Author's response to reviews

Title: S-carboxymethylcysteine in the treatment of glue ear: quantitative systematic review

Authors:

Robert A Moore (andrew.moore@pru.ox.ac.uk)
Dermot Commins (dcommins@freeuk.net)
Grant Bates (grant.bates@surgery.oxford.ac.uk)
Ceri J Phillips (C.J.Phillips@swansea.ac.uk)

Version: 3 Date: 7 Mar 2001

PDF covering letter
S-carboxymethylcysteine in the treatment of glue ear: quantitative systematic review

Reply to referees

Ian Williamson

Searching was not performed according to acute otitis media as the search term. Instead we use a highly sensitive strategy that looks for the drug itself, and then evaluate the likelihood of an article being a randomised controlled trial in glue ear. We have found this to be more labour intensive than using Cochrane-style searching, but it has been shown to be highly sensitive over many dozens of reviews.

The myringotomy study was probably not randomised, and additionally we felt was far from standard current clinical practice. We did not mention that the study was not randomised in the text, and we should have done.

We agree that selection bias could be a problem, but not much of one. In the UK it has become standard that children are accepted for surgery three months after referral, referral usually comes after a period of watchful waiting or antibiotic treatment, or both, in primary care. The three months can be used for treatment with carboxymethylcysteine. There is nothing in the inclusion criteria in the original reports to suggest that children in those studies differ in any significant particular from today’s. We cover the points in the discussion.

Chris Cates

Chris seems to think that we have been sufficient silly as to aggregate all data and then calculate the relative risks. Let me assure him that we do not. We use standard equations in our own model run on Excel. This was set up some time before Revman, it was done in conjunction with the Centre for Statistics in Medicine in Oxford, and was thoroughly tested against other data using RevMan and other packages. Not only has it been used for calculating information for about a hundred systematic reviews, but is used almost daily by us and collaborators around the world. It is also used to calculate numbers for Bandolier, and while discrepancies of a minor nature are found occasionally in confidence intervals (usually due to rounding errors in programmes), major discrepancies are rare (as in Pignatoro where some of the odds ratios do not compute with data given, which was one spur to perform this systematic review.). Heterogeneity tests are not much use for detecting heterogeneity, and we reference this.
The statistical methods are given, and have been accepted in exactly the wording given in about 100 reviews, including the BMJ, BioMed, Pain, and others with good statistical reviewers. Moreover, many of those reviews also exist as Cochrane reviews, and no discrepancies have been found.

For these and other reasons we think that the statistics are described adequately. We do not believe that the calculation of an overall relative risk applied to particular situations is helpful here. We understand the point, but have no intention of doing it.

We understand the comments about ears and children. Which is why, of course, we undertook sensitivity analysis and report ears and children separately in Table 2. The results were the same. Children may have one or both ears affected, though surgery is less likely with one good and one bad ear. From the point of view of surgery avoided, it doesn't make much difference.

Matthias Egger

Matthias takes us to task on a number of points, some of which overlap with those raised by Chris.

On the QUORUM statement, we are tempted to recall David Sackett's comment that all consensus is wrong. We applaud QUORUM, and most of the QUORUM features are met in this manuscript. In particular, the fate of all retrieved and read papers is described in results, and excluded papers referenced with reasons for exclusion. There is no particular merit in performing a flow diagram that adds no more information. Those abstracts not retrieved were numerous, and it has never been our policy to waste time on irrelevance.

The statements in the introduction are referenced, and are derived from the superb dissertation on glue ear from the people at CRD (reference 1). One might put the reference number behind each sentence, but we deemed this something of an overkill.

The way in which we extracted improvement data comes is explicitly laid out in the methods section:

"We were interested in outcomes (including adverse effects) for children, or failing that for individual ears, which could be interpreted as saving a child from an operation for grommet insertion. Ideally this would be a tympanogram type B becoming type A. Where this was not available (because several studies were quite old) we chose outcomes which would today influence the making of the clinical decision not to operate. Definitions used were:
A normal tympanogram
Reversion to normal audiometry
Normal otoscopic findings
Parental view (improvement/lack of problem)."

We fail to see how we can do more, especially having agonised over this to ensure that we ended up combining information irrelevant to today.

We do take Mattias' arguments about the Oxford score for trial quality. However, it is now probably the single most used scoring system, and has the advantage of being brief and not falling into the trap of trying also to score for validity, which is often a very different problem. It is also useful because we know from several publications by Alex Jadad that scores lower than 2 are subject to bias. Matthias is right to chastise us for omitting to reference these, and not fully to discuss the possibility of bias in the discussion. We have rectified this. However, the quality and validity scores, set up to measure different things, gave results with overlapping confidence intervals, and with the perennial problem of sensitivity analysis that with every cut the numbers get smaller.

Stratifying by key domains of methodological quality would be interesting, but with so few trials a not very useful exercise. Even with about 90 RCTs for aspirin were unable to show validity of operational factors to make much of a difference in trials of minimally high quality. We know that poor quality is problematical already.

The statistical issues are referred to above. The method is referenced. Ears and children is discussed above. The intention was to calculate NNH if data presented. It didn't but perhaps we need to say that better than to say that only one adverse event was reported in the three trials that mention it.

Forest plots and relative risk are unhelpful. L'Abbé plots are better. The difference between studies is simple an issue of size and the random play of chance. We reference this.

We dislike the idea of a combined relative benefit used to calculate an NNT, as we described above. All the information is there if others want to do it, but we dismissed it as artificial.

The Porru trial has that data available that was useable based on our inclusion criteria. We decided to include it. If we exclude papers with fewer than 10 patients per group we have been criticised for that. If we include them we get criticised. All the information is available in the tables for anyone to make their own assessment if they so want.
The Commins paper is published. The offending sentence will be removed.