

Factual errors in the Cochrane Review (Webster et al, 2023) on Betahistine in Ménière's disease

Robert Gürkov, MD

The Cochrane review (Webster et al, 2023) on betahistine in Ménière's Disease (Hydropic Ear Disease) represents a problematic handling of evidence. The existing clinical trials in this field are extremely heterogeneous. Furthermore, there is a clear correlation between the quality of trial and its result: low-quality trials typically claim to find a significant therapeutic effect of betahistine in Hydropic Ear Disease. High-quality trials usually do not find a therapeutic effect.

When summarizing the evidence from such extremely heterogeneous trials, it is obvious at first sight that a meta-analysis of several trials is neither feasible nor useful. Therefore, a narrative systematic review is warranted. It allows to clearly assess individual trial quality and preserves the informational content of the original studies and hence is the most valuable and useful approach to provide the decision-makers (physicians, healthcare providers, guideline developers) with the best available evidence as a basis for their decisions.

However, the Cochrane 2023 review chose another path: it chose to blend extremely heterogeneous studies into one melting pot and transformed the original data structure by an analysis method using aggregation of primary data. This aggregation leads to a substantial alteration of the primary data structure and a significant loss of information. For example, the longitudinal data which have been collected daily during 9 months (!) of continuous treatment in the BEMED trial have been truncated to a single cross-sectional data point at one single time point. Furthermore, this aggregation of primary data fails to yield the desired positive effect of enabling a meta-analysis because the data are still too heterogeneous across trials. It merely leads to an artificial weakening of very robust and precise results with a high statistical power and consequently to a misleading perpetuation of "uncertainty" where in fact clarity exists.

Apart from this unsuitable method of analysis which distorts the extremely valuable primary data, this Cochrane review suffers from numerous factual errors. The amount and severity of these errors warrant a major revision of the Cochrane review. These errors are explained here in order to provide decision makers with the necessary information for interpretation of this problematic review.

Errors are grouped by error category and original study. Page numbers refer to the Cochrane review.

I. Errors in basic study characterisation

BEMED trial (Adrion et al., 2016)

1. Incorrect reporting of betahistine dosage

Page 42

Error: The original trial randomized patients to a high dose of (144 mg/day); The wrong dosage, only 72 mg/day, is reported in the Cochrane review.

2. Incorrect classification of a fully reported prespecified primary endpoint as selective reporting

Page 44

Error: The primary endpoint was prespecified for months 7–9 and fully reported; months 10–12 were follow-up only. The Cochrane review misclassifies this as selective reporting bias.

Albu trial (Albu et al., 2016)

3. Incorrect classification of the comparator

Page 45

Error: Classified as “betahistine versus placebo/no treatment”, whereas the actual comparison was intratympanic dexamethasone + betahistine vs intratympanic dexamethasone + placebo.

4. Incomplete description of the intervention regimen

Page 45

Error: Repeated intratympanic dexamethasone re-treatments “on demand” were part of the published study design but are not reported in the Cochrane review.

5. Failure to report timing and number of repeat intratympanic dexamethasone treatments

Page 45

Error: Timing and number of re-injections is not reported despite clinical relevance.

6. Asymmetric reporting of repeat intratympanic dexamethasone treatments

Page 47

Error: Additional intratympanic treatments are mentioned only for the placebo group although they occurred in both groups.

7. Mischaracterization of intratympanic dexamethasone as identical background treatment

Pages 45 and 18 and 5 and 19 and 96

Error: Unequal exposure between groups is not captured when describing the background intervention.

8. Ignoring incorrect percentage calculations for numbers of patients with repeat intratympanic dexamethasone treatments

Error: Percentages reported in the study publication correspond to the randomized population, not the analyzed population. This contradiction is ignored.

9. Misleading “triple-blind” classification

Page 44

Error: Primary outcome was patient-reported vertigo diaries; no independent blinded outcome assessor is described.

Mira trial (Mira et al, 2003)

10. Ignoring author communications reported in previous Cochrane review

Error:

The Cochrane review 2001 states:

“In Mira 2003 two randomized lists (one for Ménière's disease and one for BPPV) were generated centrally by the Medical Department of the pharmaceutical company that supplied the drug and placebo tablets, using Fisher and Yates random number tables. The study investigators assigned the study admission numbers "corresponding to the progressive number reported in the related randomization list (i.e. according to the two diagnoses) to the participants (personal communication, Mira 2003). “

However, the Cochrane 2023 review describes the randomization as:

„Comment: indicative of block randomization, stratified by condition (MD or PPV). However there is insufficient information about sequence generation. “

Schmidt trial (Schmidt et al., 1992)

11. Incorrect statement that baseline attack frequency was not reported

Page 67

Error: Baseline vertigo data are reported in the original publication in the text and in the results figures. The reported vertigo score is a combination of vertigo attack frequency, duration and severity. In their methods, the Cochrane review authors state that, concerning the primary outcome of interest for their review, on page 10,

"Vertigo symptoms comprise a variety of different features, including frequency of episodes, duration of episodes and severity/intensity of the episodes. Where possible, we included data for the vertigo outcomes that encompassed all of these three aspects (frequency, duration and severity/intensity of symptoms)." Hence, it is an error to ignore the baseline vertigo score which is reported in the original study.

12. Incorrect statement that randomization numbers were unclear

Page 69

Error: The original study clearly reports 20 betahistine and 20 placebo patients as being randomized.

13. Incorrect assessment of blinding of study personnel

Page 69

Error:

The original publication describes maintained blinding and reports specific unblinding circumstances, e.g. when "the code was broken" at the occasion of drop out patients with possible major adverse reactions to the study treatment. Also, the study reports in the results section that "once the code was broken, the drug that was used in the period of preference was known." This represents evidence that blinding was upheld in a general (double-blind) manner, not only for the patients.

Furthermore, on page 9 of their introduction, under the subheading "Requirements of clinical drug trials in Meniere's disease", the original study describes blinding as essential for a clinical trial in MD:

"A trial has to be double-blind. Neither doctor nor patient should know which of the medications to be studied is being given." This represents further evidence for double-blinding in this study.

On page 46, the original study mentions again that "The effect of sustained-release betahistine dihydrochloride was studied in a placebo-controlled, randomized, double-blind, cross-over trial in patients suffering from Meniere's disease".

On page 52, the original study states again that "Capsules containing 24 milligrams of sustained-release betahistine dihydrochloride were used as the active drug (betahistine SR) in the placebo-controlled, randomized, double-blind, cross-over trial in patients with Meniere's disease".

On page 64, the original study clearly describes possible events that would lead to "breaking the code", making clear that the code was not broken before the end of the trial.

The methods section and the discussion of the original study describes it as double-blind on several occasions. It also specifically discusses the difference to single-blinding.

On page 159 of the discussion, the original study states that "Double-blinding can only be guaranteed as long as neither the patient nor the investigator can tell the difference between active medication and placebo."

In the Cochrane review, however, this is erroneously treated as unclear blinding.

14. Reduction of longitudinal outcome data to a single time point

Page 24

Error: Monthly longitudinal repeated measurements are erroneously represented as a single measurement at the 16-week time point.

15. Incorrect maximum vertigo score

Page 24

Error: Reported maximum score (63) is incompatible with the score definition in the original study (maximum 270).

16. Failure to incorporate previously published study details

Pages 67–69

Error: Information previously obtained from study authors and published in an earlier Cochrane review is not incorporated, e.g. number of randomized patients in each group.

II. Errors related to eligibility criteria and diagnosis

Mira trial (Mira et al., 2003)

17. Inclusion despite violation of stated diagnostic eligibility criteria (definite or probable Menière's disease)

Page 56

Error: The original study included patients with possible Menière's disease.

18. Incorrect summary statement on diagnostic reporting

Page 56

Error: Diagnosis is presented as “not reported” despite being clearly specified in the original study.

Duphar study (Duphar, year not reported)

19. Inclusion despite missing diagnostic classification

Page 50

Error: Definite/probable diagnosis is stated as not reported, yet the study is included.

III. Errors in analysed population and attrition

Mira trial (Mira et al., 2003)

20. Ignoring previously clarified analyzed sample size

Pages 56–57

Error: Earlier clarification by author communication in the previous Cochrane review that only 72 patients (instead of 81) were analyzed is not carried forward. Instead, the Cochrane 2023 review states that " We therefore assume that complete follow-up was achieved".

Khan trial (Khan et al., 2011)

21. Ignoring contradictory dropout numbers

Page 55

Error: Multiple incompatible dropout numbers are reported in the original study.

22. Ignoring mismatch between analyzed population and baseline table

Page 53

Error: Baseline characteristics are reported for only 104 patients and not 106 analyzed patients in the original study.

23. Ignoring unclear analyzed population for vertigo outcomes

Pages 54–55

Error: Responder denominators are inconsistent with multiple different reported drop out numbers and imply unexplained exclusions in the original study. (106 analyzed patients, but each of the 3 groups contains 35 patients = 105 in total). The analyzed population in the Cochrane review is therefore ill-defined. The Cochrane review uncritically imputes implausible numbers into their analysis.

IV. Errors in outcome definition and reporting

Khan trial (Khan et al., 2011)

24. Incorrect statement that the definition of improvement was unclear

Page 54

Error: Improvement is clearly defined in the original publication: " Overall improvement as defined as improvement in vertigo. Improvement in vertigo meant decrease in number or severity of attacks or both."

In contrast, the review states: " It is also unclear how participants would have been judged to have 'improved' if there had been improvement in one measure but not the other (e.g. if the frequency had reduced, but the severity increased)."

25. Ignoring unresolved inconsistency between text and figure

Page 54

Error: Text and figure report different responder numbers (23 vs 24 patients with improved symptoms) in the original study.

26. Incorrect assessment of baseline age comparability as "not excessively similar"

Page 54

Error: Age distributions are nearly identical in the original study. Age was 51.7 vs 50.2 years (mean) and 51 vs 51 years (median).

Mira trial (Mira et al, 2003)

27. Ignoring internal contradiction for baseline vertigo attack frequency

Error: The baseline table reports a monthly attack frequency of about 1.0 in the betahistine group. But the results report a baseline attack frequency of 6.7. The review ignores this contradiction.

28. Ignoring internal contradiction for decrease in attack frequency

Error: The results figure in the original publication shows a 63% decrease in attack frequency, whereas the text reports a 69% decrease in attack frequency. The analysis in the Cochrane review ignores this internal contradiction.

Ricci trial (Ricci et al., 1987)

29. Failure to report a major baseline imbalance

Page 61

Error: Marked baseline imbalance across treatment groups is not addressed.

Pre-treatment vertigo-frequency (calculated from the treatment duration) was markedly higher in the placebo group: the attack-free interval was 31,2 days for the betahistine group and 21 days for the placebo group, corresponding to a 50% difference; this difference is not mentioned in the review.

30. Unresolved inconsistency between baseline and outcome data

Page 61

Error: Baseline and outcome values are internally inconsistent. Since the mean treatment duration (defined as 10 times higher than the average interval between vertigo attacks before the therapy) in the BH group was 10m, and in PL group only 7m, it must be assumed that the mean pre-trial vertigo-frequency was higher in the PL group: vertigo-free interval was 31,2 days for BH and 21 days for PL. However, in the baseline data table, the mean vertigo-free interval was given as 26 for BH group and 29 for PL group. The review ignores this internal inconsistency.

V. Misclassification of implausible results

31. Misclassification of Albu trial:

Error:

- a) The patient numbers and percentages reported for patients with additional Dexamethasone injections are contradictory to the analysed patient numbers. This is an implausible result.
- b) The number of drop outs reported are contradictory, they are reported as 3 in one location, and 4 in another location. This is an implausible result.

c) The Kaplan–Meier analyses are not appropriate for the study’s treatment structure. Kaplan–Meier curves are a descriptive survival tool, but their use in comparative inference implicitly assumes that:

- treatment exposure is effectively fixed (or at least definable) over time,
- censoring is non-informative, and
- post-baseline changes that are outcome-dependent do not confound the hazard comparison.

In Albu et al., these assumptions are violated because intratympanic Dexamethasone (ITD) is:

- repeated as part of the protocol,
- triggered by clinical course (vertigo persistence/recurrence),
- and delivered unequally across arms,
- with unknown timing.

This creates a feedback loop:

Vertigo recurrence (the Kaplan Meier “failure”) influences subsequent ITD exposure, and ITD exposure influences future vertigo recurrence.

Under such conditions, a Kaplan–Meier (KM) “time-to-failure” curve cannot be interpreted as reflecting the causal effect of betahistine (or even the causal effect of the initial randomization). Furthermore, the visit schedule versus the apparent time resolution creates a conflict.

The protocol states patients were seen every two months. That implies re-treatment decisions could only occur at discrete two-month intervals. Yet the KM plots display time in months, and the curves appear to change on a finer timescale.

Given these violations, this is an implausible result.

However, the review ignores these implausible results in the “research integrity checklist” and misclassifies the study as “The study is free from implausible results”.

32. Misclassification of Mira trial:

Error:

- a) Baseline table and results report contradicting numbers for vertigo attack frequency at baseline. This is an implausible result.
- b) The original study reports that the intensity score of vertigo was more frequently improved in BE-treated patients than in the PL-treated group. However, the methods section it states that the patients reported only the NUMBER of vertigo attacks per months (not the intensity). (intensity and duration was only assessed within the GISFaV questionnaire, and these results are reported separately). This is an implausible result.
- c) The original study reports that the duration of attacks was improved significantly more by BH than by the PL by the 2nd month. However, the methods section it states that the patients reported only the NUMBER of vertigo attacks per months (not the duration). (intensity and duration was only assessed within the GISFaV questionnaire, and these results are reported separately). This is an implausible result.
- d) In results table 2, some data are missing and marked as "NS", which is explained in the table caption as "statistically not significant". This is an implausible result.

However, the review ignores these implausible results in the “research integrity checklist” and misclassifies the study as “No implausible results were noted”.

33. Misclassification of the Ricci trial

Error:

Since the mean treatment duration in BH was 10m, and in PL group only 7m, it must be assumed that the mean pre-trial vertigo-frequency was HIGHER in the PL group: vertigo-free interval was 31,2 days for BH and 21 days for PL, that is a 50% difference. BUT: in the baseline data table, the mean vertigo-free interval was given as 26 for BH group and 29 for PL group. This is an implausible result.

However, the review ignores these implausible results in the “research integrity checklist” and misclassifies the study as “No implausible results were noted”.

34. Misclassification of the Khan trial:

Error:

- a) The duration of the study is entirely unclear. There are only 2 mentions of a time span in the whole original study:
The abstract states that: " The patients were reviewed after 06 weeks."
The methods section states that: "All Patients were followed for a period of one year."
This is an implausible result.
- b) Baseline data report 104 patients, but results report 106 patients.
This is an implausible result.
- c) The number of drop outs is reported as 13, or 14 or 15 in a contradictory manner without explanation.
This is an implausible result.
- d) The responders in the BH group are reported as 24 (text) vs 23 (figure).
This is an implausible result.
- e) The reported p-value of 0.021 is not possible with the reported Chi square test.
This is an implausible result.

However, the review ignores these implausible results in the "research integrity checklist" and merely states that

"Baseline characteristics are only reported for age and gender. These variables are not excessively similar across groups" (which is another error in itself, see above).

and

"Some loss to follow-up is reported".

VI. Errors in statistical reporting and interpretation

Khan trial (Khan et al., 2011)

35. Acceptance of a p-value inconsistent with the stated statistical method

Page 55

Error: The reported $p = 0.021$ is incompatible with the stated two-sided chi-square test. The Cochrane review ignores this contradiction.

36. Ignoring Implicit use of a one-sided Fisher exact test without reporting

Page 55

Error: Statistical significance relies on an unreported one-sided test. This is ignored by the review.

37. Failure to address sensitivity to minimal changes in contradictory responder numbers

Page 55

Error: The responder number in the Betahistin group is reported in a contradiction between text (24) and figure (23) in the original study (see also other error above). The Cochrane review uncritically adopts the "better" number of 24 responders into their statistical analysis. However, if 23 (responders in the betahistine group) is the true number, then the statistical significance reported by the original study ($p = 0.021$) is now non-existent ($p = 0.0921$; chi square test as specified in the methods section). Furthermore, if the Cochrane review would impute the number of 23 (responders in the betahistine group) into their analysis, their p-value would also change from 0.06 (borderline) to 0.101 (clearly non-significant). The Cochrane review, however, ignores this contradiction and statistic fragility.

VII. Internal consistency / trustworthiness

38. Inconsistent application of trustworthiness screening

Pages 13–14, 31–32, 105

Error: Trustworthiness concerns are handled inconsistently.

When applying the trustworthiness tool, as pre-specified in the review protocol, the Cochrane authors found that all the betahistine studies - except Adrion 2016 - had concerns. Then they "attempted to contact authors to clarify these issues, but we either received no reply, or the authors were unable to access the original trial data to clarify our queries." Their next step was then to reason that "the trustworthiness screening tool may be excessively sensitive" and to include the studies with unresolved concerns into the review nonetheless.

39. Inconsistent and incorrect application of the Risk of Bias tool

Error: The review states that the Risk of Bias 2 (RoB 2) tool was used. However, the terminology and domain structure applied throughout the review correspond to the original Risk of Bias tool (RoB 1), including domains such as "incomplete outcome data", "selective reporting", and the category "unclear risk". These concepts do not exist in RoB 2.

The mixing of RoB 1 terminology with a claimed RoB 2 assessment constitutes a methodological inconsistency and results in systematically more severe bias judgements than would be justified under a correct RoB 2 application.

Declaration of interest:

RG is a co-author of the BEMED-trial (Adrion et al 2016).