Dear Reviewers and *The BMJ* Review Committee,

We thank the editorial committee and the three reviewers for their thorough review of our manuscript and for their helpful recommendations, to which we respond point by point below.

On the following pages, we address each of the comments by *The BMJ* Committee, and Reviewers #1, #2 and #3, in that order.

Reply to Comments from the meeting

(Manuscript committee, comment #1)
The outcomes in the tables look different from the outcomes listed in the methods. For example the primary outcome is given in the text as "the individual attack rate standardized on a 30 day period", but it seems to be given as "decay rate" in table 3. Secondary outcomes start with "the median duration and median severity of evaluated Menière attacks during months 7 to 9 within the 9-month treatment period". That outcome seems not to be given in table 4.

Authors' Response:
Primary efficacy outcomes: We improved the methods section by focusing on the two approaches applied to quantify and compare treatment effects.
There are two types of approaches. First, we follow the general principles of longitudinal cohort studies when estimating and reporting incidence rates. We used a GLMM (Generalized linear mixed effect model) which contains individual effects (random effects) and fixed effects. The fixed effects describe a time dependent incidence rate of the unexposed (in our case: the placebo group) and relative risks (rate ratios, RR) which transform the time-dependent incidence rate into the time-dependent incidence rates for the exposed groups (i.e. both betahistine groups LD and HD). In the BEMED trial, the time-dependent incidence rate is given by a continuous decay from the baseline incidence. The decay rate per month was estimated and reported. The respective RRs for the exposed groups were also estimated and reported.
Second, based on the random and fixed effects estimated from the GLMM, we are able to estimate time-dependent individual (patient-specific) incidence rates (in the literature called conditional estimates) and can take their mean over all subjects of our study which represents a *population-based* mean incidence rate (in the literature also called a *marginal* estimate; for a discussion refer to chapter 13 of Fitzmaurice GM, Laird NM, Ware JH (2004) *Applied longitudinal analysis.* New York: Wiley. We also reported the "marginal" estimates for attack rates per 30 days in order to quantify the effect within a treatment arm by a number which a patient can translate to his/her own experience (number of attacks per month).
Therefore, both quantitative primary efficacy results are reported in our manuscript.

Secondary outcomes: We agree that the results concerning the two diary-based secondary outcomes attack duration and severity were not reported in a detailed manner. For clarity and completeness, the resulting parameter estimates of the cumulative odds model are now given in a new Table 4a. Please also refer to our answer to Manuscript committee's comment #9.

(Manuscript committee, comment #2)

We expect all outcomes to be reported in the order and as defined at the registration website. If there are any discrepancies these need to be explained and justified, in the response to reviewers and editorial comments as well as in the manuscript. This goes for the primary outcome as well as all secondary outcomes.

Authors' Response:

According to the trial protocol, "objective hearing loss" – determined by acoustic evoked potentials (AEP) − was a further secondary outcome. However, during the blind data review we realized that it was not possible to analyze these data due to extremely poor data quality and a huge amount of missing data (the main reason for the latter was that, too often, this examination was simply not performed, or not performed in a standardized manner).

From a clinical point of view, AEPs do not capture Menière's-specific cochlear dysfunctions, and so would not have been a meaningful secondary outcome anyway.

We have included a new short section in the Supplement S3.3.

Since our manuscript focused on the 9-month treatment period, we did not present results concerning the 3-month post-treatment follow-up period (months 9 to 12) since this would have gone far beyond the scope of our manuscript. The results of the post-treatment follow-up period will be published separately.

(Manuscript committee, comment #3)

Can you say a bit more about this drug and its presumed mechanism, and also say whether it or something similar is used in other countries in the world, such as the US?

Authors' Response:

Thank you for this suggestion. A few sentences were included in the introduction section as requested.

Betahistine is a histamine-like drug (partial H1 histaminergic receptor agonist, and H3 receptor antagonist (Wang & Dutia 1995; Dutia 2000; Arrang JM 1985; Bertlich 2015). Its mode of action is not yet fully understood.

1) Animal models have shown that one possible mode of action could be a dose-dependent increase in cochlear blood flow. It seems that the metabolites of betahistine (like aminoethylpyridine) are even more effective. (Bertlich 2014, Bertlich 2015, Ihler 2012)

2) Imbalanced activity in the vestibular nuclei complexes leads to the activation of the central histaminergic system (Horii et al, 1993; Lacour et al, 2001). Therefore another working point could be an improvement of central vestibular compensation (Tighilet et al. 2007, 2005; Barresi et al. 2005; Dutia 2000; Lacour & Tighilet 2000; Lozada et al. 2004), most likely via its H3-mediated action on GABA-release in the commissural system (Bergquist et al. 2006, 2008).

3) Another possible mode of action could be histamine mediated reduction of resting activity of ampullar hair cells (Mira et al 2003).

Betahistine has been used worldwide for decades in the field of vertigo or dizziness, and is the only licensed product for „Menière's-like symptoms". In the US betahistine is not FDA-approved, and the standard therapy for MD is primarily Gentamycin injection followed by surgical approaches.

(Manuscript committee, comment #4)
P-values in table 1 should be removed. There is no need to 'test' for balance - this is a randomised controlled trial.

Authors' Response:
Thanks for pointing this out. In the revised version, we have removed the last column of Table 1.

(Manuscript committee, comment #5)
+/- should be removed throughout with standard deviations given in brackets.

Authors' Response:
In Table 1 and 2, the symbol ± was removed, and the SD is now given in brackets.

(Manuscript committee, comment #6)
Page 17. The main result you report (first) is the mean attack rate over the nine 30-day time intervals for the placebo group. This isn't really an outcome per se, (as no treatment comparisons are being made), but rather a description of what is happening in one arm (placebo) of the trial. We thought the main outcomes are the rate ratios over the 90 day period at the end of the treatment period.

Authors' Response:
We believe this is now clarified in the revised methods section. Furthermore, some changes were made in the Result section concerning the presentation of the model-based primary efficacy results. We report fixed effects of the incidence rates from a NB GLMM (i.e. the rate ratios together with the decay rate for the placebo group), as well as the population-based mean incidence rates for the last 3 months of the treatment period (=pre-specified assessment period).

The rate ratios together with the decay rate refer to the whole 9-month treatment period and result from the longitudinal model which considers all data available.

(Manuscript committee, comment #7)
The sample size calculation is nicely detailed but complex, and it would benefit from a re-writing to make it clear what (and why) you are doing.

Authors' Response:
Thank you for this helpful suggestion. We realize that the sample size section is lengthy and involves multiple aspects – in an attempt to make the methods easier to follow we have modified the corresponding section in the revised version of our manuscript.

(Manuscript committee, comment #8)

There is missing data for many of the secondary outcomes (table 4) and a complete case was carried out. Why not impute? It was indicated in the methods (page 15) that you would impute for secondary outcomes.

Authors' Response:

The editors make an important point. In the revised version, we have replaced the results of the complete case ANCOVA analyses and now present the results using multiple imputation techniques as indicated in the methods section (p. 15). To be more precisely, the multiple imputations by chained equations (MICE) approach was applied (m=21 imputed datasets created, presenting pooled estimates according to Rubin's rule).

The original Table 4 showing the results from the complete case ANCOVA has been removed from the main text, and is now included in the Web Supplement (please refer to the new Supplement section S3.4). In the new Table 4b of the manuscript, we now report differences in mean change for the treatment groups (LD vs. PL, HD vs. PL) together with pooled P values according to Meng & Rubin (1992).

Meng X-L, Rubin DB. Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika* 1992;79(1):103-11.

(Manuscript committee, comment #9)

Attack severity and duration (secondary outcomes) reported on page 20, should for clarity and completeness be reported in Table 4 with the other outcomes. Only p-values are reported in the text for these outcomes without and quantification of severity or duration.

Authors' Response:

Thank you for requesting this additional clarification. Following this recommendation, we now provide more details on the resulting estimates from the proportional odds (cumulative logit) model. We have now included this additional information concerning the ordinal secondary outcomes attack duration and severity in a new Table 4a. The reference to this table is given on manuscript page 20.

Table 4b consists of the results from all other secondary outcomes which were measured during office visits (absolute change from baseline and 9-month visit).

REFEREE COMMENTS

Before we address each comment in turn, we would like to provide some general comments:

At baseline visit, "relevant" medical history (in particular, vestibular or neurological diseases) was documented in the CRF. That is, data for diseases that bother the patient or needed/ had to be treated up to 5 years prior to inclusion were available.

To answer the questions concerning "disease duration" prior to enrolment and "age at onset of the disease", we made an effort to retrospectively obtain this information from the study sites on the basis of the patients' medical records − if feasible −, since these medical history data were not documented on the CRF and therefore not part of the official BEMED study database.
For all study participants, we defined "time of onset" as the first documented vertiginous symptom according to the patients' medical records.

Reply to Reviewer #1

1. ***Please state the duration of MD prior to enrollment for each treatment group. It would be desirable to see a balanced duration of symptoms of MD in each group prior to enrollment, such that one group did not have a proportionately earlier or later stage MD when starting the trial.***

Authors' Response:
After contacting 14 study sites, a date (i.e. at least: a year) for time of onset was documented in the patients' records for about 70% out of 221 patients randomized. We defined time of onset as first occurrence of vertiginous symptoms.
Apparently, the treatment groups were well balanced for disease duration prior to enrolment, with a median of 2 years for all three treatment groups (mean (SD) was 4.1 (4.2) ys for PL, 4.0 (4.6) ys for LD, 5.1 (6.7) ys for HD).
Due to a relatively high proportion of missing values and partly imprecise documentation of the time of diagnosis or first occurrence of vertiginous symptoms (e.g. only year available, date or even year "estimated" or only rough time interval given) leading to poor data quality, we decided not to include these data (which were not documented on the CRF) in Table 1 of our manuscript. Instead, we included the following sentence at the end of the section "Participants' baseline characteristics" of the results section: "*Moreover, groups were well balanced with regard to disease duration and age at onset of vertiginous symptoms (data not shown).*"

According to a study by Pyykkö I et al. (BMJ open 2013) a reliable estimation of the "disease onset" is very challenging due to the very variable onset of MD (i.e. monosymptomatic besides MRI-documented endolymphatic hydrops) and the disease progression. According to Pyykkö I et al, for more than 20% of the patients the time delay in assigning a diagnosis of probable MD was more than 5 years. The rigid diagnostic criteria for MD (according to the 1995 AAO-HNS classification) include

mostly cases with an affection of the vestibular and cochlear function, hence, the later stages. A clear concensus does not exist what should be taken as possible "onset" for MD.
Pyykkö I et al. stated that MD more frequently started with vestibular symptoms than with hearing loss. Therefore, we decided to search for the first documented vertiginous symptoms documented in the patients' records and defined this time point as "disease onset".

2. ***How did the authors screen for retrocochlear disease in their subjects (MRI, ABR)? There are cases of small intracanalicular or even intralabyrinthine schwannomas that mimic MD, and routine clinical care for all patients with asymmetric hearing loss is the consideration of retrocochlear or intracochlearlesions.***

Authors' Response:
We thank you very much for this important comment. It is absolutely correct that intracanalicular or even intralabyrinthine schwannomas could cause Menière's-like symptoms.
As stated in the manuscript, two of the inclusion criteria were diagnosis of "definite MD" and "other causes excluded". So the accuracy of the diagnosis was up to the investigators. Therefore, in most of the cases a MRI should have been performed prior to inclusion of a patient. However, it was not requested to document this information on the CRF, and therefore these data are not available in the BEMED study database.
In order to answer this question, we tried to get this information retrospectively by contacting trial sites: For approx. 30% of all patients randomized an MRI was documented in the patient's records. However, this does not necessarily mean that MRI imaging has not more often been performed in all the other cases.
We did not perform ABR at our site (i.e. the center of the coordinating investigator, N=86 patients randomized) for differentiation as it is less sensitive than an MRI scan. ABR was not an obligatory examination at the screening visit.

3. ***How was "postural vertigo" differentiated from BPPV in your analysis? There is an increased incidence of BPPV in MD and I would assume these positional attacks would be excluded if they are caused by canalithiasis.***

Authors' Response:
We agree that there is a higher incidence of BPPV in MD patients.
Postural vertigo can easily be differentiated from BPPV attacks due to their duration and their dependency on body-/ head-movement or -position. As patient-reported vertigo attacks documented in the paper-based diary were reviewed on a regular basis by trained clinicians, it was possible to differentiate these two entities. Hence, BPPV attacks were not considered as evaluated Menière's attacks used for primary efficacy analyses.

4. How complete were the daily diaries for vertigo reporting? Did the patients report their symptoms regularly each day or wait and fill in days at a time? In my experience with diary studies, it is important that the subjects enter data on a regular basis.

Authors' Response:

Thank you for requesting this clarification.
The patients were instructed to report their MD-induced symptoms every day at home in a paper-based diary which they received at the baseline visit.
This all-dominant data collection scheme seeking to accurately record the subject's daily experience has already been addressed in the submitted version of our manuscript, e.g., on page 8 or page 10 (last paragraph: "...the patient's vertigo status provided by the daily diary recordings"). Furthermore, we have now included the word "daily" in the discussion section, page 24, and "every day" on page 8.
To directly check compliance to fill the diaries over the whole study duration, patients were asked to bring their diaries to each scheduled clinic visit at months 1, 4, 6, and 9 (see page 8). During these clinic visits their raw diary recordings were monitored by the investigator for data completeness and it was checked whether the patient understood how to document their symptoms.
As written in the manuscript (page 8), telephone contact was scheduled at the midpoint between the clinic visits in months 2, 3, 5, 7, 8 in order to remind the patients to document their vertigo symptoms on a daily basis and to avoid incomplete documentation.

To visualize the proportion of missings over time, we included a so-called missingness map in the Supplement (Section S2, Figure S1). The proportion of monotone (and intermittent) missings was not higher than expected for trials assessing the ability of an intervention to provide symptomatic relief from the condition (e.g. trials for chronic pain trials or migraine prophylaxis trials).

5. How were missing data points handled? For instance, if a subject skipped five days in a month in their diary, were these days counted as "failures=vertigo" or just not analyzed. Missing data points is always a struggle in large trials like this one and the proper management of missing data is critical. Please expand in the manuscript how these missing data were handled.

Authors' Response:
We agree that this is an important point.
Throughout the trial including the planning phase, we were aware of this ever-present problem of missing data which is crucial for long-term symptomatic trials. In particular, this methodological issue was accounted for in the development of the statistical analysis plan for the BEMED trial:

As mentioned in the Methods section, the primary analysis was a mixed effects modeling approach that assumes that missingness is at random (MAR) for both permanent (i.e. study dropouts) and intermittent missing data patterns (e.g. missing diary pages, undocumented intervals). That is, the mixed model assumes that, given the statistical model (i.e. conditional upon the covariates in the analysis) and given the observed values of the primary outcome 'number of evaluated attacks per 30 days', the probability of missingness does not depend on the unobserved outcomes of the dependent variable.

In this longitudinal model, we accounted for the number of evaluated days (i.e. the number of days with non-missing information concerning attack status) by including an offset term. This offset term

can be interpreted as some measure of *exposure* ("observation window") within a certain time interval $t = 1,2,3, \ldots ,9$, and the observation window is allowed to vary for each time unit of a patient i. For example, if patient's diary recordings were available for only 25 days out of the maximum number of 30 days for a certain 30-day time interval, the offset term is shortened and set to 25 days, using all available information for the likelihood-based analysis.

Following the widely accepted concept proposed by, e.g. White IR et al. (2012) or White et al. (2011), or the recommendations issued in the NRC report (2010), the new guidance on handling missing data in clinical trials that was commissioned by the FDA, or the EMA (2010) guideline on missing data, the main analysis is valid under the MAR assumption about the missing data and uses all observed data.

The main model under MAR is based on the assumption that no post-randomization variable will be predictive of the partially observed outcome. No multiple imputation techniques were performed for the primary efficacy analysis which is based on an "*all observed data approach*", and therefore is optimally statistically efficient (see, e.g., to White IR et al. (2012)).

Furthermore, several sensitivity analyses were performed to explore the impact of departures from the assumptions made in the main analysis: In our manuscript, we presented the results of a sensitivity analysis which only used patients with a total number of evaluated days larger than 0 across the assessment period, i.e. month 7, 8, and 9 (for further details of this methods are included in the Supplement S2.1.1).

References:

- National Research Council. *The Prevention and Treatment of Missing Data in Clinical Trials*. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. The National Academies Press, Washington, DC, 2010.
- White IR, Horton NJ, Carpenter J, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. *BMJ* 2011;342:d40.
- White IR, Carpenter J, Horton NJ. Including all individuals is not enough: lessons for intention-to-treat analysis. *Clin Trials* 2012;9(4):396-407.
- CHMP. Guideline on Missing Data in Confirmatory Clinical Trials. European Medicines Agency, Committee for Medicinal Products for Human Use (CHMP), London, UK, 2010. EMA/CPMP/EWP/1776/99 Rev. 1. URL http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf.

Reply to Reviewer #2

The study design, participants, methods, results, interpretations and references are beyond criticism. I strongly recommend the article to The BMJ.

Authors' Response: Thank you very much for your extremely positive appraisal of our BEMED study and our manuscript.

Reply to Reviewer #3

1. ***Patients with MD have a high prevalence of migraine. Vestibular migraine may overlap with MD is some cases. How did the author consider migraine and vestibular symptoms? Was migraine one of the exclusion criteria? If so, it should be stated. If this was not the case, this information should be added on Table 1. The most common adverse effect was headache and it would be interesting to know if patients reporting betahistine-associated headache had in fact a previous history of migraine.***

Authors' Response:
We agree with the reviewer that symptoms of vestibular migraine (VM) and MD could be similar and that there is an overlap of both diseases.
Patients with diagnosed VM were not included in the BEMED trial. As stated on page 7, VM was one of the exclusion criteria.
In the BEMED trial, history of migraine headache was not an exclusion criterion. However, we were aware of the problem that a "simple" migraine headache may convert into a migraine with vertigo symptoms. Therefore - if patients suffered from migraine - we attempted to figure out at the screening visit if there were any migraine (related) symptoms during or correlated with vertigo attacks. If there was a reasonable suspicion of expansion into a vestibular migraine, patients were not included.

The number of BEMED patients with a history of migraine or headache (according to the MedDRA coding system version 17.0) up to five years prior to enrolment is now added in Table 1.
In total, 39 patients had a history of migraine headaches or headaches up to 5 years prior to enrolment (22 headache, 1 cluster headache, 11 migraine, 1 migraine with aura, 2 migraine without aura, 2 tension headache).
9 patients with a previous history of migraine reported 11 betahistine-associated headaches AEs (PL: 4, LD: 0, HD: 5 patients (2 of these had 2 AEs)). A "betahistine-associated" headache AE was defined as an AE with a drug-event relationship reported as 'possible' or 'probable' according to the safety database.

2. ***Patients with MD also have a higher comorbidity of autoimmune diseases (Gazquez 2011; Tyrell 2014). On Page 7: in the exclusion criteria, complex diseases that might confound treatment assessment were included. Autoimmune background is probably involved in a subset of patients with MD. Did the authors exclude patients with systemic autoimmune disorders such as rheumatoid arthritis or SLE? This should be also included in the exclusion criteria.***

Authors' Response:

Thank you for requesting this clarification.
Patients with "other serious illness, or a complex disease that might confound treatment assessment" were not eligible to be enrolled. This means that patients that - in all likelihood – would not be able to take study medication regularly, attend study visits or who risk suffering from other symptoms that cannot be clearly distinguished from possible side effects, were not included. Cancer was an exclusion criterion.
The pathophysiology of MD is still unknown; genetics, autoimmunological, infectious and environmental causes and activators may contribute to the onset and course of the disease.
Patients with systemic autoimmunological disorders were included.

Possible systemic autoimmune disorders according to the BEMED study database (MedDRA coded medical history) were as follows:
1 patient with rheumatoid arthritis, 1 patient with systemic lupus erythematosus, and 1 patient with ankylosing spondylitis.
Other possibly autoimmunological concomitant diseases according to the BEMED study database were as follows:
In total, 47 patients suffered from disorders related to the thyroid gland (1 patient with "autoimmune thyroiditis", 1 "Basedow's disease", 1 "hyperthyroidism", 43 patients with "hypothyroidism", and 1 patient with "thyroid disorder"). Unfortunately we cannot state if this was caused by an autoimmune process.
Other possible autoimmunological (or triggered) diseases should be named briefly:
psoriasis (N=4), neurodermatitis (N=5), immune thrombocytopenic purpura (N=1), autoimmune hepatitis (N=1), colitis ulcerative (N=1), alopecia areata (N=1), vitiligo (N=1) and spondylodiscitis (N=1).
In total, 12 patients suffered from two of the above-mentioned diseases, one from three different diseases.

3. Table 1 does not include the age of onset of the disease on each group. This could be a potential bias in case of differences among groups.

Authors' Response:
As stated above (general comments; Reply to Reviewer #1, question 1), these data were not recorded on the BEMED CRF. Therefore, for this revision, we made an effort to request this additional information by contacting all 14 trial sites.

Defining "time of onset" as the date of the first vertiginous symptoms documented on the patients' records, the treatment groups apparently were well balanced for age at disease onset:
mean (SD) age at disease onset was 51.9 (12.5) ys for the PL, 52.6 (12.1) ys for the LD, and 52.5 (15.3) ys for the HD group.

Due to a relatively high proportion of missing values and partly imprecise documentation of the time of first occurrence of vertiginous symptoms (e.g. only year available, date or even year "estimated" or only rough time interval given) leading to insufficient data quality, we decided not to include these data (which were not documented on the CRF) in Table 1 of our manuscript. To make a compromise, we included the following sentence at the end of the section "Participants' baseline characteristics"

(results section): "*Moreover, groups were well balanced with regard to disease duration and age at onset of vertiginous symptoms (data not shown).*"

For limitations concerning the a reliable 'estimation' of the disease onset please refer to our reply to reviewer #1, comment #1.

□

In the revised version of our manuscript there are a small number of additional minor edits, all indicated in track changes.

In Figure 3 the title was improved, and this figure now displays the attack rates per month (not per day) in order to be consistent with the main text.

The Supplement was extended in order to further clarify some methods as requested by the committee meeting or the reviewers. Among other things, an additional Figure S2 was included which visualizes the new Table 3b of the manuscript. We have also uploaded a copy of the Supplement with changes highlighted.

Once again, thank you for the opportunity to submit our manuscript to *The BMJ* and for offering provisional acceptance.

Please let us know if there are any issues that require further clarification.

Yours sincerely,

Christine Adrion, MPH

Biostatistician

Institute for Medical Informatics, Biometry und Epidemiology (IBE)
University of Munich
Marchioninistr. 15
81377 Munich, GERMANY

Tel: +49 89 4400 77486
Fax: +49 89 4400 77491
email: adrion@lmu.de