

Do Certain Countries Produce Only Positive Results? A Systematic Review of Controlled Trials

Andrew Vickers, Niraj Goyal, Robert Harland, and
Rebecca Rees

Research Council for Complementary Medicine, London, UK (A.V., R.R.); and Queen
Mary & Westfield College, London, UK (N.G., R.H.)

ABSTRACT: Objective: To determine whether clinical trials originating in certain countries always have positive results. Data sources: Abstracts of trials from Medline (January 1966–June 1995). Study selection: Two separate studies were conducted. The first included trials in which the clinical outcome of a group of subjects receiving acupuncture was compared to that of a group receiving placebo, no treatment, or a nonacupuncture intervention. In the second study, randomized or controlled trials of interventions other than acupuncture that were published in China, Japan, Russia/USSR, or Taiwan were compared to those published in England. Data extraction: Blinded reviewers determined inclusion and outcome and separately classified each trial by country of origin. Data synthesis: In the study of acupuncture trials, 252 of 1085 abstracts met the inclusion criteria. Research conducted in certain countries was uniformly favorable to acupuncture; all trials originating in China, Japan, Hong Kong, and Taiwan were positive, as were 10 out of 11 of those published in Russia/USSR. In studies that examined interventions other than acupuncture, 405 of 1100 abstracts met the inclusion criteria. Of trials published in England, 75% gave the test treatment as superior to control. The results for China, Japan, Russia/USSR, and Taiwan were 99%, 89%, 97%, and 95%, respectively. No trial published in China or Russia/USSR found a test treatment to be ineffective. Conclusions: Some countries publish unusually high proportions of positive results. Publication bias is a possible explanation. Researchers undertaking systematic reviews should consider carefully how to manage data from these countries. *Controlled Clin Trials* 1998;19:159–166 © Elsevier Science Inc. 1998

KEY WORDS: *Randomized controlled trial, publishing, China, Russia, Taiwan, Japan, acupuncture, Medline, meta-analysis, periodicals*

INTRODUCTION

The Cochrane Collaboration, an international network of individuals committed to preparing, maintaining, and disseminating systematic reviews of

Address reprint requests to: Andrew Vickers, Research Council for Complementary Medicine, 60 Great Ormond Street, London WC1N 3JF, UK.

Received July 31, 1996; revised September 15, 1997; accepted October 20, 1997.

clinical trials [1], has recently established a complementary medicine “field.” This group of researchers has various functions connected with supporting those who undertake Cochrane reviews, such as developing a register of randomized controlled trials, encouraging reviewers who are examining particular health problems (such as back pain) to undertake reviews of trials of complementary therapies, and offering methodologic advice where appropriate.

Because the field wants to promote systematic reviews of acupuncture, we are interested in obtaining data from China; however, researchers have suggested that, for reasons unknown, clinical trials from China always find acupuncture superior to the control intervention. Since this finding would have implications for systematic reviews, we wanted to learn whether it was true. We analyzed the results of trials referenced on Medline as a preliminary means of investigating whether clinical trials originating in any particular country had unusually high proportions of positive results. We are unaware of any similar published study.

METHODS

The study comprised two separate searches and analyses. First, we retrieved and analyzed a set of trials on acupuncture that had been published worldwide. To test hypotheses resulting from this study, we then retrieved and analyzed clinical trials of interventions other than acupuncture that had been published in five specified countries.

The first search, made on Medline (January 1966–May 1995), used “exp acupuncture” as a focus term and “human” and “abstract” as limits. This search strategy retrieves those papers indexed as investigating the use of acupuncture for which abstracts are available. The second search, made on Medline (January 1991–June 1995), retrieved all papers classed as “randomized controlled trial” or “controlled clinical trial” in the “publication type” field and published in China, Taiwan, Japan, or Russia/USSR, some of the countries that published a high proportion of positive results in our first study. We chose the cut-off year of 1991 because at the time of our study, the publication type field did not provide reliable data for studies published earlier. The 330 most recent randomized or controlled clinical trials published in England were retrieved as a comparison group. The figure of 330 was the greatest number of trials we expected to find published in our test countries.

Selection and data extraction of trials were made solely on the basis of their abstracts. The reviewers (N. G. and R. H.) were blinded to any data that might have provided information about the origin of a study, such as authors’ names or affiliations, in order to minimize bias. In the second study, the reviewers were not only blinded but were also deliberately misled as to the possible origin of the abstracts by the suggestion that the studies came from a large number of different countries instead of from just five. It was not always possible to maintain blinding, as the abstracts from particular countries used characteristic words and sentence structures. In addition, a few abstracts identified the site of research. We did not formally assess the degree of blinding.

Inclusion Criteria

The first study included papers in which the clinical outcome of a group of subjects receiving acupuncture was compared to that of a group receiving placebo, no treatment, or a nonacupuncture intervention. The second study included papers in which the clinical outcome of a group of subjects receiving a treatment other than acupuncture was compared to that of a group receiving a control intervention. Trials with healthy volunteers were excluded unless a symptom, for example, pain, was artificially induced. We also excluded trials that examined only the adverse effects of treatment. For trials that compared two or more treatments, we included the paper only if it was clear that at least one treatment provided the control condition. In order to determine the test treatment and control, we looked for words and phrases such as "active," "control," "placebo," "new regimen," "standard care," and, in the case of acupuncture, "superficial treatment" or "inappropriate point location." We included all trials with English abstracts, regardless of the language of the full publication.

Outcome

The outcome of each included trial was determined as follows. The trial was classed "test treatment superior to control" if the author made a clear statement to this effect; or, in the absence of such a statement, if at least one test intervention was described as statistically superior ($p \leq 0.05$) to the control intervention for at least one outcome measure; or, in the absence of inference data, at least one test intervention was described as superior to the control intervention (irrespective of the size of the difference) for at least one outcome measure. Trials involving more than one control group were classed as "test treatment superior to control" only if the treatment was superior to placebo/sham technique or in the absence of a placebo group, if treatment was superior to no intervention. All other trials were classed as "test treatment equal or inferior to control."

Country

In the first study, we determined the country in which the research was conducted by examining the "institution" field of Medline, which gives the affiliation of the first author. Where this information was unavailable or inconclusive; we obtained a copy of the paper. We assumed that research published in a language spoken only in one country (e.g., Bulgarian) was conducted in that country. In the second study, we used the Medline "country of publication" field, as we had found that trials published in one of the four test countries had always been conducted in that country.

The study organizers (A.V. and R.R.) conducted an unblinded review of a random sample of abstracts to check for errors.

Table 1 Results of Controlled Clinical Trials of Acupuncture by Country of Research

Country	Total Trials Analyzed	Favoring Test Treatment	
		Number	Percentage
USA	47	25	53
China	36	36	100
Sweden	27	16	59
UK	20	12	60
Denmark	16	8	50
Germany	16	10	63
Canada	11	3	27
Russia/USSR	11	10	91
Austria	9	8	89
Italy	9	8	89
Australia	6	1	17
France	6	5	83
Taiwan	6	6	100
Japan	5	5	100
Finland	4	2	50
Hong Kong	3	3	100
Netherlands	3	1	33
New Zealand	3	2	67
Poland	3	2	67
Switzerland	3	1	33
Bulgaria	2	2	100
Brazil	1	1	100
Croatia	1	1	100
Israel	1	1	100
Nigeria	1	1	100
Sri Lanka	1	0	0
Vietnam	1	1	100
Total	252	171	68

RESULTS

Study 1: Acupuncture

The initial Medline search located 1085 papers of which 252 were eligible for analysis. Overall, 171 trials (68%) were classed as “acupuncture superior to control” and 80 (32%) as “acupuncture equal or inferior to control.”

Acupuncture research was conducted in 27 different countries with the United States (47 trials), China (36 trials), Sweden (27 trials), and the United Kingdom (20 trials) contributing the greatest number of studies. Table 1 presents the raw outcome data for each country. Confidence intervals are not given because the size of the total population of trials is unknown. If, as seems likely, the trials reported on Medline comprise a significant fraction of the total number conducted, we would need to apply the finite population correction to the standard error. This correction cannot be calculated without good estimates of the size of the total population of trials ever conducted that would be eligible for this review.

Research conducted in certain countries, notably those in East Asia and eastern Europe, had particularly high proportions of results positive for acupuncture. For example, acupuncture was reported superior to control in all 36 trials conducted in China and in 10 of 11 trials conducted in Russia/USSR.

Table 2 Results of Controlled Clinical Trials of Interventions Other Than Acupuncture by Country

Country of Publication	Abstracts Screened	Abstracts Included	Favoring Test Treatment	
			Number	Percentage
China	196	109	108	99
England	329	107	80	75
Japan	317	120	107	89
Russia/USSR	180	29	28	97
Taiwan	78	40	38	95
Total	1100	405	361	89

Study 2: Interventions Other Than Acupuncture

A total of 1100 abstracts were downloaded from Medline; of these, 414 trials met the inclusion criteria for the study. Table 2 presents data on the number of trials screened and included and the outcome for trials published in different countries.

Trials published in England provided the comparison group. Of these trials, 75% reported the test treatment to be superior to control. In all four countries examined, results favoring the test treatment were published more frequently than in England. Particularly high rates of positive results were seen in China (99%) and Russia/USSR (97%). These two countries published no trials in which the test treatment was not reported effective. In the one study published in Russia/USSR classed as “test treatment equal or inferior to control,” nitrendipine was found to be equally effective as standard antihypertensives such as propranolol. Similarly, in the one Chinese trial that did not favor the test treatment, twice-weekly rifapentine was found superior, although not significantly so, to rifampicin given daily. Both drugs were said to be effective in more than 96% of cases. Even though the proportion of studies favoring the test treatment was higher in papers published in Taiwan (95%) and Japan (85%) than in those published in England, these countries did occasionally publish data showing that a treatment was not of benefit.

The results of randomized trials from study 1 (those described as such in the Medline “publication type” field or where mention of random allocation was made in the little or abstract) were combined with those from the second study. Table 3 presents these combined data.

Table 3 Results Combining Randomized Trials from Both Studies

Country of Publication	Abstracts Included	Favoring Test Treatment	
		Number	Percentage
China	121	120	99
England	118	88	75
Japan	120	107	89
Russia/USSR	29	28	97
Taiwan	45	43	96
Total	433	386	89

DISCUSSION

Medline-indexed clinical trials conducted or published in certain countries favor acupuncture over control much more frequently than those conducted and published elsewhere in the world. All 51 trials conducted in China and East Asia and listed on Medline in the period studied favored acupuncture, suggesting that clinical trials of acupuncture originating in certain countries have an unusually high proportion of positive results. One possible explanation for this finding is that acupuncture is more effective in countries where it is traditionally practiced. Perhaps, for example, Chinese doctors are more skilled at acupuncture or are better able to predict which patients may benefit.

The results of our second study, however, suggest that unusually high proportions of positive results are not restricted to acupuncture trials and that explanations in terms of a higher efficacy of acupuncture treatment do not hold. There are several possible/alternative explanations for our findings.

1. The sample of trials may not have been representative. Our search would not have located all published trials. We chose Medline because it provides a sufficiently representative set of the best-quality trials, particularly in acupuncture [2, 3], but trials are published in many journals not indexed by Medline. Possibly, negative trials originating from eastern Europe and Asia are found solely in non-Medline journals. We believe that this is unlikely; Medline might be expected to be a conservative source of information on unconventional therapies. Moreover, anecdotal evidence, such as the results presented at acupuncture conferences, does not suggest any considerable number of negative results published in non-Medline journals. For example, of the many hundreds of trials reported at the third World Conference on Acupuncture [4], we were unable to locate any studies originating in East Asia that showed acupuncture to be equal or inferior to a control procedure.

2. The abstracts may not have accurately reflected the results of trials. Pocock et al, for example, found that abstracts were much more likely to mention endpoints that showed statistical significance than those that did not [5]. The differences among countries are large. We find it unlikely that overenthusiastic reporting in abstracts could be solely responsible for such differences, especially because a positive abstract could not describe a clearly negative trial. Our figure from the second study of 75% of English trials favoring the test treatment is identical to that reported [6] for trials published in the *Lancet* and similar to that reported by Easterbrook [7] for research conducted in Oxford in the mid-1980s. This gives support to our methods.

3. Our judgements of whether the test treatment was superior to control were, in some cases, subjective. Where the authors of a paper failed to provide an explicit statement of their conclusions or a statistical analysis, we made a judgement of outcome that was, arguably, subjective. We reanalyzed the Chinese data from the second study and found that in only 5 of the 109 studies was such a judgment made. Excluding these studies would not appreciably change the results: 103 positive results from 104 trials is little different from 108 out of 109.

4. Trials conducted in certain countries may involve more outcome measures and "data dredging." Both of these maneuvers would increase the chance that the study met our criterion that the acupuncture intervention be statistically superior to the control intervention for at least one outcome measure. As a

check, we conducted an informal *post hoc* review of our data; however, we saw no evidence that trials from countries we found to produce exclusively positive results reported numerous outcome measures more often than did those from other countries. Moreover, we did not find appreciable numbers of papers with “mixed statistics,” where there were differences between some outcome measures but not others. Again, we believe it unlikely that a different approach to outcome measures could cause such large differences among groups. In particular, even if the use of multiple outcomes *inflates* the likelihood of a positive result, it still seems unlikely that *all* studies from a country would report positive findings for at least one outcome measure in the absence of some other factor.

5. Trials may have been conducted with different levels of methodologic rigor. Schulz et al [8] have observed that methodologically rigorous trials show smaller differences between experimental and control groups than do those conducted with less rigor. If trials from certain countries involved, say, insufficient blinding or inadequate concealment or randomization, this might explain, at least in part, the greater proportion of positive trials. If our results are due to low methodologic rigor, then the implications for systematic review are not as obvious as they might appear. A discerning reader can often, but not always, identify poor methodology. For example, the method of allocation concealment is unclear in the great majority of studies [8, 9]. Most reports state only that random assignment was used but do not give sufficient information to allow a judgment of whether or not it was conducted properly. Systematic reviewers need to consider the cost-effectiveness and feasibility of translating, say, a Chinese trial and then entering into a Chinese-language correspondence with the authors to clarify methodologic details.

6. Publication bias may be greater in some countries than in others. Publication bias is the tendency for individuals to submit or publish trials depending on the direction or strength of the findings. Dickersin [6], for example, found that clinical trials were much more likely to be published if there were statistically significant differences among treatment groups. If there were national variations in publication bias, one would expect that the overall proportion of positive trials would tend to be higher in countries with the greatest publication bias. If our results were indeed due to publication bias, there would be a number of implications for the science of systematic review. The usual method by which systematic reviewers circumvent publication bias is to undertake extensive searches for unpublished material. It is unclear whether this would be feasible in China and Russia/USSR. Furthermore, if the lack of negative findings is evidence that trials conducted in these countries are insufficiently rigorous, such extensive literature-searching may not be a good use of resources, particularly given the very high costs of translation. One practical solution might be for reviewers to include data from Chinese and Russian studies when they are readily available but to undertake sensitivity analyses to see whether the results of a review are robust to the exclusion of these trials. Reviewers might also use a variety of techniques to determine the existence of and adjust for publication bias. L'Abbé et al [10], for example, suggest that reviewers calculate the size of a hypothetical, unpublished, negative trial which, when added to the meta-analysis, would reduce the difference among groups so that it no longer reached statistical significance.

Alternatively, reviewers can calculate what number of small, unpublished, negative trials would change the conclusions of a review. A more sophisticated set of methods for identifying and estimating publication bias has been described by Begg [11, 12]. The problem with these methods is that they work best for large meta-analyses. Their power for interpreting the results of reviews with fewer than ten studies is questionable. Moreover, they can be used only where results of different studies are statistically combined, something that is not always justified in a systematic review. As such, in a review of, say, six trials, two of which are from China, a sensitivity analysis is likely to be more illuminating than an estimate of publication bias.

In a study published after we had finished data collection, Moher and colleagues compared completeness of reporting for randomized trials published in English to those published in French, German, Spanish, and Italian [9]. Although they interpreted their results as providing "evidence for inclusion of all trial reports, irrespective of the language in which they are published," they also stated that "we do not know the extent to which [our] results would be true for other languages (e.g., Russian, Japanese, and Chinese)." Our findings give further weight to their call for others "to replicate [our] study for such languages."

In conclusion, we have shown that Medline-indexed published clinical trials conducted in China and Russia/USSR almost never report an experimental treatment to be equal or inferior to control. Researchers undertaking systematic literature reviews should carefully consider how to manage data from these countries.

REFERENCES

1. Bero L, Rennie D. The Cochrane Collaboration. *JAMA* 1995;274:1935-1938.
2. Kleijnen J, Ter Riet G, Knipschild P. Acupuncture and asthma: a review of controlled trials. *Thorax* 1991;46:799-802.
3. Ter Riet G, Kleijnen J, Knipschild P. Acupuncture and chronic pain: a criteria-based meta-analysis. *J Clin Epidemiol* 1990;43:1191-1199.
4. World Federation of Acupuncture-Moxibustion Societies. The Third World Conference on Acupuncture. Kyoto, Japan; 20-23 November 1993.
5. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. *N Eng Med* 1987;317:426-432.
6. Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990;263:1385-1389.
7. Easterbrook PJ, Berlin JA, Gopalan R, et al. Publication bias in clinical research. *Lancet* 1991;337:867-872.
8. Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias. *JAMA* 1995; 273:408-412.
9. Moher D, Fortin R, Jadad AR, et al. Completeness of reporting of trials published in languages other than English. *Lancet* 1996;347:363-366.
10. L'Abbé KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. *Ann Intern Med* 1987;107:224-233.
11. Begg CB. A measure to aid in the interpretation of published clinical trials. *Stat Med* 1985;4:1-9.
12. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994;50:1088-1101.