Introduction

Clinical experience suggests that acupuncture reduces pain, but a causal relationship cannot be established by clinical experience alone. This is because clinical experience cannot determine whether pain relief was due to spontaneous improvement of the ailment, natural fluctuations in symptoms, the patient's expectation that acupuncture should work, or the patient exaggerating their report of the effect of acupuncture to please the therapist. Controlled clinical trials help to establish whether treatment effects are due to the therapeutic agent of the treatment or due to some other confounding factor. With so many clinical trials and systematic reviews of clinical trials of medical treatments, expert groups have emerged to interpret the research and provide simple statements about effectiveness. The purpose of this paper is twofold. First, to encapsulate the clinical research facing practitioners trying to establish the ability of acupuncture to relieve pain, and secondly, to critically review the processes used to establish one-sentence statements about effectiveness. However, it is important at the outset to demonstrate why it is imperative to attempt to establish clinical effectiveness and for practitioners to act on the outcome.

Fake treatments and their unintentional use

In 2004, Dr Copperfield, columnist in the British newspaper The Times, Body and Soul supplement wrote: ‘Most of us [GPs] realise that our job is to keep patients distracted while they get better for themselves.’ This ‘tongue in cheek’ comment raises an interesting ethical debate. One could in principle distract a patient through deception, by giving them a fake treatment that was known to have no therapeutic agent (active ingredient). This could be in the form a placebo pill or a sham procedure. It would ‘tick’ the patient into believing that they were receiving a treatment with a therapeutic agent when in fact they...
Education, practice and debate

It is imperative that attempts are made to establish clinical effectiveness to avert accusations that a treatment with unproven clinical effectiveness is inert. Interestingly, it has been shown that some practitioners continue to use their favoured treatments even if the bulk of scientific evidence suggests that the treatment is not clinically effective.

Criticalisms of acupuncture

Acupuncture has not been immune from criticism about its effectiveness. In the UK, the BMA approval of acupuncture was met with disdain in some quarters, with claims that 'The BMA report is quite simply wrong.' In the USA, the National Council Against Health Fraud (NCAHF) which is 'a private non-profit, voluntary health agency that focuses upon health misinformation, fraud, and quackery,' produced a position statement on acupuncture in 1990, which is still prominent on its web site today.

It states that:

'Acupuncture is an unproven modality of treatment; its theory and practice are based on primitive and fanciful concepts of health and disease that bear no relationship to present scientific knowledge; research during the past 20 years has failed to demonstrate that acupuncture is effective against any disease; perceived effects of acupuncture are probably due to a combination of expectation, suggestion, counter-irritation, operant conditioning, and other psychological mechanisms; the use of acupuncture should be restricted to appropriate research settings; insurance companies should not be required by law to cover acupuncture treatment; and licensure of lay acupuncturists should be phased out.'

NCAHF advises physicians that: 'acupuncture should not be offered without full informed consent, reminding patients that acupuncture is experimental, has not been proven more effective than a placebo, and has some risk of complications.' Acupuncturists would, quite rightly, argue that these emotive statements do not reflect current scientific and clinical knowledge on acupuncture. Clearly, there is a discrepancy in the interpretation of the 'knowledge', so it is necessary to assess how knowledge on clinical effectiveness is gathered.

Evidence-based medicine, clinical effectiveness and active ingredients

Evidence-based medicine (EBM) integrates clinical experience with systematic clinical research to produce a position statement on acupuncture in 1990, which is still prominent on its web site today. It states that:

'Acupuncture is an unproven modality of treatment; its theory and practice are based on primitive and fanciful concepts of health and disease that bear no relationship to present scientific knowledge; research during the past 20 years has failed to demonstrate that acupuncture is effective against any disease; perceived effects of acupuncture are probably due to a combination of expectation, suggestion, counter-irritation, operant conditioning, and other psychological mechanisms; the use of acupuncture should be restricted to appropriate research settings; insurance companies should not be required by law to cover acupuncture treatment; and licensure of lay acupuncturists should be phased out.'

NCAHF advises physicians that: 'acupuncture should not be offered without full informed consent, reminding patients that acupuncture is experimental, has not been proven more effective than a placebo, and has some risk of complications.' Acupuncturists would, quite rightly, argue that these emotive statements do not reflect current scientific and clinical knowledge on acupuncture. Clearly, there is a discrepancy in the interpretation of the 'knowledge', so it is necessary to assess how knowledge on clinical effectiveness is gathered.
determine best practice. A cornerstone of EBM is establishing clinical effectiveness and exposing the unintentional use of interventions with no therapeutic value. A treatment is clinically effective when beneficial effects can be attributed to its therapeutic agent and this is achieved by distinguishing the clinical effects associated with receiving the treatment from those associated with the therapeutic agent of the treatment. Therapeutic agents give treatments their uniqueness and identity, and for drugs it is often a unique chemical structure. For complex technique-based interventions underpinned by different philosophical principles the therapeutic agent may be difficult to define.

Determining the identity and effectiveness of acupuncture

The uniqueness and identity of acupuncture appears to be the process of inserting needles in the skin at chosen points (Latin *acu* ‘with a needle’, Latin *punctura*, from pungere ‘to prick’). The Oxford dictionary defines acupuncture as ‘a system of complementary medicine in which fine needles are inserted in the skin at specific points along supposed lines of energy.’ If needles are not ‘inserted’ or ‘specific points’ are not used, then, by strict definition, the treatment is not acupuncture (though a modern definition should obviously include specific points in accordance with known physiological or anatomical rationale).

Two key questions arise:
(i) Do you have to insert needles in the skin to produce a clinically meaningful effect?
(ii) Does it matter where in the skin these needles are inserted to produce a clinically meaningful effect?

If the answer to either of these questions is no, then it would start to bring into question the value of acupuncture, in its strictest definition, as a clinically effective treatment.

Absolute effectiveness

To establish absolute effectiveness for acupuncture it would be necessary to demonstrate that inserting needles in the skin at specific points produced clinically meaningful effects over and above everything except inserting needles in the skin at specific points, i.e. ‘pretending’ that you are inserting needles in the skin at specific points when in fact you are not. Pretend or bogus interventions (sham interventions) and substances with no therapeutic effect (placebo interventions) are used in clinical research to isolate the effects associated with the therapeutic agent of the treatment. Sham and placebo controls help to control for natural fluctuations in symptoms over time and the patients’ expectations about the treatment and its effects. This is important because we know that in medicine ‘Doing nothing causes things to happen.’ Sham acupuncture could include administering needles that do not pierce the skin or administering needles at non-specific acupuncture points. Ideally the sham acupuncture intervention should consist of both.

Relative effectiveness

In clinical practice where practitioners need to select from a range of potential treatments, information about relative effectiveness is more useful because it compares the effectiveness of treatments ‘head-to-head’. However, clinical trials that attempt to establish the relative effectiveness of acupuncture against other treatments often score low on methodological quality rating scales because it is notoriously difficult to blind trial participants to treatment groups. Despite the difficulties in designing and executing effectiveness trials on acupuncture the amount of published clinical research available overwhelms most practitioners.

Interpreting clinical research on acupuncture

A broad free text search on PubMed using the key word ‘acupuncture’ conducted on 7 March 2006, identified 10257 hits. Restricting the search to ‘acupuncture’ and ‘pain’ reduced to 2674 hits, of which 392 were clinical trials, 259 RCTs and 25 meta-analyses. It is impossible for practitioners to read and interpret such a volume of information so expert groups have emerged to interpret the research for them. These expert groups include Bandolier, The Cochrane Collaboration and the National Health Service Centre for Reviews and Dissemination (CRD).

Bandolier

Bandolier is an independent journal written by Oxford scientists and it claims to provide ‘Evidence based thinking about health care.’ Bandolier’s *Little Book of Pain* is a pocket sized evidence based guide
to treatments and uses ‘Clinical Bottom Lines’ to aid quick and easy decision making. The clinical bottom lines for acupuncture include:

- Acupuncture is ineffective for back pain (p365);
- ‘There is no convincing evidence demonstrating that acupuncture is more effective than placebo for the relief of back or neck pain’ (p366);
- ‘There is no evidence for efficacy of acupuncture in fibromyalgia’ (p366);
- ‘There is no evidence that acupuncture is more effective than sham/placebo acupuncture for the relief of joint pain due to osteoarthritis (OA)’ (p366).

In addition, the following Clinical Bottom Lines appear on Bandolier’s web site:

- ‘Based on a very small number of trials, there is no convincing evidence for the effectiveness of acupuncture in relieving dental pain.’
- ‘There is no evidence from high quality trials that acupuncture is effective for the treatment of migraine and other forms of headache.’
- ‘...bottom line is that the use of acupuncture for chronic pain is unsupported by any evidence of quality. Consumers and providers should beware.’

These statements should alarm acupuncturists because Bandolier, and its Little Book of Pain, has much influence within medical and allied health care professions.

Cochrane Collaboration

The Cochrane Collaboration Library claims to be ‘The reliable source of evidence in Health Care’. A search of the Cochrane database of systematic reviews on 7 March 2006 using the key word ‘acupuncture’ in all fields, yielded 172 hits (121 reviews and 51 protocols). Restricting the search to ‘acupuncture’ and ‘pain’ in all fields yielded 105 hits (76 reviews and 29 protocols), and further restriction to ‘acupuncture’ in record title and ‘pain’ in all fields yielded 19 hits (13 reviews and six protocols).

Of these 13 reviews, six examined pain outcomes. Two reviews concluded that acupuncture was effective, one for idiopathic headache, and one for chronic low back pain up to three months. One review found acupuncture to be ineffective for improving symptoms of rheumatoid arthritis. The remaining reviews reported insufficient evidence to make definitive conclusions for shoulder pain, lateral elbow pain, and primary dysmenorrhoea, although these three reviews found poor quality evidence hinting that acupuncture may be of benefit. The non-pain reviews captured by the search reported that acupuncture was effective for preventing nausea and vomiting, and ineffective for the treatment of cocaine dependence. There was insufficient evidence to make definitive conclusions for smoking cessation, induction of labour, symptoms of depression, acute stroke, and Bell’s palsy. However, this is not an exhaustive list of reviews on acupuncture for non-pain conditions as the search was restricted to pain in all fields.

Interestingly, one review of acupuncture and pain has changed in its conclusion over time, as new and better quality RCTs became available. A review on acupuncture for low back pain originally concluded that: ‘The evidence summarised in this systematic review does not indicate that acupuncture is effective for the treatment of back pain’ [Reviewers Conclusion]. An update of this review was more positive: ‘For chronic low-back pain, results show that acupuncture is more effective for pain relief than no treatment or sham treatment, in measurements taken up to three months’ [Plain Language Summary]. The conclusion of the update is consistent with a recently published meta-analysis on acupuncture for low back pain which was conducted by a different team of investigators. It is likely that this trend of ‘shifting knowledge’ will continue as better quality primary research is conducted.

National Health Service Centre for Reviews and Dissemination (CRD)

With so much information, the National Health Service Centre for Reviews and Dissemination (CRD) produced a Health Care Bulletin in 2001 that reviewed all available systematic reviews and meta-analyses. The CRD claims to ‘Promote(s) the use of research based knowledge in health care’, and it concluded that:

- ‘Acupuncture appears to be effective for postoperative nausea and vomiting in adults, chemotherapy related nausea and vomiting, and for postoperative dental pain.’
- ‘Current evidence suggests that acupuncture is unlikely to be of benefit for smoking cessation and tinnitus.’
- ‘For most other conditions the available evidence is insufficient to guide clinical decisions. The most..."
problematic area is chronic pain where there is a large body of data open to conflicting interpretations.’

The large body of data reviewed by the CRD consisted of 19 systematic reviews and/or meta-analyses. There were three reviews of chronic pain, 35-37 seven reviews of neck and back pain, 38-44 five reviews of rheumatic diseases, 45-49 and four reviews of headache. 50-53 When taken as a whole, the CRD concluded that acupuncture reduced pain when compared to no treatment or waiting list controls in most studies. 33 However, ‘RCTs comparing acupuncture to a sham technique were more evenly balanced between those that did and those that did not show statistically significant differences between groups.’ 33 This finding was consistent with an editorial in the Lancet a decade earlier: ‘...the better the study technique, the less likely that acupuncture surpasses the placebo effect.’

Expert opinion is also available from authors of numerous textbooks, chapters and tutorial reviews on acupuncture. In their book Acupuncture: a Scientific Appraisal, Ernst and White evaluated systematic reviews on acupuncture as conclusively positive for dental pain, low back pain and nausea and vomiting, and conclusively negative for smoking cessation and weight loss. 55 Evidence was inconclusive for experimental pain, neck pain, headache/migraine, osteoarthritis, inflammatory rheumatic diseases, stroke, addictions and asthma. Further critique of textbook authors is of limited value in this article because a clear picture is emerging. When compared to other treatments the number of systematic reviews on acupuncture is not only staggering, but more importantly their conclusions are conflicting, confusing and often confounded by methodological inadequacies in the primary research. However, some of the uncertainty about the evidence-base for acupuncture has also been due to a reliance on a review process which is itself open to bias.

Critically appraising interpretations of clinical research on acupuncture

Systematic reviews and meta-analyses use scientific methods in an attempt to remove human prejudice from the evaluation process. However, scientific method is grounded in probability theory so answers (conclusions) will have a degree of uncertainty. Moreover, human judgments are critical in the systematic review process, although the effect of this is often be ignored in the quest for single sentence conclusions, clinical bottom lines or plain language summaries that say more than ‘the results are inconclusive’.

Shortcomings of systematic reviews

A critical review of clinical research into acupuncture by White et al revealed shortcomings of the systematic review process which may distort review conclusions. 56 Imprecise inclusion criteria used by reviewers led to comparisons of heterogeneous study populations within the same review. Imprecise definitions of acupuncture allowed trials to be included in reviews which acupuncturists would not consider true acupuncture. For example, laser acupuncture, which has been included in systematic reviews, does not ‘puncture’ the skin in the same way as a needle and is unlikely to elicit the same physiological effects. Reviews also included control groups that were potentially ‘active’ and generated clinically meaningful outcomes in the patient groups. All of these shortcomings are likely to bias study outcome towards a negative outcome (ie no difference between acupuncture and control). Such methodological shortcomings need to be screened using robust inclusion and exclusion criteria, or accounted for in sub-group analysis, as they compromise review validity.

Discrepancies in vote counting

Linde and Willich highlighted differences in the outcomes of eight systematic reviews on the effectiveness of acupuncture for pain. 57 They revealed variations in the number of RCTs included in reviews on similar topics because of the use of different inclusion criteria. They also found that variations in systematic review methodology caused discrepancies in trial vote counting between similar reviews on acupuncture and low back pain. Using votes to summarise the outcome of a primary trial is common practice when trial data are heterogeneous and cannot be pooled for statistical analysis. However, vote counting requires interpretation of trial data by the reviewers and is open to abuse, especially when trials use multiple outcome measures presenting with different results. Judgements about what constitutes clinically relevant effects and the criteria used to
Education, practice and debate

compile votes can differ between review groups and can alter review outcome. For example, similar reviews on low back pain used three item scales (positive, negative and neutral) and two item scales (positive and negative) leading to discrepancies in the vote counting process between the reviewers.44 Inconsistency in judgements about trial outcome in reviews that dichotomise complex trial data as positive or negative have been described for transcutaneous electrical nerve stimulation (TENS), and dichotomised interpretations of trial reports tend to be systematically biased towards the reviewers' conclusion.44 In their reappraisal of acupuncture reviews, Linde and Willich concluded that: 'Readers should be aware that apparently minor decisions in the review process can have a major impact [on review outcome]' [abstract p1].

Problems with methodological scoring and trial validity
Credence is given to trials that score high on methodological quality determined by scoring methods such as the Cochrane 60-item scale and the Jadad 5-item scale. These scales give weight to randomisation and blinding because failure to do so can exaggerate treatment effects by up to 40% and 17% respectively.44 However, trials can achieve high methodological quality scores even if they lack validity or credibility due to insufficient dosing or inappropriate treatment technique. This is problematic where optimal treatment technique is not known or the intervention is administered as part of a combination therapy as is often the case with acupuncture.

Smith et al developed an instrument called the Oxford Pain Validity Scale (OPVS) to measure trial validity.44 OPVS is a 5-item instrument using a 16-point scale to score blinding, sample size, outcomes, baseline pain and trial sensitivity, and data analysis. OPVS was piloted as part of a systematic review of 13 RCTs on acupuncture and back and neck pain with scores ranging from 4 to 14.44 Original trial authors concluded that acupuncture was effective in five RCTs and ineffective in eight. However, Smith et al claimed that trial authors drew conclusions that were not consistent with trial data. A re-analysis of trial data revealed that higher validity scores were associated with negative findings and Smith et al concluded that there was no convincing evidence for the analgesic efficacy of acupuncture for back or neck pain. This scenario demonstrates that authors of primary research may misinterpret trial data and that systematic reviewers may alter the conclusions of primary research. Interestingly, the OPVS does not account for adequacy of the treatment intervention in terms of frequency, duration, timing and technique. In other words, a trial could achieve a high OPVS despite under-dosing treatment during the trial.

The importance of adequacy and appropriateness of treatment intervention
The importance of ensuring that treatment interventions are adequate in dosage and technique was revealed in a meta-analysis on the effectiveness of TENS for postoperative analgesic consumption pain.57 A subgroup analysis demonstrated a significantly better outcome for TENS when applied using adequate (optimal) stimulation techniques when compared to non-adequate stimulation techniques. Optimal TENS techniques were defined as an intensity that was strong enough to generate a strong but comfortable paraesthesia, a pulse frequency that was below 150 pulses per second and electrodes applied at the site of the operative scar. Surprisingly, only 11 of 21 published RCTs met these criteria and they produced a significantly larger reduction in analgesic consumption of over 30% when compared to TENS applied inadequately.

The importance of adequate dosing and appropriate technique in trials of technique based interventions like acupuncture cannot be over emphasised and the problem has been recognised. Ezzo et al examined the effectiveness of acupuncture for chronic pain and reported that there were 21 positive outcome studies, three negative and 27 neutral.59 Positive results were associated with low-quality methodology and negative results with a risk of type II errors (false negatives) due to inadequate sample size. Interestingly, six or more acupuncture treatments were associated with positive outcomes even when methodological quality had been taken into account. However, attempts to develop a method for assessing the adequacy of acupuncture techniques used in controlled clinical trials has met with limited success. White and Ernst used six acupuncture experts to rate adequacy of acupuncture treatment and the level of information provided in 15 controlled clinical trials in which dry needles were inserted into
the skin for back pain. Inter-rater reliability was low, with ‘considerable variation’ in scores awarded by the six acupuncture experts, although good correlation was seen within four of the experts. Interestingly, determination of adequacy did not depend on whether needle sensation was elicited, or how many needles were used. The authors concluded that this method of assessing adequacy ‘may not be reliable in other situations because adequacy scores were inconsistent, possibly indicating problems with the scale’. The difficulty of assessing the adequacy of interventions in clinical trials of acupuncture has not prevented the publication of Standards for Reporting Interventions in Controlled Trials of Acupuncture (STRICTA) which were modified from the consolidated standards for reporting trials (CONSORT). The STRICTA recommendations describe six items that should be included in trial reports. These are acupuncture rationale; needling details; treatment regimen; co-interventions; practitioner background; and control interventions. Needling details include statements of points used, numbers of needles inserted, depths of insertion, responses elicited, needle stimulation, needle insertion time and needle type. By providing this level of detail in trial reports, readers and referees will be able to comment on the appropriateness of interventions, even if the optimal procedures for acupuncture are still disputed. When adequate criteria for acupuncture become available it will be possible to analyse published work retrospectively. However, two RCTs recently published in prestigious journals did not adhere to STRICTA recommendations and omitted important detail about the acupuncture intervention. Vas et al reported that acupuncture plus diclofenac was better than sham acupuncture plus diclofenac for osteoarthritis of the knee in a trial on 97 outpatients, with nine dropouts. Distal and local points were used with local points given electroacupuncture (EA), although there was no justification of this approach. The authors claimed to have achieved adequate blinding, although this would have been compromised during EA because of electrical paraesthesia. Details of the electrical characteristics of EA and whether manual acupuncture needles were ‘twirled’ was absent because of editorial pressure to restrict space. The authors were happy to supply the information in subsequent correspondence. Journals publishing trials of acupuncture have a duty to allow full reporting of trial methodology.

Accounting for the physiological intention of treatment interventions

Investigators also need to consider and report the physiological intention of acupuncture and whether this was achieved. For example, the physiological intention of TENS is to stimulate the skin in order to activate selectively, non-noxious afferent fibres (ie A-beta), as this is known to generate segmental antinociception. If this is achieved it is more likely that a patient will obtain pain relief. When non-noxious afferent fibres are activated the TENS user will report experiencing ‘a strong but comfortable electrical TENS paraesthesia’ and this is the criterion used to ensure adequate technique. It is often stated that the physiological intention of acupuncture is to stimulate small diameter nerve fibres arising from deep and superficial tissue. It is not known whether criteria could be established to monitor whether this has been achieved during treatment. One possibility, which is a matter of some debate, is whether the sensation of de qi could be used to monitor adequate stimulation and possibly used as a sensory predictor of patient response. The acupuncturist in the trial by Vas et al determined the patient’s sensation of de qi as an elicitation of needle sensation to check that the puncture was performed in the ‘correct site’. Unfortunately, the investigators did not record the number of patients in active and sham groups who experienced de qi.

Conclusion

It is still relatively easy for advocates to cite primary research and reviews that show acupuncture in a positive light and for sceptics to cite primary research and reviews that show acupuncture in a negative light. Therefore, it was hardly surprising that the BMA’s approval of acupuncture was challenged and defended with such vigour. The conclusions of systematic reviews and meta-analyses depend on the quality of the primary research. If you put garbage in to a systematic review you will get garbage out. Systematic reviewers need to be attentive to the adequacy of the acupuncture technique used in clinical trials and whether sufficient dosage was used. Recently, the publication of better quality clinical
Education, practice and debate

trials coupled with greater awareness of methodological issues affecting systematic reviews of acupuncture have resulted in shifts in review conclusions, in favour of chronic back pain. It is hoped that these improvements in clinical trial design and review methodology can be rolled out across other pain conditions. For the immediate future, however, you can be certain that the uncertainty about the effectiveness of acupuncture will remain.

Reference list
1. Dr Copperfield. Inside the Mind of a GP.
The clinical effectiveness of acupuncture for pain relief – you can be certain of uncertainty

Mark I Johnson

Acupunct Med 2006 24: 71-79
doi: 10.1136/aim.24.2.71