ELSEVIER

Journal of Clinical Epidemiology ■ (2008) ■

Journal of Clinical Epidemiology

REVIEW ARTICLE

The conclusions on the effectiveness of homeopathy highly depend on the set of analyzed trials

R. Lüdtke[a,*], A.L.B. Rutten[b]

[a]*Karl und Veronica Carstens-Stiftung, Essen, Germany*

[b]*Commissie Methode en Validering, Artsenvereniging voor homeopathie VHAN (Association of Dutch homeopathic physicians), Breda, The Netherlands*

Accepted 18 June 2008

Abstract

Objective: Shang's recently published meta-analysis on homeopathic remedies (*Lancet*) based its main conclusion on a subset of eight larger trials out of 21 high quality trials out of 110 included trials. We performed a sensitivity analysis on various other meaningful trial subsets of all high quality trials.

Study Design: Subsets were defined according to sample size, type of homeopathy, type of publication, and treated disease/condition. For each subset, we estimated the overall odds ratios (ORs) from random effect meta-analyses.

Results: All trials were highly heterogeneous ($I^2 = 62.2\%$). Homeopathy had a significant effect beyond placebo (OR = 0.76; 95% CI: 0.59−0.99; p = 0.039). When the set of analyzed trials was successively restricted to larger patient numbers, the ORs vary moderately (median: 0.82, range: 0.71−1.02) and the *P*-values increased steadily (median: 0.16, range: 0.03−0.93), including Shang's results for the eight largest trials (OR = 0.88, CI: 0.66−1.18; $P = 0.41$).

Shang's negative results were mainly influenced by one single trial on preventing muscle soreness in $N = 400$ long-distance runners.

Conclusions: The meta-analysis results change sensitively to the chosen threshold defining large sample sizes. Because of the high heterogeneity between the trials, Shang's results and conclusions are less definite as they had been presented.

Keywords: Homeopathy; Randomized clinical trials; Meta-analysis; Selection bias; Heterogeneity; Sensitivity analysis

1. Introduction

Homeopathy is a complementary medicine system which has been controversially discussed for more than 200 years. Recently, this discussion was taken up by a systematic review of homeopathic trials published in *Lancet* [1]. Here, Shang and coauthors performed a meta-analysis on 110 randomized trials comparing homeopathic medicines with placebo. This analysis was supplemented by a similar analysis of 110 matched placebo-controlled trials from conventional medicine. Both analyses showed that the trial results depended on a number of external parameters, such as the methodological quality of the trial, the publication language, the type of publication, and the precision of the effect estimator, hereby confirming previous results on randomized clinical trials in homeopathy [2]. These findings were interpreted as the presence of multiple bias and further analyses were restricted to a subset of 21 homeopathic trials which were of high methodological quality. Finally, from these 21 trials, a subset of eight trials which had included large patient numbers was analyzed. Here, the overall treatment effect (in terms of odds ratios [ORs]) was estimated at 0.88, its 95% confidence interval (CI) ranging from 0.65 to 1.19. As this interval covered the 1.00 (identity of verum and placebo), the effectiveness of homeopathic medicines could not be proved significantly (at a level of 5%). Based on these figures, the authors concluded that ''the effects seen in placebo-controlled trials of homeopathy are compatible with the placebo hypothesis'' [1].

Shang's analysis has been criticized to be prone to selection bias, especially when the set of 21 high quality trials was reduced to those eight trials with large patient numbers. In a letter to the *Lancet*, Fisher et al. posed the question: ''to what extend the meta-analysis results depend on how the threshold for 'large' studies was defined [3]. The present article addresses this question. We aim to investigate how Shang's results would have changed if other thresholds had been applied. Moreover, we extend our

* Corresponding author. Am Deimelsberg 36, D-45276 Essen, Deutschland. Tel.: +49 (0)201 56305-16; fax +49 (0)201 56305-30.

E-mail address: r.luedtke@carstens-stiftung.de (R. Lüdtke).

0895-4356/08/$ − see front matter
doi: 10.1016/j.jclinepi.2008.06.015

What is new?

1 The results of placebo-controlled clinical trials on homeopathic remedies are highly heterogeneous.

2 The heterogeneity cannot completely be explained by publication bias, methodological quality, or other external factors.

3 Meta-analyses on homeopathic remedies therefore yield completely different results depending on which trials were included, even if one restricts to only large, high-quality trials.

4 Previously published results on the (lack of) effectiveness of homeopathy are less conclusive than reported.

analyses to other meaningful subsets of the 21 high quality trials to investigate other sources of heterogeneity, an approach that is generally recommended to be a valuable tool for meta-analyses [4,5].

2. Methods

2.1. Data extractions

We reviewed the original publications of all 21 trials [6–26] which had been classified as "high quality" by Shang et al. [27]. All relevant data were extracted strictly following Shang's specifications which outcome parameter was used.

2.2. Data analysis

The data were processed and analyzed with methods identical or equivalent to Shang's analysis. All results were expressed as ORs and continuous outcomes were converted to ORs using standard formulas [28]. Here, ORs below 1.0 indicate a superiority of the homeopathic medicines, whereas ORs above 1.0 indicate superiority of placebo.

As a first step of analysis, the ORs were plotted on a log-scale against their standard errors in a funnel-plot [29]. Its asymmetry was explored by a weighted linear regression analysis (meta-regression) which modeled the log OR as a function of its standard error [30]. Weights were chosen inversely to the squared standard error. From this model the log of the asymmetry coefficient (AC) was estimated as the slope of the regression line and a predicted OR was obtained for trials with a standard error as small as the smallest observed standard error of all included trials.

Out of the 21 trials, the eight trials [6-13] with the highest patient numbers were selected. In these trials, a random effects meta-analysis was performed [31] and the pooled OR was estimated. Heterogeneity between trials was assessed by the I^2 coefficient [32]. I^2 measures the percentage of total variation across studies due to true heterogeneity rather than chance. I^2 values from different reviews cannot be compared directly.

Subsequent analyses followed exactly the same statistical methods: for each of the various subsets (see below), we performed a meta-regression and a random effects meta-analysis from which we estimated the predicted and pooled ORs.

All analyses were performed with SAS/STAT software, release 9.1 (SAS Institute Inc., Cary, NC, USA).

2.3. Definition of subsets

From the 21 trials, we defined and analyzed various subsets according to the following criteria

- *Sample size*: we ordered the trials according to the number of analyzed patients and then successively excluded one trial after another starting with the trial with the smallest patient numbers. This refers to our main questions whether the choice of the sample size threshold influences the results.
- *Treated condition/disease*: we separated those four trials on the prevention or treatment of muscle soreness after physical activities from the others; muscle soreness was the most frequent health condition in Shang's analysis.
- *Type of publication*: we included only those trials which were published in English or in MEDLINE listed journals.
- *Quality of statistical analysis*: we located those trials which were analyzed according to the intention-to-treat principle.
- *Type of homeopathic treatment*: we defined subsets depending on whether the patients were treated with classical homeopathy (the use of a single medicine prescribed according to the individual's presentation and history), fixed homeopathic treatment (the use of the same single agent for a group of patients), or complex homeopathic treatment (more than one medicine used concurrently).
- *Homeopathic dilution*: homeopathic dilutions are prepared by a process of serial dilution with vigorous shaking. Such dilutions are known as ultramolecular in that they are diluted to such a degree that not even a single molecule of the starting substance is likely to remain. For our analyses, we defined subsets depending on whether the homeopathic medicines were of ultramolecular dilution or not. If several homeopathic medicines were given, we defined the treatment only as ultramolecular if each medicine was ultramolecular.

2.4. Sensitivity analyses

To determine how sensitive Shang's analysis is to the results of a single trial, we successively omitted each of the

21 high quality trials and analyzed the remaining 20 with random effects meta-analyses. We did the same for the eight largest high quality trials and analyzed the remaining seven trials. This might give an impression if the overall results are supported by homogenous results of all trials or if they are governed by one or two single trials.

3. Results

Our data extraction seemed to work fairly good: when the 21 ORs and their confidence intervals were graphically displayed, the resulting figures matched those of Shang et al. Moreover, a random effects meta-analysis of the eight trials with highest patient numbers provided an overall treatment effect at OR = 0.88 (95% CI: 0.66−1.18) which only slightly differs from Shang's original results. Details of the data extracted are given in Table 1.

Fig. 1 displays the funnel plot of all 21 high quality trials. There were three outliers which do not match the general impression: two trials were found on the lower left side, indicating imprecise studies with very high treatment effects [19,25]. One trial was located at the lower right side, referring to a small trial with high effects in favor of placebo [26]. Due to those three outliers, the funnel plot seemed to be slightly skewed. This is corroborated by

Table 1
Extracted results for the 21 top quality placebo controlled trials from Shang, ordered by patient numbers (ORs < 1.00 indicate superiority of verum)

First author [reference]	Condition	Sample size	OR	95% Confidence interval
Rottey [6]	Influenza-like disease	501	0.77	0.56−1.06
Vickers [7]	Muscle soreness	400	1.38	0.97−1.96
Papp [8]	Influenza-like disease	334	0.56	0.36−0.85
Schmidt [9]	Weight loss	208	0.91	0.56−1.49
Labrecque [10]	Plantar warts	162	1.29	0.61−2.70
Jacobs [11]	Diarrhea	116	0.52	0.27−1.02
Weiser [12]	Sinusitis	104	0.67	0.33−1.36
Walach [13]	Headaches	98	1.74	0.83−3.69
Jacobs [14]	Diarrhea	81	0.45	0.20−0.99
Jacobs [15]	Otitis media	75	0.54	0.18−1.60
Hart [16]	Hysterectomy	73	1.22	0.52−2.81
Wiesenauer [17]	Pollinosis	72	0.31	0.11−0.87
Zell [18]	Sprains	69	0.31	0.11−0.87
Böhmer [19]	Sports injuries	67	0.06	0.01−0.32
Vickers [20]	Muscle soreness	57	1.39	0.53−3.63
Jawara [21]	Muscle soreness	50	1.29	0.45−3.63
Chapman [22]	Traumatic brain injuries	50	1.06	0.38−3.00
Tveiten [23]	Muscle soreness	46	0.68	0.23−2.01
Stevinson [24]	Hand surgery	42	0.93	0.34−3.34
Lepaisant [25]	Premenstrual syndrome	36	0.04	0.00−0.88
Chapman [26]	Premenstrual syndrome	10	2.25	0.18−28.17

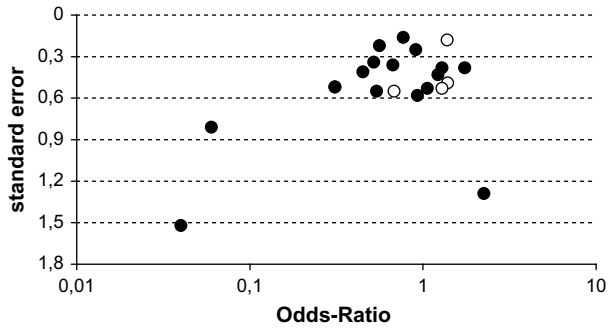

Fig. 1. Funnel-plot of 21 high quality trials comparing homeopathic remedies with placebo (ORs plotted on a log scale, standard errors in reverse order; open circles refer to the four trials on preventing or treating muscle soreness).

a meta-regression analysis which yielded a substantial (but statistically not significant) asymmetry coefficient (AC) of 0.40 ($P = 0.17$). No asymmetry could be found with the eight largest trials (AC = 1.15; $P = 0.94$).

For all 21 high quality trials the pooled OR, as estimated from random effects meta-analysis, was 0.76 (CI: 0.59−0.99; $P = 0.039$). This estimate changed to some extent when the number of analyzed trials was successively restricted according to sample size. As can be taken from Fig. 2, the OR slightly increased, and the respective confidence intervals broadened if fewer trials were included. The highest OR (1.02) was found for only 2 included studies (corresponding to a threshold of $N = 400$), followed by 5 (OR = 0.91, threshold at $N = 162$) and 8 included trials (OR = 0.88, threshold at $N = 98$). If 14 or more trials were included (threshold at $N = 69$), the OR was always significant (with one exception at 17 trials and a threshold at $N = 50$).

In most of these analyses there was a moderate, but statistically not significant (each $P > 0.05$) asymmetry in the funnel plots, the AC ranged from 0.13 (4 trials, threshold at $N = 208$) to 1.97 (5 trials, threshold at $N = 162$), its median was 0.37. Consequently, the predicted ORs (by means of

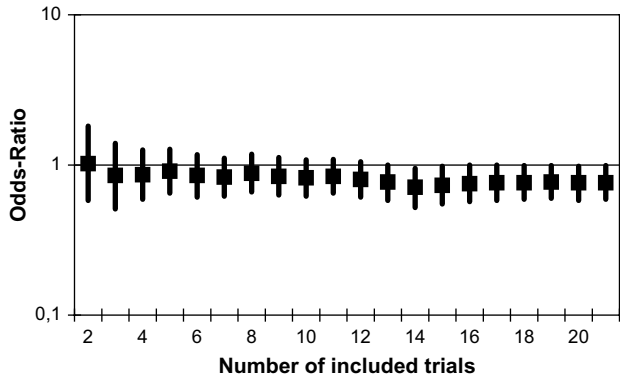

Fig. 2. Plot of overall ORs (from random effect meta-analyses) and respective confidence intervals against the number of included trials, successively drawn from the 21 high quality trials in descending order of patient numbers.

meta-regression) were always close to unity, indicating no difference between homeopathy and placebo (Fig. 3).

Interestingly, independent from the number of included trials, we found a substantial heterogeneity between the trials. The respective I^2 coefficients varied from 55.6% (11 trials included) to 82.8% (2 trials included).

These findings suggest that there are some unknown factors responsible for this heterogeneity. Presumably, one factor could be the respective disease/condition. Four of the 21 trials dealt with preventing or treating muscle soreness. A subset meta-analysis showed that homeopathic medicines are probably not helpful in muscle soreness (OR = 1.30; CI: 0.96−1.76; $P = 0.093$; $I^2 = 0\%$), as already had been suggested by others before [33]. When we restricted our analyses to the remaining 17 trials, we found an overall statistically significant effect (OR = 0.68; CI: 0.52−0.90; $P = 0.007$; $I^2 = 52.1\%$) which sustained when we further restricted to the eight largest remaining trials (OR = 0.75; CI: 0.58−0.96; $P = 0.025$; $I^2 = 37.1\%$).

These results were partly corroborated by meta-regression. The predicted ORs were OR = 1.41 (CI: 1.08−1.82; $P = 0.010$; AC = 0.48) for the muscle soreness trials, OR = 0.80 (CI: 0.58−1.11; $P = 0.19$; AC = 1.69) for the 17 non muscle soreness trials, and OR = 0.72 (CI: 0.49−1.05; $P = 0.087$; AC = 0.46) for the 8 largest non muscle soreness trials.

There are other meaningful subsets that may help to explain the heterogeneity between the 21 high quality trials. As can be seen from Table 2, with one exception, none of these provided any finding which substantially differed from the overall results. The only exception were the six trials with low (molecular) homeopathic dilutions [12,17−19,24,25] where we found a significant effect in favor of homeopathy with random effects meta-analysis.

Above we showed that treating/preventing muscle soreness may be one major factor which could partly explain the heterogeneity between trials. This finding was corroborated by our sensitivity analyses: Shang's overall results are essentially affected by Vickers' trial on the effectiveness of Arnica D30 in preventing muscle soreness in long distance runners [7]. If this trial was omitted from the analysis of all 21 high quality trials the overall OR, as estimated from random effects meta-analysis, reduced slightly from 0.76 (see above) to 0.73 (CI: 0.56−0.93; $P = 0.013$, Table 3). More interestingly, the I^2-coefficient decreased considerably, indicating that this trial accounted substantially for the heterogeneity between the trials.

If the Vickers' trial was omitted from the Shang's analysis of the eight largest high quality trials (Table 4), the overall OR reduced from 0.88 (see above) to 0.80 but remained statistically not significant (CI: 0.61−1.05; $P = 0.11$). In contrast, there was a significant result when the Vickers' trial was omitted and the sample size threshold was changed, either to $N = 80$ (eight trials, OR = 0.75, CI: 0.58−0.96, $P = 0.025$) or to $N = 100$ (six trials, OR = 0.73, CI: 0.59−0.91, $P = 0.005$). The last subset was one of our analyses which showed the smallest heterogeneity between trials ($I^2 = 12.7\%$).

In contrast, it can be shown that the positive results from all 21 trials (see above) was mainly determined by the two trials on influenza-like diseases [6,8]. Compared with the analysis of all 21 trials (see above), the ORs remained nearly constant if one of them was omitted from the analysis but the P-values substantially increased and exceeded the 5% significance limit (Table 3).

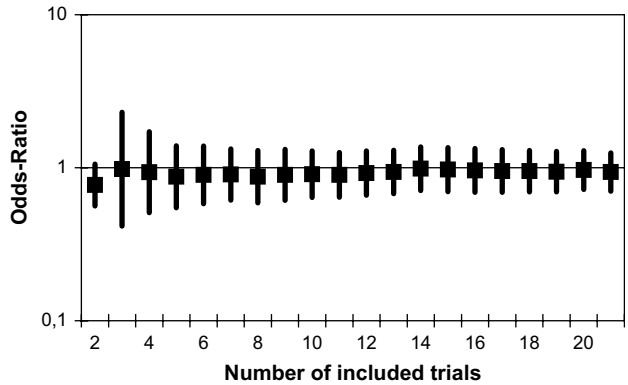

Fig. 3. Plot of predicted ORs (from meta-regression) and respective confidence intervals against the number of included trials, successively drawn from the 21 high quality trials in descending order of patient numbers.

4. Discussion

In our study, we performed a large number of meta-analyses and meta-regressions in 21 high quality trials comparing homeopathic medicines with placebo. In general, the overall ORs did not vary substantially according to which subset was analyzed, but P-values did.

From a Bayesian point of view, these figures mainly confirm the results of Shang, the a posteriori probability, that homeopathic remedies are effective, essentially stays stable. However, neither Shang nor others discussed the original results from a Bayesian but rather from a frequentist's point of view, strictly adhering to an arbitrarily chosen cut-off point of $\alpha = 5\%$ to decide whether homeopathy works or not. In the frequentist's perspective, the conclusions substantially depend on the point of view one takes. Shang et al. in their article arbitrarily defined one subset of eight trials which provided an overall negative result for homeopathy. Our article shows that the choice of other meaningful subsets could lead to the opposite conclusion. For example, there is no external criterion why a ''large'' trial should have $N = 98$ or more patients as defined by Shang, other thresholds are as valid and meaningful as this one: for example, splitting the data set at a threshold of $N = 66$, the median sample size of all 110 homeopathic trials, there had been a significant effect in favor of homeopathy with random effects meta-analysis.

This result can be interpreted differently. Following Shang's perspective it can be explained by small study bias

Table 2
Number of trials (N), coefficients of hetrogeneity (I^2) and asymmetry (AC), overall ORs and 95% confidence intervals (CI), and P-values (from random effect meta-analyses and meta-regression) for various subsets of the 21 high quality trials

				Meta-analysis		Meta-regression	
Subset	N	I^2	AC	OR (CI)	P-value	OR (CI)	P-value
MEDLINE listed	14	55.5	0.14	0.83 (0.60−1.14)	0.25	1.29 (0.91−1.84)	0.16
English language	15	39.6	1.02	0.93 (0.72−1.19)	0.56	0.94 (0.68−1.30)	0.69
Intention-to-treat principle	8	64.2	0.20	0.69 (0.36−1.32)	0.26	1.13 (0.60−2.15)	0.70
Ultramolecular dilutions[a]	14	45.2	1.11	0.89 (0.71−1.13)	0.34	0.88 (0.66−1.18)	0.39
Molecular dilutions[a]	6	59.8	0.06	0.34 (0.16−0.74)	0.006	0.66 (0.33−1.31)	0.24
Classical homeopathy	6	45.8	2.70	0.78 (0.46−1.32)	0.36	0.70 (0.37−1.31)	0.26
Fixed homeopathy	13	50.7	0.60	0.88 (0.67−1.17)	0.39	0.94 (0.70−1.26)	0.68

[a] One trial in classical homeopathy [35] could not be classified into high or low potencies because the details were missing regarding which treatments were actually given.

(which includes publication bias). In contrast, one may hypothesize that Shang's result is falsely negative. According to Linde and Jonas the chance to find a false-negative result increases when the trials are highly heterogeneous, for example, when a therapeutic system under study works in some but not all indications [34]. This interpretation was rejected by Shang when noting that essentially the same approach did not lead to false-negative results in conventional medicine [35].

Meta-analysis is defined as "a statistical analysis which combines or integrates the results of several independent clinical trials considered by the analyst to be 'combinable'" [36]. This definition makes clear that it is somewhat subjective as to which trials are considered combinable and which are not. This is especially true when trials are extremely heterogeneous not only in results but also in the interventions and health conditions under study. The Cochrane Handbook for Systematic Reviews recommends that "Meta-analysis should only be considered when a group of trials is sufficiently homogeneous in terms of participants, interventions and outcomes to provide a meaningful summary" [4]. But, up to date there is no guideline to explain the meaning of "sufficiently" or "meaningful" in this context except that Higgins mentions I^2 values beyond 50% as a substantial heterogeneity [37].

Similarly, there is no guideline which tells a researcher when to prefer meta-regression to random effects meta-analysis or vice versa. As the statistical test for asymmetry only has a small power, Egger suggests to perform meta-regressions when the respective P-value falls below 0.10 [29]. Applying this criterion there seemed to be no need to perform a meta-regression in most of the subsets we analyzed. On the other hand, our meta-regression results

Table 3
Number of trials (N), coefficients of hetrogeneity (I^2) and asymmetry (AC), overall ORs and 95% confidence intervals (CI), and P-values (from random effect meta-analyses and meta-regression) if one of the 21 high quality trials was omitted from the analysis

				Meta-analysis		Meta-regression	
Omitted trial	N	I^2	AC	OR (CI)	P-value	OR (CI)	P-value
Rottey [6]	20	58.7	0.28	0.76 (0.57−1.01)	0.057	1.03 (0.73−1.45)	0.88
Vickers [7]	20	47.7	0.60	0.73 (0.56−0.93)	0.013	0.80 (0.59−1.10)	0.17
Papp [8]	20	55.4	0.33	0.78 (0.60−1.03)	0.077	1.02 (0.76−1.38)	0.87
Schmidt [9]	20	58.8	0.40	0.75 (0.57−0.99)	0.041	0.93 (0.68−1.26)	0.64
Labreque [10]	20	57.6	0.37	0.74 (0.57−0.97)	0.027	0.93 (0.69−1.24)	0.60
Jacobs [11]	20	57.2	0.41	0.78 (0.60−1.02)	0.070	0.95 (0.71−1.28)	0.74
Weiser [12]	20	58.6	0.40	0.77 (0.59−1.01)	0.055	0.94 (0.70−1.27)	0.69
Walach [13]	20	54.9	0.36	0.73 (0.56−0.94)	0.017	0.92 (0.70−1.21)	0.55
Jacobs [14]	20	56.8	0.43	0.79 (0.61−1.02)	0.073	0.94 (0.71−1.26)	0.70
Jacobs [15]	20	58.4	0.41	0.77 (0.59−1.01)	0.055	0.94 (0.70−1.26)	0.66
Hart [16]	20	58.2	0.37	0.74 (0.57−0.97)	0.030	0.93 (0.70−1.25)	0.63
Wiesenauer [17]	20	55.5	0.45	0.79 (0.61−1.03)	0.077	0.93 (0.70−1.24)	0.64
Zell [18]	20	55.5	0.57	0.79 (0.61−1.03)	0.077	0.93 (0.70−1.24)	0.64
Böhmer [19]	20	47.7	0.35	0.81 (0.64−1.02)	0.076	0.91 (0.69−1.19)	0.48
Vickers [7]	20	57.9	0.35	0.74 (0.57−0.97)	0.027	0.94 (0.70−1.25)	0.65
Jawara [21]	20	58.3	0.36	0.75 (0.57−0.97)	0.030	0.94 (0.70−1.25)	0.67
Chapman [26]	20	58.7	0.39	0.75 (0.58−0.98)	0.035	0.94 (0.70−1.26)	0.67
Tveiten [23]	20	58.8	0.36	0.77 (0.59−1.00)	0.048	0.94 (0.70−1.26)	0.67
Stevinson [24]	20	58.7	0.50	0.75 (0.58−0.98)	0.036	0.94 (0.70−1.26)	0.68
Lepaisant [25]	20	55.4	0.30	0.78 (0.61−1.00)	0.052	0.91 (0.68−1.23)	0.54
Chapman [26]	20	58.4	0.40	0.76 (0.58−0.98)	0.034	0.97 (0.72−1.29)	0.81

Table 4
Coefficients of hetrogeneity (I^2) and asymmetry (AC), overall ORs and 95% confidence intervals (CI), and *P*-values (from random effect meta-analyses and meta-regression) if one of the eight large high quality trials was omitted from the analysis

				Meta-analysis		Meta-regression	
Omitted trial	*N*	I^2	AC	OR (CI)	*P*-value	OR (CI)	*P*-value
Rottey [6]	7	65.7	0.49	0.91 (0.64−1.30)	0.61	0.98 (0.58−1.66)	0.95
Vickers [7]	7	43.6	3.86	0.80 (0.61−1.05)	0.11	0.69 (0.48−1.00)	0.051
Papp [8]	7	54.6	1.02	0.96 (0.71−1.29)	0.78	0.96 (0.65−1.41)	0.83
Schmidt [9]	7	67.5	1.13	0.88 (0.63−1.24)	0.47	0.87 (0.56−1.35)	0.54
Labreque [10]	7	65.7	0.64	0.85 (0.62−1.17)	0.32	0.89 (0.58−1.36)	0.59
Jacobs [11]	7	62.4	2.51	0.94 (0.69−1.27)	0.68	0.86 (0.58−1.28)	0.47
Weiser [12]	7	66.4	1.82	0.91 (0.66−1.25)	0.57	0.87 (0.57−1.32)	0.50
Walach [13]	7	60.2	0.38	0.83 (0.62−1.11)	0.20	0.90 (0.61−1.33)	0.61

considerably differed from our meta-analysis results: no single predicted OR could be shown to differ significantly from unity. On the first glance, this seems to support Shang's thesis that homeopathy is not better than placebo. But, from our point of view this perspective might be misleading. First, the asymmetry of funnel-plots is not necessarily a result of bias. It can also occur when smaller studies show larger effect just because they were done in a condition with high treatment effects, and thus requiring smaller patient numbers. Moreover, meta-regression predicts the OR at an extremal value (the minimal standard deviation observed). From mathematical statistics, it is well known that these predictions are imprecise, especially when the number of observations is small and the estimate of the regression line is unstable. This is the case in our analyses. For example, the funnel plot of the four largest trials was negatively skewed (AC = 0.13), whereas it was positively skewed for the five largest trials (AC = 1.97). Thus, our meta-regression analyses generally suffer from small statistical power.

In various studies, it has been shown that the results of meta-analyses depend on the selection criteria applied. The exclusion of gray literature from meta-analyses can lead to exaggerated estimates of intervention effectiveness [38,39]. Especially in complementary medicine, language restrictions may lead to biased estimates [40] but not in conventional medicine [41], although in general for randomized controlled trials the quality of reports on complementary medicine is comparable with those of conventional medicine [42]. Thus, it is important to note that with our analyses, we completely relied on the data provided by Shang et al.; we took the data as they were. For example, we did not question whether Shang et al. were able to identify all relevant placebo-controlled trials on homeopathy. Such analyses would have been far beyond the purpose of this article.

Similarly, we did not question Shang's quality ratings which lead to a subset of 21 high quality trials. There may be other quality ratings (based on the assessments of Linde [43] or Dean [44], for example) which could change the subset of high-quality trials and therefore affect the results. For example, the trials of de Lange-de-Klerk et al. [45], Reilly et al. [46,47], and Hofmeyr et al. [48] have been classified as low quality by Shang but as high quality by others [43].

Moreover, we did not question the outcome data Shang et al. extracted from the reports. When we re-extracted the data, we found it easy to understand the numbers that had been used to calculate the ORs. But, in some trials we could not replicate *why* exactly these numbers have been extracted. In Schmidt's trial on obesity [9], for example, the data were extracted from weight loss at day 1, but day 2 had been defined as the main outcome parameter, a decision which favored the results toward homeopathy. In contrast, in Chapman's trial on traumatic brain injuries [22], he reports on substantial effects in favor of homeopathy in 2 out of 3 outcome criteria, but Shang et al. extracted the data from the one criterion which showed no effect between the active and control groups. There are other examples [13,19,24] which show that data extraction was (necessarily) somewhat arbitrary. This shows that results from meta-analyses in general are superimposed by some extravariability which comes from the data extraction process.

As already noted by Shang, the results of placebo-controlled trials in homeopathy were extremely heterogeneous. In our analyses, we were not able to explain this heterogeneity sufficiently: even if the four trials on treatment/prevention of muscle soreness were excluded, there remained a substantial residual heterogeneity: more than half of the variation across studies was because of true systematic heterogeneity rather than chance. Similar results were obtained for most of our analyses. Reviewing the literature, these values must be regarded as unusually high: less than 20% of all published meta-analyses have I^2 values $> 50\%$ [37].

Heterogeneity seems to be intrinsic in placebo-controlled trials on homeopathy. This suggests, that the main underlying scientific hypothesis "Are the effects of homeopathy placebo?" does not make sense. Similarly it does not make sense to ask whether conventional medicine is placebo. From our point of view, also in homeopathy the hypothesis should be more specific and include both a specific definition of the homeopathic intervention and a clear definition of the disease/health condition.

Despite all efforts for achieving objectivity, there "remain considerable areas of subjectivity in carrying out

a meta-analysis'' [5]. This includes decisions on how to search the literature, which trials to include or exclude, and how to direct the statistical analysis, especially how to investigate sources of heterogeneity. Moreover, there is subjectivity involved, when we rate the effectiveness of a therapy under study: ''We sometimes accept the randomized trial evidence and discard the theory but at other time we stick to the theory and dismiss the 'facts''' [49]. In philosophy, this perspective is known as the Duhem-Quine hypothesis: Any theoretical claim can consistently be retained in the face of contrary evidence by making adjustments elsewhere in one's web of beliefs [50]. In an editorial to Shang's article Vandenbroucke states that its analyses and conclusions on homeopathy only gain meaning on the background, that the mechanisms of action of highly diluted homeopathic substances is completely implausible [51]. This seems to be the core of the problem: Until such a mechanism of action is not established, the a priori credibility for homeopathy is low, and this should cause a long lasting and ongoing discussion about the use of research in complementary and alternative medicine [52].

5. Conclusions

Our results do neither prove that homeopathic medicines are superior to placebo nor do they prove the opposite. This, of course, was never our intention, this article was only about how the overall results—and the conclusions drawn from them—change depending on which subset of homeopathic trials is analyzed. As heterogeneity between trials makes the results of a meta-analysis less reliable, it occurs that Shang's conclusions are not so definite as they have been reported and discussed.

6. Conflicts of interest

We declare that we have no financial conflicts of interests in connection with this article. R Lüdtke works for an independent non-profit foundation which is dedicated to research funding in homeopathy. Lex Rutten is a homeopathic doctor.

References

[1] Shang A, Huwiler-Muntener K, Nartey L, Juni P, Dorig S, Sterne JA, et al. Are the clinical effects of homoeopathy placebo effects? Comparative study of placebo-controlled trials of homoeopathy and allopathy. Lancet 2005;366:726−32.

[2] Linde K, Scholz M, Ramirez G, Clausius N, Melchart D, Jonas WB. Impact of study quality on outcome in placebo-controlled trials of homeopathy. J Clin Epidemiol 1999;52:631−6.

[3] Fisher P, Berman B, Davidson J, Reilly D, Thompson T. Are the clinical effects of homoeopathy placebo effects? [letter]. Lancet 2005;366:2082−3.

[4] 4.2.4 [updated March 2005]. In: Higgins J, Green S, editors. Cochrane handbook for systematic reviews of interventions. Chichester, UK: John Wiley & Sons, Ltd.; 2005.

[5] Thompson SG. Why and how sources of heterogeneity should be investigated. In: Egger M, Smith GD, Altman D, editors. Systematic Reviews in Health Care. London: BMJ Publishing Group; 2001. p. 157−75.

[6] Rottey EED, Verleye GB, Liagre RLP. Het effect van een homeopathische bereiding van micro-organismen bij de preventie von griepsymtomen—Een gerandomiseerd dubbel-blind onderzoek in de huisartspraktijk. Tijdschr Integ Geneeskunde 1995;11:54−8.

[7] Vickers AJ, Fisher P, Wyllie SE, Rees R. Homeopathic Arnica 30X is ineffective for muscle soreness after long-distance running—a randomized, double-blind, placebo-controlled trial. Clin J Pain 1998; 14:227−31.

[8] Papp R, Schuback G, Beck E, Burkard G, Bengel J, Lehrl S, et al. Oscillococcinum in patients with influenza-like syndromes: a placebo controlled double-blind evaluation. Br Homeopath J 1998;87(2): 69−76.

[9] Schmidt JM, Ostermayr B. Does a homeopathic ultramolecular dilution of Thyroidinum 30CH affect the decrease of body weight reduction in fasting patients?—A randomised placebo-controlled double-blind trial. Homeopathy 2002;91:197−206.

[10] Labrecque M, Audet D, Latulippe L, Drouin J. Homeopathic treatment of plantar warts. Can Med Assoc J 1992;146:1749−53.

[11] Jacobs J, Jiménez LM, Malthouse S, Chapman E, Crothers D, Masuk M, et al. Homeopathic treatment of acute childhood diarrhea—results from a clinical trial in Nepal. J Alternat Complement Med 2000;6(2): 131−9.

[12] Weiser M, Clasen BPE. Randomisierte plazebokontrollierte Doppelblindstudie zur Untersuchung der klinischen Wirksamkeit der homöopathischen Euphorbium compositum-Nasentropfen S bei chronischer Sinusitis. Forsch Komplementärmed 1994;1:251−9.

[13] Walach H, Haeusler W, Lowes T, Mussbach D, Schamell U, Springer W, et al. Classical homeopathic treatment of chronic headaches. Cephalalgia 1997;17:119−26.

[14] Jacobs J, Jiménez LM, Gloyd SS, Gale JL, Crothers D. Treatment of acute childhood diarrhea with homoeopathic medicine: a randomized clinical trial in Nicaragua. Pediatrics 1994;93:719−25.

[15] Jacobs J, Springer DA, Crothers D. Homeopathic treatment of acute otitis media in children: a preliminary randomized placebo-controlled trial. Pediatr Infect Dis J 2001;177−83.

[16] Hart O, Mullee MA, Lewith G, Miller J. Double-blind, placebo-controlled, randomized clinical trial of homoeopathic arnica C30 for pain and infection after total abdominal hysterectomy. J Royal Soc Med 1997;90:73−8.

[17] Wiesenauer M, Häussler S, Gaus W. Pollinosis-Therapie mit Galphimia glauca. MMW Fortschr Med 1983;101:811−4.

[18] Zell J, Connert WD, Mau J, Feuerstake G. Behandlung von akuten Sprunggelenksdistorsionen—Doppelblindstudie zum Wirksamkeitsnachweis eines homöopathischen Salbenpräparats. MMW Fortschr Med 1988;106(5):96−100.

[19] Böhmer D, Ambrus P. Behandlung von Sportverletzungen mit Traumeel-Salbe—Kontrollierte Doppelblindstudie. Biol Medizin 1992; 21(4):260−8.

[20] Vickers AJ, Fisher P, Smith C, Wyllie SE, Lewith GT. Homoeopathy for delayed onset muscle soreness—a randomised double blind placebo controlled trial. Brit J Sports Med 1997;31:304−7.

[21] Jawara N, Lewith GT, Vickers A, Mullee MA, Smith C. Homoeopathic Arnica and Rhus toxicodendron for delayed onset muscle soreness—a pilot for a randomized, double-blind, placebo-controlled trial. Brit Hom J 1997;86(1):10−5.

[22] Chapman EH, Weintraub RJ, Milburn MA. Homeopathic treatment of mild traumatic brain injury—a randomised, double-blind, placebo-controlled clinical trial. J Head Trauma Rehabil 1999;14: 521−42.

[23] Tveiten D, Bruset S, Borchgrevink CF, Norseth J. Effects of the homeopathic remedy Arnica D30 on marathon runners: a randomized, double-blind study during the 1995 Oslo Marathon. Complem Ther Med 1998;6(2):71−4.

[24] Stevinson C, Devaraj VS, Fountain-Barber A. Homeopathic arnica for prevention of pain and bruising: randomized placebo-controlled trial in hand surgery. J R Soc Med 2003;96:60−5.
[25] Lepaisant C. Essai thérapeutique en homéopathie: traitement des tensions mammaires et mastodynies du syndrome prémenstruel. Rev Fr Gynécol Obstét 1995;90(2):94−7.
[26] Chapman EH, Angelica J, Spitalny G, Strauss M. Results of a study of the homoeopathic treatment of PMS. J Am Inst Hom 1994;87: 14−21.
[27] Egger M. Study characteristics of homoeopathy studies. Available from: http://www.ispm.unibe.ch/downloads/Lancet_Homoeopathy 2005;. [cited 06.03.2006].
[28] Hasselblad V, Hedges L. Meta-analysis of screening and diagnostic tests. Psychol Bull 1995;117:167−78.
[29] Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. British Med J 1997;315:629−34.
[30] Thompson SG, Sharp SJ. Explaining heterogeneity in meta-anaysis: a comparison of methods. Statist Med 1999;18:2693−708.
[31] Fleiss J. The statistical basis of meta-analysis. Stat Meth Med Res 1993;2:121−45.
[32] Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21:1539−58.
[33] Ernst E, Barnes J. Are homoeopathic remedies effective for delayed-onset muscle soreness?—a systematic review of placebo-controlled trials. Perfusion 1998;11:4−8.
[34] Linde K, Jonas W. Are the clinical effects of homoeopathy placebo effects?. [letter]. Lancet 2005;366:2081−2.
[35] Shang A, Jüni P, Sterne JAC, Huwiler-Müntener K, Egger M. Author's reply. Lancet 2005;366:2083−5.
[36] Egger M, Smith GD, Phillips AN. Meta analysis: principles and procedures. BMJ 1997;315:1533−7.
[37] Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Br Med J 2003;327:557−60.
[38] McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? Lancet 2000;356:1228−31.
[39] Chan A, Hjobartsson A, Haahr M, Gotzsche P, Altman D. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA 2004;291:2457−65.
[40] Pham B, Klassen TP, Lawson ML, Moher D. Language of publication restrictions in systematic reviews gave different results depending on whether the intervention was conventional or complementary. J Clin Epidemiol 2005;58:769−76.
[41] Moher D, Fortin P, Jadad AR, Juni P, Klassen T, Le Lorier J, et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. Lancet 1996;347:363−6.
[42] Klassen TP, Pham B, Lawson ML, Moher D. For randomized controlled trials, the quality of reports of complementary and alternative medicine was as good as reports of conventional medicine. J Clin Epidemiol 2005;58:763−8.
[43] Linde K, Clausius N, Ramirez G, Melchart D, Eitel F, Hedges L, et al. Are the clinical effects of homoeopathy placebo effects?—a meta-analysis of placebo-controlled trials. Lancet 1997;350:834−43.
[44] Dean ME. The trials of homeopathy. Essen: KVC-Verlag; 2004.
[45] de Lange-de-Klerk ESM, Blommers J, Kuik DJ, Bezemer PD, Feenstra L. Effects of homoeopathic medicines on daily burden of symptoms in children with recurrent upper respiratory tract infections. BMJ 1994;309:1329−32.
[46] Reilly DT, Taylor MA, Beattie NGM, Campbell JH, McSharry C, Aitchison TC, et al. Is evidence for homoeopathy reproducible? Lancet 1994;344:1601−6.
[47] Reilly DT, Taylor MA, McSharry C, Aitchison T. Is homoeopathy a placebo response?—Controlled trial of homoeopathic potency—with pollen in hayfever as model. Lancet 1986;2:881−6.
[48] Hofmeyr GJ, Piccioni V, Blauhof P. Postpartum homoeopathic Arnica montana—a potency-finding pilot study. Br J Clin Pract 1990;44: 619−21.
[49] Vandenbroucke JP, de Craen AJ. Alternative medicine: a ''mirror image'' for scientific reasoning in conventional medicine. Ann Intern Med 2001;135:507−13.
[50] Quine WV. Ontological relativity and other essays. New York: Columbia University Press; 1969.
[51] Vandenbroucke JP. Homoeopathy and ''the growth of truth. Lancet 2005;366:691−2.
[52] Gezondheidsraad. Alternatieve Behandelwijzen en Wetenschappelijk Onderzoek [Alternative treatments and scientific research]. Den Haag; 1993.