Original Paper

European Neurology

Eur Neurol 2004;51:89–97
DOI: 10.1159/000076535

Received: September 15, 2003
Accepted: December 5, 2003
Published online: January 28, 2004

The First Placebo-Controlled Trial of a Special Butterbur Root Extract for the Prevention of Migraine: Reanalysis of Efficacy Criteria

H.C. Diener[a] V.W. Rahlfs[b] U. Danesch[c]

[a]Department of Neurology, University of Essen, Essen, [b]idv Data Analysis and Study Planning, Gauting, and [c]Weber & Weber GmbH, Inning, Germany

Key Words
Migraine · Migraine prophylaxis · Butterbur

Abstract
This is an independent reanalysis of a randomised, placebo-controlled parallel-group study on the efficacy and tolerability of a special butterbur root extract (Petadolex®) for the prophylaxis of migraine. The original protocol and analysis had a number of major shortcomings. In order to follow regulatory requirements, an independent reanalysis of the original data was performed. Following a 4-week baseline phase, 33 patients were randomised to treatment with two capsules 25 mg butterbur twice a day and 27 to placebo. The mean attack frequency per month decreased from 3.4 at baseline to 1.8 after 3 months ($p = 0.0024$) in the verum group and from 2.9 to 2.6 in the placebo group (n.s.). The responder rate (improvement of migraine frequency $\geq 50\%$) was 45% in the verum group and 15% in the placebo group. Butterbur was well tolerated. This small trial indicates that butterbur may be effective in the prophylaxis of migraine.

Introduction

Between 1993 and 1995, an investigator-driven, randomised, placebo-controlled trial of a special butterbur root extract for the prophylaxis of migraine was performed and published [1]. The protocol, primary study report and the publication of this study had a number of major shortcomings. For example, the final study report and the publication did not contain appropriate information on baseline characteristics and medication for the treatment of acute migraine attacks during the study. Although statistical analyses were carried out, the primary sample for evaluation was the per-protocol population only, and not the intention-to-treat (ITT) population. Tests for significance were one-sided with $\alpha = 0.05$. Efficacy parameters were not defined precisely and a priori in the protocol. The t test used was not appropriate for the type of data set. Apparent baseline differences were not considered as a covariate in the primary analysis of efficacy. No comments or adjustments were made with regard to multiple testing.

Therefore, and to assure the confirmatory character of the study for regulatory purposes, an independent investigator and an independent third-party statistical institute were commissioned to prepare a complete re-evaluation of the efficacy data under the requirements of the current

KARGER

Fax + 41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0014–3022/04/0512–0089$21.00/0

Accessible online at:
www.karger.com/ene

Hans-Christoph Diener, MD
Department of Neurology, University of Essen
Hufelandstrasse 55
DE–45122 Essen (Germany)
Tel. +49 201 723 2460, Fax +49 201 723 5901, E-Mail h.diener@uni-essen.de

International Conference on Harmonisation Guideline E9 and the principles of state-of-the-art statistics. The methods applied for the reanalysis of the efficacy criteria avoid a bias by post hoc selections.

Methods

Study Design

This double-blind, randomised, placebo-controlled, parallel-group study comparing the clinical effect of a special butterbur root extract (Petadolex®, Weber & Weber) and placebo was conducted between January 1993 and February 1995 at a neurological department of a German hospital. Following a 4-week run-in period without trial medication, patients were treated with 2 × 50 mg of a special butterbur root extract or placebo for 12 weeks. Eligible patients were randomly allocated to receive study medication according to a computer-generated randomisation sequence. The randomisation numbers were issued on a sequential basis, and the code was withheld from the investigators until completion of the study. The double-blind study medication was identified by a patient number according to the randomisation code.

Patients

Male and female out-patients, aged 18–60 years, with and without aura as diagnosed according to the criteria of the International Headache Society were enrolled [2]. Inclusion and exclusion criteria were also according to the criteria of the International Headache Society [3]. In short, a history of migraine of at least 1 year as well as a minimum of 3 migraine attacks per month within the last 3 months and a minimum of 2 attacks in the 4-week run-in phase without study medication were required for recruitment into the active treatment phase. Exclusion criteria were among others: treatment with agents known to have an effect on migraine within 4 weeks prior to the start of the run-in phase and regular consumption of analgesics for more than 12 days per month. Patients were seen at 4-week intervals. At each visit, study medication and a diary for the next 4-week treatment period was issued. Number, intensity (as assessed using a visual analogue scale from 1 to 10), duration of migraine attacks, accompanying symptoms as well as the use of acute migraine medication were recorded in the diary. A total of 60 patients (28 males, 32 females) were enrolled; 33 patients were allocated to the active group and 27 patients to the placebo group. Compliance was checked by pill count.

Ethics

Patients provided informed consent prior to the study entry. This study was conducted in accordance to the ethical standards of good clinical practice and the regulations of the Declaration of Helsinki.

Statistical Analysis

Data entry was completely repeated based on the original case report forms under good clinical practice conditions (double data entry, plausibility and consistency checks). In the original analysis, neither the primary efficacy criteria nor the time point were defined. Since an a posteriori choice of efficacy criteria or single time points for primary analysis are not appropriate, all 4 primary efficacy criteria of the patient's diary (number of migraine attacks per month, number of days with migraine per month, mean duration of migraine attacks per month, mean intensity of migraine attacks per month) and all 3 follow-up visits were evaluated equally weighted and analysed by means of a multivariate technique, controlling the multiple level of α. This procedure is not data driven and avoids a bias by post hoc selections.

Since many of the outcome measures were either not continuous, or might have outliers, or else might have skewed distributions, we chose to use the Wilcoxon-Mann-Whitney rank test. All primary efficacy criteria were tested as an ensemble for superiority of the test drug using the non-parametric directional test of the Wei-Lachin procedure. This method is a multivariate generalization of the Wilcoxon-Mann-Whitney test, which takes account of the correlation among univariate Mann-Whitney tests for each outcome to produce an overall average estimate of benefit and test for treatment differences [4–6]. If the combined global test gives a significant result, then efficacy is proven in a confirmatory way (at least one null hypotheses of a null difference cannot be true). In addition to the p values of the Wilcoxon-Mann-Whitney test, the related effect size measures with their confidence intervals were calculated, a requirement of the International Conference on Harmonisation Guideline E9.

The Mann-Whitney estimator was used as a measure of relevance of group differences, being the associated effect size measure for the Wilcoxon-Mann-Whitney test. It is defined as the probability that a randomly selected patient from the test group is better off than a randomly selected patient from the reference group. A robust estimator of this probability is implemented by comparing all pairs of patients in which 1 patient is given the test treatment and 1 the comparator. Score 1 is assigned if the actively treated patient has a better response; rank 0.5, if the 2 patients are tied, and 0, if the comparator patient has a better response. Then the sum of these scores is divided by the total number of pairs compared. A Mann-Whitney estimator of 0.5 means that patients do equally well under both treatment conditions. Mann-Whitney estimators greater than 0.5 are evidence of some benefit from the test treatment. The closer the Mann-Whitney estimator to 1, the more reliable the evidence. Well-known benchmarks for small-, medium-, and large-sized superiority/inferiority are 0.56/0.44, 0.64/0.36, and 0.71/0.29, respectively [7].

All calculations (p values, Mann-Whitney estimators and their confidence intervals) were performed within the framework of the Wei-Lachin procedure using the validated computer program SmarTest from idv-Data Analysis and Study Planning in Gauting/Munich, Germany.

For baseline adjustment, the percent changes from baseline were processed as first-line analysis since the efficacy criteria showed a proportional decrease. In addition, the analysis of absolute values as well as the analysis of changes from baseline were performed in the same way (multivariate directional test) as sensitivity analyses. These analyses should also be statistically significant in order to keep the quasi-confirmatory character of the reanalyses and to avoid bias by post hoc selection of the method of baseline adjustment. Further analyses included contingency tables (shift tables), scattergrams and cumulative distribution functions.

Baseline homogeneity was analysed with univariate Mann-Whitney estimators, two-sided 90% confidence intervals for equivalence and a Wei-Lachin global unidirectional test for difference with two-sided $\alpha = 0.05$. Confirmatory analyses of efficacy were performed as one-sided directional tests for superiority with $\alpha = 0.025$ and will be presented as Mann-Whitney estimators and their 97.5% confidence intervals; as a benchmark for relevance, an estimator of 0.36 (medium-sized inferiority) has been applied. For 'proven' superiority,

Table 1. Baseline characteristics and test statistics of baseline criteria verum vs. placebo: univariate and combined results (exploratory analysis, ITT data)

a Demographics and medical history (prior to run-in period)

Criterion	Verum	Placebo	MW	LB-CI	UB-CI	p value
Age (mean ± SD, years)	29 ± 9.26	29.1 ± 9.46	0.4988	0.3773	0.6204	0.9874
Gender			0.5202	04139	0.6265	0.7546
Female, %	51	55				
Male, %	49	45				
Age at first attack (mean ± SD, years)	17.6 ± 4.82	19.7 ± 5.15	0.3562	0.2344	0.4781	0.0523
Attacks per month (mean ± SD)	3.4 ± 1.06	3.1 ± 0.85	0.5673	0.4505	0.6842	0.3433
Previous therapy (mean ± SD, months)	13.8 ± 17.23	13.1 ± 18.51	0.5481	0.3689	0.7273	0.6590
b Primary efficacy variables (4-week run-in period)						
Attacks per month (mean ± SD)	3.4 ± 1.48	2.9 ± 1.15	0.4091	0.2895	0.5287	0.2112
Days with attacks per month (mean ± SD)	3.6 ± 1.93	3.0 ± 1.27	0.4186	0.2986	0.5387	0.2649
Duration of attacks per month (mean ± SD, h)	9.4 ± 3.32	9.3 ± 3.94	0.4461	0.3255	0.5667	0.4625
Intensity of attacks per month (mean ± SD)	3.9 ± 0.91	3.6 ± 0.73	0.4287	0.3081	0.5494	0.3313
Combined criteria	–	–	0.4256	0.3521	0.4992	0.0965
c Acute migraine attack treatment (4-week run-in period)						
Attacks with acute medication (mean ± SD, %)	20.6 ± 31.51	12.8 ± 25.41	0.4056	0.3026	0.5087	0.1321

MW = Mann-Whitney estimator; LB-CI = lower bound of the two-sided 90% confidence interval for equivalence of the Mann-Whitney estimator; UB-CI = upper bound of the two-sided 90% confidence interval for equivalence of the Mann-Whitney estimator; p value = directional test for difference using the summarizing Wei-Lachin procedure.

the lower bound (worst case) of the 97.5% confidence interval of the Mann-Whitney estimator has to be above the benchmark for equality (0.5).

All analyses were performed with the ITT data set (all randomised patients who took the study medication at least once). Missing values were replaced by the 'last value carried forward' method.

Safety Evaluation

Each patient underwent a complete physical and neurological examination including blood pressure measurement and serum chemistry [total blood cell count, alanine aminotransferase (ALT), aspartate aminotransferase (AST) and bilirubin] before and at the completion of the study. Medical history was recorded at the beginning of the study. Adverse events, if any, were reported every 4 weeks at each visit.

Results

Baseline Characteristics

Baseline characteristics were separated into demographic data and prior history, efficacy criteria and criteria for acute migraine attack medication (table 1). Homogeneity tests with univariate and combined Mann-Whitney estimators showed no group differences ($0.36 <$ Mann-Whitney estimator <0.64), denoting good comparability of the two patient groups (table 1).

For the criterion 'age at first attack', the Mann-Whitney estimator (0.3562) indicated a 'relevant' group difference (0.36 being the benchmark for medium-sized group differences) with lower age in the verum group.

The potential influence of the group difference with regard to the criterion age at first migraine attack upon the efficacy results was analysed by means of a covariance analysis between age at first migraine attack and efficacy outcome 'number of attacks' as well as by means of stratified analyses using dichotomized age of first migraine attack as strata. Both analyses showed that the age at the first migraine attack had only a negligible influence on the efficacy results (data not shown).

Primary Efficacy Criteria

The number of migraine attacks per month was markedly reduced in the verum group compared with the placebo group. The mean group difference in the percent

Table 2. Primary criteria of efficacy per month

a Primary criteria of efficacy per month as percent change from baseline (exploratory analysis, ITT data set); mean ± SD after 4, 8 and 12 weeks of treatment, respectively

	4 weeks (visit 2)	8 weeks (visit 3)	12 weeks (visit 4)
Number of attacks			
Verum	−38.5 ± 31.88	−51.7 ± 31.27	−33.7 ± 42.43
Placebo	−8.4 ± 57.09	−8.4 ± 49.69	−8.6 ± 28.56
Number of days with attacks			
Verum	−39.6 ± 31.19	−53.0 ± 29.96	−35.3 ± 42.33
Placebo	−5.9 ± 73.27	−9.0 ± 51.17	−11.0 ± 26.08
Duration of attacks			
Verum	−10.8 ± 28.81	−26.2 ± 40.92	−16.7 ± 38.69
Placebo	+1.2 ± 36.15	−8.6 ± 25.16	−1.6 ± 28.81
Intensity of attacks			
Verum	−23.2 ± 22.39	−35.7 ± 38.17	−19.1 ± 46.87
Placebo	−9.4 ± 35.20	−9.0 ± 28.17	−6.3 ± 24.33

b Primary criteria of efficacy per month (absolute values; exploratory analysis, ITT data set); mean ± SD at baseline and after 4, 8 and 12 weeks of treatment

	Baseline (visit 1)	4 weeks (visit 2)	8 weeks (visit 3)	12 weeks (visit 4)
Number of attacks				
Verum	3.4 ± 1.48	1.9 ± 0.82	1.4 ± 0.97	1.8 ± 0.95
Placebo	2.9 ± 1.15	2.2 ± 0.70	2.4 ± 1.05	2.6 ± 1.09
Number of days with attacks				
Verum	3.6 ± 1.93	2.0 ± 1.15	1.4 ± 1.00	1.8 ± 0.95
Placebo	3.0 ± 1.27	2.3 ± 0.76	2.4 ± 1.09	2.6 ± 1.15
Duration of attacks, h				
Verum	9.4 ± 3.32	8.0 ± 2.59	6.3 ± 3.46	7.2 ± 3.14
Placebo	9.3 ± 3.94	8.7 ± 2.50	7.9 ± 2.41	8.4 ± 2.02
Intensity of attacks				
Verum	3.9 ± 0.91	2.9 ± 1.01	2.5 ± 1.75	3.1 ± 1.73
Placebo	3.6 ± 0.73	3.1 ± 0.76	3.3 ± 1.17	3.4 ± 1.08

changes from baseline at the end of the therapy was 25.1% in favour of the verum group (table 2a). The analysis of the absolute values (table 2b) and the analysis of the changes from baseline confirmed those findings. The mean group difference calculated as changes from baseline at the end of the therapy was 1.3 attacks. Figure 1 shows the time course of the percent changes from baseline of the number of attacks per month with classical statistics (mean and standard deviation). The shaded area reflects the range of deterioration (values ≥ 0%). Cumulative distribution function analysis was done to determine the proportion of patients with an improvement regarding the primary efficacy variable (as percent change from baseline) of at least 50 or 20%, respectively. Figure 2 shows that at the end of the therapy, there were marked differences between the two groups. A percent change of at least −50.0% (responder) was reached by 45% of the patients in the verum group and by 15% of the patients in the placebo group (rate difference with regard to marked therapeutic response: 30%).

The mean group difference in the percent changes from baseline at the end of the therapy with the variable 'days with attack per month' was 24.3% in favour of the verum group (table 2); the mean group difference in the changes from baseline at the end of the therapy was 1.4 days with attack. Responders (percent change of at least −50.0%) were 48% of the patients in the verum group and 15% of the patients in the placebo group.

The duration of migraine attacks and the intensity of the attacks were also reduced by the treatment with but-

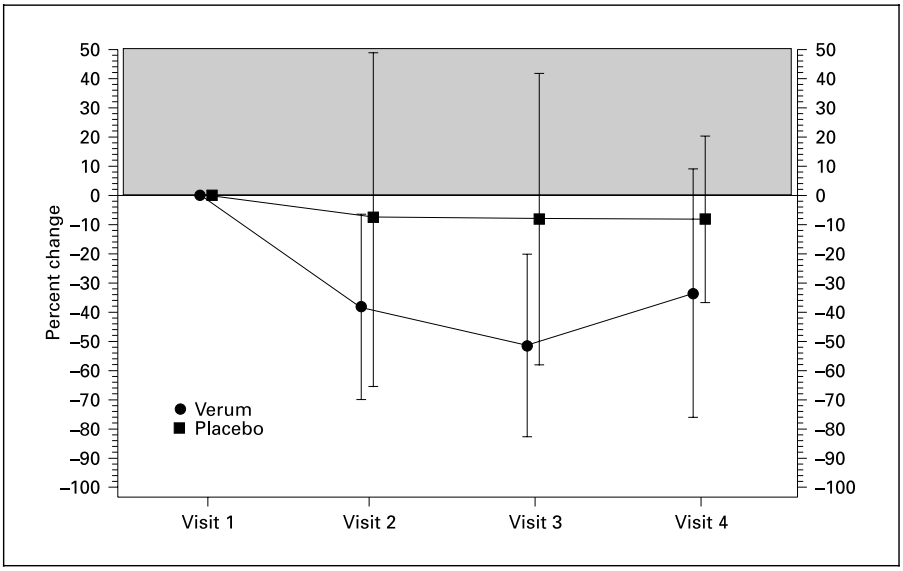

Fig. 1. Number of attacks per month (exploratory analysis, ITT data set); mean ± SD after 4, 8 and 12 weeks of treatment, respectively. The shaded area reflects the range of deterioration (values ≥0%).

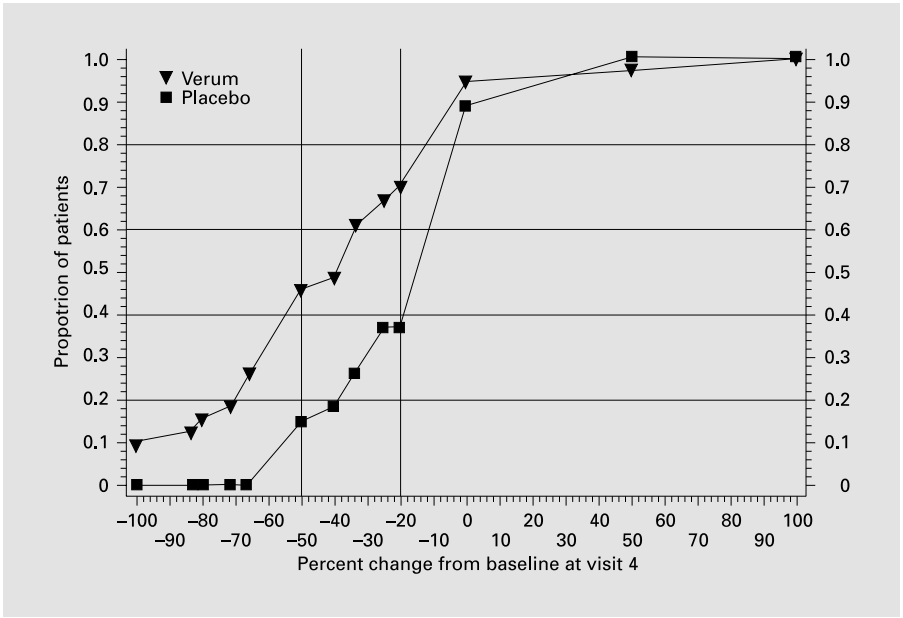

Fig. 2. Cumulative distribution function of the number of attacks per month after 3 months of treatment between the verum and placebo group (ITT data set).

terbur compared with placebo (table 2). The mean group difference in the percent changes from baseline after 3 months was 15.1% and 12.8%, respectively, in favour of the verum group.

Medication for Acute Migraine Attacks

While in the verum group the number of patients with medication for acute migraine attacks was reduced by more than half between the baseline month and the final month (from 44 to 18%), the number of patients changed only marginally in the placebo group (from 27 to 26%). Figure 3 shows the percentages of patients with medication for acute migraine attacks at the different visit intervals. After 2 months of treatment (visit 3), the number of patients with acute medication is identical in both groups; at the end of the trial, fewer patients took medication to treat acute migraine attacks in the verum group compared with the placebo group.

The mean percentage of migraine attacks with acute medication in the verum group was 20.6, 16.7, 11.1 and 7.1% at baseline and after 4, 8 and 12 weeks of treatment, respectively. The figures for the placebo group were 12.8,

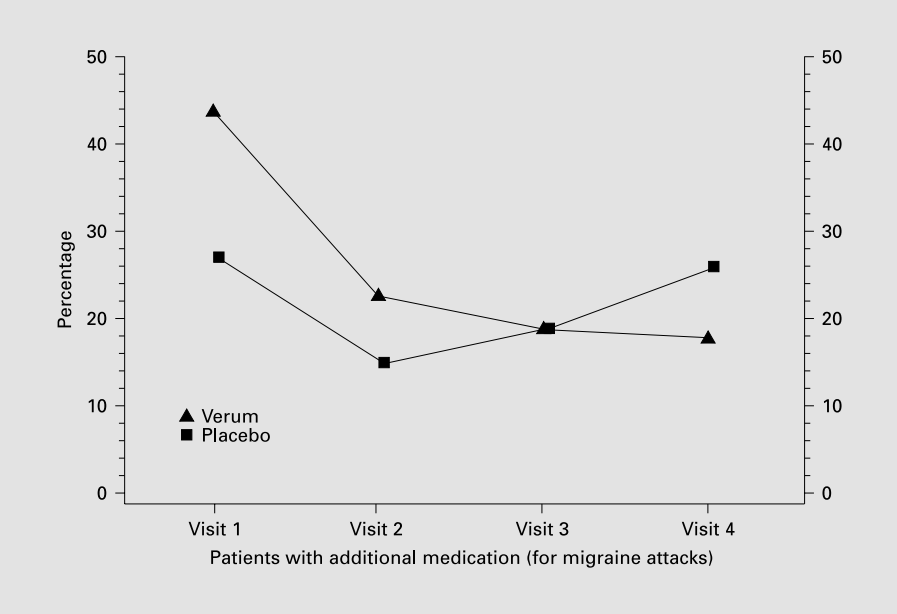

Fig. 3. Patients with additional medication at the different visit intervals (ITT data set).

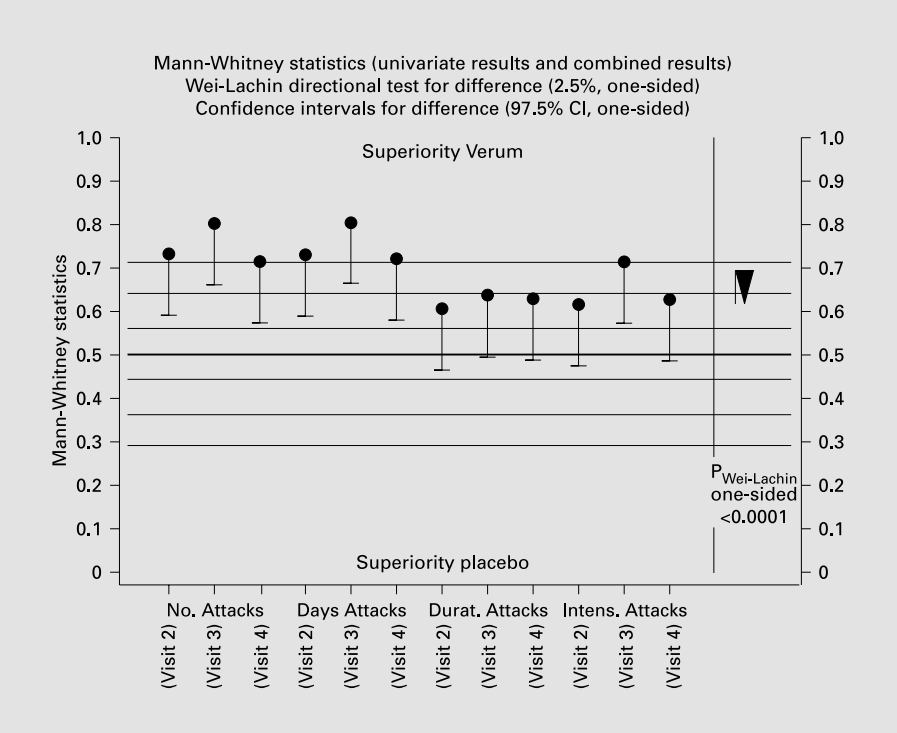

Fig. 4. Test statistics of primary efficacy criteria at visit 2, visit 3 and visit 4 (percent changes from baseline, verum vs. placebo; confirmatory analysis, ITT data set). 0.29/0.71 = Large difference; 0.36/0.64 = medium-sized difference; 0.44/0.56 = small difference; 0.5 = equality. Durat. = Duration; Intens. = intensity.

7.4, 6.2 and 11.7%, respectively. In the verum group, the mean percentage of migraine attacks with acute medication was reduced by more than half between the baseline month and the final month (from 20.6 to 7.1%); however, the mean percentage changed only marginally in the placebo group (from 12.8 to 11.7%).

Confirmatory Multivariate Efficacy Analysis

In order to avoid post hoc bias, the confirmatory analysis is based on the summarizing evaluation of the 4 primary efficacy criteria of the patients' diary, and all 3 follow-up visits were evaluated and equally weighted. Figure 4 and table 3 show the Mann-Whitney estimators and con-

Table 3. Test statistics of primary efficacy criteria at visit 2, visit 3 and visit 4, verum vs. placebo (ITT data set)

Criterion	Percent changes from baseline			Absolute values			Changes from baseline		
	MW	97.5% LB-CI	p_{exact}	MW	97.5% LB-CI	p_{exact}	MW	97.5% LB-CI	p_{exact}
Number of attacks per month									
Visit 2	0.7312	0.5923	0.0006	0.6257	0.4974	0.0274	0.7155	0.5834	0.0007
Visit 3	0.7980	0.6597	<0.0001	0.7514	0.6201	<0.0001	0.7856	0.6543	<0.0001
Visit 4	0.7104	0.5706	0.0016	0.6902	0.5578	0.0024	0.7205	0.5864	0.0006
Number of days with attacks per month									
Visit 2	0.7262	0.5862	0.0008	0.6229	0.4943	0.0306	0.7071	0.5728	0.0013
Visit 3	0.8002	0.6635	<0.0001	0.7452	0.6132	0.0001	0.7710	0.6362	<0.0001
Visit 4	0.7172	0.5784	0.0011	0.6908	0.5584	0.0024	0.7194	0.5833	0.0008
Mean duration of attacks per month (hours)									
Visit 2	0.6044	0.4608	0.0772	0.5881	0.4442	0.1150	0.6027	0.4588	0.0809
Visit 3	0.6330	0.4918	0.0325	0.6341	0.4937	0.0306	0.6195	0.4764	0.0509
Visit 4	0.6274	0.4853	0.0394	0.6150	0.4736	0.0555	0.6190	0.4760	0.0514
Mean intensity of attacks per month (Visual Analogue Scale)									
Visit 2	0.6128	0.4697	0.0611	0.6044	0.4637	0.0729	0.6072	0.4644	0.0706
Visit 3	0.7121	0.5694	0.0018	0.6768	0.5357	0.0070	0.7177	0.5749	0.0014
Visit 4	0.6257	0.4840	0.0411	0.5657	0.4231	0.1834	0.6291	0.4888	0.0357
Combined result									
(Wei-Lachin)	0.6915	0.6159	<0.0001	0.6509	0.5835	<0.0001	0.6845	0.6080	<0.0001

Benchmarks: 0.29: large inferiority, 0.36: medium-sized (relevant) inferiority, 0.44: small inferiority, 0.50: equality, 0.56: small superiority, 0.64: medium-sized (relevant) superiority, 0.71: large superiority. MW = Mann-Whitney estimator; LB-CI = lower bound of the one-sided 97.5% confidence interval; p value = directional test for difference using the summarizing Wei-Lachin procedure.

fidence intervals for the comparison of the verum group versus the placebo group. The confirmatory test – with control of the multiple level of α – is statistically significant ($p_{Wei\text{-}Lachin}$ one-sided <0.0001). The p values are below the level of significance ($\alpha = 0.025$ one-sided corresponding to a level of $\alpha = 0.05$ two-sided). Thus, the efficacy of verum is confirmatorily proven. The measure of relevance (Mann-Whitney) of the combined result indicates more than medium-sized (relevant) superiority of verum (benchmark 0.64). The lower bound of the one-sided 97.5% confidence interval (worst case) is clearly above the benchmark for small superiority (Mann-Whitney estimator = 0.6159 > 0.56). Not only superiority but also more than small superiority is proven for butterbur in patients suffering from migraine attacks.

The result of the sensitivity analyses (table 3) with absolute values and with changes from baseline, respectively, is also statistically significant ($p_{Wei\text{-}Lachin}$ one-sided <0.0001). The measure of relevance (Mann-Whitney) of the combined result indicates more than medium-sized (relevant) superiority of verum (benchmark 0.64). The lower bound of the one-sided 97.5% confidence interval (worst case) is above the benchmark for small superiority (absolute values: Mann-Whitney estimator = 0.5835 > 0.56; changes from baseline: Mann-Whitney estimator = 0.6080 > 0.56).

As may be seen in figure 4 as well, all univariate effect sizes are located above the benchmark of equality (>0.5), thus demonstrating superiority of butterbur for all efficacy criteria and all points in time. All 12 univariate effect sizes are above the benchmark of small superiority (benchmark 0.56). Furthermore, 7 of the 12 effect sizes show more than medium-sized (benchmark 0.64, relevant) or large superiority (benchmark 0.71). For 7 of the 12 single criteria, the lower bound of the one-sided 97.5% confidence interval (worst case) is above the benchmark of equality, thus proving superiority of butterbur.

Safety and Tolerability

Two patients dropped out of the verum group; one because of a suspected pregnancy, the other was unwilling to complete the study without giving any reasons. No significant changes from baseline were reported for vital signs and physical examination results. Three patients treated with petasites displayed increases slightly above normal ranges for ALT and AST. Beside ALT and AST,

mean changes from baseline were also statistically significant for bilirubin and erythrocyte count in the petasites group. None of these changes were rated to be of clinical relevance.

Discussion

Petasites hybridus (butterbur) is a perennial shrub native to Europe, northern Africa, and southwestern Asia. It has been used traditionally for medicinal purposes for many years. Clinical applications include migraine and asthma, among other conditions [1, 8]. From a regulatory perspective, the special butterbur root extract is considered a food product in the United States and the UK. In Germany, the petasites extract is a pharmacological agent under full regulatory supervision by the German Health Authority.

The leaves, rhizome and roots of butterbur contain a mixture of sesquiterpenes, either eremophilane lactones or petasins like petasin and isopetasin [9]. The scientific focus has been on the petasins, which have been shown in various in vitro systems to have spasmolytic actions on smooth muscle cells [10, 11]. Both petasins have also been reported to exert an anti-inflammatory effect by inhibiting leukotriene synthesis [12–15]. Another possible site of action involves an effect on calcium channels as demonstrated in vascular smooth muscle and trachea [16, 17].

The special butterbur root extract used in this study (Petadolex) is a standardized special lipophilic extract (petasitidis extractum e radice, spissum 28–44:1, 25 mg) of the rhizome of *P. hybridus* with a minimum content of 15% petasins. The extract is obtained by high-pressure liquid carbon dioxide extraction in a standardized and patented procedure that also removes pyrrolizidine alkaloids, which are present in the plant and are potentially hepatotoxic and carcinogenic.

A randomised double-blind placebo-controlled clinical study of 60 patients treated with the special butterbur root extract had been published earlier [1]. The results suggested that the extract was an effective preventive treatment for migraine. However, the statistical analysis was flawed in several aspects. The reanalysis of this trial, performed by a third-party biometrical institute, showed that the active treatment group demonstrated superiority vs. the placebo group for the ensemble of 12 primary variables of efficacy (4 criteria at 3 points in time) for the percent changes from baseline as well as for the absolute values and for the changes from baseline. Statistical significance was demonstrated with control of the multiple level of α and, in addition, a clinically relevant treatment effect with effect sizes above the benchmark of medium-sized (relevant) superiority. Even with regard to the very conservative Bonferroni approach, 7 single variables were still statistically significant. All single variables showed more than small superiority in favour of the treatment with butterbur (petasites extract). While the number of patients in need of additional medication during migraine attacks changed only marginally in the placebo group between baseline and the final visit, this number was reduced by more than half in the verum group.

No adverse events were reported in the trial report, which seems to be unusual. This might be explained by the fact that the study population of this trial consisted of young and otherwise healthy people (mean age 29 years).

We are aware of the flawed first data analysis published by Grossmann and Schmidramsl [1]. Therefore, we took great care in using statistical methods for the second data analysis that reduce any bias as much as possible and applied conservative statistics for the primary endpoints. The superiority of the special butterbur extract over placebo as shown in this analysis is confirmed by a larger international, randomised and controlled multicentre trial in 245 patients [18].

References

1 Grossmann M, Schmidramsl H: An extract of *Petasites hybridus* is effective in the prophylaxis of migraine. Int J Clin Pharmacol Ther 2000; 38:430–435.

2 Headache Classification Committee of the International Headache Society: Classification and diagnostic criteria for headache disorders, cranial neuralgias and facial pain. Cephalalgia 1988;8(suppl 7):1–93.

3 Tfelt-Hansen P, Block G, Dahlof C, Diener HC, Ferrari MD, Goadsby PJ, Guidetti V, Jones B, Lipton RB, Massiou H, Meinert C, Sandrini G, Steiner T, Winter PB, for the International Headache Society Clinical Trials Subcommittee: Guidelines for controlled trials of drugs in migraine: Second edition. Cephalalgia 2000;20:765–786.

4 Wei LJ, Lachin JM: Two-sample asymptotically distribution-free tests for incomplete multivariate observations. J Am Stat Assoc 1984;79: 653–661.

5 Thall PF, Lachin JM: Analysis of recurrent events: Nonparametric methods for random interval count data. J Am Assoc 1988;83:339–347.

6 Lachin JM: Some large-sample distribution-free estimators and tests for multivariate partially incomplete data from two populations. Stat Med 1992;11:1151–1170.

7 Colditz GA, Miller JN, Mosteller F: Measuring gain in the evaluation of medical technology: The probability of a better outcome. Int J Technol Assess Health Care 1988;4:637–642.

8 Ziolo G, Samochewiec L: Study on clinical properties and mechanism of action of Petasites in bronchial asthma and chronic obstructive bronchitis. Pharm Acta Helv 1998;72: 359–380.
9 Aebi A, Buechi J, Waaler T, Eichenberger E, Schmutz J: Inhaltsstoffe von *Petasites hybridus* (L) Fl Wett I. Pharm Acta Helv 1955;29:277–279.
10 Bucher K: Über ein antispastisches Prinzip in *Petasites officinalis Moench.* Arch Exp Pathol Pharmakol 1951;213:69–71.
11 Ko W, Lei C, Lin Y, Chen C: Relaxant effects of petasins in isolated guinea pig trachea and their structure-activity relationships. Planta Med 2000;66:650–652.
12 Thomet OA, Wiesmann UN, Blaser K, Simon HU: Differential inhibition of inflammatory effector functions by petasin, isopetasin and neopetasin in human eosinophils. Clin Exp Allergy 2001;31:1310–1320.
13 Brune K, Bickel D, Peskar BA: Gastro-protective effects by extracts of *Petasites hybridus:* The role of inhibition of peptido-leukotriene synthesis. Planta Med 1993;59:494–496.
14 Scheidegger C, Dahinden C, Wiesmann U: Effects of extracts of individual components from Petasites on prostaglandin synthesis in cultured skin fibroblasts and on leucotriene synthesis in isolated human peripheral leucocytes. Pharm Acta Helv 1998;72:359–380.
15 Thomet OA, Wiesmann UN, Schapowal A, Bizer C, Simon H: Role of petasin in the potential anti-inflammatory activity of a plant extract of *Petasites hybridus.* Biochem Pharmacol 2001;61:1041–1047.
16 Ko W, Lei C, Lin Y, Chen C: Mechanisms of relaxant action of S-petasine and S-isopetasin, sesquiterpenes of *Petasites formosanus,* in isolated guinea pig trachea. Planta Med 2001:67: 224–229.
17 Wang G-J, Shum AY-C, Lin Y-L, Liao J-F, Wu X-C, Ren J, et al: Calcium channel blockade in vascular smooth muscle cells: Major hypotensive mechanism of S-petasine, a hypotensive sesquiterpene from *Petasites formosanus.* J Pharm Exp Ther 2001;297:240–246.
18 Lipton RB, Göbel H, Wilks K, Mauskop A: Efficacy of petasites (an extract from *Petasites rhizoma*) 50 and 75 mg for prophylaxis of migraine: Results of a randomized, double-blind, placebo-controlled study. Neurology 2002;58(suppl 3):A472.