Homeopathy (2008) 97, 169–177
© 2008 The Faculty of Homeopathy
doi:10.1016/j.homp.2008.09.008, available online at http://www.sciencedirect.com

ORIGINAL PAPER

The 2005 meta-analysis of homeopathy: the importance of post-publication data

ALB Rutten[1,*] and CF Stolper[2]

[1]*Homeopathic physician, former general practitioner, Breda, The Netherlands*
[2]*General practitioner, homeopathic physician, Machteldskamp 19, 8181 ZN Heerde, The Netherlands*

Background: **There is a discrepancy between the outcome of a meta-analysis published in 1997 of 89 trials of homeopathy by Linde et al and an analysis of 110 trials by Shang et al published in 2005, these reached opposite conclusions. Important data were not mentioned in Shang et al's paper, but only provided subsequently.**
Questions: **What was the outcome of Shang et al's predefined hypotheses? Were the homeopathic and conventional trials comparable? Was subgroup selection justified? The possible role of ineffective treatments. Was the conclusion about effect justified? Were essential data missing in the original article?**
Methods: **Analysis of post-publication data. Re-extraction and analysis of 21 higher quality trials selected by Shang et al with sensitivity analysis for the influence of single indications. Analysis of comparability. Sensitivity analysis of influence of subjective choices, like quality of single indications and of cut-off values for 'larger samples'.**
Results: **The quality of trials of homeopathy was better than of conventional trials. Regarding smaller trials, homeopathy accounted for 14 out of 83 and conventional medicine 2 out of 78 good quality trials with $n < 100$. There was selective inclusion of unpublished trials only for homeopathy. Quality was assessed differently from previous analyses. Selecting subgroups on sample size and quality caused incomplete matching of homeopathy and conventional trials. Cut-off values for larger trials differed between homeopathy and conventional medicine without plausible reason. Sensitivity analyses for the influence of heterogeneity and the cut-off value for 'larger higher quality studies' were missing. Homeopathy is not effective for muscle soreness after long distance running, OR = 1.30 (95% CI 0.96–1.76). The subset of homeopathy trials on which the conclusion was based was heterogeneous, comprising 8 trials on 8 different indications, and was not matched on indication with those of conventional medicine. Essential data were missing in the original paper.**
Conclusion: **Re-analysis of Shang's post-publication data did not support the conclusion that homeopathy is a placebo effect. The conclusion that homeopathy is and that conventional is not a placebo effect was not based on comparative analysis and not justified because of heterogeneity and lack of sensitivity analysis. If we confine ourselves to the predefined hypotheses and the part of the analysis that is indeed comparative, the conclusion should be that quality of homeopathic trials is better than of conventional trials, for all trials ($p = 0.03$) as well as for smaller trials ($p = 0.003$).** *Homeopathy* (2008) **97**, 169–177.

Keywords: homeopathy; meta-analysis; comparative analysis; quality bias; selection bias; cut-off value; adverse effects

*Corresponding author. ALB Rutten, Aard 10, 4813 NN Breda, The Netherlands.
E-mail: lexrtn@concepts.nl
Received 11 January 2008; revised 6 August 2008; accepted 11 September 2008

Introduction

The discussion about proof for homeopathy is in part, a meta-discussion about proof. Several meta-analyses of randomised controlled trials (RCTs) – in 1991, 1997 and

2000 – indicate a specific effect of homeopathy.[1–3] Both homeopathic and conventional meta-analyses have been criticised.[4–6] Some authors suggest that there is no difference between proof requirements for homeopathy and for conventional methods.[1,7] However, the implausibility of homeopathy's mechanism of action seems to have led to an amalgamation of bias. Sterne, Egger and Smith concluded that the role of low quality in small studies was neglected in Linde's meta-analysis.[8] Commenting on the analysis of homeopathy by Shang et al published in August 2005, and which referred to the 'small low quality study' hypothesis, the editor of the Lancet advised ''doctors need to be bold and honest with their patients about homeopathy's lack of benefit''.[9] Vandenbroucke concluded that this meta-analysis showed higher sensitivity for potential bias for homeopathic than for allopathic trials.[10]

The Cochrane Handbook for Systematic Reviews states ''Reliable conclusions can only be drawn from analyses that are truly pre-specified before inspecting the trials' results''.[11] Such pre-specification is more difficult because most homeopathy trials have been analysed in earlier meta-analyses. The Cochrane Handbook further recommends ''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''. Pooling of results is thus questionable if homeopathy works for some conditions and not for others.[12] Egger stated ''If subgroup analyses are to be done, they need to be as complete as possible and should involve commonly defined subgroups and outcomes across all the trials in the subgroup''.[13] Meta-analysis is a subjective procedure, Boden warns that it can easily become a weapon instead of a tool.[14]

The hypotheses predefined mentioned in the introduction of Shang et al's paper were: ''Bias in conduct and reporting of trials is a possible explanation for positive findings of placebo-controlled trials of both homeopathy and allopathy (conventional medicine)''; and: ''These biases are more likely to affect small than large studies; the smaller a study, the larger the treatment effect necessary for the results to be statistically significant, whereas large studies are more likely to be of high methodological quality and published even if their results are negative''.

Shang et al's analysis was criticised because the authors failed to include essential data to support their conclusion.[15–17] Four months later the missing data were revealed (www.ispm.ch). The missing data were 1. Excluded trials. 2. The trials regarded as of higher quality. 3. The trials (8 homeopathy, 6 conventional medicine) that led to the final conclusion.

Questions

More or less the same set of homeopathy trials has been re-analysed several times. The contradiction between Linde's conclusion based on 89 trials, and Shang et al's conclusion, based on 110 trials seems odd. Shang et al's analysis was presented as a comparative analysis matching 110 homeopathy trials with 110 conventional trials by indications. The conclusion was based on 8 homeopathy trials and 6 conventional trials.

The post-publication data enabled us to reconstruct the analysis, although data were presented as graphs, not as raw numbers. In our recent paper 'The conclusions on the effectiveness of homeopathy highly depend on the set of analysed trials' we re-analysed the data from the original articles, did sensitivity analyses and estimated the influence of heterogeneity.[18] The large amount of heterogeneity suggests that this factor was not considered at all. We found no reasonable explanation for the choice of cut-off value for 'larger trials'.

After these basic conclusions several questions remain:

1. What was the outcome of the pre-specified hypotheses?
2. Were the two methods comparable?
3. Was subgroup selection justified?
4. What is the influence of ineffective treatments?
5. Was the final conclusion justified?
6. Were essential data missing from the original article?

Methods

We analysed the subsequently disclosed data and investigated which hypotheses were tested. The ISPM website presented graphs, but no data about effect sizes and confidence intervals were given. We reconstructed the odds ratios and confidence intervals of the 21 higher quality homeopathy studies from the original articles. Data were processed and analysed with methods identical or equivalent to those of Shang et al's analysis. We checked the results with Shang et al's data, then focused on these 21 higher quality studies because the conclusion was based on larger higher quality studies. For these trials a random effects meta-analysis was performed and the pooled odds ratio was estimated. We estimated odds ratios and confidence intervals for some of the trials excluded by Shang et al, but regarded as good quality by Linde et al. We performed meta-analyses for other eligible sets of trials. We tested comparability and matching of trials. We compared this analysis with referenced publications to check predefinition of hypotheses. We assessed the influence of some subjective choices, like quality and cut-off values for sample size and performed sensitivity analysis to check for the influence of separate indications. SAS/Stat®, release 9.1 statistical software was used.

Results

Shang et al presented their study and their conclusion as a comparison of homeopathy and conventional medicine. To reconstruct their work we had to make several hypotheses that were not predefined by Shang et al, to arrive at their conclusions. In this process comparibility of the homeopathic and conventional groups was lost.

The predefined hypotheses

The first predefined hypothesis (quality in homeopathy is worse than in conventional medicine) was falsified by

Shang et al. Median sample sizes were the same: 65.5 in homeopathy, 65 in conventional medicine. Effects of homeopathy and conventional medicine were similar; 95% of the odds ratios (or) were from 0.12 to 1.65 for homeopathy and from 0.13 to 1.52 for conventional medicine. According to Shang et al ''Most odds ratios indicated a beneficial effect of the intervention''. In the homeopathy group (including unpublished trials) 21 (19%) of the trials were of higher quality, in the conventional group 9 (8%). Overall quality in homeopathy studies was better than for conventional medicine ($p = 0.03$).

Quality in small studies. Shang et al referred, for their second predefined hypothesis, to Sterne, Egger and Smith stating that quality bias is mainly influenced by quality of small studies.[8] Effects of treatment could in their view truly be larger in high quality smaller trials because of better selection of patients. On the other hand effects are over-estimated if quality is low. In both cases we see asymmetry in the funnel plot, but in the first case this does not indicate bias and if larger trials with poorer patient selection then indicate no effect the conclusion that the therapy is placebo is not justified.

Post-publication data showed which studies were regarded as of higher quality. We chose $n < 100$ as cut-off value for smaller studies, Shang chose $n < 98$ for homeopathy and $n < 146$ for conventional medicine (see below). There were 14 homeopathy studies of higher quality out of 83 trials (16.9%) with $n < 100$. There were two conventional studies of higher quality out of 78 trials (2.6%) with sample size <100.[19,20] The hypothesis that low quality small studies are therefore responsible for the positive findings in homeopathy is mostly falsified ($p = 0.003$, Fisher exact probability test). There is statistically significant difference in quality of smaller studies in favour of homeopathy.

Since the quality of conventional and homeopathic studies was not comparable, comparison of effects of the two methods was not valid. The underlying hypothesis for Shang et al's analysis was that results cannot be compared if quality is different. As so much emphasis was laid on the relation between quality and result we will nevertheless continue with our observations concerning this relation, although we did not compare effects of homeopathy and conventional medicine.

Comparability

The comparison of the two methods was somewhat flawed by publication bias. The 110 homeopathy trials were matched on indication with 110 conventional trials. But all conventional trials were published as journal articles while 16 (15%) of the trials in the homeopathy group were unpublished. According to Chan et al the odds of publishing results in conventional medicine are greater if results were significant (pooled odds ratio 2.4, 95% Confidence Interval [CI] 1.4–4.0).[21] So, in comparing effects homeopathy is disadvantaged by the selective inclusion of unpublished trials. But it also affects comparison of quality. Shang et al reported (in post-publication data) that none of the 16 unpublished homeopathy trials were of higher quality. The ratio of higher quality trials in published trials was 22% (21 out of 94) instead of 19% in the original paper.

We did not further investigate possible selection bias by excluding trials, but we were surprised by the exclusion of Wiesenauer's trial on chronic polyarthritis.[22] This was a larger trial ($n = 176$), of good quality according to Linde, with positive results.[2] This trial would have contributed positively to the outcome of the larger higher quality trials. Shang excluded this trial because no matching trial could be found.

Subgroups were selected on quality. This selection further influenced matching on indication, and therefore comparability. The homeopathy group contained 21 'higher quality' studies, the conventional group 9. At this point only 4 homeopathy studies were matched on indication by conventional studies (19%). From this point onward Shang's study consisted in fact of two incomparable meta-analyses of effects, one about homeopathy, one about conventional medicine.

Differences in effect between methods can no longer be evaluated if the matching is disrupted. This can be shown by comparing results for muscle soreness. The post-publication data show that neither homeopathy nor conventional medicine is effective for this indication, see Figure 1. But the homeopathy studies are of 'higher quality' while the conventional studies are not. This difference was of fundamental importance in the subset that led to the final conclusion.

The indication 'muscle soreness' has the largest influence on the results of homeopathy and on the comparison between homeopathy and conventional medicine because four homeopathy studies were classified as higher quality against none for conventional medicine. One of the homeopathy trials was also large and therefore higher in the funnel plot. This trial inclines the funnel plot to the right (towards OR = 1.0), while the smaller trials for this indication inclined it to the left because the pivot point is above these trials. There is a strong influence of chance in such a limited number of indications.

We did not consider clinical relevance, but one could wonder about the inclusion of treatments that may not be used because of serious adverse effects. Shang et al mentioned in the discussion that a limitation of their study was its disregard of adverse effects. They highly valued larger studies as a measure of quality and extrapolated effects towards the largest studies. This extrapolation is questionable if the largest studies involve treatments that are not available because of serious adverse effects. In a larger trial of higher quality on weight loss homeopathy had no effect.[23] The matched conventional study showed a considerable positive effect of Dexfenfluramine,[24] but Dexfenfluramine for weight loss was withdrawn by the American Food and Drug Administration in 1997 because of serious cardiac complications.[25] Two other larger studies, Deladumone (androgen–estrogen) in breastfeeding and Piroxicam for soft tissue injury suffered from the same problem.[26,27] These two treatments were also withdrawn because of adverse effects.[28,29] There might be other

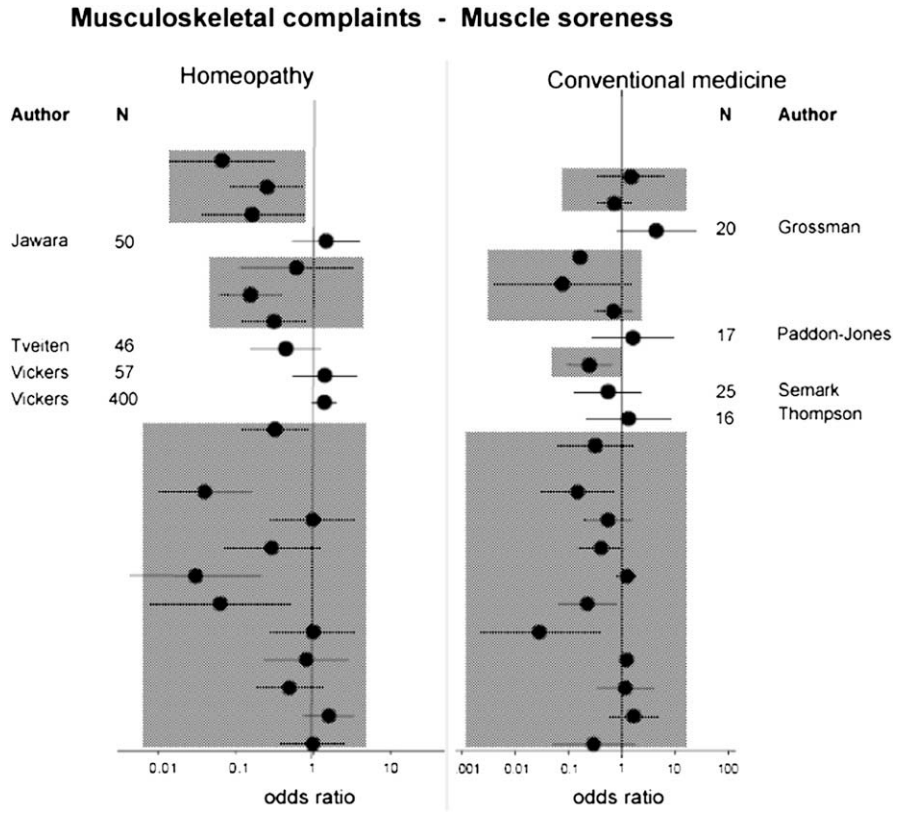

Figure 1. The effects of homeopathy and conventional medicine on 'muscle soreness' compared. The other trials in the group 'Musculoskeletal complaints' are disregarded. The four studies concerning muscle soreness for both methods are indicated by author names. *N* = trial size. Source www.ispm.ch.

treatments which are hard to compare because of safety, such as Tamoxifen for pre-menstrual syndrome.[30]

Possible bias in subgroup selection

Shang et al's conclusion was based on comparison of 'larger higher quality trials'. Possible pitfalls here are incomparability, heterogeneity and subjective criteria for quality and sample size.

The subjectivity of interpreting quality is demonstrated by differences between authors of meta-analyses. The following studies were not classified as good quality by Shang et al, although they are among the quality top 10 of Linde's meta-analysis[2]: de Lange-de Klerk,[31] Reilly 1986,[32] Hofmeyr,[33] Reilly 1994,[34] see Table 1.

If we add these four studies to Shang et al's pooled Odd Ratio (OR) of 25 trials becomes 0.74 (95% CI: 0.59–0.94). Why should these studies not be valued as of higher quality? Schultz et al showed that inadequate concealment of treatment allocation is the most important quality factor, associated with 41% (95%CI: 27–52%) exaggeration of effect.[35] Other quality factors have less effect; sequence generation 15%, (95%CI: 12–19%), double blinding 17% (95%CI: 4–29%). The Jadad score for quality used by Linde, does not consider allocation concealment. These 4 trials in Linde's meta-analysis had maximum Jadad scores, and as far as we can tell also had adequate allocation concealment. We cannot estimate the influence of effect of choices regarding quality because some new trials were published after Linde's study and some of Linde's lower quality trials were regarded as of higher quality. But, in combination with an unclear definition of 'larger' sample size, this subjectivity in defining quality opens a variety of possible subgroups which could be considered 'larger higher quality' trials.

Pooled odds ratio of all higher quality studies. We reconstructed the ORs and confidence intervals of the 21 higher quality studies selected by Shang et al. [36–55] The pooled OR using random effects analysis for all 21 higher quality studies in homeopathy is 0.76 (95% CI 0.59–0.99), which is not compatible with the placebo hypothesis.[18]

Cut-off value for sample size. Cut-off values for sample size were not mentioned or explained in Shang el al's analysis. Why were eight homeopathy trials compared with six conventional trials? Was this choice predefined or post-hoc? Post-publication data showed that cut-off values for larger higher quality studies differed between

Table 1. The four best studies according to Linde et al, arranged by sample size

First author	*Indication*	*Sample size*	*OR*	*95% CI of OR*
de Lange-de Klerk[31]	Upper respiratory tract infection	170	0.85	0.47–1.53
Reilly[32]	Pollinosis	144	0.43	0.22–0.85
Hofmeyr[33]	Childbirth	122	1.03	0.40–2.64
Reilly[34]	Asthma	24	0.08	0.02–0.40

the two groups. In the homeopathy group the cut-off value was $n = 98$, including eight trials (38% of the higher quality trials). The cut-off value for larger conventional studies in this analysis was $n = 146$, including six trials (66% of the higher quality trials). These cut-off values were considerably above the median sample size of 65. There were 31 homeopathy trials larger than the homeopathy cut-off value and 24 conventional trials larger than the conventional cut-off value. We can think of no criterion that could be common to the two cut-off values. This suggests that this choice was post-hoc.

Effect of larger higher quality trials. Shang et al decided that, based on this subset, homeopathy is a placebo response. The studies that constitute the evidence for the conclusion of the authors are listed in Table 2.

The two sets of trials are incomparable and heterogeneous with a pooled OR = 0.88 (95% CI: 0.65–1.19) for homeopathy. Only two homeopathy studies[38,40] could be matched with conventional studies.[56–58] The homeopathy group consisted of 8 trials on 8 different indications. Egger warned ''Opinions will often diverge on the correct method for performing a particular meta-analysis. The robustness of the findings to different assumptions should therefore always be examined in a thorough sensitivity analysis''.[59] Our sensitivity analysis showed that if Vickers's trial on muscle soreness is omitted from the eight largest higher quality homeopathy trials the overall odds ratio reduces from 0.88 to 0.80, but remains statistically not significant (95% CI: 0.61–1.05).

Ineffective unusual treatment

Sensitivity analysis of the higher quality studies showed one indication with four studies: homeopathic *Arnica* for muscle soreness after long distance running.[37,49,50,52] The pooled effect of those studies was in favour of placebo, OR = 1.30 (95% CI: 0.96–1.76). As treatment of healthy individuals is very rare in homeopathic practice this outcome has low external validity to judge the effect of homeopathy as a method. The fact that conventional medicine is also ineffective for this indication (see Figure 1) is omitted due to disrupting of matching on indication.

Table 2. Larger higher quality studies, according to Shang et al

Indication	*Homeopathy*	*Conventional medicine*
Diarrhoea	Jacobs.[40] *N* = 116	Kaplan.[56] *N* = 256
Treatment of influenza	Papp.[38] *N* = 334	Nicholson.[57] *N* = 319 de Flora.[38] *N* = 248
Prevention of influenza	Rottey.[36] *N* = 501	
Plantar warts	Labrecque.[39] *N* = 162	
Weight loss.	Schmidt.[23] *N* = 208	
Muscle soreness	Vickers.[37] *N* = 400	
Headaches.	Walach.[42] *N* = 98	
Sinusitis	Weiser.[41] *N* = 104	
Stroke (venous)		Horn. *N* = 454
Post operative infection		Crowley. *N* = 273
Pollinosis		Möller. *N* = 146

The final conclusion

The final conclusion that homeopathy is a placebo response (and conventional medicine is not) was flawed on several grounds:

1. Homeopathy and conventional trials were not comparable.
2. Heterogeneity disallows conclusions about effect.
3. Sensitivity analysis was missing.
4. The cut-off value for larger trials was decisive.

How essential were the missing data?

In the Box 1 we summarise the conclusions that could only be drawn from post-publication data. These data provided all answers to the questions we mentioned above except one: the fact that overall quality was better in homeopathy trials.

Another possible outcome

Often subjective choices must be made in meta-analyses. We evaluated the influence of some such choices in this case: the indication 'muscle soreness', the cut-off value and the interpretation of quality. We did not consider exclusion of trials, publication bias, quality bias or other possible bias. Table 3 and Figure 2 show the influence of cut-off values after exclusion of the trials on muscle soreness. In Table 3 some pooled OR and confidence intervals for 'larger trials' are given, with and without Linde's trials discarded by Shang et al. If we choose the overall median sample size ($n = 65$) as cut-off we disregard half of all trials. This seems a reasonable cut-off value for larger studies.

Figure 2 shows the cumulated-pooled OR (including the Linde trials omitted by Shang et al) if we increase step by step the number of included higher quality studies, starting with the largest two.

Adding the four 'Linde trials' does not change effects, but shifts the most unfavourable cut-off value from the 7th to the 10th trial. Discarding the indication 'muscle soreness' lowers the pooled OR from 0.88 to 0.80. Depending on the choice of cut-off value the OR varies between 0.72 and 0.80. Cochrane reviews are typically based on 8–10 studies [60] and are homogenous as to indication. Linde concluded that there was insufficient evidence for only single condition.[2] But Shang et al's 110 trials included 8 of homeopathy for acute upper respiratory tract infection, with no evidence of quality bias and a considerable effect size, OR = 0.36, 95% CI: 0.26–0.50. For muscle soreness after marathon running homeopathic *Arnica* is clearly not effective.

Box 1. Conclusions that could only follow from the subsequently disclosed data

- Comparison of quality and effect is flawed by inclusion of unpublished trials only for homeopathy. Restraining to published trials, quality was higher in 22% instead of 19% as mentioned by Shang.
- Judgement of quality is different from other analyses in at least four trials.
- The predefined hypothesis that positive results of homeopathy could be explained by quality bias in smaller trials was falsified ($p = 0.003$).
- If only higher quality trials are considered, the placebo hypothesis for homeopathy is falsified, OR = 0.76 (95% CI 0.59–0.99).
- The final conclusion was not based on comparative analysis, there was no matching on indication between homeopathy and conventional medicine.
- The conclusive subgroup analysis was not rectified because of heterogeneity, it considered 8 trials for 8 different indications.
- Cut-off values for larger trials were unexplainably different for homeopathy ($n = 98$) and conventional medicine ($n = 146$). This suggests post-hoc hypothesizing.
- Sensitivity analysis showed that one indication and the chosen cut-off value for larger trials explained the final conclusion of statistically non-significant effect.
- At least one larger higher quality homeopathy trial with positive result was excluded on unclear grounds.
- Comparative extrapolation of effects was questionable because of publication bias, selection bias, difference in quality and sample size and difference in safety.

Discussion

We found indications that Shang et al's hypothesis and hence its conclusion was sensitive to subjective choices and the influence of one indication and that the subsets on which the conclusions were based were not comparable. The missing data were of crucial importance, exploring these data seriously undermines the conclusion that homeopathy is a placebo response.

We calculated a number of possible pooled odds ratios for the effect of homeopathy as a method, excluding one indication for which homeopathy is ineffective, to show the isolated position of Shang et al's hypothesis. Taking all pooled odds ratios indicates an effect, but in some cases the confidence interval includes 1.0, depending on the definition of 'larger trial'. The quality of the whole set ($p = 0.03$) and quality in small studies ($p = 0.003$, Fisher exact test) are better in homeopathy than in conventional medicine. The placebo hypothesis is also falsified if only higher quality studies are considered. The comparison of homeopathy and conventional medicine was flawed by the inclusion of unpublished trials only in the homeopathy group and possibly by excluding trials. The conclusion of Shang et al is based on one subgroup of 8 trials on 8 different indications, not on a comparative analysis. Our sensitivity analysis showed one indication and a specific cut-off value play a decisive role in the final conclusion. Our addition of Linde's best trials did not alter Shang's overall results, but increased the number of (larger) higher quality trials, and the effect of the eight largest higher quality trials became significant (OR = 0.73; 95%CI: 0.59–0.91).

Small effects can be clinically relevant. In a meta-analysis of statin treatment and the occurrence of haemorrhagic stroke Vergouwen et al found an effect of OR = 0.88. This is the same OR as in Shang's final subset.[61] In the case of statins the 95% confidence interval is below 1.0 because the pooled sample size is large. Insufficient sample size and heterogeneity could easily lead to type II error (false negative) if OR = 0.88.

Shang et al compared homeopathy and conventional medicine in terms of quality and effect, but originally matched trials on indication. Quality and effect are

Table 3. The influence of cut-off values for 'larger studies', excluding 'muscle soreness', with or without Linde's best studies, using random effects analysis

Hypothesis	*Cut-off value*	*Outcome (95% CI)*	*P*
Largest higher quality trials,	6 trials, cut-off $n = 104$	OR = 0.73 (0.59–0.91)	0.0051
Shang's quality criteria,	7 trials, cut-off $n = 98$	OR = 0.80 (0.61–1.05)	0.1087
without muscle soreness	8 trials, cut-off $n = 81$	OR = 0.75 (0.58–0.96)	0.0246
	13 trials, cut-off overall median	OR = 0.66 (0.49–0.89)	0.0058
Largest higher quality trials,	8 trials, cut-off $n = 116$	OR = 0.73 (0.54–0.98)	0.0336
without muscle soreness, + 3	9 trials, cut-off $n = 104$	OR = 0.72 (0.56–0.94)	0.0158
of Linde's best quality studies	10 trials, cut-off $n = 98$	OR = 0.78 (0.59–1.03)	0.0776
	11 trials, cut-off $n = 81$	OR = 0.75 (0.57–0.98)	0.0340
	16 trials, cut-off overall median	OR = 0.78 (0.57–0.97)	0.0273

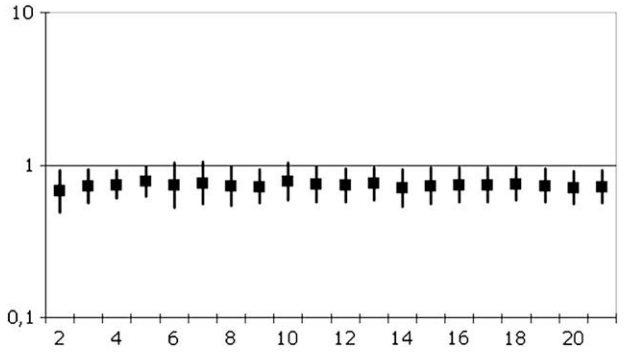

Figure 2. Indicates that all but 3 of 20 possible cut-off values lead to a significant effect if we consider all higher quality trials. The exceptions are $n = 144$ (including 6 trials), $n = 122$ (7 trials) and $n = 98$ (10 trials). If the cut-off point is below $n = 98$ the pooled results of higher quality studies is consistently statistically significant.

interrelated; better quality trials show less effect. Comparing effects when quality is not matched is thus questionable. They subsequently selected subgroups matched for quality, but that disrupts matching by indication. In the end the conclusion about the effect of homeopathy was based on meta-analysis of a selection of trials. Discarding trials for some indication because no comparable trials could be found causes selection bias. This selection could have had more influence on the final conclusion (positive or negative) than our re-analysis of 21 trials could detect. The insufficient matching on indication in the final subsets of 8 homeopathy and 6 conventional trials did not allow any comparison of effects. We assumed matching on indication when we tested the hypothesis of quality bias in smaller trials. The subgroups of 78 (homeopathy) and 83 (conventional medicine) smaller trials are largely but not fully (82%) matched on indication.

Shang et al made the choice to disregard safety. This decreased the relevance of the comparison of effects of homeopathy and conventional medicine. Some conventional treatments in this analysis are not available because of serious adverse effects.

We also performed meta-regression analysis on the 21 good quality trials and found asymmetry in the funnel plot. If we extrapolate the odds-ratios by meta-regression we see no difference between homeopathy and placebo at extreme sample numbers. We think that this is irrelevant to this discussion. Sterne, Egger and Smith stated that asymmetry in good quality trials is not caused by bias but by stronger effects in smaller trials.[8] This could be interpreted as proof that the asymmetry in the set of 110 homeopathy trials is not caused by bias. However, from mathematical statistics it is well known that such meta-regressions are imprecise, especially when the number of observations is small. Asymmetry of homeopathy trials and conventional trials cannot be compared, because there is a significant difference in the number of smaller good quality trials between homeopathy and conventional medicine. Different size in matched trials also plays a role in asymmetry of the funnel plot: both homeopathy and conventional medicine are ineffective for muscle soreness, but homeopathy is higher in the funnel plot because the trials are larger. The influence of this indication on asymmetry is opposite for homeopathy and conventional medicine. One could also question the role of drugs with strong effects but with serious adverse effects on asymmetry. Three conventional treatments, which have been withdrawn because of serious adverse effects, had large effect sizes and small standard errors and therefore considerable positive influence on the asymmetry and the extrapolated effect of the funnel plot of conventional medicine. Difference in publication bias has also influence on the position of the funnel plot.

We did not investigate the influence of subjective choices on the OR of conventional trials. We think that there are methodological objections against comparing effects in this analysis. Moreover, the clinical relevance of such a comparison is low. Homeopathy is mostly used after conventional medicine failed, so the indication for use is different. Homeopathy is highly valued for its safety. The scientific relevance of Shang's comparative analysis lies in the comparison of quality. Quality of trials was an important issue in the discussion about proof and implausibility.

The fact that the pooled effect of homeopathy excluding the indication 'muscle soreness' is positive does not mean that homeopathy is effective for all other indications. If the trials on influenza or Jacobs' trials on diarrhoea are excluded results become statistically not significant. As our re-analysis is post-hoc we cannot draw conclusions regarding efficacy.

The clinical relevance of trials is not considered in this analysis, but doctors must be interested in Shang's finding that eight trials showed a substantial effect of homeopathy in acute upper respiratory tract infections (OR = 0.36, 95% CI: 0.26–0.50), without indications of bias.

Conclusion

A review of data provided after publication of Shang et al's analysis did not support the conclusion that homeopathy is a placebo effect. There was intermingling of comparison of quality and comparison of effects, and thus matching was lost. The comparison of effects was also flawed by subjective choices and heterogeneity. The result in the subgroup from which the conclusion was drawn was further influenced by the choice of cut-off value for 'larger' trials. If we confine ourselves to the predefined hypotheses and the part of this analysis that is consistent with the comparative design, the only legitimate conclusion is that quality of homeopathy trials is better than of conventional trials, for all trials ($p = 0.03$) as well as for smaller trials with $n < 100$ ($p = 0.003$).

Acknowledgement

We thank Rainer Lüdtke for assisting with the statistics.

References

1 Kleijnen J, Knipschild P, ter Riet G. Clinical trials of homeopathy. *BMJ* 1991; **302**: 316–323.

2 Linde K, Clausius N, Ramirez G, *et al*. Are the clinical effects of homeopathy placebo effects? A meta-analysis of placebo-controlled trials. *Lancet* 1997; **350**: 834–843.
3 Cucherat M, Haugh M, Gooch M, Boissel J. Evidence of clinical efficacy of homeopathy - A meta-analysis of clinical trials. *Eur J Clin Pharmacol* 2000; **56**: 27–33.
4 Morrison B, Lilford RJ, Ernst E. Methodological rigour and results of clinical trials of homeopathic medicines. *Perfusion* 2000; **13**: 132–138.
5 Ernst E. A systematic review of systematic reviews of homeopathy. *J Clin Pharmacol* 2002; **54**: 577–582.
6 Ezzo J, Bausell B, Moerman DE, Berman B, Hadhazy V. Reviewing the reviews. How strong is the evidence? How clear are the conclusions? *Int J Technol Assess Health Care* 2001 Fall; **17**(4): 457–466.
7 Vandenbroucke JP. Medical journals and the shaping of medical knowledge. *Lancet* 1998; **352**: 2001–2006.
8 Sterne JAC, Egger M, Smith GD. Investigating and dealing with publication and other biases in meta-analysis. *BMJ* 2001; **323**: 101–105.
9 Shang A, Huwiler-Müntener K, Nartey L, *et al*. Are the clinical effects of homeopathy placebo effects? Comparative study of placebo-controlled trials of homeopathy and allopathy. *Lancet* 2005; **366**: 726–732.
10 Vandenbroucke JP. Homeopathy and the growth of truth. *Lancet* 2005; **366**: 691–692.
11 Higgins JPT, Green S (eds). Cochrane Handbook for Systematic Reviews of Interventions 4.2.5. *The Cochrane Library* 2005;(Issue 3). Chichester, UK: John Wiley & Sons, Ltd [updated May 2005].
12 Jonas WB, Kaptchuk TJ, Linde K. A critical overview of homeopathy. *Ann Intern Med* 2003; **138**: 393–399.
13 Egger M, Smith GD, Altman DG. *Systematic reviews in health care: meta-analysis in context*. 2nd edition. BMJ Books, 2001.
14 Boden WE. Meta-analysis in clinical trials reporting: has a tool become a weapon? *Am J Cardiol* 1992; **69**: 681–686.
15 Walach H, Jonas W, Lewith G. Letter to the editor. *Lancet* 2005; **366**: 2081.
16 Linde K, Jonas WB. Letter to the editor. *Lancet* 2005; **366**: 2081–2082.
17 Fisher P, Berman B, Davidson J, Reilly D, Thompson T, *et al*. Letter to the editor. *Lancet* 2005; **366**: 2082–2083.
18 Lüdtke R, Rutten ALB. The conclusions on the effectiveness of homeopathy highly depend on the set of analysed trials. *J Clin Epid* 2008; 10.1016/j.jclinepi.2008.06.015.
19 Gade J, Thorn P. Paraghurt for patients with irritable bowel syndrome: a controlled clinical investigation from general practice. *Scand J Prim Health Care* 1989; **7**: 23–26.
20 Humphrey RG, Bartfield MC, Carlan SJ, O'Brien WF, O'Leary TD, Triana T. Sulindac to prevent recurrent preterm labor: a randomized controlled trial. *Obstet Gynecol* 2001; **98**: 555–562.
21 Chan AW, Hjobartsson A, Haahr MT, Gotzsche PC, Altman DG. Emperical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004; **291**: 2457–2465.
22 Wiesenauer M, Gaus W. Wirksamkeitsnachweis eines Homöopathikums bei chronischer Polyarthritis. Eine randomisierte Doppelblindstudie bei niedergelassenen Ärzten. *Aktuelle Rheumatologie* 1991; **16**: 1–9.
23 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**(4): 197–206.
24 Enzi G, Crepaldi G, Inelmen EM, Bruni R, Baggio B. Efficacy and safety of dexfenfluramine in obese patients: amulticenter study. *Clin Neuropharmacology* 1995; **12**: S173–S178.
25 Available from: <http://www.fda.gov/CDER//news/phen/fenphenpr81597.htm>.
26 Louviere RL, Upton RT. Evaluation of Deladumone OB in the suppression of postpartum lactation. *Am J Obstet Gynecol* 1975; **121**: 641–642.
27 Lacey PH, Dodd GD, Shannon DJ. A double blind, placebo controlled study of piroxicam in the management of acute muscoloskeletal disorders. *Eur J Rheum Inflam* 1984; **7**: 95–104.
28 Available from: <http://www.fda.gov/ohrms/dockets/98fr/102998b.pdf>.
29 EMEA. *Press release*. London: European Medicines Agency Recommends Restricted Use for Piroxicam, 25 June 2007.
30 Grio R, Cellura A, Geranio R, Porpiglia M, Piacentino R. Efficacia clinica del tamoxifene nel trattamento della mastodynia premenstruale. *Min Ginecol* 1998; **50**: 101–103.
31 de Lange-de Klerk ESM. Effects of homeopathic medicines on children with recurrent upper respiratory tract infections. *BMJ* 1994; **309**: 1329–1332.
32 Reilly DT, Taylor MA, McSharry C, Aitchison T. Is homeopathy a placebo response? Controlled trial of homeopathic potency with pollen in hay fever as model. *Lancet* 1986;881–886.
33 Hofmeyr GJ, Picconi V, Blauhof P. Postpartum homeopathic Arnica montana: a potency-finding pilot study. *Br J Clin Pract* 1990; **44**: 619–621.
34 Reilly D, Taylor MA, Beattie NG, *et al*. Is evidence for homeopathy reproducible? *Lancet* 1994; **344**: 1601–1606.
35 Schulz KF, Chalmers I, Hayes RJ, Altman D. Emperical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; **273**: 408–412.
36 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**: 54–58.
37 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**(3): 227–231.
38 Papp R, Schuback G, Beck E, *et al*. Oscillococcinum in patients with influenza-like syndroms: A placebo controlled double-blind evaluation. *Brit Homeopath J* 1998; **87**(2): 69–76.
39 Labrecque M, Audet D, Latulippe L, Drouin J. Homeopathic treatment of plantar warts. *Can Med Assoc J* 1992; **146**(10): 1749–1753.
40 Jacobs J, Jiménez LM, Malthouse S, *et al*. Homeopathic Treatment of Acute Childhood Diarrhea - Results from a Clinical Trial in Nepal. *J Alternat Complement Med* 2000; **6**(2): 131–139.
41 Weiser M, Clasen BPE. Randomisierte plazebokontrollierte Doppelblindstudie zur Untersuchung der klinischen Wirksamkeit der homöopathischen Euphorbium compositum-Nasentropfen S bei chronischer Sinusitis. *Forsch Komplementarmed* 1994; **1**: 251–259.
42 Walach H, Haeusler W, Lowes T, *et al*. Classical homeopathic treatment of chronic headaches. *Cephalalgia* 1997; **17**: 119–126.
43 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**(5): 719–725.
44 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;177–183.
45 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 R Soc Med* 1997; **90**: 73–78.
46 Wiesenauer M, Häussler S, Gaus W. Pollinosis-Therapie mit Galphimia glauca. *MMW Fortschr Med* 1983; **101**(17): 811–814.
47 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.
48 Böhmer D, Ambrus P. Behandlung von Sportverletzungen mit Traumeel-Salbe - Kontrollierte Doppelblindstudie. *Biol Medizin* 1992; **21**(4): 260–268.

49 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**: 304–307.
50 Jawara N, Lewith GT, Vickers Aj, 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): 10–15.
51 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**: 521–542.
52 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. *Complement Ther Med* 1998; **6**(2): 71–74.
53 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**: 60–65.
54 Lepaisant C. Essai thérapeutique en homéopathie: traitement des tensions mammaires et mastodynies du syndrome prémenstruel. *Rev Fr Gynécol Obstét* 1995; **90**(2): 94–97.
55 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.
56 Kaplan MA, Prior MJ, McKonly KI, Du Pont HL, Temple AR, Nelson EB. A multicenter randomized controlled trial of a liquid loperamide product versus placebo in the treatment of acute diarrhea in children. *Clin Pediatr* 1999; **38**: 579–591.
57 Nicholson KG, Aoki FY, Osterhaus ADME, *et al*. Efficacy and safety of oseltamivir in treatment of acute influenza: a randomised controlled trial. *Lancet* 2000; **355**: 1845–1850.
58 de Flora S, Grassi C, Carati L. Attenuation of influenza-like symptomatology and improvement of cell-mediated immunity with long-term N-acetylcysteine treatment. *Eur Respir J* 1997; **10**: 1535–1541.
59 Egger M, Smith GD, Philips AN. Meta-analysis: principles and procedures. *BMJ* 1997; **315**: 1533–1537.
60 Shang A, Huwiler-Muntener K, Martey L, Jüni P, Dörig S, Sterne JAC, *et al*. Author's reply. *Lancet* 2005; **366**: 2083–2085.
61 Vergouwen MD, de Haan RJ, Vermeulen M, Roos YB. Statin treatment of hemorrhagic stroke in patients with a history of cerebrovascular disease. *Stroke* 2008; **39**: 497–502.