Evidence from RCTs on optimal acupuncture treatment for knee osteoarthritis – an exploratory review

Jorge Vas, Adrian White

Abstract
There are many differing opinions on what constitutes an optimal acupuncture dose for treating any particular patient with any particular condition, and only direct comparisons of different methods in a clinical trial will provide information on which reliable decisions can be made. This article reviews the recent research into acupuncture treatment for osteoarthritis of the knee. It is speculated that optimal results from acupuncture treatment for osteoarthritis of the knee may involve: climatic factors, particularly high temperature; high expectations of patients; minimum of four needles; electroacupuncture rather than manual acupuncture; and a course of at least 10 treatments. These factors offer some support to criteria for adequate acupuncture used in the recent review. In addition, ethnic and cultural factors may influence patients’ reporting of their symptoms, and different versions of an outcome measure are likely to differ in their sensitivity – both factors which may lead to apparent rather than real differences between studies. The many variables in a study are likely to be more tightly controlled in a single centre study than in multicentre studies.

Keywords
Optimal acupuncture, adequacy of treatment, electroacupuncture, RCT.

Introduction
Acupuncturists develop their skills and knowledge from listening to other practitioners and by reflecting on their own experience. They interpret what they observe according to their own particular beliefs about the various explanations for the effects of acupuncture. This all leads to a great variety of opinions on the optimal way to treat any particular condition, or to modify that treatment for an individual patient. We do not yet have high level evidence – by which we mean direct head to head comparisons of different treatments – on which we can be sure of the optimal acupuncture treatment. One example of this method that has been published includes a comparison of different points, different durations of needling and different frequencies of stimulation in the treatment of fibromyalgia. Two indirect methods that have been used to try to establish the optimal treatment are consensus formation and surveys of practice. A consensus of senior, experienced practitioners has been achieved by informal methods, or by a more formal Delphi-type procedure. Information on surveys of acupuncturists’ actual practice could also help, but few have been published. These methods generally do not take account of dissenting opinions, and are typically only produced for the rather narrow demands of clinical trials.

Another potential source of information about optimal acupuncture is to compare the effects of different treatments used in a range of trials. One previous example was in a systematic review of acupuncture for chronic back pain. Trials that used six or more treatments were more likely to be
positive. At the same time, the reviewers could not demonstrate that trials using at least six points, or eliciting de qi, or giving individualised treatments were more likely to be positive.

Treatment with acupuncture is a complex business involving physical stimulation of the nervous system, as well as expectation and conditioning, and there are very many factors that can lead to real or apparent differences in the outcome of treatment. These factors could be considered under the headings of: 1) setting of the study, such as the reputation of the clinic; 2) patients’ characteristics, such as severity of illness; 3) features of trial design, such as number of practitioners; 4) details of the interventions, particularly the selection of points and the amount of stimulation given; 5) details of control interventions, such as whether the sham needle penetrates skin or not; and 6) choice of outcome measures, including their sensitivity and their appropriateness for the particular patients in the studies.

In theory, the best evidence might come from examining the actual ‘specific’ treatment effects – ie the difference between the two arms in a placebo controlled trial. However, this is not possible in acupuncture since there is no inactive placebo for the acupuncture needle. Even sham needling that appears to be similar in different studies may actually be performed in subtly different ways, and so have different effects.

The publication of four high quality, sham controlled RCTs of acupuncture for osteoarthritis of the knee, 2,3,7,8 and a subsequent systematic review of the data from these and other trials, 9 have provided two opportunities to ask whether anything useful can be learned about the best treatment techniques for this particular condition. The discussion between the authors of this work has focused on two particular questions: why did one particular study find a much greater effect than the other three; 8 and does the whole data set provide any indication of the optimal acupuncture treatment? We shall consider these questions in turn in this article, but it is important that readers are aware that this kind of exercise is very speculative and cannot lead to firm conclusions for clinical practice.

Large treatment effects in the study by Vas et al

Table 1 sets out the essential trial characteristics of the four recent sham controlled studies by Berman et al, 7 Scharf et al, 3 Vas et al, 8 and Witt et al, 2 which we shall now refer to by the name of the first author only. All studies met the criteria for high quality in the review (all at least 6 points on a scale of 9). Figure 1 sets out the actual mean WOMAC (Western Ontario and McMaster Universities Osteoarthritis Index) pain scores in these trials, before treatment and at short term follow up, nearest to 12 weeks. This short term follow up took place after the completion of treatment for all studies except Berman, in which patients continued to receive treatment monthly for one year. Figure 2 gives the same information, but this time adds the changes in the sham acupuncture groups and shows the changes in the scores (pain reduction). Both figures ignore any other comparison groups, for example education or usual care.

It is obvious that patients in the study by Vas had the highest WOMAC mean pain score before acupuncture, and the lowest afterwards: they experienced a remarkable reduction of 10.7 points, or 86% reduction in pain score. These changes were

**Table 1** Characteristics of four recent sham controlled RCTs

<table>
<thead>
<tr>
<th>Reference</th>
<th>Setting; recruitment</th>
<th>Mean age (y)</th>
<th>Acupuncture group (number of sessions)</th>
<th>Control group Intervention</th>
<th>n=</th>
<th>Control group Intervention</th>
<th>n=</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berman et al 2004</td>
<td>OP clinic, USA; adverts</td>
<td>65.5</td>
<td>EA (23)</td>
<td>true sham acupuncture education groups</td>
<td>190</td>
<td>191</td>
<td></td>
</tr>
<tr>
<td>Scharf et al 2006</td>
<td>GP, Germany; self referral</td>
<td>62.8</td>
<td>MA (12.5) + physical therapy + NSAID</td>
<td>off-point superficial acupuncture conservative</td>
<td>330</td>
<td>365</td>
<td>342</td>
</tr>
<tr>
<td>Vas et al 2004</td>
<td>GP pain clinic, Spain; referrals</td>
<td>67.0</td>
<td>EA + diclofenac (12)</td>
<td>true sham acupuncture + diclofenac</td>
<td>48</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>Witt et al 2005</td>
<td>GP, Germany; adverts</td>
<td>64</td>
<td>MA (12)</td>
<td>off-point superficial acupuncture waiting list</td>
<td>150</td>
<td>76</td>
<td>74</td>
</tr>
</tbody>
</table>

Abbreviations: EA = electroacupuncture; MA = manual acupuncture; NSAID = non-steroidal anti-inflammatory; OP = out patients
so much larger than the other studies that even their 95% confidence intervals did not overlap. Although we do not show data for the functional subscale of the WOMAC, or the total WOMAC scores, the changes are essentially similar to those of the pain scale.

There are several possible factors, under the headings listed above, that we need to consider to try to explain this difference.

**Trial setting**

In the Berman study in Maryland, the climate is variable, with mild summers (26°C) and very cold winters. The climate in New York is humid, continental type. Most of the state receives abundant precipitations in the form of snow. The Schiff and Witt studies were conducted in Germany, where winters can be extremely cold, and lakes and canals

---

**Figure 1** The WOMAC pain scores at baseline and after about 12 weeks are shown, in the acupuncture groups of four recent sham-controlled RCTs of acupuncture for knee osteoarthritis.

**Figure 2** The reductions in WOMAC pain scores are shown from the treatment and control arms of four recent sham-controlled RCTs.
commonly freeze. Mean temperatures range from 2ºC in winter to 24ºC in summer. In Seville (Spain), in contrast, where the Vas study was set, the mean annual temperature is 18ºC. Winters are mild and warm, while summers are very hot with maximum temperatures exceeding 40ºC.

The climate is known to have an effect on patients with rheumatological conditions, and studies have related the appearance of osteoarthritis and the aggravation of symptoms to the presence of environmental cold and damp.\(^{10,11}\) Also, it is interesting to note that, according to traditional Chinese medicine, the pathogenesis of osteoarthritis of the knee consists of the blockage of the circulation in the meridians by pathogenic factors such as Cold and Damp, which become established in the tissues.\(^{12}\)

The classical approach in China is to recommend heat treatment (moxibustion) in cold, northerly areas and therapy with acupuncture needles in hotter, more southerly latitudes. It is interesting to note that the best responses in these studies were reported by patients treated in the warm climate areas of Seville.

**Patient characteristics**

As the clinic used in the Vas study receives patients referred from healthcare centres throughout the province and belongs to a Public Health System that generally attains a high index of satisfaction, its credibility and patients' expectations, though not specifically measured, can be assumed to be high. This effect may have been the same for experimental and control groups, but it is also possible that high expectation facilitates treatment by priming the central nervous system in some way.\(^{13,14}\) The expectations of the settings in the other studies are not known to us.

A study in patients with rheumatoid arthritis has shown cultural differences and possible ethnic influences on patients' assessment of their symptoms in clinical trials which 'may be important in interpreting differences in prognosis and outcomes of patients'.\(^{15}\) Hispanic patients scored higher symptoms that Caucasian or African American patients with objectively similar disease states.

**Trial design: single centre vs multicentre study**

The Vas study was conducted at a single centre, the Berman study at three centres, the Witt study at 28 and the Scharf study at 315. A multicentre environment is associated with greater external validity ie the degree to which the results can be generalised to the whole population. In the case of Vas, internal validity was greater but at the cost of reducing external validity, so caution is necessary in generalising the results achieved. The high internal validity would be a factor of the consistency that can be provided by a single practitioner: for example, consistency in decisions on enrolment, in the therapeutic relationship, in the details of treatment technique for acupuncture and sham acupuncture. The same consistency cannot be expected in the 315 centres that treated 1007 patients (average about three per centre) in the study by Scharf, for example, and it would not be surprising if the real and sham acupuncture were not delivered exactly as described in the protocols. Some evidence of problems with performance is provided by the fact that 33 patients were accidentally unmasked.

**Intervention 1: the experience of the acupuncture practitioner**

Some authors suggest that the acupuncturist's experience is an important component of the effect of acupuncture treatment.\(^{16}\) Objective information on an acupuncturist's skill would be difficult to determine, but might include hours of training, years of practice, and number of patients seen each week. It would be interesting to be able to analyse the treatment responses in relation to the experience of the doctor who applies the true acupuncture technique, and to the number of patients recruited for the study and subsequently treated. Moreover, acupuncture at extra points *Wai* and *Nei Xiyan* requires a special degree of skill; this needling is normally recommended to be carried out to a depth of 2.5cm at least, aiming towards *Weizhong* BL40, for optimal effect. However, among patients with osteoarthritis of the knee, this technique, if not performed correctly, may damage the articular surface of the joint, aggravating pain and stiffness and worsening the articular functioning of the knee, as well as risking haemarthrosis and joint infection. For this reason, this technique is not recommended by the British Medical Acupuncture Society.

**Intervention 2: stimulation applied**

Two studies (Vas and Berman) applied electroacupuncture (EA) to local points. Vas used EA to all local points except ST36, at the highest tolerable
intensity. Berman applied EA to Xiyan only, ie not to muscle points, and the intensity is not described. Witt used manual manipulation once per session, and Scharf, twice per session. It is therefore possible that strong EA to muscle points could be necessary for optimal acupuncture for knee osteoarthritis.

Control intervention
The studies by Berman and Vas used non-penetrating needles near the knee in the control group, whereas both Witt and Scharf studies needled skin and muscle in segments that are the same as the innervation of the knee.

Different formats of the WOMAC index
Three scale formats can be used for the WOMAC Index: five point Likert scale (LKS), 100mm visual analogue scale (VAS), and 11 point numerical rating (NRS). The VAS is classified as a continuous measure, and the other two as categorical measures. The Likert and the 100mm VAS are the most commonly used. The Likert format is much simpler than the VAS format but is less sensitive. In the pain scale, for each single step between the categories of the Likert format, there are 20 possible options in the VAS format. In addition, the non-parametric analysis generally used for categorical data is less sensitive than analyses for continuous data.

Two studies (Berman and Vas) used the LKS format, and the others (Witt and Scharf) used the VAS or NRS formats. Vas and colleagues decided upon the LKS format because, in a pilot study, it was found to be more appropriate for the characteristics of their population, which had a low educational level – over half of the patients had only completed primary education. We do not possess corresponding data for the other three studies.

The different scale formats could lead us to suspect a possible bias associated with the measuring instrument, with over- or under-estimation of the effect, a bias that could be differential if it occurred with greater significance in one of the comparison groups.

Exploring all available studies for general trends
In this section, we wish to cautiously explore whether the data provided by all the studies in the recent systematic review give any indication of what could be optimal acupuncture. The treatment details of all 13 included studies, originally extracted in duplicate for the review, are presented in Table 2. In Figure 3 we show simply the percentage reduction of pain at close to 12 weeks in the acupuncture treatment group. Two studies (Tukmachi et al, Sangdee et al) used acupuncture in two arms of their studies, and we present an average of the two scores.

It made no sense to consider the comparative treatment effects (acupuncture versus control) because of the different controls used. This analysis is highly speculative, therefore, because the different trial settings would be likely to influence the response to treatment. Also, the varied quality of studies limits the robustness of the data.

Overall, the average size of response is about 40% of baseline. The results of Vas and Yurtkuran are notably higher than this, and those of Ng et al, Petrou et al, and Takeda et al, are somewhat lower.” The results of the Yurtkuran study look impressive (96% improvement in the acupuncture group) but the OA was mild and the outcome measure insensitive, with a mean baseline score of only 1.36 on a categorical scale of 0 (no pain) to 5 (excruciating), so it is best not to give too much weight to the results of this single study.

Number of points needled: most trials used at least 6 points, except Takeda (5), Sangdee (4), Yurtkuran (4) and Ng (2). The correlation between the number of points and the size of the effect could be interpreted as a threshold effect: as long as at least four points are used, there may not be much benefit from adding more. Point selection: the data do not support the idea that needling distant points, either in the feet or the hands, is likely to improve the patient’s response.

Eliciting de qi: all studies reported obtaining de qi except Sangdee and Yurtkuran, and the good response in these two studies could be due to the use of electroacupuncture. Interestingly, the two studies of manual acupuncture with the lowest response – Petrou and Takeda – both give good descriptions of careful technique to elicit de qi.

Stimulation: there seems to be some trend supporting the view that electroacupuncture has more effect than manual stimulation. Of the seven studies using EA, all except Ng achieved a response above 40%, and Ng’s small effect may be explained by only using two needles. Of the six studies using manual acupuncture, only half showed a greater than 40% response.

Duration of treatment: it is not possible to conclude from these data that the needles should be
left in for any longer than 20 minutes, since the three studies which did this (Scharf, Tukmachi and Witt) did not have above average response.

Treatment schedule: all trials gave acupuncture at least twice weekly at the outset, except Vas who gave it once weekly and had the greatest response. The number of treatments given in the course ranged from 6 (Christensen) to 16 (Berman, both trials).

It is difficult to draw any conclusions, though the best four trials (Sangdee, Vas, Witt and Yurtkuran) gave at least 10 sessions. In normal practice, six sessions of acupuncture are often offered, and the one study that did this (Christensen) had a slightly below average result – but used manual stimulation.

Discussion

In summary, this speculative comparison of RCTs of acupuncture for osteoarthritis of the knee suggests that the following factors might contribute to optimal results from acupuncture treatment: the prevailing climatic conditions, particularly high temperature; high expectations of patients; minimum of four needles; EA rather than manual acupuncture, and particularly, strong EA to needles placed in muscle; and a course of at least 10 treatments. (Deep needling of points Wai and Nei Xiyan may be another factor, but is not recommended in Western practice.) In addition, ethnic and cultural factors may influence patients’ reporting of their symptoms, and different versions of an outcome measure are likely to differ in their sensitivity – both factors which may lead to apparent rather than real differences between studies. The many variables in a study are likely to be more tightly controlled in single centre than in multicentre studies.

No suggestions can be made about the need to elicit de qi, or about the frequency of treatment, since the studies were similar in these respects.

This information offers some support for the five criteria for the ‘adequacy’ of acupuncture that were used to decide which studies should be included in the meta-analysis within the recent systematic review:

1. At least six treatments, at least one per week, with at least four points needled for each painful knee for at least 20 minutes, and either needle sensation (de qi) achieved in manual acupuncture, or electrical stimulation of sufficient intensity to produce more than minimal sensation.

Advances in research design and methodology, in the measurement instruments available and in data processing techniques and presentation have given rise to an unprecedented amount of knowledge acquisition. In consequence, the volume and rate of appearance of published research are ever greater, thus generating a huge quantity of available data.

Table 2: Treatment details of 13 RCTs included in a recent systematic review

<table>
<thead>
<tr>
<th>Reference</th>
<th>needles n=</th>
<th>de qi elicited</th>
<th>points used</th>
<th>duration (minutes)</th>
<th>sessions per week</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berman et al 1999</td>
<td>9</td>
<td>y</td>
<td>ST35 EX32 ST36 SP9 GB34; BL60, GB39, SP6, KI3</td>
<td>20</td>
<td>16</td>
</tr>
<tr>
<td>Berman et al 2004</td>
<td>9</td>
<td>y</td>
<td>ST35 EX32 ST36 SP9 GB34; BL60, GB39, SP6, KI3</td>
<td>20</td>
<td>16</td>
</tr>
<tr>
<td>Christensen et al 1992</td>
<td>6</td>
<td>y</td>
<td>ST34 SP10 ST35 EX32 ST36; LI4</td>
<td>20</td>
<td>6</td>
</tr>
<tr>
<td>Molsberger et al 1994</td>
<td>9</td>
<td>y</td>
<td>ST34 SP10 EX31 ST35 EX32 ST36 SP9 GB34 BL40</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>Ng et al 2003</td>
<td>2</td>
<td>y</td>
<td>ST35 EX32</td>
<td>20</td>
<td>8</td>
</tr>
<tr>
<td>Petrou et al 1988</td>
<td>12</td>
<td>y</td>
<td>EX31, ST35 EX32 ST36, ST43 BL40 bilateral, L4</td>
<td>20</td>
<td>8</td>
</tr>
<tr>
<td>Sangdee et al 2002</td>
<td>4</td>
<td>n</td>
<td>ST35 EX32 joint line, LR8</td>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td>Scharf et al 2006</td>
<td>7 to 15</td>
<td>y</td>
<td>ST34, 36 Xiyan (both) SP9, GB34 + optional ah shi, 2 distant</td>
<td>20 to 30</td>
<td>12.5</td>
</tr>
<tr>
<td>Takeda &amp; Wessel 1994</td>
<td>5</td>
<td>y</td>
<td>EX31 ST35 EX32 GB34 SP9</td>
<td>30</td>
<td>9</td>
</tr>
<tr>
<td>Tukmachi et al 2004</td>
<td>9</td>
<td>MA, EA</td>
<td>a) ST36, LI4, LR3 (MA)</td>
<td>20 to 30</td>
<td>10</td>
</tr>
<tr>
<td>Vas et al 2004</td>
<td>9</td>
<td>EA</td>
<td>EX32 ST36 SP9 GB34; SP6 KI3 ST40, L4</td>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td>Witt et al 2005</td>
<td>8 to 20</td>
<td>MA</td>
<td>ST34 SP10 EX31 ST35 EX32 ST36 SP9 BL40 KI10</td>
<td>30</td>
<td>12</td>
</tr>
<tr>
<td>Yurtkuran &amp; Kocagil 1999</td>
<td>4</td>
<td>EA</td>
<td>ST34 ST35 GB34 SP9</td>
<td>20</td>
<td>10</td>
</tr>
</tbody>
</table>

*Total 23 treatments, but 16 by the time this measurement was taken.

Abbreviations: EA = electroacupuncture; MA = manual acupuncture; physio = physical therapy
The generalised use of clinical trials, the introduction of techniques of quantitative synthesis, especially systematic reviews of the literature and meta-analysis, all contribute to highlighting the current variability in clinical practice. This is particularly apparent in the case of studies aimed at assessing the efficacy or effectiveness of acupuncture. The variability identified in acupuncture treatments depends on many factors, including those inherent in the fact that the treatment is performed either in the context of a study or in standard clinical practice; those arising from the type of acupuncture that is applied; or on the empathy of the therapist, the credibility of the technique; or, the experience possessed by the person applying it. Therefore, systematic reviews should show us that the long term validity of knowledge decreases in direct proportion to its age. If we are not capable of identifying and addressing knowledge needs, our professional practice will degenerate, together with our competence.

This article highlights the lack of information we have about the optimal acupuncture treatment for knee pain, and the difficulty of obtaining that information reliably, except by direct, head to head comparisons. In addition, treatment recommendations for general use may not apply to every individual patient. Nevertheless, future studies should take account of the findings of these previous studies in deciding what form of acupuncture to use.

Reference list
Evidence from RCTs on optimal acupuncture treatment for knee osteoarthritis – an exploratory review

Jorge Vas and Adrian White

doi: 10.1136/aim.25.1-2.29

Updated information and services can be found at:

http://aim.bmj.com/content/25/1-2/29

**Email alerting service**

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Notes**

To request permissions go to:

http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:

http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:

http://group.bmj.com/subscribe/