0.07]) for 20 mg ambroxol suggesting roughly 10% more pain reduction compared to placebo after 3 hours seems small".

– Comment: I certainly agree with the author that it is difficult to interpret the primary outcome in clinical terms. However, I disagree (as I already explained in the first review of the manuscript) that the primary outcome had been evidenced to be small. - see extensive comments on this topic above.

*Since there is no universally agreed standard on what is considered a relevant effect we believe that a further discussion based on papers published for other medical conditions is fruitless. We stick to our interpretation which is in line with conclusion drawn by others e.g. Prescrire Int. 2009; 18(100):52.*
DISCUSSION: "Sore throat is usually a self-limiting condition lasting on average for 6 to 8 days with decreasing intensity [17]."

– Comment: Cochrane: "Most sore throats are of viral origin, are easy to manage and do not require antibiotics (Howie 1971; Little 1996; Del Mar 2000; Kumar 2003). They usually resolve within three to four days (50%) and it is unusual for the illness extend beyond one week (about 10%) (Del Mar 2000)." Considering the delay between the onset of symptoms and the time of seeking medical advice for their relief, there is no reason to doubt that at least part of the observed placebo effect is due to spontaneous regression of the condition and the implications thereof – see ref.14: "At least two aspects need to be taken into account when assessing the observed effects in terms of relevance. First of all, sore throat is a self-limiting condition, which heals spontaneously in most subjects over the course of 2-3 days. This is well reflected by the changes from pre-dose baseline observed in the patients treated with placebo. Secondly, the sucking of a placebo lozenge with a distinct mint-flavoured taste is not an "inactive" treatment. Increased salivation by sucking the placebo lozenge has an evident pain relieving rather than a pain amplifying effect (otherwise eventually evidenced by pain on swallowing). Hence, the observed efficacy of the ambroxol lozenges has several components: the physical actions (sucking, salivation, etc.), the psychological aspects (anticipation of relief) and an effect component, which is specifically attributable to the pharmacological actions of ambroxol. The latter is evidenced by the biostatistically separated and estimated treatment effect. Since these efficacy components are not necessarily additive (especially when the placebo response is already relatively large), the effect component which is specifically attributable to the pharmacological actions of ambroxol might have been underestimated".

We assume that the non-pharmacodynamic factors increased salivation, anticipation of relief, are the same for both lozenges. We clearly state that mint-flavored lozenges are most likely not a placebo in the discussion (Ref. 20).

DISCUSSION: "For clinically relevant treatment effects beyond 3 h we have to rely on the secondary outcome the patients’ global assessment of efficacy after one and three days. However, these data were only presented by treatment arm pooled across trials [13] which prevents the application of appropriate meta-analytic methods which requires an analysis stratified by study". – Comment: I presume that this refers to ref.14 rather than ref.13. It is indeed correct that this was only reported in the summary paper (ref.14) and not in the individual reports.

We have corrected that.

However, both ref.14 and the present meta-analysis invariably indicate a quite high level of consistency among the studies with regard to the SPIDNORM-effects i.e. with regard to the primary criteria of pain relieving efficacy. This appears also to apply to this secondary criterion; as reported in ref.14 (across all studies) "at the end of the 3rd day, 78-84% of the patients treated with 20 mg Ambroxol scored the efficacy as 'very good' or 'good' vs. 22-16% of the patients who scored it as 'not so good' or 'poor'; with placebo in contrast, only 55-57% scored it as 'good' or 'very good' vs. 45-43% who scored it as 'not so good' or 'poor'".
Considering these narrow ranges, there is no reason therefore to doubt the reported pooled efficacy scores on the sole ground that they cannot be reconstituted by meta-analytic methods.

You confirm that the published aggregated data does not allow to perform a stratified meta-analysis. We do not imply it is very likely that a stratified analysis would change the evaluation dramatically. However, Simpson’s paradoxon is one of the reasons why stratified analysis and reporting is considered as standard for reporting summarized results. Our statement is factual and does not need to be changed.

DISCUSSION: "The reported differences for patient satisfaction after 1, 2 and 3 days varying from 13 to 16 percentage points [13] need therefore be interpreted cautiously because of methodological shortcomings. Also the effect appears to be rather small with more than 50% reported very good or good efficacy with both treatments after 1 day".

– Comment: I again presume that this refers to ref.14 rather than ref.13. As pointed out above, "patient satisfaction" was neither investigated nor reported. In ref.14 it is unclear why the reported global efficacy assessment outcome should be doubted on the sole ground that it was not derived by meta-analytical methods (see above). Even if such restraint were to be exercised, such caution is clearly not adhered to by the authors who repeatedly call upon these data in order to claim the effects to have been small or "unclear" in terms of their relevance. Also, the simplified argument that "more than 50% reported very good or good efficacy with both treatments after 1 day" is misleading while incomplete; as reported in ref.14: “at the end of the 3rd day, 78-84% of the patients treated with 20 mg Ambroxol scored the efficacy as 'very good' or 'good' vs. 22-16% of the patients who scored it as 'not so good' or 'poor'; with placebo in contrast, only 55-57% scored it as 'good' or 'very good' vs. 45-43% who scored it as 'not so good' or 'poor". From a clinical perspective both the differences with regard to the 'good' or 'very good' scores on the one hand and the 'not so good' and 'poor' efficacy scores matter to "clarify" whether the effect is relevant.

We apologize for the technical problem with the automatic reference system. Please see also comment above regarding non stratified analysis. Data needs to be interpreted and interpretation is matter to subjective judgment. We have already pointed out previously that our interpretation of the data is shared by others, e.g. the independent drug review bulletins arzneimitteltelgram and prescrire.

DISCUSSION: "The effect of local treatments naturally wears off after a few hours and data on repeated use of ambroxol or mint-flavoured lozenges is not provided which further complicated interpretation of the reported figures". – Comment: The course of the mean effect was shown to last for at least 3 hours after a first lozenge (see Figure 1 in ref.14). Data on repeated use of ambroxol are available: as specified in ref.14, the patients’ efficacy scores after 1, 2 and 3 days of treatment relate to ambulatory treatment with the assigned medication under conditions of relatively free use (up to 6 lozenges per day with a minimum interval of 0:30 h between consecutive lozenges).

We have clarified our issue that a conclusion on relationship between continued use of ambroxol lozenges and the effectiveness cannot be drawn by the provide data.
DISCUSSION: "In patients with associated systemic symptoms like arthralgia, headache and chills, over the counter analgesic medications with systemic action like Paracetamol (Acetaminophen) or Ibuprofen might to be a better treatment option." – Comment: as already pointed out in the first review of the manuscript, I fail to understand on why the authors claim that over the counter analgesic medications with systemic action like Paracetamol (Acetaminophen) or Ibuprofen might to be a better treatment option. This was not investigated in any of the publications analysed. To my knowledge there is no evidence to support any such claim.

Many patients presenting with sore throat have concomitant symptoms like headache and arthralgia. We are not aware that you or anybody else claimed effectiveness for those symptoms. When discussing alternative treatment options to ambroxol it seems appropriate to refer to a cheaper medication with proven efficacy against pain associated with sore throat and the frequently associated symptoms (Reference 4).

DISCUSSION: "Concomitant use of such medication in the reported trials could not be ruled out."
– Comment: ref. 14: "Most of the patients did not report having used other medication. No distinction could be made between the treatments in this regard". Additionally, it is self-evident that the use of possibly confounding medication (also if over-the-counter) was prohibited and that violation of this protocol directive was seen as major.

Ref. 14 Section 3.3 page 6 you wrote most patients did not report use of other medications. Therefore we conclude some did. You did not report that you excluded those patients.

DISCUSSION: "Adverse effects related to systemic ingestion of Ambroxol have been reported [22, 23, 24], however in this large sample no serious side effects from mainly topical application of ambroxol were observed."
– Comment: although having local anaesthetic properties, ambroxol is systemically well bioavailable also when administered by sucking a lozenge. The safety and tolerability of systemic ambroxol is extensively and sufficiently detailed in the Summary of Product Characteristics (SmPC) of the various ambroxol forms. Since ambroxol is a medication that is available for systemic use since several decades and since the related SmPC are subject to regulatory control and mandatory periodic update it would have been appropriate and sufficient to take reference upon these SmPC. Instead the authors introduce (de novo in this revision) references to highly unusual casuistic findings that are inadequate and doubtful on many accounts, with regard to the presumed causality of ambroxol in particular.

We essentially agree that ambroxol is safe and we do not see why we should not cite available information beyond the SmPC. We have added the word case report in the manuscript.

DISCUSSION: "The RCTs were all sponsored by the manufacturer and did not meet current standards of reporting."
– Comment: see above

See response above.
DISCUSSION: "A patient flow chart as stipulated by the CONSORT-statement [16] was only available for two trials [15]."
– Comment: see above

See response above.

DISCUSSION: "The settings where patients were recruited are not sufficiently described".
– Comment: incorrect - see above

See response above. As stated above our concerns about the setting based on the reports remains unchanged.

DISCUSSION: "Selection bias is very likely since none of the included trials report the number of screened patients for eligibility ".
– Comment: incorrect – see above

As stated above our concern about the unusual high inclusion rate of screened patients has not changed. We still do not know how many patients were screened overall since you do not provide data on patients who were screened and did not fulfill inclusion criteria, e.g. subjects with suspicion of bacterial throat infection.

DISCUSSION: "There were only few dropouts and it is unclear whether patients received some kind of incentive for participation and completion of the study".
– Comment: the trials were designed and conducted as conventional clinical phase-III trials; for such trials, incentives are highly uncommon. There is no reason to presume that the reported trials might have been exceptions to this rule. No incentives were provided or paid.

It is difficult to incorporate personal unpublished communication about facts like incentives. We agree that phase-III studies in Germany are usually not paid, but we know little of the setting and circumstances in the RSA and two studies of which the country where the studies were done remains unpublished. We have changed the sentence from unclear to not reported, what makes it a factual statement.

CONCLUSIONS: "Ambroxol is slightly more effective in relieving pain in acute sore throat than mint flavoured lozenges over a period of 3 h".
– Comment: the authors fail to provide any reasonable explanation why they qualify the observed effect as "slight".

Please see comments above.

CONCLUSIONS: "However, the additional benefits of ambroxol beyond three hours, though clinically important, remain unclear given that more than 50% of patients using mint flavoured lozenges for pain relief reported satisfactory results after 1 day".
– Comment: the studies have not investigated nor have they reported "satisfaction". The authors fail to provide any reasonable explanation or evidence why the patients’ efficacy scores should be reported in such oversimplified and therefore possibly misleading fashion – see above.
We have now changed the term to the exact words you have chosen in your publication in the entire manuscript.

CONCLUSIONS: "In patients with associated systemic symptoms over the counter analgesic medications might be a better option".
– Comment: as already pointed out in the first review of the manuscript, I fail to understand why the authors claim that over the counter analgesic medications with systemic action like Paracetamol (Acetaminophen) or Ibuprofen might to be a better treatment option. This was not investigated in any of the publications analysed. To my knowledge there is no evidence to support any such claim.

We have commented on that previously and above. We do not claim that you have studied that but believe it is the task of a review to put things into perspective.

In summary: the revision failed to address and/or to resolve several of the concerns raised in my review of the first version of the manuscript. The publication presents a pooled analysis of data that already were reported extensively 4 years ago (ref 14); the present publication does not lead to effect estimates that are essentially different from those already reported; however, the present paper is deficient in several important methodological aspects, particularly those pertinent to the evaluation of the clinical relevance of the observed effects and the assessment of the quality of the available study reports; additionally, the paper concludes ill-substantiated claims for instance with regard to the likely superiority of alternative medications, although such claims were neither investigated nor reported in any of the publications that were analysed.

We disagree. We have corrected factual errors and improved the precision of our meta-analysis thanks to your meticulous review.