# PROGRAMME GRANTS FOR APPLIED RESEARCH

## Acupuncture for chronic pain and depression in primary care: a programme of research

Hugh MacPherson, Andrew Vickers, Martin Bland, David Torgerson, Mark Corbett, Eldon Spackman, Pedro Saramago, Beth Woods, Helen Weatherly, Mark Sculpher, Andrea Manca, Stewart Richmond, Ann Hopton, Janet Eldred and Ian Watt

**NHS**

*National Institute for Health Research*

# Acupuncture for chronic pain and depression in primary care: a programme of research

Hugh MacPherson,[1]* Andrew Vickers,[2] Martin Bland,[1] David Torgerson,[1] Mark Corbett,[3] Eldon Spackman,[4] Pedro Saramago,[4] Beth Woods,[4] Helen Weatherly,[4] Mark Sculpher,[4] Andrea Manca,[4] Stewart Richmond,[1] Ann Hopton,[1] Janet Eldred[1] and Ian Watt[5]

[1]Department of Health Sciences, University of York, York, UK
[2]Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA
[3]Centre for Reviews and Dissemination, University of York, York, UK
[4]Centre for Health Economics, University of York, York, UK
[5]Department of Health Sciences/Hull York Medical School, University of York, York, UK

*Corresponding author

This report should be referenced as follows:

MacPherson H, Vickers A, Bland M, Torgerson D, Corbett M, Spackman E, *et al*. Acupuncture for chronic pain and depression in primary care: a programme of research. *Programme Grants Appl Res* 2017;**5**(3).

# Programme Grants for Applied Research

**Criteria for inclusion in the *Programme Grants for Applied Research* journal**
Reports are published in *Programme Grants for Applied Research* (PGfAR) if (1) they have resulted from work for the PGfAR programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

## Programme Grants for Applied Research programme
The Programme Grants for Applied Research (PGfAR) programme, part of the National Institute for Health Research (NIHR), was set up in 2006 to produce independent research findings that will have practical application for the benefit of patients and the NHS in the relatively near future. The Programme is managed by the NIHR Central Commissioning Facility (CCF) with strategic input from the Programme Director.

The programme is a national response mode funding scheme that aims to provide evidence to improve health outcomes in England through promotion of health, prevention of ill health, and optimal disease management (including safety and quality), with particular emphasis on conditions causing significant disease burden.

For more information about the PGfAR programme please visit the website: http://www.nihr.ac.uk/funding/programme-grants-for-applied-research.htm

## This report

# Abstract

## Acupuncture for chronic pain and depression in primary care: a programme of research

Hugh MacPherson,[1]* Andrew Vickers,[2] Martin Bland,[1]
David Torgerson,[1] Mark Corbett,[3] Eldon Spackman,[4]
Pedro Saramago,[4] Beth Woods,[4] Helen Weatherly,[4] Mark Sculpher,[4]
Andrea Manca,[4] Stewart Richmond,[1] Ann Hopton,[1] Janet Eldred[1]
and Ian Watt[5]

[1]Department of Health Sciences, University of York, York, UK
[2]Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center,
 New York, NY, USA
[3]Centre for Reviews and Dissemination, University of York, York, UK
[4]Centre for Health Economics, University of York, York, UK
[5]Department of Health Sciences/Hull York Medical School, University of York, York, UK

*Corresponding author  hugh.macpherson@york.ac.uk

**Background:** There has been an increase in the utilisation of acupuncture in recent years, yet the evidence base is insufficiently well established to be certain about its clinical effectiveness and cost-effectiveness. Addressing the questions related to the evidence base will reduce uncertainty and help policy- and decision-makers with regard to whether or not wider access is appropriate and provides value for money.

**Aim:** Our aim was to establish the most reliable evidence on the clinical effectiveness and cost-effectiveness of acupuncture for chronic pain by drawing on relevant evidence, including recent high-quality trials, and to develop fresh evidence on acupuncture for depression. To extend the evidence base we synthesised the results of published trials using robust systematic review methodology and conducted a randomised controlled trial (RCT) of acupuncture for depression.

**Methods and results:** We synthesised the evidence from high-quality trials of acupuncture for chronic pain, consisting of musculoskeletal pain related to the neck and low back, osteoarthritis of the knee, and headache and migraine, involving nearly 18,000 patients. In an individual patient data (IPD) pairwise meta-analysis, acupuncture was significantly better than both sham acupuncture ($p < 0.001$) and usual care ($p < 0.001$) for all conditions. Using network meta-analyses, we compared acupuncture with other physical therapies for osteoarthritis of the knee. In both an analysis of all available evidence and an analysis of a subset of better-quality trials, using aggregate-level data, we found acupuncture to be one of the more effective therapies. We developed new Bayesian methods for analysing multiple individual patient-level data sets to evaluate heterogeneous continuous outcomes. An accompanying cost-effectiveness analysis found transcutaneous electrical nerve stimulation (TENS) to be cost-effective for osteoarthritis at a threshold of £20,000 per quality-adjusted life-year when all trials were synthesised. When the analysis was restricted to trials of higher quality with adequate allocation concealment, acupuncture was cost-effective. In a RCT of acupuncture or counselling compared with usual care for depression, in which half the patients were also experiencing comorbid pain, we found acupuncture and counselling to be clinically effective and acupuncture to be cost-effective. For patients in whom acupuncture is inappropriate or unavailable, counselling is cost-effective.

## ABSTRACT

**Conclusion:** We have provided the most robust evidence from high-quality trials on acupuncture for chronic pain. The synthesis of high-quality IPD found that acupuncture was more effective than both usual care and sham acupuncture. Acupuncture is one of the more clinically effective physical therapies for osteoarthritis and is also cost-effective if only high-quality trials are analysed. When all trials are analysed, TENS is cost-effective. Promising clinical and economic evidence on acupuncture for depression needs to be extended to other contexts and settings. For the conditions we have investigated, the drawing together of evidence on acupuncture from this programme of research has substantially reduced levels of uncertainty. We have identified directions for further research. Our research also provides a valuable basis for considering the potential role of acupuncture as a referral option in health care and enabling providers and policy-makers to make decisions based on robust sources of evidence.

**Trial registration:** Current Controlled Trials ISRCTN63787732.

**Funding:** The National Institute for Health Research Programme Grants for Applied Research programme.

# Contents

# List of tables

# List of figures

# List of abbreviations

| | | | |
|---|---|---|---|
| AACP | Acuncture Association of Chartered Physiotherapists | EULAR | European League Against Rheumatism |
| AAOS | American Academy of Orthopaedic Surgeons | EVPI | expected value of perfect information |
| ACR | American College of Rheumatology | GP | general practitioner |
| ACUDep | Acupuncture, Counselling or Usual Care for Depression | HAQ | Health Assessment Questionnaire |
| AIMS | Arthritis Impact Measurement Scale | HERC | Health Economics Research Centre |
| ANCOVA | analysis of covariance | HRQoL | health-related quality of life |
| ANOVA | analysis of variance | ICER | incremental cost-effectiveness ratio |
| ATC | Acupuncture Triallists' Collaboration | IPD | individual patient data |
| AUC | area under the curve | MCII | minimal clinically important improvement |
| AWR | adopt with research | MCMC | Markov chain Monte Carlo |
| BAcC | British Acupuncture Council | MCS | mental component summary |
| BAWMA | British Academy of Western Medical Acupuncture | MPCI | minimal perceptible clinical improvement |
| BDI-II | Beck Depression Inventory-II | NICE | National Institute for Health and Care Excellence |
| BEAM | Back pain Exercise And Manipulation | NMES | neuromuscular electrical stimulation |
| BMAS | British Medical Acupuncture Society | NRS | numerical rating scale |
| BMI | body mass index | NSAID | non-steroidal anti-inflammatory drug |
| CAM | complementary and alternative medicine | OARSI | Osteoarthritis Research Society International |
| CARE | Consultational and Relational Empathy | OIR | only in research |
| CI | confidence interval | OLS | ordinary least squares |
| CLAD | censored least absolute deviations | PCS | physical component summary |
| CMS | Constant–Murley Score | PEMF | pulsed electromagnetic field |
| CRD | Centre for Reviews and Dissemination | PES | pulsed electrical stimulation |
| CrI | credible interval | PHQ-9 | Patient Health Questionnaire-9 items |
| DIC | deviance information criterion | QALY | quality-adjusted life-year |
| EQ-5D | EuroQol-5 Dimensions | RCT | randomised controlled trial |
| | | SD | standard deviation |

| | | | |
|---|---|---|---|
| SE | standard error | TOIB | Topical or Oral IBuprofen |
| SF-12 | Short Form questionnaire-12 items | TTH | tension-type headache |
| SF-36 | Short Form questionnaire-36 items | VAS | visual analogue scale |
| SMD | standardised mean difference | WOMAC | Western Ontario and McMaster Universities Osteoarthritis Index |
| TENS | transcutaneous electrical nerve stimulation | | |

# Plain English summary

Increasing numbers of people are seeking acupuncture treatment for various conditions. However, we do not yet know how effective acupuncture is or if it provides value for money. By studying all of the available research, we should be able to reduce uncertainty about these issues and provide clear evidence for decision-makers. We aimed to provide high-quality evidence about the use of acupuncture for chronic pain from existing data and to develop fresh evidence on the use of acupuncture for depression.

We used systematic processes to combine and interpret data from high-quality clinical trials of acupuncture involving approximately 18,000 patients. The types of pain investigated in these trials were neck and lower back pain, osteoarthritis of the knee, headache and migraine. We found acupuncture to be more than simply a placebo as it was more effective than sham acupuncture, with sham acupuncture consisting of needling that did not penetrate the skin or needling at the wrong points. Acupuncture was also found to be better than standard medical care for all of these chronic pain conditions. In another analysis, we compared acupuncture with other physical therapies for osteoarthritis of the knee and found acupuncture to be one of the more clinically effective therapies and, when based on high-quality trial evidence, also cost-effective. When all trials were analysed, transcutaneous electrical nerve stimulation was cost-effective. We also conducted a large-scale clinical trial of acupuncture or counselling for depression in which we found both interventions to be clinically effective and acupuncture to be cost-effective. Counselling was also cost-effective where acupuncture was contraindicated or unavailable.

Drawing together this evidence provides high-quality information on the effect of acupuncture for patients with chronic pain or depression, offers directions for further research and gives health-care providers and policy-makers an opportunity to consider if wider access to acupuncture is appropriate and cost-effective.

# Scientific summary

## Background

Acupuncture is widely practised in the UK, with an estimated 4 million treatments provided a year, primarily for chronic pain conditions such as musculoskeletal pain and headache or migraine. In addition to chronic pain conditions, psychological distress, including depression, has been a common reason that people have consulted acupuncturists. The evidence base on acupuncture has been patchy, with some controversy regarding the clinical benefits and specifically the extent that acupuncture is more than simply a placebo. Clinical questions remain as to the effect of acupuncture in everyday practice and there are unanswered questions regarding whether or not acupuncture is cost-effective. One of the challenges in the field has been the quality of the available evidence. For most conditions, systematic reviews of acupuncture have had a limited ability to draw any definitive conclusions regarding effectiveness. As with research into many other physical therapies for chronic pain, there have been too many underpowered trials conducted and the methodological standards have been insufficiently robust. In this climate of uncertainty, however, there has been a recent and dramatic increase in the quality and quantity of trials of acupuncture, especially for chronic pain conditions, which has provided this programme of research a unique opportunity to substantially reduce the uncertainty. In this context, our primary aim was to use high-quality methods and the best evidence available to determine the clinical and economic impact of acupuncture for chronic pain and depression.

## Acupuncture for chronic pain

Although acupuncture is widely used for chronic pain, there is limited understanding of how it works, which in turn fuels some of the uncertainty as to its potential role as a treatment modality. The uncertainty is associated with a concern that acupuncture might simply be a theatrical placebo and the suggestion that when issues of bias are taken into account, the size of any effect might vanish. To address the uncertainty, in our first study we set out to determine the effect size of acupuncture for chronic pain, whether acupuncture is compared with sham acupuncture or with a non-acupuncture control, based on direct trial data on these comparisons. On building a collaboration of key triallists, the Acupuncture Triallists' Collaboration (ATC), we were able to conduct an individual patient data (IPD) meta-analysis. To establish eligible studies we identified randomised controlled trials (RCTs) of acupuncture for the four chronic pain conditions in which allocation concealment was determined unambiguously to be adequate. Data from 29 of 31 eligible RCTs, with a total of 17,922 patients, were analysed. Our primary result, when including all eligible RCTs, was that acupuncture was statistically significantly more effective than both sham controls and non-acupuncture controls for all four pain conditions ($p < 0.001$). When an outlying set of RCTs that strongly favoured acupuncture was excluded, the effect size across each pain condition was similar. Patients receiving acupuncture had less pain than those receiving sham controls, with effect sizes based on standardised mean differences (SMDs) of 0.23 [95% confidence interval (CI) 0.13 to 0.33], 0.16 (95% CI 0.07 to 0.25) and 0.15 (95% CI 0.07 to 0.24) for back and neck pain, osteoarthritis and chronic headache, respectively. In the comparison between acupuncture and non-acupuncture controls, effect sizes were 0.55 (95% CI 0.51 to 0.58), 0.57 (95% CI 0.50 to 0.64) and 0.42 (95% CI 0.37 to 0.46), respectively. A variety of sensitivity analyses was conducted, including those related to publication bias, with little impact on the main findings of the study. To conclude, in these high-quality studies, a clinically relevant and statistically significant difference was found between acupuncture and non-acupuncture controls, suggesting that acupuncture is a suitable referral option for chronic pain. Statistically significant differences between true and sham acupuncture indicate that acupuncture is more than a placebo.

## Physical therapies for osteoarthritis

Many systematic reviews have evaluated individual types of physical treatments for osteoarthritis of the knee, but no study has compared these different physical treatments against each other in the same analysis. In a second study we used a network meta-analysis to address the question of how effective physical treatments for osteoarthritis of the knee are, when compared with each other, for relieving pain. We reviewed the literature up to January 2013, which involved searching 17 electronic databases. We identified RCTs of physical treatments in patients with osteoarthritis of the knee in which pain was reported as an outcome. In a network meta-analysis, both direct and indirect evidence was synthesised to compare the effectiveness of acupuncture with that of other relevant physical treatments for alleviating pain. In total, 114 trials (covering 22 treatments and 9709 patients) were included in the analysis. Many trials were classed as being of poor quality with a high risk of bias in several domains. Eight interventions statistically significantly outperformed standard (usual) care, producing an improvement in pain compared with standard care: interferential therapy, acupuncture, transcutaneous electrical nerve stimulation (TENS), pulsed electrical stimulation, balneotherapy, aerobic exercise, sham acupuncture and muscle-strengthening exercise. The better-quality studies, most commonly of acupuncture (11 trials), muscle-strengthening exercise (nine trials) and sham acupuncture (eight trials), were included in a sensitivity analysis. Acupuncture was statistically significantly more effective than muscle-strengthening exercise and sham acupuncture (SMD 0.49, 95% CI 0.00 to 0.98; and 0.34, 95% CI 0.03 to 0.66, respectively). Acupuncture was also statistically significantly better than standard care (SMD 1.01, 95% CI 0.61 to 1.43). To conclude, in a network meta-analysis, acupuncture was found to be one of the more effective physical treatments for alleviating osteoarthritis knee pain in the short term. However, given that much of the evidence was of poor quality, there is uncertainty about the effectiveness of many physical treatments.

## Towards evaluating the cost-effectiveness of acupuncture for osteoarthritis

To address the question of the cost-effectiveness of acupuncture for chronic pain, we have benefited from the availability of the data from the ATC repository. In this third study we conducted a network meta-analysis of the individual patient-level data from the ATC repository. By synthesising all of the data, including indirect data, we derived treatment effect estimates, a key parameter for cost-effectiveness analyses to be used for resource allocation decisions. In this primarily methodological study, new Bayesian methods for analysing multiple individual patient-level data sets reporting heterogeneous continuous outcomes were developed. An essential step towards a cost-effectiveness analysis is a preference-based measure of health-related quality of life, such as the EuroQol-5 Dimensions (EQ-5D). To synthesise heterogeneous outcomes, we used mapping to convert and compare heterogeneous outcomes on to the EQ-5D summary index scale. The models developed entailed a Bayesian random-effects network meta-analysis specification, including exchangeable pain type interaction effects. The implications for cost-effectiveness analysis were also demonstrated. The methods were illustrated using a case study of acupuncture for chronic pain in primary care, including headache/migraine, musculoskeletal pain and osteoarthritis of the knee. By using the same evidence base as in the first study, our analysis included approximately 17,500 patients from 28 trials (we were unable to use IPD from one trial), in which we compared acupuncture, sham acupuncture and usual care with each other. The synthesis of mapped EQ-5D estimates found that acupuncture was effective compared with usual care, with median treatment effects estimated as 0.056 [95% credible interval (CrI) 0.021 to 0.092] for headache/migraine, 0.082 (95% CrI 0.047 to 0.116) for musculoskeletal pain and 0.079 (95% CrI 0.042 to 0.114) for osteoarthritis of the knee. The EQ-5D benefit of acupuncture over sham acupuncture was smaller and more uncertain (headache: 0.004, 95% CrI –0.035 to 0.042; musculoskeletal pain: 0.023, 95% CrI –0.007 to 0.053; osteoarthritis of the knee: 0.022, 95% CrI –0.014 to 0.060). Although not all relevant interventions were compared for decision-making purposes, cost-effectiveness results suggest that when acupuncture is compared with usual care alone it is cost-effective, with incremental cost-effectiveness ratios (ICERs) ranging from £7000 to £14,000 per quality-adjusted life-year (QALY) across the pain types of headache/migraine, musculoskeletal

pain and osteoarthritis of the knee. In this methodologically oriented case study we showed that the mapping of heterogeneous outcomes on to the EQ-5D summary index provides a useful step towards conducting a cost-effectiveness analysis.

## Cost-effectiveness of non-pharmacological treatments for osteoarthritis

In a fourth study within this programme of research, we conducted an economic evaluation to evaluate the cost-effectiveness of non-pharmacological treatments used to reduce chronic pain in osteoarthritis of the knee. We used novel network meta-analysis methods to synthesise RCT evidence for 17 active interventions and three control interventions. Data were obtained from the systematic review carried out in the second study, which were available for 88 RCTs including 7507 patients. IPD from the first study were available for five of the RCTs, including 1329 patients. A wide range of health-related quality-of-life outcomes were reported in the trials. As the ultimate objective of the analysis was to inform resource allocation decisions in the UK, data from these instruments were mapped to EQ-5D preference weights prior to synthesis, extending methods developed in the third study. Sensitivity analyses were conducted to address potential bias associated with poor study conduct and to explore the importance of the time point of reporting for the study results. Resource use associated with the interventions was estimated from trial data, expert opinion, the literature and information obtained from NHS trust websites. Non-intervention resource use related to changes in EQ-5D was obtained from another UK trial. Outcomes and costs were synthesised using an area-under-the-curve cost-effectiveness model with a time horizon of 8 weeks. When all trials were included in the synthesis, TENS was cost-effective with an ICER of £2690 per QALY compared with usual care. When the analysis was restricted to trials with adequate allocation concealment, acupuncture was cost-effective with an ICER of £13,502 per QALY compared with TENS. There were limited data regarding the long-term effects of many non-pharmacological interventions used to treat osteoarthritis of the knee. The active and control interventions in the trials informing this analysis were subject to heterogeneity in the method, duration and intensity with which they were administered. These results are subject to some decision uncertainty and the expected value of perfect information is relatively high, suggesting that additional research may be cost-effective.

## Acupuncture or counselling for depression

Depression is a significant cause of morbidity. Many patients have communicated an interest in non-pharmacological therapies to their general practitioners. Systematic reviews of acupuncture and counselling for depression in primary care have identified limited evidence. The aim of the fifth study was to evaluate acupuncture compared with usual care and counselling compared with usual care for patients who continue to experience depression in primary care. Moreover, a cost-effectiveness analysis is needed to understand whether or not such therapies should be considered a good use of limited health resources. In total, 755 patients with depression (Beck Depression Inventory-II score of ≥ 20) were recruited to a RCT carried out in 27 primary care practices in the north of England. Allocation was to one of three arms using a ratio of 2 : 2 : 1 to acupuncture ($n = 302$), counselling ($n = 302$) and usual care alone ($n = 151$). The difference in mean Patient Health Questionnaire-9 items (PHQ-9) score at 3 months was the primary outcome. Patients were followed up over 12 months and analysis was by intention to treat. Additional quantitative and qualitative substudies were conducted. Data on PHQ-9 scores were available for 614 patients at 3 months and 572 patients at 12 months. A mean of 10 sessions was attended for acupuncture and a mean of nine sessions was attended for counselling. There was a statistically significant reduction in mean PHQ-9 depression score at 3 months for acupuncture compared with usual care (–2.46, 95% CI –3.72 to –1.21) and counselling compared with usual care (–1.73, 95% CI –3.00 to –0.45), and at 12 months for acupuncture (–1.55, 95% CI –2.41 to –0.70) and counselling (–1.50, 95% CI –2.43 to –0.58) compared with usual care. When controlling for time and attention, no significant differences in clinical outcome were found between acupuncture and counselling. No serious treatment-related adverse events were reported. The trial was not designed to separate

out specific from non-specific effects. Acupuncture and counselling were found to have higher mean QALYs and costs than usual care. In the base-case analysis, acupuncture had an ICER of £4560 per additional QALY and was cost-effective with a probability of 0.62 at a cost-effectiveness threshold of £20,000 per QALY. A scenario analysis of counselling compared with usual care, excluding acupuncture as a comparator when inappropriate or unavailable, resulted in an ICER of £7935 and a probability of being cost-effective of 0.91. To summarise, patients with ongoing depression in primary care who received a short course of either acupuncture or counselling experienced statistically significant reductions in depression compared with usual care alone.

## Conclusion

In conclusion, this programme of research has provided the most substantive evidence to date on acupuncture and its potential impact. Drawing on the existing data from RCTs, we used an IPD meta-analysis and found acupuncture to be clinically effective across the chronic pain conditions of headache and migraine, back and neck pain, and osteoarthritis of the knee. Our evidence from this data set suggests that acupuncture is a statistically significantly more effective intervention than placebo. In a network meta-analysis, our evidence across a range of physical therapies for osteoarthritis of the knee suggested that acupuncture is associated with more high-quality trials than any of the other therapies and is also one of the most effective therapies. Acupuncture was also cost-effective if only high-quality trials were analysed. When all trials were included in the synthesis, including both low- and high-quality trials, we found TENS to be cost-effective. In the trial that we conducted on depression, we found that acupuncture and counselling were statistically significantly better than usual care and that acupuncture was also cost-effective.

There remains some uncertainty regarding our results. For example, with regard to the cost-effectiveness analyses of acupuncture for musculoskeletal pain and headache and migraine, not all competing therapies relevant to decision-making were included in the analysis. There are few data regarding the long-term effects of many non-pharmacological interventions used to treat osteoarthritis of the knee, and sensitivity analyses suggested that the cost-effectiveness model results may be sensitive to the magnitude of these effects. When comparing acupuncture with usual care for depression, there was no control for non-specific effects, although we did control for time and attention when comparing acupuncture with counselling. Nevertheless, our programme of research has used high-quality methods to provide the best possible data on the clinical effectiveness and cost-effectiveness of acupuncture for chronic pain and depression. Robust evidence is in the interests of all stakeholders and our results will be used to inform patients, practitioners, policy-makers and commissioners of services.

## Trial registration

The trial of acupuncture and counselling for depression is registered as ISRCTN63787732.

## Funding

# Chapter 1 Introduction

Around 4 million acupuncture treatments are provided each year in the UK according to a national survey conducted in 2012.[1] Approximately one-third of the 4 million treatments were reported as being provided within the NHS and the remainder in the independent or not-for-profit sectors.[1] Reasons for consultation were dominated by musculoskeletal complaints (59%) and neurological conditions (9%), primarily headaches and migraine. Younger people predominantly consulted for back pain and headaches, whereas older people were proportionately more likely to consult for knee pain. These data reflect a steady increase over time in the utilisation of acupuncture: in 2001, a survey reported an estimate of the total number of acupuncture treatments in the UK per year of 3 million, with a similar proportion, namely one-third, provided within the NHS.[2] The provision of acupuncture within the NHS appears to be patchy, with limited access for patients with an interest in receiving acupuncture. Despite general practitioners (GPs) being, in general, supportive of the idea of wider acupuncture provision,[3] the funding of acupuncture clinics in primary care has been difficult.[4] This situation has led to many patients turning to the independent sector for acupuncture treatment.[5,6] The common conditions treated by independent acupuncturists showed a marked correspondence with the conditions that GPs in primary care have acknowledged that they are not fully effective in treating, especially musculoskeletal conditions, depression and chronic pain.[7] Across Europe, a similar pattern has been reported whereby patients are most commonly seeking help for painful conditions.[8]

The provision of acupuncture in the UK involves practitioners who are primarily members of the four main professional organisations that regulate acupuncture in the UK: the Acupuncture Association of Chartered Physiotherapists (AACP), British Acupuncture Council (BAcC), British Academy of Western Medical Acupuncture (BAWMA) and British Medical Acupuncture Society (BMAS). The AACP is a clinical interest group of the Chartered Society of Physiotherapy with membership ($n = 5600$) requiring a minimum of 80 hours of training. The BAcC is the leading self-regulatory body for independent acupuncturists in the UK, with membership ($n = 2600$) open to graduates of acupuncture courses based on 3 years of full-time study or equivalent, with most courses awarding a university degree. Most members ($n = 300$) of BAWMA are nurses or other health-care professionals, and they have received approximately 100 hours of training delivered over eight weekends, leading to an Academy Licentiate Certificate. Practising members ($n = 2400$) of the BMAS are primarily doctors with minimum training taking place over two weekends, which may be extended, with accreditation based on completion of 100 hours of training along with provision of a series of case histories.

Unlike the pharmaceutical agents within conventional medicine, acupuncture has had no regulatory gatekeeper controlling its therapeutic activity before being made available on the open market. Treatments have therefore been used widely before researchers have evaluated their effect, whether in terms of safety, efficacy, clinical effectiveness or cost-effectiveness. It has been argued that all of these issues are important in the context of the 'uncontrolled' provision of acupuncture, especially when widely used outside the national health-care system, and therefore it is the question of patient safety that needs to be most urgently addressed.[9] This concern about the risks of acupuncture was highlighted some time ago by a systematic review that documented case reports of six deaths that may have been caused by acupuncture.[10] However, case reports are a limited source of evidence on risk, as there is no sense of the frequency or rate with which such events occur. More robust research led to data on adverse event rates involving two prospective surveys of practitioners in the UK, with data involving members of three professional associations: one study involving practitioners of the BAcC[11] and the other involving practitioners from both the BMAS and AACP.[12] These studies covered > 30,000 treatments each and reported no serious adverse events, leading to the conclusion that acupuncture is safe in qualified hands.[13] A subsequent prospective UK survey[14] of patient reports of adverse events over a 3-month period found results that were largely consistent with the aforementioned practitioner surveys. For most patients it has been found that their experience of benefit following acupuncture appears to outweigh the perceived adverse reactions to treatment.[15] These data on safety have been reinforced by two independent prospective surveys in Germany. One covered adverse events associated with 760,000 acupuncture treatments, leading to reports of two cases of pneumothorax, one case of exacerbation of depression, one acute hypertensive crisis, one vasovagal reaction,

and one asthma attack with hypertension and angina.[16] Another covered 2.2 million consecutive acupuncture treatments provided for 229,230 patients, with two patients found to have had a pneumothorax (life-threatening for neither patient) and one to have had a lower limb nerve injury that persisted for 180 days.[17] Taken together, these prospective surveys provide evidence that serious adverse events associated with acupuncture are extremely rare.

Although questions of safety have been largely addressed over the last decade or so, questions of the physiological mechanisms of acupuncture have been more widely debated. A feature of acupuncture research has been the concentrated effort to understand how acupuncture works. For example, numerous acupuncture-related biomarkers have been identified, including antinociceptive endogenous opioids,[18,19] immune system markers,[20,21] cardiovascular activity,[22] gastrointestinal function[23] and functional magnetic resonance imaging-detected brain activity.[24] Biomarker outcomes, however, are more revealing of correlations (i.e. when needling occurs, changes can be detected) than mechanisms or causal pathways. Reviews of acupuncture from China[25] and the West[26] continue to use the overarching term 'mechanism' but focus almost entirely on 'correlates'.

Beyond correlations between acupuncture and biomarkers, research has focused on the search for acupuncture-related biochemical, physiological and anatomical mechanisms. These research efforts include a focus on elucidating the nature of acupuncture points and meridian pathways, and the neurological 'signals' that they may carry. For example, some researchers are exploring to what extent the stimulation of the underlying neural pathways accounts for the physiological effects of and clinical responses to acupuncture in patients.[27] The experimental recording of neural activity associated with needle insertion, along with correlations between acupuncture-induced pain and endogenous opioids, led to models that proposed a set of pathways of acupuncture analgesia involving both the peripheral and the central nervous systems.[18,19] These models showed how the effects of acupuncture could be mapped onto the nervous system,[27,28] an understanding that has received some support from neuroimaging research.[24,29] The neural hypothesis has been used to explain the existence of meridians, given the observations that many acupuncture points and sections of meridians overlie the major peripheral nerves.[27]

There is also an emerging view that loose connective tissue (fascia) provides an alternative biomedical explanation for the role of acupuncture points and meridians.[30] For example, a study has shown that a large proportion of traditional acupuncture points are located at sites where the underlying nerve–vessel bundles are wrapped in a loose sheath of connective tissue that penetrates the fascia to reach the outer dermal layers.[31] A subsequent study demonstrated a high correspondence between the sites of acupuncture points and the location of loose connective tissue planes.[30] Indications are that the superficial fascia provides an initial 'response element' to needle stimulation, which may explain the 'needle grasp' phenomenon of acupuncture practice.[32] These studies provide a viable alternative to the prevailing neurobiological models based on the emerging evidence on loose connective tissue anatomy and its relationship with the acupuncture system.

Another line of physiological research has explored electrodermal activity at acupuncture points. For example, in an early study, acupuncture points were found to be at the summits of individually contoured conductivity fields.[33] The experimental and physiological confounders to such measurements have been highlighted in recent reviews.[34] In a recent narrative review, the relation of acupuncture point electrodermal activity to pathology has been described.[35] In this review one blinded study found that electrodermal activity at auricular acupuncture points could be used to distinguish which patients had recent or prior cardiopathology and which were healthy control subjects.[36] The observation that traditional pathways of acupuncture meridians correspond to ultrasound images of connective tissue planes[30] has been followed by the insight that these meridian-oriented collagenous structures are associated with lower electrical impedance.[37] Lower electrodermal activity along acupuncture meridians was also reported in seven of nine studies at both subcutaneous and intermuscular depths.[38]

Our understanding of how acupuncture might have a pain-relieving effect has been informed by basic science research. One mechanism to explain acupuncture is called the 'gate theory'.[39] This suggests that the pathway associated with acupuncture involves the A delta fibres entering the dorsal horn of the spinal cord, which inhibit pain impulses that are carried in the slower, unmyelinated C fibres. Descending inhibition of C fibre pain impulses is also enhanced through neural connections in the midbrain.[40] Additional mechanisms have been proposed for acupuncture's effect on pain, which are not necessarily in contradiction to the gate theory; for example, acupuncture stimulates release of endogenous opioids and other neurotransmitters. Interest in acupuncture and endogenous opioids was sparked in the 1970s by research into acupuncture's analgesic effects, for example research showing that acupuncture could induce analgesia in mice, which was blocked by naloxone.[18] After receiving acupuncture, levels of endogenous opioids in the cerebrospinal fluid have been directly observed to increase in humans.[41] Neuroimaging research has also provided insights into neurological changes associated with acupuncture when used to treat for pain. The research using functional magnetic resonance imaging has demonstrated that acupuncture elicits changes in the brain that appear to correlate with the presumed clinical effects of the points used. For example, a study of acupuncture for carpal tunnel syndrome has shown how acupuncture elicits neural plasticity in the somatosensory area of the brain that correlates with clinical benefits.[42]

These physiological studies are important in the context of efficacy research and the need to design a sham acupuncture needle to be used as a control in clinical trials. The lack of a clear understanding of the physiological events that are initiated by acupuncture needling is problematic. This is compounded by confusion surrounding the concept and definition of a 'placebo', and the difficulty in interpreting 'placebo effects' in a clinical context. This arises because placebo effects vary with different treatments, different settings, different coloured pills, varying patient and/or practitioner expectations, whether a drug or a device is used as a placebo, how placebo effects are explained at the outset of a trial, the extent that placebo effects might interact with concurrent treatment and whether or not placebo effects can be satisfactorily separated out from within a complex intervention.[43] Separating out placebo effects from other effects, such as the natural history of the condition or regression to the mean, is not straightforward. In a Cochrane review, larger effects of placebo interventions were associated with physical placebo interventions, such as sham acupuncture, and with patient-reported outcomes, for example as commonly used to monitor pain-related conditions.[44] Moreover, sham needling is implemented in a number of ways, each with its associated physiological effects. Two approaches to sham needling are commonly used: either needles are used at 'incorrect' locations but to penetrate the skin or non-penetrating (stage dagger type) needles are used at the 'correct' locations.[45] Some argue that a placebo should be physiologically inert, whereas others suggest that a placebo intervention is acceptable as long as it looks and feels the same as the active intervention that it is controlling for but does not trigger any of the physiological activity elicited by the active intervention. One aspect of this debate has clearly emerged, namely that there is no agreement that the physiological activity of sham acupuncture has been fully characterised.

Beyond the uncertainty of what a sham needle actually triggers at a physiological level, there is considerable agreement that it is useful in principle to determine if an active intervention has a 'specific' effect over and above what are commonly called 'non-specific effects', a term often used synonymously with 'placebo effects'. The use of the term 'non-specific effect' is preferred as it bypasses the confusion associated with the concept of the placebo that has been discussed above. In this programme of research, we have reviewed many trials that have designated 'sham' acupuncture arms and which have made the assumption that any difference in effect that we might observe between true and sham acupuncture in these trials provides the best available assessment of whether or not acupuncture outperforms a placebo.

A major focus in this programme was on questions regarding the efficacy and clinical effectiveness of acupuncture's putative benefits. Understandably, these are essential questions for the field, whether for patients, practitioners or commissioners. First, the widespread utilisation of acupuncture, as discussed above, raises public health issues related to patient safety. In addition, concerns about safety are often linked to questions regarding the risk/benefit ratio; for example, it can be argued that 'if there is no benefit, any risk is too much'.[46] Given that an overview of systematic reviews in 2010 concluded that

'numerous reviews have produced little convincing evidence that acupuncture is effective in reducing pain',[47] this is an understandable concern. Another concern is over questions of bias, given the tendency of trials with a greater risk of bias to deliver more positive results. In this context, a 2009 review of trials of acupuncture for pain stated that 'the effects of acupuncture cannot be clearly distinguished from bias and that it is unclear whether needling at acupuncture points relieves pain independently of the psychological impact of the treatment ritual'.[48] The question of interest here is whether or not true acupuncture, when adequately assessed, has an effect above and beyond that of a placebo or whether or not the effect becomes negligible as bias is reduced, potentially leading to all effect finally vanishing in the sands of placebodom.[49] As with many emerging fields, and especially those fields that are outside the many regulatory structures of modern medicine, the need for rigorous and unbiased research is an essential requirement so that fair judgements can be made regarding a role in our national health-care system. An unbiased assessment of the evidence base on acupuncture is necessary to inform decisions made by patients, practitioners and policy-makers.

It is useful to distinguish between randomised controlled trials (RCTs) of efficacy and RCTs of effectiveness. The term 'effectiveness' is used to measure the overall impact of an intervention on outcome, as would be expected to occur in usual care, with an emphasis on generalisability. The term 'efficacy' is used to measure the impact of an intervention on outcome in as ideal conditions as possible, with an emphasis on controlling for placebo effects. In both cases there is a need to limit bias as much as possible, although the challenges of doing so vary somewhat between the two types of trials. When evaluating effectiveness, a comparative or pragmatic design is commonly used in which acupuncture is compared with another active treatment, usual care or no treatment. It should be noted that many acupuncture trials have three-arm designs that include both a sham and another comparator treatment, because the researchers are attempting to address the questions of efficacy and effectiveness in the same trial. The effectiveness/ efficacy dichotomy is a useful perspective as it provides a framework for understanding the different types of acupuncture trials in the field, the different questions that they seek to answer and the different ways that the results of these trials need to be interpreted. In reality, many trials are hybrids that contain features of both effectiveness and efficacy trials.[50]

With regard to questions of the efficacy of acupuncture for a number of pain-related conditions, there has been considerable uncertainty, especially prior to the mid- to late 2000s. The state of the evidence to that time had been drawn together in a number of systematic reviews of acupuncture trials that included at least one sham-controlled arm, thereby enabling assessment of the difference between true and sham acupuncture. These reviews raised as many questions as answers, with authors identifying a range of mixed outcomes, some positive and some negative. These included data on chronic pain,[51] osteoarthritis of the knee,[52–54] headache and migraine,[55,56] and lower back pain.[57,58] A common feature identified in these reviews was a concern about the relatively small numbers of patients in many of the included trials. As an example, a review of acupuncture for chronic pain, published in 2000,[51] included 51 RCTs of acupuncture for a variety of conditions. Typical of these early reviews, there was a low sample size, with a median of 18 patients per trial arm, and weak methodology, with 68% of the trials defined as being of poor quality and only three of the 51 studies receiving a maximum quality score; in addition, there was a typical final conclusion of 'inconclusive evidence' on whether or not acupuncture is more effective than a placebo.[51]

The state of the evidence when considering the question of the effectiveness (rather than efficacy) of acupuncture, and specifically the evidence on acupuncture compared with usual care, has appeared more clear-cut. The aforementioned reviews published during the 2000s were generally more positive, with effect sizes considerably larger than when acupuncture was compared with sham acupuncture and with the differences between acupuncture and usual care more commonly being statistically significant. Nevertheless, these trials were also subjected to criticism because of the absence of a sham control, the argument being that there might be bias introduced because of unblinded practitioners, increasing the relative effect of the acupuncture, or bias because of resentful demoralisation in participants in the usual care arm who enrolled into a trial because they wanted to receive acupuncture.[59] Two pragmatic trials of

acupuncture for chronic pain conducted in the UK were of particular relevance to discussions of effectiveness.[60,61] Both trials were pragmatic in nature, recruiting everyday patients from within primary care and providing the approach to acupuncture that was as near as possible to normal practice, with one evaluating acupuncture for headache and migraine[60] and one evaluating acupuncture for lower back pain.[61] The cost-effectiveness analyses of these two trials[62,63] turned out to be of importance in terms of subsequent decisions on clinical guidance related to policy and practice. In reviewing these and other trials evaluating acupuncture against usual care for chronic pain in the mid- to late 2000s, authors of meta-analyses typically found that acupuncture was more effective than non-acupuncture controls (comprising waiting list, usual care or no treatment) for the conditions of lower back pain,[58] migraine/headache[64,65] and osteoarthritis of the knee.[52] Nevertheless, there remained some uncertainty about the clinical relevance of the effect size. As with the efficacy comparison between acupuncture and sham acupuncture, the studies that included an evaluation of effectiveness at this time were dominated by small trials of questionable methodological quality.

Towards the end of the 2000s it was becoming clear that the landscape of research into the efficacy and effectiveness of acupuncture was undertaking a remarkable shift in terms of the number of completed trials, the number of participants in the trials and the methodological quality of the trials. Of particular note was the funding by insurance companies of a series of trials in Germany, some of which had patient numbers in the thousands. These included the cluster of German Acupuncture Randomized Trials published in 2005 and early 2006,[66–69] which recruited around 300 patients in each of four separate trials on osteoarthritis,[69] chronic lower back pain,[66] migraine[68] and chronic tension headache.[67] Conducted at the same time were the GERman ACupuncture (GERAC) trials, including around 1000 patients with osteoarthritis,[70] chronic lower back pain[71] and migraine,[72] and 400 patients with chronic tension headache.[73] The third group of trials, the Acupuncture in Routine Care (ARC) trials, had even larger sample sizes, with ≥ 3000 patients in each of three separate pragmatic trials of back pain,[74] neck pain[75] and chronic headache,[76] and 700 patients in a trial of osteoarthritis arthritis.[77] The combination of this remarkable set of trials along with various other larger and higher-quality trials provided an extraordinary opportunity towards the end of the 2000s to provide a robust meta-analysis to provide the necessary clarity to supersede the prevailing uncertainty of the times.

Our programme of research consisted of a series of studies, some of which utilised this unusually large and recently completed set of acupuncture trial data related to chronic pain. The first of these was an individual patient data (IPD) meta-analysis, funded primarily by the US National Institutes of Health, which we report in *Chapter 2*. In this study we examined the clinical effectiveness of acupuncture for managing chronic pain in conditions including back and neck pain, osteoarthritis of the knee, and chronic headache and migraine. IPD meta-analysis is the most powerful method to synthesise research data. One of the founders of the Cochrane Collaboration, Ian Chalmers, has been quoted as saying that IPD meta-analysis is the 'yardstick' by which meta-analyses should be compared.[78] There are a number of advantages of using IPD compared with the summary methods of traditional meta-analyses, which analyse only the published summary data.[79] IPD meta-analysis allows for standardisation within the analysis of different types of outcome measures, for example allowing for the combination of continuous change scores with those reporting only percentage response rates. There is greater power to address subgroups within the population and explore if patient characteristics such as age, sex or baseline severity might impact on outcome. Prior to combining data sets, reanalysis of all trials should be carried out, which will ensure that the data are of a high quality. Finally, an IPD meta-analysis has greater statistical power and consequently more precision in estimating clinical effects. In *Chapter 2* we use this method for the first time in acupuncture research to evaluate outcomes from the large number of high-quality trials of acupuncture for chronic pain.

Acupuncture has been recommended by the National Institute for Health and Care Excellence (NICE) for the treatment of chronic headache and migraine,[80] and lower back pain,[81] but not chronic pain associated with osteoarthritis of the knee.[82] This last decision in part reflected concerns regarding the available evidence.[83] In *Chapter 3* we address the question of how competing physical therapies compare for osteoarthritis of the knee, a condition for which there is some uncertainty with regard to the effectiveness of

acupuncture.[84] A powerful method to compare the outcomes from a range of interventions for the same condition is a network meta-analysis, which has several advantages over comparing interventions using only pairwise meta-analysis. A network meta-analysis, which is also referred to as a multiple treatment comparison meta-analysis, can be used to compare interventions that may or may not have been evaluated directly against each other. Although direct evidence can come from head-to-head trials, a network meta-analysis can incorporate indirect evidence, which adds strength to the analysis as it allows the effects to be compared between interventions that have not been investigated head to head in RCTs and uses both direct and indirect evidence to inform estimates of effect.[85] For many comparisons, the network meta-analysis may yield more reliable and definitive results than a pairwise meta-analysis would.[86] Clinical practice guidelines that inform the decisions about optimal care need to rely on evidence-based evaluation of often many treatment options. As it provides comparisons across multiple interventions, a network meta-analysis is an approach that can inform a cost-effectiveness analysis, which in turn can inform clinical decision-making. As a step towards providing evidence on competing physical therapies, including acupuncture, for osteoarthritis of the knee, we report a network meta-analysis in *Chapter 3*.

Given that the current study has been designed to improve evidence around the costs and effects of acupuncture, in *Chapter 4* we report a synthesis of the IPD from the RCTs that provided the data set in *Chapter 2*, which evaluated acupuncture for headache/migraine and musculoskeletal and osteoarthritis pain. We used a network meta-analysis to leverage all available evidence to inform estimates of relative treatment effects when acupuncture was compared with usual care or sham acupuncture, or when both control interventions were compared with each other. The availability of IPD for all studies expanded the set of feasible analyses and allowed development of de novo methods to fully exploit the benefits of access to these data. Although evidence of effectiveness is important, policy-makers faced with difficult resource allocation decisions require estimates of the costs and effects of alternative treatment options. These estimates should reflect all relevant data and treatments should be compared using a metric that can be used across clinical areas – in the UK the quality-adjusted life-year (QALY) is typically used. Synthesising all relevant evidence to produce comparable estimates of costs and effects generates a series of challenges as the available evidence base rarely captures all costs and effects of treatment (because of the nature of data collection or the duration of follow-up), and often requires evidence to be generalised from different populations. The available trial evidence may compare different sets of treatments and in many instances the health-related quality-of-life (HRQoL) data required to estimate QALYs directly are not available. The mapping to convert heterogeneous outcome data on to the EuroQol-5 Dimensions (EQ-5D) summary index scale forms a key component of *Chapter 4*. In turn, this enabled us to estimate the cost-effectiveness of acupuncture for chronic pain conditions, with the caveat that not all other possible interventions have informed this analysis.

In *Chapter 5* we report a cost-effectiveness analysis that was conducted by systematically identifying and synthesising outcome data on a wide range of adjunct non-pharmacological interventions for osteoarthritis of the knee. This allowed an assessment of value for money to be made for all available alternatives at this point in the treatment pathway. We mapped the available HRQoL data to EQ-5D preference weights, producing a common statistic for synthesis. Network meta-analysis was used to synthesise data reported at both the individual-patient and aggregate level. Estimates of effect from a network meta-analysis of EQ-5D outcomes were used to estimate QALYs within a decision-analytic model, which also incorporated cost data from a range of sources. As well as allowing us to estimate the expected costs and effects associated with a wide range of treatments, the decision-analytic methods used allowed us to quantify both the nature and extent of uncertainty, and the value of further research. The synthesis in *Chapter 5* therefore combines the use of the IPD related to osteoarthritis of the knee presented in *Chapter 2*, the data and network meta-analysis methods used to compare competing treatments in *Chapter 3* and the mapping methods as set out in *Chapter 4*.

The focus on acupuncture for chronic pain conditions is central to this programme of research for two related reasons. First, much of the basic research into acupuncture is related to its pain-relieving effects. For example, acupuncture-induced analgesia caught the public imagination in the 1970s and led to

research findings showing how acupuncture analgesia is mediated in part by endogenous opioids.[87] More recently, acupuncture neuroimaging research on pain has not only led to a better understanding of how acupuncture might work,[88] but also has informed biomedical understanding of neuroplasticity.[89] Second, there has been a remarkable growth in the utilisation of acupuncture for chronic pain after its transmission from East Asia to the West. In part because of media attention and insurance-related reimbursement, the leading indication for acupuncture utilisation is for pain, whether provided in Europe,[8] the USA[90] or Australia.[91] Moreover, increased acceptance of acupuncture by physicians and allied health-care specialists is paralleled by proportionately more provision for chronic pain within biomedical settings, as is the case for the one-third of the 4 million annual acupuncture sessions that the UK provided within the NHS by doctors and physiotherapists.[1]

The final study in this programme of research moved on from chronic pain and focused on acupuncture as a potential treatment for depression. In a previous study exploring the clinical areas in which GPs experience themselves to be not fully effective, described as 'effectiveness gaps', GPs reported that depression was the second most common effectiveness gap after musculoskeletal problems.[7] Moreover, patients with psychological problems, including depression, make up the second most common group treated by acupuncture practitioners after those with chronic pain, with much of the provision resulting from patients seeking help from independent acupuncturists.[1] Acupuncture is rarely available as a referral option within NHS mental health services or primary care.[2] Pain and depression often appear to be experienced concurrently, with around 50% of individuals who are diagnosed with and treated for depression also presenting with painful symptoms.[92] Although these data formed a basis for further investigation, the evidence base from systematic reviews in the mid-2000s suggested that there was insufficient evidence to draw conclusions.[93,94] The evidence for pharmacological antidepressant treatment also raised some concern at the time, with pharmacological antidepressant treatment being associated with up to 33% of patients not showing an adequate response.[95] Moreover, 30% of patients have been found to not adhere to their medication regime.[96] An over-reliance on prescribed antidepressant medications has also been identified by patients, who also report that they are interested in being offered more of a range of possible treatment choices.[97]

The focus of the work in *Chapter 6* was the evaluation of the clinical effectiveness and cost-effectiveness of acupuncture or counselling for depression when offered in primary care as an adjunct to usual GP care. It is accepted that the question of whether or not acupuncture is more than simply a placebo is important; however, we were reluctant to use a form of sham acupuncture as a control for reasons addressed above regarding the lack of an adequate understanding of the mechanism of acupuncture, leading to a difficulty in interpretation. Moreover, the feasibility of implementing a sham acupuncture arm would be challenging, given the lack of institutional support if acupuncture was to be delivered in the field. For these reasons we opted for a pragmatic design that built on our pilot RCT[98] and used non-directive counselling as an active control. Our rationale for this was based on the following: (1) counselling is a credible and widely used intervention for patients with depression; (2) there is structural equivalence between acupuncture and non-directive counselling in terms of contact time (1-hour sessions) with empathetic practitioners and therefore if acupuncture performs better than counselling the difference is unlikely to be because of the effects of time or quality of attention; (3) this trial design would help inform patients, decision-makers and providers of the relative merits of counselling compared with acupuncture; and (4) the most recent Cochrane systematic review at the time proposed the wider use of non-placebo comparative designs when evaluating acupuncture for depression – future studies may need to consider the use of comparative designs using medication or structured psychotherapies (cognitive–behavioural therapy, psychotherapy, counselling) or standard care, due to the ethics of administering this intervention to this study population'.[94] It is this design that formed the basis for the RCT described in *Chapter 6*.

The overarching aim of this programme of research was to use high-quality methods, and innovative ones if necessary, to develop the evidence base on acupuncture. The widespread utilisation of acupuncture combined with insufficient confidence regarding outcomes and decision-making provides a research imperative that is in the public interest. An important question has been asked regarding the extent that

acupuncture is simply a remarkably effective placebo as opposed to a physiologically active and scientifically proven intervention. This programme of research provides the latest evidence from high-quality trials that have been carefully designed to answer this question. Innovative research has been conducted, in particular the IPD meta-analysis (see *Chapters 2*, *4* and *5*) and the network-meta-analysis (see *Chapters 3–5*). As for any research endeavour, not all of the concerns and questions are answered within this report. For example, we do not directly address the placebo question with regard to acupuncture for depression (see *Chapter 6*); however, we do address this question rigorously, and in some depth, when conducting reviews of the literature on the evaluation of acupuncture for chronic pain in IPD meta-analyses (see *Chapter 2*) and for osteoarthritis of the knee in a network meta-analysis (see *Chapter 3*). We did not take into account all competing therapies for the cost-effectiveness analyses of acupuncture for musculoskeletal pain and for headache and migraine (see *Chapter 3*), although we do so for osteoarthritis of the knee (see *Chapter 5*). Overall, our focus has been on delivering results that inform patients, providers and decision-makers, which required us to assess whether or not acupuncture outperforms a placebo, if there is a clinically relevant change in clinical status and whether or not there is a sufficiently robust economic case, ideally based on comparisons with all other competing therapies.

# Chapter 2 Acupuncture for chronic pain: an individual patient data meta-analysis

## Background

An estimated 4 million acupuncture treatments are provided each year in the UK and the most common reasons for consulting are related to chronic pain.[1] Despite this widespread use, there remains uncertainty regarding the clinical effectiveness of acupuncture, and particularly the effectiveness of acupuncture over and above that of sham acupuncture. Many RCTs of acupuncture for chronic pain have been conducted. Most of these trials are methodologically poor in quality, which in turn leads to difficulty in interpreting their results in meta-analyses. Moreover, there has been some controversy regarding the role of sham acupuncture and concerns have been raised that the differences found in these trials between acupuncture and sham acupuncture have been either negligible or sufficiently small to be of little value. Indeed, some commentators have suggested that acupuncture is entirely a theatrical placebo,[99] whereas others claim that any putative differences between acupuncture and sham acupuncture are vanishingly small, tending to zero when issues of bias are fully addressed.[49] It is in this climate of uncertainty that the opportunity has arisen to resolve important questions regarding the true effect of acupuncture.

The recent growth in the number[100] of RCTs of acupuncture, and improvement in quality, have provided a further rationale regarding the timing of this project to establish more robust evidence. Much of the clinical trial-based research has evaluated the effectiveness of acupuncture for typical chronic pain conditions that commonly occur in primary care.[101] Our method of choice to synthesise these data was an IPD meta-analysis, which is a superior method to conventional meta-analysis using summary data. In the words of Iain Chalmers, one of the founders of the Cochrane Collaboration, using IPD in a meta-analysis is the 'yardstick' by which all meta-analyses should be measured.[78] Compared with traditional reviews, which analyse summary data that have already been published, the advantages of using IPD are as follows:[102]

1. Standardisation is possible between different analytical approaches. Some trials of acupuncture have reported mean change in pain whereas others have reported 'response rates' relating to the proportion of patients who experienced a threshold reduction in pain (e.g. 33%). These results cannot be combined without access to the raw data, which allows conversion from one type of analysis to another.
2. Greater power is available in the application of statistical methods. In a typical meta-analysis, the investigator records the means and standard deviations (SDs) for the acupuncture and control groups separately. This does not allow the application of techniques such as analysis of covariance (ANCOVA), which have greater statistical power than unadjusted analysis.[103,104]
3. When analysing associations between patient-level characteristics and outcomes, IPD analyses have far greater power to investigate questions such as whether age or baseline symptom severity influence outcome. As an example, if there were four trials with 250 patients in each, analysis of published data would attempt to correlate four values of a predictor (e.g. mean age in each trial) with four values of an outcome (e.g. difference between mean pain scores). Analysis of IPD would be able to create a model with 1000 data points.
4. With regard to data quality, the process of combining data from different sources requires careful data scrutiny by an independent investigator, which provides an opportunity to identify and correct errors in the data set.

The updating of results is an issue of particular importance for trials with longer-term outcomes as data continue to accrue on a daily basis after publication. It is possible that acupuncture triallists have data from long-term follow-up that have yet to be published. Because this method of meta-analysis uses superior

statistical methods, it leads to greater precision in the results. In summary, we have optimised the synthesis of existing acupuncture trials by conducting an IPD meta-analysis, including only the highest-quality trials to enhance the quality of the resulting evidence.

To this end, the Acupuncture Triallists' Collaboration (ATC) was established to manage this project. Collaborators included a group of triallists, statisticians and other researchers with the goal of sharing raw data and developing, in partnership, a set of research questions and associated analytical strategies. The group was motivated to work together to help break down the oppositional culture of competing triallists, to share data in a robust scientific collaboration and to help translate clinical trial findings into patient benefit. Lead investigators from each of the eligible trials contributed raw data, which then were combined into a single data set. This data set was then analysed to address questions concerning the management of chronic pain conditions. The full protocol of the meta-analysis has been published.[105] The study was conducted in three phases: (1) identification of eligible RCTs; (2) collection, checking and harmonisation of raw data; and (3) the IPD meta-analysis.

Our primary objectives, which are addressed in this chapter, were identified as follows:

1. To conduct a systematic review to identify high-quality trials of acupuncture for common chronic pain conditions and then establish a single individual patient-level database of raw data from these trials. This database provided the opportunity to address several key questions in acupuncture research. Our plan is to then publish the database for the benefit of the acupuncture research community.
2. To determine whether or not real acupuncture is superior to sham acupuncture for the treatment of common chronic pain conditions and, if so, to determine the effect size. 'Real acupuncture' was defined as the acupuncture intervention that is designed to have activity against pain. 'Sham acupuncture' was defined as a comparator intervention that is designed to mimic real acupuncture, with the patient not knowing whether he or she has received real acupuncture, and which ideally has no acupuncture-specific effects.
3. To determine whether or not real acupuncture is superior to non-acupuncture controls for the treatment of common chronic pain conditions and, if so, to determine the effect size. 'Non-acupuncture controls' were defined to include care, such as medication 'as needed', that is also received by the acupuncture group. Non-acupuncture controls were sometimes described as waiting-list controls, usual or standard care controls or controls receiving no additional treatment. Attention control, in which patients receive general education and advice, was also included in this category.

Within this chapter we also address two secondary objectives in two substudies. In the first substudy we determined the influence of the control group on the effect size of acupuncture. We first identified the variations in types of sham and non-sham controls used and then analysed their impact on effect size. This substudy will inform the design of trials that evaluate acupuncture, as the choice of control will help inform aspects of the design such as sample size. In the second substudy we analysed the data set to determine whether there are characteristics of acupuncture or acupuncturists that act as effect modifiers for treatment outcome.

By meeting both primary and secondary objectives, it is hoped that the evidence generated by this collaboration will have important implications for both clinical practice and research. IPD meta-analysis of high-quality trials provides the most reliable basis for treatment decisions. Analyses of the impact of different sham techniques, styles of acupuncture or frequency and duration of treatment sessions can be expected to guide future clinical trials of acupuncture.

## Methods

The methods related to the primary objectives for this study are described in three phases below, after which the methods for the two substudies that address the secondary objectives are described.

### Phase I: systematic review to identify eligible trials

## Trial quality criteria for trial eligibility

In terms of methodological quality, unconcealed allocation is the most important source of bias in RCTs.[106,107] To be included in the study, a key criterion therefore was the requirement that RCTs of acupuncture for chronic pain conditions had to have unambiguously concealed allocation of the randomisation sequence. When this was not clear from the published paper, we contacted the trial authors for further information concerning the exact logistics of the randomisation process. We considered allocation to be adequately concealed if both of the following two conditions held: (1) the researchers were unable to predict the group to which a patient would be randomised until the patient was explicitly registered on study and (2) the researchers were unable to change a patient's allocation after a patient was randomised. Allocation concealment was considered inadequate if participants or investigators enrolling participants could possibly foresee or modify assignments and, thus, introduce selection bias. Researchers had to have established clear procedures to ensure that these two conditions were met. For example, there should have been procedures to prevent investigators resealing and reusing an envelope after it was opened (e.g. envelopes were held by an independent party).

## Patient criteria for trial eligibility

Trials were eligible if the patient population was recruited on the basis of pain conditions related to osteoarthritis, chronic or recurrent headaches (e.g. tension or migraine headaches), specific and non-specific shoulder pain, and non-specific back or neck pain. Trials were excluded when the back or neck pain was associated with specific pathologies (e.g. osteoporotic fracture). Trials of shoulder pain were included when the pain was associated with specific pathologies (e.g. rotator cuff tendonitis, frozen shoulder or bursitis). As the main analyses were conducted separately by indication, we did not expect to identify more than one or two eligible trials for other pain conditions. For osteoarthritis and headache pain we did not require a specific pain duration, as both are chronic in nature. Back, neck and shoulder pain are commonly episodic conditions and we used the frequently employed criterion for chronicity that the current episode must be of at least 4 weeks' duration.

## Intervention criteria for trial eligibility

Trials were included provided patients in at least one trial arm received acupuncture in the form of penetrating needles at either acupuncture points or trigger points. Trials were classed as ineligible if patients in the acupuncture group, but not the control group, were protocolled to receive medication (conventional or otherwise), surgery or physical therapy. With regard to control groups, eligible trials needed to have included at least one group receiving sham acupuncture or a non-acupuncture control intervention. Sham acupuncture was defined as any intervention designed to prevent the patient from knowing whether he or she received real acupuncture but which was thought by researchers to have minimal activity against pain. Variations of sham acupuncture included variations of superficial needle insertion; needle insertion at non-acupuncture points or at points not indicated for the condition under study; 'placebo' needles such as the Streitberger needle,[108] which act like stage daggers, appearing to penetrate the skin but which do not do so; techniques such as tapping on a guide tube, designed to feel like needle penetration; and non-needle methods, such as detuned lasers or deactivated transcutaneous electrical nerve stimulation (TENS) devices. It is worth noting that we did not consider these controls to be equivalent a priori; possible differences between sham procedures were analysed as one of our objectives.

Trials with non-acupuncture control groups were included provided the care received in the control group was defined as any of the following: trials with a waiting-list control; trials in which patients received usual clinical care in both arms of the trial, for example a study in which the effects of a course of physiotherapy plus acupuncture were compared with the effects of physiotherapy alone; trials in which the intervention in the control group involved general advice, education and support (sometimes described as 'attention control'); and trials in which the control group, but not the acupuncture group, received recommendations for guideline care, although no specific treatment plan was mandated and no treatment was provided by the trial. As with sham acupuncture, we did not expect these different types of non-acupuncture controls

to have equivalent effects, but we included these different types of control in our analyses and we also investigated differences between them. Trials were excluded if the control groups received a specific programme of treatment such as medication, massage or physical therapy in addition to sham acupuncture or treatments also available in the true acupuncture group.

## Outcome criteria for trial eligibility

For eligibility, trials were required to have a measure of pain. The primary end point must have been measured > 4 weeks after the end of the initial acupuncture treatment. There was no restriction on eligibility because of the type of end point.

## Trial size and language for trial eligibility

For inclusion, there was no restriction on the size of the trial. We also had no language exclusions. All papers in languages other than English were translated into English and the English text made available to all collaborators.

## Search strategy for identification of trials

We searched MEDLINE, Cochrane Central Register of Controlled Trials and the citation lists of systematic reviews. The search strategy used (detailed in *Appendix 1*) was the same as for the previous reviews of headache,[109] back pain[57] and osteoarthritis[110] (each of which was coauthored by one or more members of the ATC) with the addition of the following terms: 'neck', 'shoulder', 'cervical' and 'musculoskeletal'. Searching established databases for trials conducted in China or published in the Chinese language was expected to have involved very poor precision as few of these studies are of sufficient quality to merit inclusion.[111–114] Accordingly, Chinese trials were identified by a separate process: Jianping Liu of the Chinese Cochrane Centre in Beijing used that institution's resources to identify trials of acupuncture for chronic pain that involved full allocation concealment.

## Inclusion of trials

All retrieved references were scanned by one of two investigators to remove any clearly inappropriate titles. Hard copies of all remaining papers were then obtained and read by both investigators to remove any for which there was no possibility of eligibility. Inclusion criteria for the remaining papers were applied by two reviewers separately (no reviewer assessed a trial on which he or she was listed as a coauthor). Disagreements about study inclusion were resolved by consensus. Authors of trials were contacted, if necessary, to clarify details such as allocation concealment if this was not clear. All retrieved trials that were excluded from the review were given a reason for exclusion as follows: not a randomised trial; allocation unclear or inadequate; not acupuncture; inappropriate control; not pain; only short-term measurement of pain; or not an osteoarthritis, headache, back, neck or shoulder pain trial.

## Quality assessment

With regard to potential bias, the most important quality criteria for a RCT concerned the quality of randomisation, blinding and exclusions and dropouts.[106,107] The quality of randomisation was an inclusion criterion for this study: only trials with full allocation concealment were included in the analysis. Exclusions and dropouts were dealt with by multiple imputation in the statistical analysis. Hence, our quality assessment focused on blinding. For all studies involving sham acupuncture, assessment of blinding followed that in previous Cochrane reviews, with grading as A, B or C. In this categorisation, A represented a low likelihood of bias: either the adequacy of blinding was checked by direct questioning of patients, for example with a credibility questionnaire, and no important differences were found between groups or a blinding method was used that had previously been validated as being able to maintain blinding (e.g. the Streitberger sham device[108]). A categorisation of C represented a high likelihood of bias: there were clear reasons to believe that blinding was broken, for example differential responses to a credibility questionnaire or an obviously non-credible sham technique was used. Between these two categorisations, category B represented an intermediate likelihood of bias: a trial that did not meet the criteria for a grade of either A or C. Quality assessment was conducted by two reviewers separately with disagreements resolved by consensus.

### Phase II: collection, checking and harmonisation of data

## Development of the database

We sought IPD for all of the included trials, which we entered into a single database. Data were obtained for all randomised patients, regardless of whether they received treatment or provided post-randomisation data. Trial-level data were then added to the individual patient records. For example, a data set for a trial might have an indicator variable for acupuncture compared with control. This was replaced by several variables indicating the type of acupuncture and control as described in the trial report. When raw data were not available for a trial, we conducted sensitivity analyses to determine if inclusion of the trial might alter the results.

## Initial data manipulation

The raw data were saved in their original format and then converted to a Stata format (version 11; StataCorp LP, College Station, TX, USA). Three blank statistical programs were then saved: one to undertake preliminary checks on the data, one to rename and label the variables and one to replicate statistics reported in the trial publication. All files were saved using a standard notation: 'raw data [descriptor]', 'initial import [descriptor]', 'initial set up [descriptor]', 'initial data checks [descriptor]' and 'replication [descriptor]', where '[descriptor]' is a unique label for each data set (e.g. 'Linde 2005 migraine').

## Annotation checks

Statistical code was written for the 'initial set up' program. Each variable in the raw data set was renamed to a standard notation (e.g. 'age_at_randomisation' became 'age') and given a standard label (a label is a text description of the variable, such as 'combined headache score at 60 days', that is stored by the statistical software). Variables unique to a particular data set, for example a Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) score[115] in an arthritis trial, were then identified and labelled. Any variables that could not be identified, or which were ambiguous, were documented and appropriate clarification was sought from the original investigator.

## Checking for erroneous or missing data

Statistical code was written for the 'initial data checks' program. First, the number of missing observations for each variable was calculated and checked against data available in the original publication. Any inconsistencies, or variables for which information on rates of missing data were not available in the trial publication, were brought to the attention of the original investigator for clarification. Second, 'range' checks were conducted on all variables to determine whether or not all values were reasonable. As a trivial example, a visual analogue scale (VAS) of 123, or an age of 567 years, immediately suggested an error. Third, we checked categorical variables by tabulation. For instance, if 200 patients were categorised as having stage I disease, 220 categorised as having stage II disease and one as having stage IIa disease, the investigator would be queried as to the accuracy of the IIa categorisation.

## Replication

The third program 'replication [descriptor]' was then written. This replicated as far as possible every number reported in the trial publication. Replications included baseline characteristics such as age, sex and duration of disease within each group; outcome data such as pain scores within each group at each follow-up time; and comparisons, such as the difference in pain scores between groups at the post-treatment follow-up. In each case, we used the statistical methods reported by the authors and derived the statistics given in the publication. For example, if a mean and SD for baseline pain score were given in the trial publication, we similarly calculated the mean and SD and, if the difference between groups was calculated by linear regression with baseline score and duration of disease as covariates, we used exactly this method to see if we could obtain the same difference between groups, 95% confidence interval (CI) and *p*-value. Any discrepancies between our results and those reported in the published papers were brought to the attention of the investigators for clarification. We considered that any data set that had gone through these checks – independent labelling of every variable, assessment of prima facie errors

and replication of all reported statistics – could be considered valid for inclusion in an independent patient data meta-analysis. Across all trials, the variable names were harmonised.

## Phase III: statistical methods

### Principal end point

For each trial we identified the primary outcome defined by the study authors in terms of both the scale (e.g. WOMAC) and time point (e.g. 6 months after randomisation). We kept end points on a continuous scale. For example, in some studies the primary end point was defined in terms of the proportion of patients who had at least a 35% reduction in the number of days with headache pain at 6 months' follow-up. In this case, the primary end point was specified as the number of days with pain at 6 months. If multiple criteria were considered in the primary outcome, or if the primary outcome was inherently categorical, we used a continuous measure of pain measured at the same time point as the original primary end point. For example, if a trial's primary outcome was a response to treatment defined as a given degree of improvement on a pain scale or a function scale, we selected the pain scale for inclusion in our primary analysis. If there were multiple pain measurements we selected one according to the outcome measures preference list (see *Appendix 1*). For analyses that included trials with different primary end points, we created a standardised primary end point by dividing by the SD.

### Primary analysis: analysis of the effect size of acupuncture

Each trial was reanalysed by ANCOVA with the standardised principal end point as the dependent variable and the baseline principal end point and variables used to stratify randomisation as covariates. This approach has been shown to have the greatest statistical power for trials in general with baseline and follow-up measures,[104] and also when specifically applied to acupuncture research.[103] For trials in which randomisation was stratified by centre or practitioner, this stratification was included in the analysis only if there were ≤ 20 sites and there was a mean of at least 20 patients per site, with at least one patient in each arm at each site. In trials in which there was more than one acupuncture group, for example trials in which patients were randomised to local points, distal points or sham points, results from all real acupuncture groups were combined (local points and distal points in this case). The standardised mean between-group difference (effect size) for acupuncture from each trial [i.e. the coefficient and standard error (SE)] was then entered into a meta-analysis; the meta-analytic statistics were created by weighting each coefficient by the reciprocal of the variance, summing and dividing by the sum of the weights. Meta-analysis was accomplished using the *metan* command in Stata 11.

Our primary analysis was a fixed-effects model. Our rationale for a fixed-effects analysis was that it constituted a valid test of the null hypothesis of no treatment effect. Moreover, we have taken the view that the use of a fixed-effects model does not imply an assumption that all trials are estimating the same effect, but that the robustness of the fixed-effects approach is likely to lead to a more accurate estimate. Nonetheless, we also report the results of the random-effects analysis. In addition, we report heterogeneity statistics.[116] We computed effect sizes separately for comparisons of acupuncture with sham and non-acupuncture controls. Comparisons between acupuncture and sham controls omitted trials graded as category C (high likelihood of bias) because of concerns regarding blinding. These analyses were conducted separately for each pain condition (specific shoulder conditions, musculoskeletal pain, osteoarthritis, headache) and then within each pain condition (neck pain and back pain or chronic tension headache and migraine).

### Secondary analyses

We repeated the analyses of effect size for the secondary end points of pain intensity, pain frequency, functional impairment, combined measures of pain and functional impairment, mental well-being [e.g. Short Form questionnaire-36 items (SF-36) mental health], physical well-being (e.g. SF-36 physical health), overall quality of life (e.g. global assessment), range of motion or stiffness, health change, satisfaction with care and medication use. If a trial reported more than one end point that could be placed in a particular category, the outcome measures preference list (see *Appendix 1*) was consulted to select the most

appropriate measure. On occasion, this could have involved taking a mean score of two end points. For example, if a trial reported both a daytime and a night-time VAS score, we calculated the average for each patient and the combined score was then entered into the analysis. Note that this demonstrates a key advantage of IPD meta-analysis: such a data manipulation would not be possible with summary-level data. As different measurement scales were used in the different trials, we used the standardised mean difference (SMD) as the meta-analytic statistic. Time was always measured from randomisation. For our data, we used end points of 1, 2, 3, 6 and 9 months and 1 year. For outcomes with these exact time points (or the equivalent in another unit of time: 13 weeks = 3 months), no time point standardisation was required. Otherwise, the time point closest to the selected scheme was adopted. For example, if for a trial there was no measurement at 6 months but there was for 24 weeks, this time point was selected and relabelled appropriately. Numerical rating scale (NRS) scores were converted to a 0–100 point scale by appropriate multiplication.

## Sensitivity analyses

The first sensitivity analysis involved multiple imputation for missing data, following the approach used in the analysis of the NHS trial of acupuncture for headache.[60] The second sensitivity analysis was for publication bias. Although we did not believe that there were many unpublished adequately concealed acupuncture trials large enough to have an important weight in the meta-analysis, we included scenarios that could change the study results. For example, if we found a statistically significant difference between acupuncture and sham, we estimated the parameters for the following scenarios that, if added to the meta-analysis, would change the $p$-value to 0.05: (1) the number of trials with 50 patients per group and no differences between groups; and (2) the number of trials with 50 patients per group and an effect size of 0.25 in favour of the control. The third sensitivity analysis omitted subsets of trials based on trial quality. We omitted trials graded as category B for blinding from the comparison of acupuncture with sham. Our final sensitivity analysis involved adding the results of studies for which we did not receive individual patient-level data. We calculated an estimate of the difference between groups and the resulting SE from published summary data.

## Methods for substudy 1: influence of control group on effect size

### Types of sham acupuncture control

In the included trials with a sham acupuncture control group we assessed whether or not a sham needle was used, whether or not a sham needle that penetrated the skin was used, whether sham needling was performed on true acupuncture points or non-acupuncture points, and whether a sham needle insertion was deep or superficial. Information on acupuncture characteristics was obtained from the trial manuscript supplemented by a questionnaire sent to triallists.

When trial authors reported using either a penetrating or a non-penetrating needle for sham acupuncture, the trial was classified as using a 'needle sham'. Trials using non-needle methods of sham acupuncture, such as an inactivated laser or a TENS device, were classified as 'non-needle sham'. Needle sham trials were further classified as using penetrating needles, which were almost always inserted at locations away from true acupuncture points (thereby investigating point location), or using non-penetrating sham needles, which were applied either at the same points as in the true acupuncture group (testing exclusively skin penetration and not location) or at non-acupuncture points (investigating penetration and location simultaneously).

We had hoped to include two other features of sham controls: whether the depth of insertion for penetration was categorised by triallists as superficial or deep and whether sham acupuncture was applied at or away from true acupuncture points; however, only one trial reported using deep insertion in sham acupuncture.[117] For point location, there was strong collinearity with sham technique, with only techniques avoiding skin penetration using true acupuncture points. As a sensitivity analysis, we reanalysed the data excluding four trials that were determined by consensus among external reviewers to have an 'intermediate likelihood of unblinding'.[72,117–119] However, after excluding these trials, only one remaining trial used non-needle sham acupuncture, limiting our ability to use metaregression.

## Types of non-sham control

We categorised trials that included controls without sham acupuncture into two types: 'routine care' and 'protocolled care'. In 'routine care' trials, both treatment and control groups had access to non-specified care as needed, such as rescue medications or other conventional care, but the use of such treatment was at the discretion of patients and doctors, with no specification in the protocol as to what treatments patients could receive. If protocols prescribed some treatments such as surgery but did not make specific recommendations as to allowable treatments, trials were defined as 'routine care'. Control groups in which treatment consisted of information or education given to a patient ('attention control') were also considered to be routine care control groups.

In 'protocolled care' trials, the care in the control group was specified in the study protocol. This was typically when the acupuncture group and the usual care control group both received an additional non-acupuncture treatment that was specifically indicated as part of the trial protocol. For example, trials that studied the effect of acupuncture and physical therapy compared with physical therapy alone were categorised as protocolled care.

## Statistical methods related to effect of control group

To test the effect of each characteristic of sham acupuncture on the main effect estimate, we used random-effects metaregression with the Stata command *metareg*. This command was also used to run a random-effects metaregression to test the effect of routine compared with protocolled care on the main effect estimate for usual care control groups. The main effect estimate of each trial was determined using linear regression, and the coefficient and SE for each trial were entered as the dependent variable in the random-effects metaregression.

All analyses were conducted using Stata 12. We excluded three trials by Vas *et al.*[120–122] in a sensitivity analysis. As described in *Meta-analysis*, these trials had very much larger effect sizes than average and their exclusion resulted in heterogeneity becoming non-significant in the comparisons between acupuncture and sham acupuncture. More detail on this substudy's methods is reported separately.[45]

## *Methods for substudy 2: characteristics of acupuncture*

## Data included at the trial and patient level

Data on the trial and patient-level characteristics of the acupuncture interventions were obtained directly from responding triallists by use of a questionnaire and are presented in *Appendix 2*. Characteristics investigated included the style of acupuncture, which was defined as based on traditional Chinese theory or contemporary Western acupuncture or a mixture of both approaches. Point prescriptions were defined as fixed, flexible or individualised. Trials were categorised as using a flexible needle formula if triallists indicated that acupuncture was semistandardised, a flexible formula with fixed points or both fixed and flexible formulas. Triallists reported whether or not their trial allowed electrical stimulation to be added to the needles during acupuncture sessions and whether or not moxibustion was allowed. Trial-level information was reported on whether or not acupuncturists attempted to elicit deqi and whether it was felt by the acupuncturist or the patient. Triallists reported on whether or not acupuncture-specific interactions between the patient and the acupuncturist were allowed, for example through explanations of treatment, advice, support and suggestions about helpful lifestyle changes. These interactions, when driven by acupuncture theory, are considered 'specific' to the acupuncture treatment. Triallists reported the minimum number of years of practice as an acupuncturist required to participate in their trial. The maximum number of acupuncture treatment sessions allowed during the trial period was reported by each triallist and these data were analysed per five-session increments. The frequency of sessions was recorded and analysed continuously as a weekly average (i.e. typical number of sessions per week). The duration of sessions was reported as the average length of a session in minutes among the patients receiving acupuncture. Patient-level data were used when available by taking the mean duration of patients' sessions. Trials were not included in this analysis if treatment was individualised and no individual-level data were available. The duration of sessions was included as a continuous variable in the analyses and the results reported per

5-minute increments. Triallists were asked to report the average number of needles used per treatment session. Trials were excluded from this analysis if the number of needles used was unknown. If patient-level data were available an average was included. The average number of needles used was analysed as a continuous variable, with the coefficient reported per five-needle increments. The placement of acupuncture needles was categorised as local (at or near the location of pain), distal to the location of pain or both.

## Statistical methods for analysis of characteristics

We identified the primary outcome as defined by the study authors in terms of both the scale and the time point. We kept end points on the continuous scale. For analyses that included trials with different primary end points, we created a standardised primary end point by dividing by the SD. We conducted analyses separately for sham and non-acupuncture controls using Stata 12.

We used random-effects metaregression for trial-level analyses to test the effect of each characteristic on the main effect estimate using the Stata command *metareg*. We first calculated the effect size and SE for each trial as described in *Statistical methods related to effect of control group*. For each documented treatment characteristic, we entered the effect size and SE for each trial into a metaregression along with the trial-level average for that characteristic. The coefficients obtained from these analyses are estimates, in SDs, of the effect of each acupuncture characteristic on the main treatment effect.

In the patient-level analyses we were able to use the number of sessions, the number of needles and the age and sex of the acupuncturist for a subset of the trials. For each trial we created a linear regression using random effects as for the main analysis of effect size, but included the characteristic and an interaction term between the characteristic and treatment allocation. The coefficient and SE for the interaction term represents the change in the outcome score in SDs associated with the acupuncture characteristic in the acupuncture treatment group. This was then entered into a meta-analysis using the Stata command *metan*.

We excluded a set of outlying trials, all by the same team[120–122] and which had very much larger effect sizes than other trials, as a pre-planned sensitivity analysis. Further details of the methods are reported separately.[123]

# Results

## Results from the main study

### Eligible studies

In our initial search we identified and assessed 83 RCTs for eligibility (*Figure 1*), of which 31 were eligible (for further details of the 29 studies included in the patient-level meta-analysis, see *Appendix 2*). Eleven studies were sham controlled, 10 had non-acupuncture controls and 10 were three-armed studies with both sham and no-acupuncture control arms. A second search for studies was requested by the *Archives of Internal Medicine* prior to publication and therefore subsequent to conducting the meta-analysis; we identified an additional four eligible studies, which were used in a sensitivity analysis.[124–127]

### Data extraction and quality assessment

From the 31 eligible RCTs, usable raw data were obtained from 29 trials including a total of 17,922 patients from the USA, UK, Germany, Spain and Sweden (*Table 1*). For two studies the raw data were unavailable: the study database had become corrupted in one trial[128] and in another, despite approval for data sharing being obtained from the principal investigator, the trial statisticians failed to respond to repeated enquiries.[129]

In terms of design, the 29 RCTs included 18 comparisons of acupuncture with non-acupuncture controls (14,597 patients) and 20 comparisons of acupuncture with sham acupuncture controls (5230 patients). Patients in all RCTs had access to analgesics and other standard treatments for pain. Four sham-controlled
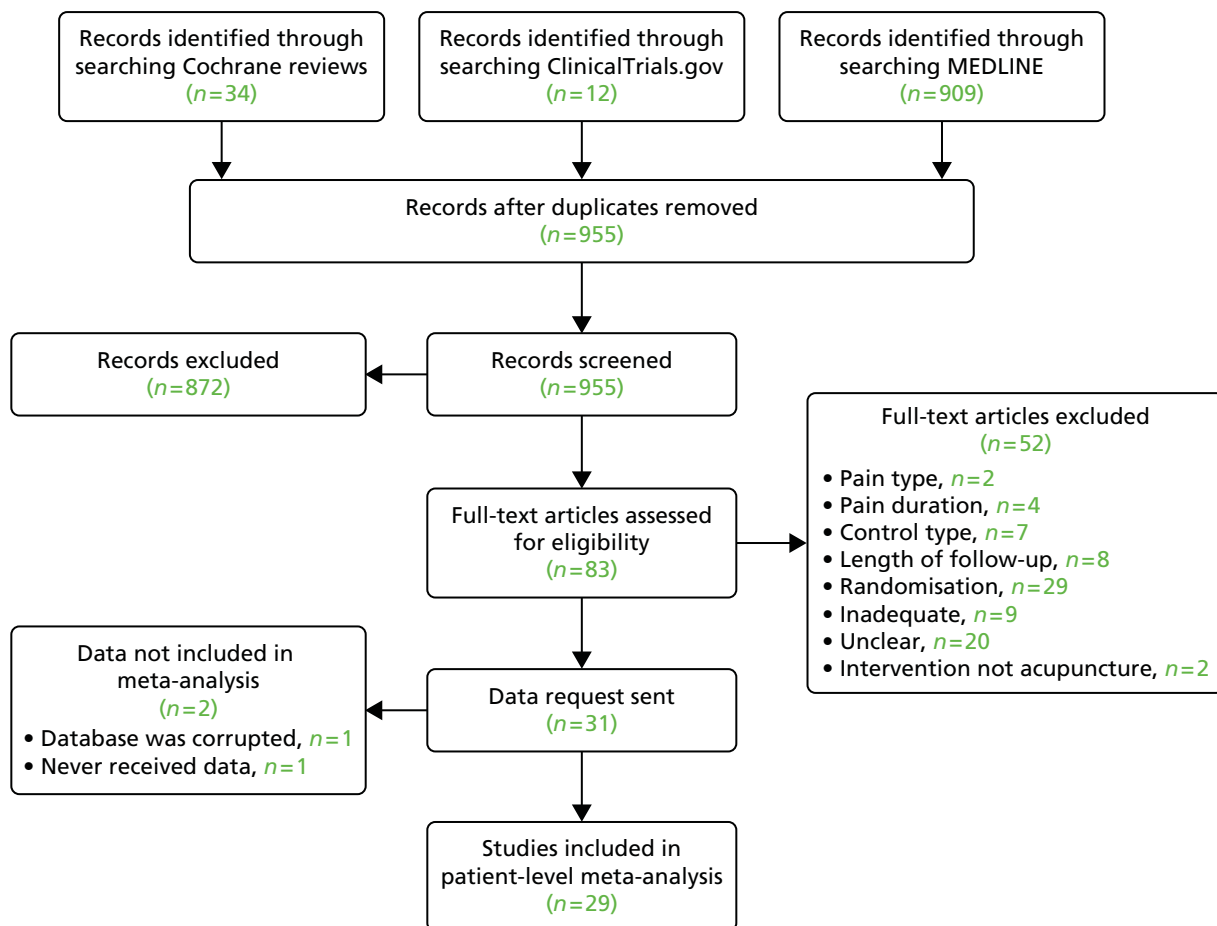
**FIGURE 1** Systematic review Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram.

RCTs were determined to have an intermediate likelihood of bias from unblinding;[72,117–119] the 16 remaining sham RCTs were graded as having a low risk of bias from unblinding. Dropout rates were low on average (weighted mean of 10%), with rates being > 25% for only four RCTs. For the trials by Molsberger et al.[127,129] (dropout rates 27%[129] and 33%[129]), the raw data were not received and neither RCT was included in the main analysis; the other two trials with a dropout rate of > 25% were those by Carlsson and Sjölund[119] (46%, RCT excluded in a sensitivity analysis for blinding) and Berman et al.[134] (31%). This RCT had a high dropout rate among non-acupuncture control subjects (43%); dropout rates were close to 25% in the acupuncture and sham groups. The RCT by Kerr et al.[117] had a large difference in dropout rates between groups (acupuncture 13%, control 33%) but was excluded in the sensitivity analysis for blinding.

Clinical heterogeneity between studies was identified as being related to the control groups. For sham RCTs, the type of sham control varied, including acupuncture needles inserted superficially,[72] sham acupuncture devices with needles that retract into the handle rather than penetrate the skin,[137] and non-needle approaches such as deactivated electrical stimulation[132] or a detuned laser.[118] The cointerventions also varied, with no additional treatment other than analgesics in some RCTs[66] and both acupuncture and sham groups receiving a course of additional treatment, such as exercise led by physical therapists, in other RCTs.[135] For the trials with non-acupuncture control groups, there was variation in usual care, for example control group patients being merely advised to 'avoid acupuncture';[62] an attention control, such as group education sessions;[134] and guidelined care, in which, for example, patients were given advice on specific drugs and doses.[72]

**TABLE 1** Summary of characteristics of included studies

| Indication (n = 35) | Pain type | Control group | Primary outcome measure | Time point |
|---|---|---|---|---|
| Chronic headache (n = 7) | Migraine (n = 2[68,72]); tension-type headache (n = 3[67,73,128]); both (n = 2[60,76]) | Sham (n = 4[67,68,72,73]); no acupuncture control (n = 6) – ancillary care[a] (n = 1[128]), usual care[b] (n = 4[60,67,68,76]), guidelined care[c] (n = 1[68]) | Severity score (n = 2[60,128]); days with headache (n = 1[73]); migraine days (n = 3[67,72,76]); days with moderate to severe pain (n = 1[68]) | 1 month (n = 1[128]); 3 months (n = 3[67,68,76]); 6 months (n = 2[72,73]); 12 months (n = 1[60]) |
| Non-specific musculoskeletal pain (back and neck) (n = 15) | Back (n = 10[61,66,71,74,117,119,124,129–131]); neck (n = 5[75,118,121,132,133]) | Sham (n = 10[66,71,117–119,121,124,129,130,132]); no acupuncture control (n = 9) – ancillary care[a] (n = 1[129]), usual care[b] (n = 6[61,66,74,75,124,133]), non-specific advice[d] (n = 1[131]), guidelined care[c] (n = 1[71]) | VAS (n = 7[66,117–119,121,129,132]); Roland Morris Disability Questionnaire (n = 3[124,130,131]); neck pain and disability (n = 1[75]); Hannover Functional Questionnaire (n = 1[74]); Northwick Park Neck Pain Questionnaire (n = 1[133]); von Korff pain score (n = 1[71]); SF-36 bodily pain (n = 1[61]) | 1 month (n = 4[117,121,131,132]); 2 months (n = 3[66,124,131]); 3 months (n = 5[74,75,129,130,133]); 6 months (n = 2[71,119]); 24 months (n = 1[61]) |
| Osteoarthritis (n = 9) | | Sham (n = 6[69,70,120,125,134,135]); no acupuncture control (n = 8) – ancillary care[a] (n = 3[70,125,135]), usual care[b] (n = 3[69,77,126]), non-specific advice[d] (n = 2[134,136]) | Oxford Knee Score questionnaire (n = 1[136]); WOMAC (n = 2[69,77]); WOMAC pain subscore (n = 6[70,120,125,126,134,135]) | 2 months (n = 2[69,136]); 3 months (n = 4[77,120,125,126]); 6 months (n = 3[70,134,135]) |
| Shoulder pain (n = 4) | | Sham (n = 4[122,127,137,138]); no acupuncture control (n = 1) – usual care[b] (n = 1[127]) | Constant–Murley score (n = 2[122,137]); VAS (n = 2[127,138]) | 1 month (n = 2[122,137]); 6 months (n = 2[127,138]) |

a  Ancillary care: programme of care received by both acupuncture and no acupuncture control groups (e.g. trial comparing physiotherapy plus acupuncture with physiotherapy alone).
b  Usual care: protocol did not specify treatments received in control group (e.g. trials with 'waiting list control').
c  Guidelined care: patients in the control group received care according to national guidelines.
d  Non-specific advice: patients in the control group receive general advice and support ('attention control').
**Note**
More information on the included trials is provided in *Appendix 2*.

## Meta-analysis

The comparisons of acupuncture against no acupuncture controls and against sham acupuncture are shown separately for each of the four pain conditions in forest plots (*Figures 2* and *3*, respectively). Meta-analytic statistics are shown in *Table 2*. Acupuncture was found to be statistically superior to all types of control intervention for all analyses ($p < 0.001$). Effect sizes were larger for the comparison between acupuncture and non-acupuncture controls than for the comparison between acupuncture and sham controls (0.37, 0.26 and 0.15 for comparison with sham controls vs. 0.55, 0.57 and 0.42 for comparison with non-acupuncture controls for musculoskeletal pain, osteoarthritis and chronic headache, respectively).

(a)

| Trial | Year | n | | Coefficient (95% CI) | Weight |
|---|---|---|---|---|---|
| Berman | 2004[134] | 250 | | 0.51 (0.31 to 0.72) | 12 |
| Witt | 2005[69] | 212 | | 1.36 (1.08 to 1.64) | 6 |
| Scharf | 2006[70] | 624 | | 0.51 (0.37 to 0.65) | 26 |
| Witt | 2006[77] | 561 | | 0.72 (0.60 to 0.84) | 37 |
| Foster | 2007[135] | 213 | | −0.01 (−0.23 to 0.20) | 11 |
| Williamson | 2007[136] | 108 | | 0.37 (0.12 to 0.62) | 8 |
| Overall (fixed-effects estimate) | | | | 0.57 (0.50 to 0.64) | 100 |
| Overall (random-effects estimate) | | | | 0.57 (0.29 to 0.85) | |

−0.50 −0.25 0.00 0.25 0.50 1.00

Acupuncture worse     Acupuncture better

(b)

| Trial | Year | Indication | n | Coefficient (95% CI) | Weight |
|---|---|---|---|---|---|
| Vickers | 2004[60] | Both | 301 | 0.27 (0.13 to 0.42) | 10 |
| Jena | 2008[76] | Both | 2776 | 0.46 (0.41 to 0.52) | 72 |
| Linde | 2005[68] | Migraine | 197 | 0.55 (0.30 to 0.81) | 3 |
| Diener | 2006[72] | Migraine | 507 | 0.09 (−0.06 to 0.24) | 9 |
| Melchart | 2005[67] | TTH | 181 | 0.58 (0.38 to 0.78) | 5 |
| Overall (fixed-effects estimate) | | | | 0.42 (0.37 to 0.46) | 100 |
| Overall (random-effects estimate) | | | | 0.38 (0.22 to 0.55) | |

−0.25 0.00 0.25 0.50 1.00

Acupuncture worse     Acupuncture better

**FIGURE 2** Forest plots comparing acupuncture with no acupuncture controls (see *Table 2* for references): (a) osteoarthritis; (b) chronic headache; and (c) musculoskeletal pain. No forest plot is available for shoulder pain as there were fewer than three trials and therefore no meta-analysis was performed. TTH, tension-type headache. (*continued*)

(c)

| Trial | Year | Indication | n | | Coefficient (95% CI) | Weight |
|---|---|---|---|---|---|---|
| Cherkin | 2001[131] | Back | 171 | | 0.24 (0.00 to 0.48) | 2 |
| Brinkhaus | 2006[66] | Back | 214 | | 0.92 (0.62 to 1.22) | 1 |
| Thomas | 2006[61] | Back | 182 | | 0.34 (0.03 to 0.65) | 1 |
| Witt | 2006[74] | Back | 2565 | | 0.43 (0.38 to 0.49) | 46 |
| Haake | 2007[71] | Back | 732 | | 0.56 (0.43 to 0.70) | 7 |
| Salter | 2006[133] | Neck | 21 | | 0.11 (−0.57 to 0.78) | 0 |
| Witt | 2006[75] | Neck | 3118 | | 0.68 (0.63 to 0.74) | 42 |
| Overall (fixed-effects estimate) | | | | | 0.55 (0.51 to 0.58) | 100 |
| Overall (random-effects estimate) | | | | | 0.51 (0.36 to 0.67) | |

−0.50  −0.25  0.00  0.25  0.50      1.00

Acupuncture worse        Acupuncture better

**FIGURE 2** Forest plots comparing acupuncture with no acupuncture controls (see *Table 2* for references): (a) osteoarthritis; (b) chronic headache; and (c) musculoskeletal pain. No forest plot is available for shoulder pain as there were fewer than three trials and therefore no meta-analysis was performed. TTH, tension-type headache.

(a)

| Trial | Year | n | | Coefficient (95% CI) | Weight |
|---|---|---|---|---|---|
| Berman | 2004[134] | 283 | | 0.27 (0.07 to 0.47) | 18 |
| Vas | 2004[120] | 88 | | 1.21 (0.92 to 1.49) | 9 |
| Witt | 2005[69] | 218 | | 0.41 (0.18 to 0.63) | 14 |
| Scharf | 2006[70] | 678 | | 0.11 (−0.02 to 0.25) | 41 |
| Foster | 2007[135] | 220 | | −0.07 (−0.28 to 0.14) | 17 |
| Overall (fixed-effects estimate) | | | | 0.26 (0.17 to 0.34) | 100 |
| Overall (random-effects estimate) | | | | 0.37 (0.03 to 0.72) | |

−0.25  0.00  0.25  0.50      1.00

Acupuncture worse        Acupuncture better

**FIGURE 3** Forest plots comparing acupuncture with sham acupuncture (see *Table 2* for references): (a) osteoarthritis; (b) chronic headache; (c) musculoskeletal pain; and (d) shoulder pain. TTH, tension-type headache. (*continued*)

FIGURE 3 Forest plots comparing acupuncture with sham acupuncture (see *Table 2* for references): (a) osteoarthritis; (b) chronic headache; (c) musculoskeletal pain; and (d) shoulder pain. TTH, tension-type headache.

**TABLE 2** Primary analyses of effect sizes

| Indication | n | Fixed effects (95% CI) | Random effects (95% CI) | p-value for overall effect |
|---|---|---|---|---|
| *Acupuncture vs. sham acupuncture* | | | | |
| Non-specific musculoskeletal pain (back and neck)[66,71,117–119,121,130,132] | 8 | 0.37 (0.27 to 0.46); heterogeneity: $p < 0.001$ | 0.52 (0.14 to 0.90) | 0.001 |
| Osteoarthritis[69,70,120,134,135] | 5 | 0.26 (0.17 to 0.34); heterogeneity $p < 0.001$ | 0.37 (0.03 to 0.72) | 0.001 |
| Chronic headache[67,68,72,73] | 4 | 0.15 (0.07 to 0.24); heterogeneity: $p = 0.3$ | 0.15 (0.05 to 0.24) | 0.001 |
| Shoulder pain[122,137,138] | 3 | 0.62 (0.46 to 0.77); heterogeneity: $p = 0.4$ | 0.62 (0.46 to 0.77) | 0.001 |
| *Acupuncture vs. no acupuncture control* | | | | |
| Non-specific musculoskeletal pain (back and neck)[61,66,71,74,75,131,133] | 7 | 0.55 (0.51 to 0.58); heterogeneity: $p < 0.001$ | 0.51 (0.36 to 0.67) | 0.001 |
| Osteoarthritis[69,70,77,134–136] | 6 | 0.57 (0.50 to 0.64); heterogeneity: $p < 0.001$ | 0.57 (0.29 to 0.85) | 0.001 |
| Chronic headache[60,67,68,72,76] | 5 | 0.42 (0.37 to 0.46); heterogeneity: $p < 0.001$ | 0.38 (0.22 to 0.55) | 0.001 |
| Shoulder pain | 0 | No trials | | |

**Note**
Effect sizes are standardised differences.

The test for heterogeneity was statistically significant for five of the seven analyses. In the case of comparisons with sham acupuncture, the RCTs by Vas *et al.*[120–122] are clear outliers. For example, the effect size in the RCTs by Vas *et al.* for neck pain is about five times greater than the meta-analytic estimate. One effect of excluding these RCTs in a sensitivity analysis (*Table 3*) was that there was no significant heterogeneity in the comparisons between acupuncture and sham acupuncture.

Moreover, the effect size for acupuncture became relatively similar for the different pain conditions: 0.23, 0.16 and 0.15 against sham control and 0.55, 0.57 and 0.42 against non-acupuncture control for back and neck pain, osteoarthritis and chronic headache, respectively (fixed effects; results similar for the random-effects analysis).

To understand what these effect sizes mean in real terms, a baseline pain score on a 0–100 scale for a typical RCT might be 60, for example. Given a SD of 25, follow-up scores would average 43 in a no-acupuncture group, 35 in a sham-acupuncture group and 30 in patients receiving true acupuncture. If average responses were defined in terms of a pain reduction of $\geq 50\%$, response rates would be approximately 30%, 42.5% and 50%, respectively.

There is evidence of heterogeneity in the comparisons with non-acupuncture controls, which can be explained by the differences in the types of control groups used. In the case of osteoarthritis, the largest effect size was in the study by Witt *et al.*,[69] in which patients in the waiting-list control group received only rescue pain medication, and the smallest was in the study by Foster *et al.*,[135] which involved a programme of exercise and advice led by physical therapists. For the musculoskeletal analyses, heterogeneity was driven by two very large RCTs[74,75] ($n = 2565$[74] and $n = 3118$[75]) for back and neck pain. If only back pain is considered, heterogeneity is dramatically reduced and is again driven by one RCT, that by Brinkhaus *et al.*[66] with a waiting-list control. In the headache meta-analysis, the study by Diener *et al.*[72] reported much smaller differences between groups. This RCT involved providing drug therapy according to national guidelines in the non-acupuncture control group, including initiation of beta-blockers as migraine

**TABLE 3** Sensitivity analyses

| Sensitivity analysis | Indication | n | Fixed effects (95% CI) | Random effects (95% CI) | p-value for overall effect |
|---|---|---|---|---|---|
| *Acupuncture vs. sham acupuncture* | | | | | |
| Exclusion of Vas *et al.*[120–122] trials | Non-specific musculoskeletal pain | 7 | 0.23 (0.13 to 0.33); heterogeneity: $p = 0.51$ | 0.23 (0.13 to 0.33) | 0.001 |
| | Osteoarthritis | 4 | 0.16 (0.07 0.25); heterogeneity: $p = 0.15$ | 0.17 (0.00 to 0.35) | 0.001 |
| | Shoulder pain | | Fewer than three trials | | |
| Separate pain types | Back pain | 5 | 0.20 (0.09 to 0.31); heterogeneity: $p = 0.4$ | 0.20 (0.09 to 0.32) | 0.001 |
| | Neck pain | 3 | 0.83 (0.64 to 1.01) heterogeneity: $p < 0.001$ | 0.82 (–0.11 to 1.75) | 0.001 |
| *Acupuncture vs. no acupuncture control* | | | | | |
| Separate pain types | Back pain | 5 | 0.46 (0.40 to 0.51); heterogeneity: $p = 0.004$ | 0.49 (0.33 to 0.64) | 0.001 |
| | Neck pain | 0 | No trials | | |

prophylaxis. There was disagreement within the collaboration about whether or not this constituted an active control. Excluding this RCT reduced evidence of heterogeneity ($p = 0.04$) but had little effect on the effect size (0.42–0.45).

Pre-specified sensitivity analyses found no substantive effect on our main estimates of either restricting the sham RCTs to those with a low likelihood of unblinding or adjusting for missing data. There was also little impact of including summary data from RCTs for which raw data were not obtained (two RCTs)[128,129] or which were published recently (four RCTs),[124–127] either for the primary analysis or the analysis with the outlying RCTs by Vas *et al.*[120–122] excluded (data not shown).

To address whether or not publication bias might be involved, we entered all RCTs into a single analysis and compared the effect sizes from small and large studies.[139] We found some evidence that small studies had larger effect sizes overall for the comparison with sham acupuncture ($p = 0.02$) but not for the comparison with non-acupuncture controls ($p = 0.72$). However, these analyses were influenced by the outlying RCTs by Vas *et al.*,[120–122] which were smaller than average, and by indication, because the shoulder pain RCTs were small and had large effect sizes. Tests for asymmetry were non-significant when we excluded the RCTs by Vas *et al.*[120–122] and shoulder pain studies ($n = 15$; $p = 0.07$), and when small studies were also excluded ($n < 100$ and $n = 12$; $p = 0 30$). Nonetheless, we repeated our meta-analyses excluding the RCTs with a sample size of $n < 100$. This had essentially no effect on our results. We also considered the possible effect on our analysis if we had failed to include high-quality unpublished studies in terms of publication bias. There would need to be 47 unpublished RCTs with $n = 100$ patients showing an advantage of acupuncture over sham acupuncture of 0.25 SDs before the difference between acupuncture and sham acupuncture would lose its significance.

The effect of pooling different end points measured at different periods of follow-up was the focus of a further sensitivity analysis. We repeated our analyses including only pain end points measured at 2–3 months after randomisation. We found no material effect on the results: effect sizes increased by 0.05–0.09 SDs for musculoskeletal and osteoarthritis RCTs, and were otherwise stable.

We compared sham control with no-acupuncture control in an exploratory analysis. In a meta-analysis of nine RCTs,[66–72,134,135] the effect size for sham control was 0.33 (95% CI 0.27 to 0.40) and 0.38 (95% CI 0.20 to 0.56) for fixed- and random-effects models, respectively (for tests of both effect and heterogeneity, $p < 0.001$).

## Results for substudy 1: influence of the control group on effect size

### Sham acupuncture controls

Twenty trials with 5230 patients included a control arm in the form of sham acupuncture (*Table 4*) and the trial-level characteristics for these trials are described in *Table 5*. The majority of sham-controlled trials (80%) used needle-based sham acupuncture. The number of trials using penetrating or non-penetrating needles was similar: seven trials used non-penetrating needles and nine trials used penetrating needles. All trials using penetrating needles placed these outside the true acupuncture points, whereas only one of seven trials using non-penetrating needles did so.

The effect sizes for sham-controlled acupuncture trials, as categorised by the type of sham, are shown in *Table 6*. Acupuncture was significantly superior to sham irrespective of the type of sham control, both in the main analysis and in a sensitivity analysis excluding outlying studies. This table also includes the results of the primary sensitivity analysis that excluded the Vas *et al.*[120–122] trials as outliers. Overall, we found that larger effect sizes were associated with acupuncture compared with non-penetrating sham needles (0.43, 95% CI 0.01 to 0.85) than with penetrating sham needles (0.17, 95% CI 0.11 to 0.23), although the difference between groups did not reach conventional levels of statistical significance.

Comparisons between types of sham control are provided in *Table 7*, which shows the results of the random-effects metaregression for sham-controlled trials. Although trials that used needles did not differ significantly from trials that used a non-needle sham control ($p \geq 0.2$ for all comparisons), there was clear evidence of a greater effect size when acupuncture was compared against non-penetrating sham than when compared against penetrating sham. Trials using a penetrating needle had an effect size that was –0.21 (95% CI –0.41 to –0.01) SDs lower than trials that did not use a needle sham ($p = 0.036$). Trials that used penetrating needles for sham control had smaller effect sizes than those with a non-penetrating sham control or sham control without needles. The difference in effect size was –0.45 (95% CI –0.78 to –0.12; $p = 0.007$). For the sensitivity analysis that excluded the Vas *et al.*[120–122] trials, this effect size reduced to –0.19 (95% CI –0.39 to 0.01; $p = 0.058$). There were no significant differences between non-penetrating needles and sham techniques that did not involve needling. Details of further sensitivity analyses are reported separately.[45]

**TABLE 4** Types of sham acupuncture control

| Needle used? | Penetrating? | True acupuncture points? | Depth of insertion? | Trials |
|---|---|---|---|---|
| Yes | Yes | No | Superficial | Linde *et al.*,[68] Melchart *et al.*,[67] Diener *et al.*,[72] Scharf *et al.*,[70] Haake *et al.*,[71] Endres *et al.*,[73] Witt *et al.*[69] and Brinkhaus *et al.*[66] |
| Yes | Yes | No | Deep | ᵃBerman *et al.*[134] |
| Yes | No | No | N/A | Vas *et al.*[122] |
| Yes | No | Yes | N/A | Foster *et al.*,[135] Guerra de Hoyos *et al.*,[138] Kennedy *et al.*,[130] Kleinhenz *et al.*,[137] Vas *et al.*[120] and Vas *et al.*[121] |
| No | No | No | N/A | Carlsson and Sjölund[119] and Kerr *et al.*[117] |
| No | No | Yes | N/A | Irnich *et al.*[118] and White *et al.*[132] |

N/A, not applicable.
a In the Berman *et al.*[134] trial, penetrating needles were used on non-acupuncture points and non-penetrating needles were used on true acupuncture points. For the main analysis, we considered this trial as penetrating-needle sham acupuncture used on non-acupuncture points.

**TABLE 5** Trial-level characteristics for trials with sham control groups

| Characteristic | n (%) |
|---|---|
| Needle used | |
| Yes | 16 (80) |
| No | 4 (20) |
| Penetrating needle used | |
| Yes | 9 (45) |
| No | 7 (35) |
| Non-needle | 4 (20) |
| True acupuncture points used | |
| Yes | 8 (40) |
| No | 12 (60) |
| Superficial or deep sham | |
| Superficial | 8 (40) |
| Deep | 1 (5) |
| Non-penetrating sham | 11 (55) |
| Pain type | |
| Lower back pain | 5 (25) |
| Migraine | 2 (10) |
| Neck | 3 (15) |
| Osteoarthritis | 5 (25) |
| Shoulder | 3 (15) |
| Tension-type headache | 2 (10) |

**TABLE 6** Effect size for acupuncture compared with sham control by type

| Type of sham control | Main analysis | | Excluding Vas et al.[120–122] trials | |
|---|---|---|---|---|
| | Number of trials | Effect size (95% CI) | Number of trials | Effect size (95% CI) |
| Needle sham | 16 | 0.42 (0.19 to 0.66) | 13 | 0.22 (0.11 to 0.33) |
| Non-needle sham | 4 | 0.38 (0.19 to 0.57) | 4 | 0.38 (0.19 to 0.57) |
| Non-penetrating needle | 7 | 0.76 (0.31 to 1.21) | 4 | 0.43 (0.01 to 0.85) |
| Penetrating needle | 9 | 0.17 (0.11 to 0.23) | 9 | 0.17 (0.11 to 0.23) |
| Non-needle and non-penetrating needle | 11 | 0.63 (0.33 to 0.94) | 8 | 0.40 (0.18 to 0.62) |

## Non-sham acupuncture controls

Eighteen trials including 14,597 patients used a non-sham control, with trial-level characteristics for these trials described in *Table 8*. The majority of these control groups (72%) were classified as 'routine care', with the rest classified as 'protocolled care'. *Table 9* provides details of the non-sham control groups by pain type. The effect size for acupuncture in trials with routine care control arms (0.55, 95% CI 0.40 to 0.70) was larger than when acupuncture was compared with protocolled care (0.29, 95% CI 0.01 to 0.58). Although the difference in effect size was large, it was not significant (difference in effect size 0.26, 95% CI –0.05 to 0.57; $p = 0.1$). Details of further sensitivity analysis are reported separately.[45]

**TABLE 7** Differences in effect size between types of sham control

| Sham control | Main analysis | | | Excluding Vas et al.[120–122] trials | | |
| --- | --- | --- | --- | --- | --- | --- |
| | Number of trials[a] | Change in effect size (95% CI) | p-value | Number of trials[a] | Change in effect size (95% CI) | p-value |
| Needle vs. non-needle sham | 16 | 0.02 (–0.49 to 0.53) | 0.9 | 13 | –0.17 (–0.43 to 0.09) | 0.2 |
| | 4 | | | 4 | | |
| Non-penetrating needle vs. non-needle sham | 7 | 0.35 (–0.28 to 0.99) | 0.3 | 4 | 0.01 (–0.45 to 0.47) | 1 |
| | 4 | | | 4 | | |
| Penetrating needle vs. non-penetrating needle | 9 | –0.57 (–0.96 to –0.18) | 0.004 | 9 | –0.19 (–0.47 to 0.08) | 0.2 |
| | 7 | | | 4 | | |
| Penetrating needle vs. non-needle sham | 9 | –0.21 (–0.41 to –0.01) | 0.036 | 9 | –0.21 (–0.41 to –0.01) | 0.036 |
| | 4 | | | 4 | | |
| Penetrating needle vs. non-needle or non-penetrating needle | 9 | –0.45 (–0.78 to –0.12) | 0.007 | 9 | –0.19 (–0.39 to 0.01) | 0.058 |
| | 11 | | | 8 | | |

a The number listed in the top row is the number of trials in the first comparison group. The number of trials listed in the bottom row is the number of trials in the second comparison group. For example, there were 16 needle sham-controlled trials and four non-needle sham-controlled trials in the main analysis.

**TABLE 8** Types of non-sham control group by trial

| Trial | Control group | Type of control group |
| --- | --- | --- |
| Foster et al. (2007)[135] | Advice and exercise: all three arms of the trial received advice and exercise. Patients received a leaflet with information on knee osteoarthritis. Patients on non-steroidal anti-inflammatory drugs were allowed to continue with a stable dose. Individualised exercises of progressive intensity for lower limb stretching, strengthening and balance (up to six 30-minute sessions over 6 weeks). Patients in the control arm did not receive verum or sham acupuncture | Protocolled |
| Linde et al. (2005),[68] Melchart et al. (2005)[67] | Waiting list control: control patients were not permitted to undergo prophylactic treatment for 12 weeks. All patients were allowed to treat acute headache as necessary (following current guidelines) | Routine |
| Thomas et al. (2006),[61] Salter et al. (2006),[133] Vickers et al. (2004)[60] | GP care: all patients received NHS treatment according to GPs' assessment and recommendations. Control patients did not receive acupuncture or any other specified interventions | Routine |
| Berman et al. (2004)[134] | Education/attention control: patients in this arm attended six 2-hour group sessions based on arthritis self-management and received periodic educational materials by mail. Patients in the acupuncture and sham-acupuncture arms did not participate in this intervention | Routine |
| Cherkin et al. (2001)[131] | Self-care education: patients in this group received a book with information about back pain, treatment, improving quality of life and coping with emotional and interpersonal issues surrounding back pain. Patients also received two professionally produced videos that addressed self-management of back pain and demonstrated exercises. Patients in the acupuncture and massage groups did not receive this educational material | Routine |

**TABLE 8** Types of non-sham control group by trial (*continued*)

| Trial | Control group | Type of control group |
|---|---|---|
| Scharf *et al.* (2006)[70] | Conservative therapy: patients in the conservative therapy group had 10 visits with a physician and received prescriptions for either diclofenac (up to 150 mg/day) or rofecoxib (25 mg/day) up to week 23. Patients in this group who had 'partially successful' results were given the option of attending an additional five visits. Patients in the verum acupuncture and sham acupuncture groups were permitted to take up to 150 mg/day of diclofenac for the first 2 weeks and a total of 1 g of diclofenac during the rest of the study. Patients in both acupuncture groups and in the conservative management group received up to six sessions of physiotherapy. All patients were prohibited from taking any analgesics other than diclofenac and rofecoxib and any corticosteroids | Protocolled |
| Diener *et al.* (2006)[72] | Standard migraine treatment: control group patients were treated according to the guidelines of the German Migraine and Headache Society. Patients had six to seven visits in which standard treatment was established. First choice of treatment was beta-blockers, followed by flunarizine and then valproic acid. Acute medication use was permitted in all groups | Protocolled |
| Haake *et al.* (2007)[71] | Conventional therapy: patients in the conventional therapy group were treated according to German guidelines. Conventional therapy patients had 10 visits with a physician or physiotherapist at which physiotherapy, exercise and/or similar treatments were offered. Patients in all three arms were permitted to take non-steroidal anti-inflammatory drugs up to the maximum daily dose | Protocolled |
| Williamson *et al.* (2007)[136] | Education and exercise: patients in the control group were told that they were in the 'home exercise' group and received an exercise and advice leaflet | Routine |
| Witt *et al.* (2005),[69] Brinkhaus *et al.* (2006)[66] | Waiting list control: patients in the waiting list control group received no acupuncture treatment for 8 weeks after randomisation. All patients were allowed oral non-steroidal anti-inflammatory drugs for pain as rescue medication. All patients were prohibited from taking corticosteroids or pain medication that acted on the central nervous system | Routine |
| Witt *et al.* (2006)[77] (osteoarthritis), Witt *et al.* (2006)[74] (lower back pain) | Conventional treatment: patients in the control group were not allowed to use any kind of acupuncture during the first 3 months. All patients were allowed to use additional conventional treatments as needed | Routine |
| Jena *et al.* (2008),[76] Witt *et al.* (2006)[75] (neck pain) | Conventional treatment: patients in the control group were not allowed to use any kind of acupuncture during the first 3 months. All patients were allowed to use additional conventional treatments as needed | Routine |

**TABLE 9** Trial-level characteristics for trials with non-sham controls

| Pain type | Routine care | Protocolled care | Total |
|---|---|---|---|
| Headache | 2 | 0 | 2 |
| Migraine | 1 | 1 | 2 |
| Tension-type headache | 1 | 0 | 1 |
| Osteoarthritis | 4 | 2 | 6 |
| Lower back pain | 3 | 2 | 5 |
| Neck pain | 2 | 0 | 2 |
| Total | 13 (72%) | 5 (28%) | 18 (100%) |

## Results for substudy 2: characteristics of acupuncture

*Table 10* provides a summary of the trial-level acupuncture characteristics and *Table 11* provides a summary of the patient-level acupuncture characteristics. Fuller details of the characteristics of each individual trial are presented in *Appendix 2*. The acupuncture in the majority of trials was based on traditional Chinese acupuncture (59%) and had a flexible point prescription (55%). In all 29 trials manual needle stimulation was used in the acupuncture group, whereas only about one-quarter of trials allowed the addition of electrical stimulation ($n = 7$) and 14% allowed moxibustion ($n = 4$). Attempts to elicit deqi in the acupuncture group were made in all 25 trials that provided this information. The mean session frequency ranged from one session every 8 days to two sessions per week. The maximum number of sessions varied widely, from three to 30, as did the mean number of needles used (range 1–18) and the mean session duration (range 15–32 minutes).

None of the acupuncture characteristics evaluated in the trial-level analysis, including style of acupuncture, number or placement of needles, number, frequency or duration of sessions, patient–practitioner interactions or experience of the acupuncturist significantly modified the effect of acupuncture on pain (all $p > 0.05$) in sham-controlled trials (*Table 12*). Compared with non-sham controls, there was little evidence that these characteristics modified the effect of acupuncture (see *Table 12*). The exception was that acupuncture effects increased in comparison to non-sham controls when more needles were used.

**TABLE 10**  Trial-level acupuncture characteristics ($n = 29$)

| Characteristics | n (%) |
|---|---|
| Style of acupuncture | |
|    Traditional Chinese techniques | 17 (59) |
|    'Western' | 4 (14) |
|    Combination of traditional Chinese techniques and Western | 8 (28) |
| Point prescription | |
|    Fixed needle formula | 4 (14) |
|    Flexible formula | 16 (55) |
|    Individualised | 9 (31) |
| Location of needles | |
|    Local points only | 0 (0) |
|    Distal points only | 1 (3) |
|    Both local and distal points | 28 (97) |
| Electrical stimulation allowed | 7 (24) |
| Moxibustion allowed | 4 (14) |
| Deqi attempted ($n = 25$) | 25 (100) |
| Acupuncture-specific patient–practitioner interactions | 12 (41) |
| Minimum years of experience required | |
|    No requirement specified (0 years) | 12 (41) |
|    6 months to 2 years | 5 (17) |
|    3 years | 9 (31) |
|    5 years | 2 (7) |
|    10 years | 1 (3) |

continued

**TABLE 10** Trial-level acupuncture characteristics (*n* = 29) (*continued*)

| Characteristics | n (%) |
|---|---|
| Maximum number of sessions | |
| 3–5 | 3 (10) |
| 6–10 | 14 (48) |
| 11–15 | 7 (24) |
| 16–20 | 3 (10) |
| 21–30 | 2 (7) |
| Frequency of sessions (mean number of sessions per week) | |
| 0.88 | 1 (3) |
| 1 | 14 (48) |
| 1.5 | 7 (24) |
| 1.67 | 1 (3) |
| 2 | 6 (21) |
| Mean duration of sessions (minutes), rounded to whole numbers | |
| 15–19 | 1 (4) |
| 20–24 | 4 (16) |
| 25–29 | 6 (24) |
| 30+ | 14 (56) |
| Mean number of needles used, rounded to whole numbers | |
| 1–4 | 1 (4) |
| 5–9 | 6 (25) |
| 10–14 | 9 (38) |
| 15–20 | 8 (33) |

**TABLE 11** Patient-level acupuncture characteristics (*n* = 18,434)

| Characteristics | n (%) |
|---|---|
| Number of sessions | |
| 0 | 383 (2) |
| 1–5 | 402 (2) |
| 6–10 | 7161 (39) |
| 11–15 | 1998 (11) |
| 16–20 | 45 (< 1) |
| 21–30 | 16 (< 1) |
| Missing | 1806 (10) |
| Not reported | 6623 (36) |

**TABLE 11** Patient-level acupuncture characteristics (*n* = 18,434) (*continued*)

| Characteristics | n (%) |
|---|---|
| Average session duration (minutes) | |
| 2–15 | 166 (1) |
| 16–30 | 2552 (14) |
| 31–45 | 406 (2) |
| 46–60 | 60 (< 1) |
| 60+ | 3 (< 1) |
| Missing | 1257 (7) |
| Not reported | 13,990 (76) |
| Average number of needles | |
| 2–5 | 20 (< 1) |
| 6–10 | 610 (3) |
| 11–15 | 717 (4) |
| 16–20 | 627 (3) |
| 21–25 | 177 (1) |
| 26+ | 27 (< 1) |
| Missing | 2529 (14) |
| Not reported | 13,727 (74) |
| Age of physician/acupuncturist (years) | |
| 30–35 | 298 (2) |
| 36–40 | 2119 (11) |
| 41–45 | 2630 (14) |
| 46–50 | 2407 (13) |
| 51–55 | 1701 (9) |
| 56–60 | 872 (5) |
| 60+ | 303 (2) |
| Missing | 368 (2) |
| Not reported | 7736 (42) |
| Physician/acupuncturist sex | |
| Male | 7002 (38) |
| Female | 3626 (20) |
| Missing | 0 (0) |
| Not reported | 7806 (42) |

**TABLE 12** Results of characteristics with univariate metaregression analyses

| Characteristic | Acupuncture vs. sham control (n = 20) | | | Acupuncture vs. non-acupuncture control (n = 18) | | |
|---|---|---|---|---|---|---|
| | β[a] | 95% CI | p-value | β[a] | 95% CI | p-value |
| Style of acupuncture | | | | | | |
|     Some traditional Chinese medicine vs. Western only | 0.05 | –0.52 to 0.63 | 0.9 | 0.13 | –0.51 to 0.77 | 0.7 |
|     Traditional Chinese medicine only vs. some Western | 0.20 | –0.20 to 0.61 | 0.3 | –0.10 | –0.38 to 0.19 | 0.5 |
| Point prescription | | | | | | |
|     Fixed needle formula | | Reference | | | Reference | |
|     Flexible formula | –0.08 | –0.58 to 0.43 | 0.8 | 0.02 | –0.64 to 0.68 | > 0.9 |
|     Individualised | –0.15 | –1.16 to 0.86 | 0.8 | –0.08 | –0.74 to 0.59 | 0.8 |
| Electrical stimulation allowed | 0.34 | –0.13 to 0.80 | 0.15 | –0.19 | –0.56 to 0.17 | 0.3 |
| Manual stimulation allowed | All allowed | | | All allowed | | |
| Moxibustion allowed | All did not allow | | | –0.28 | –0.63 to 0.06 | 0.11 |
| Deqi attempted | All allowed | | | All allowed | | |
| Acupuncture-specific patient–practitioner interactions allowed | –0.22 | –0.70 to 0.26 | 0.4 | 0.06 | –0.23 to 0.35 | 0.7 |
| Minimum experience required (years) | 0.01 | –0.08 to 0.10 | 0.8 | 0.05 | –0.05 to 0.16 | 0.3 |
| Maximum number of sessions (per five sessions) | –0.14 | –0.37 to 0.08 | 0.2 | 0.02 | –0.07 to 0.12 | 0.6 |
| Frequency of sessions (per week) | –0.19 | –0.66 to 0.27 | 0.4 | 0.09 | –0.31 to 0.49 | 0.7 |
| Duration of sessions (per 5 minutes)[b] | –0.10 | –0.30 to 0.11 | 0.4 | –0.01 | –0.26 to 0.24 | 0.9 |
| Number of needles used (per five needles)[c] | –0.17 | –0.37 to 0.03 | 0.095 | 0.33 | 0.08 to 0.58 | 0.01 |

a  β is an estimate of the change in effect of acupuncture in terms of the standardised difference compared with the control for each characteristic; a positive β indicates a larger effect of acupuncture than of the control for trials with the given characteristic vs. the reference level of the characteristic.
b  Four trials excluded missing data.
c  Five trials excluded missing data.

The results of the sensitivity analysis that excluded the three outlying trials,[120–122] all sham controlled and by the same team and with very much larger effect sizes than the other trials, are presented in *Table 13*. This showed that trials allowing electrical stimulation showed a significantly stronger effect of acupuncture than of sham controls and those with a longer average treatment session duration had a smaller effect than sham controls.

In the patient-level analysis, we found that the direction of the effect of the acupuncture characteristics was unchanged (*Table 14*). As expected, the CIs were much tighter around the patient-level estimates despite the fact that fewer trials could be included in each analysis. There were no more than six trials with patient-level data available for each of the specific predictors being analysed. The patient-level analysis suggested that a higher number of acupuncture treatment sessions improves the effect of acupuncture. Further details of the results are reported separately.[123]

**TABLE 13** Results of the metaregression for acupuncture trials with sham controls, excluding the outlying trials[120–122]

| Characteristic | n | β[a] | 95% CI | p-value |
|---|---|---|---|---|
| | | | Acupuncture vs. sham control (n = 17) | |
| Style of acupuncture | 17 | | | |
|   'Western' only | | | Reference | |
|   Traditional Chinese medicine | | −0.14 | −0.46 to 0.17 | 0.4 |
| Point prescription | 17 | | | |
|   Fixed needle formula | | | Reference | |
|   Flexible formula | | −0.25 | −0.50 to 0.00 | 0.054 |
|   Individualised | | −0.11 | −0.58 to 0.36 | 0.6 |
| Electrical stimulation allowed | 17 | 0.27 | 0.03 to 0.51 | 0.027 |
| Manual stimulation allowed | 17 | All allowed | | |
| Moxibustion allowed | 17 | All did not allow | | |
| Deqi elicited (n = 26) | 17 | All allowed | | |
| Acupuncture-specific patient–practitioner interactions allowed | 17 | −0.04 | −0.28 to 0.20 | 0.8 |
| Minimum experience required (years) | 17 | 0.00 | −0.05 to 0.05 | > 0.9 |
| Maximum number of sessions (per five sessions) | 17 | −0.05 | −0.18 to 0.08 | 0.4 |
| Frequency of sessions (per week) | 17 | −0.04 | −0.29 to 0.21 | 0.8 |
| Duration of sessions (per 5 minutes) | 17 | −0.14 | −0.22 to −0.06 | 0.001 |
| Number of needles used (per five needles) | 17 | −0.08 | −0.22 to 0.05 | 0.2 |

a  β is an estimate of the change in effect of acupuncture in terms of the standardised difference compared with the control for each characteristic; a positive β indicates a larger effect of acupuncture than of the control for trials with the given characteristic vs. the reference level of the characteristic.

**TABLE 14** Results of the patient-level analysis of acupuncture characteristics

| Characteristic | Sham control | | | | Non-acupuncture control | | | |
|---|---|---|---|---|---|---|---|---|
| | n[a] | β[b] | 95% CI | p-value | n[a] | β[b] | 95% CI | p-value |
| Number of sessions (per five sessions) | 3 (646/648) | −0.76 | −1.75 to 0.22 | 0.13 | 5 (8292/9321) | 0.11 | 0.01 to 0.21 | 0.0007 |
| Duration of sessions (per 5 minutes) | 5 (2444/2482) | −0.03 | −0.08 to 0.03 | 0.3 | Fewer than three trials | | | |
| Number of needles used (per five needles) | 5 (1769/2484) | −0.11 | −0.35 to 0.14 | 0.4 | Fewer than three trials | | | |
| Age of acupuncturist (per 5 years) | Fewer than three trials | | | | 6 (9127/9446) | −0.01 | −0.04 to 0.02 | 0.5 |
| Male acupuncturist | Fewer than three trials | | | | 6 (9384/9446) | −0.07 | −0.16 to 0.02 | 0.084 |

a  Number of trials included in analysis (number of patients included in analysis/total number of patients in included trials).
b  β is an estimate of the change in effect of acupuncture in terms of the standardised difference compared with the control for each characteristic; a positive β indicates a larger effect of acupuncture than of the control for trials with the given characteristic vs. the reference level of the characteristic.

## Discussion

Our principal finding is that there are statistically significant differences between acupuncture and sham acupuncture, and between acupuncture and non-acupuncture controls for all of the pain types studied. The meta-analytic effect sizes were similar across pain conditions for both of these comparisons after excluding an outlying set of studies. The difference between acupuncture and sham acupuncture was of a lesser magnitude, ranging from 0.15 to 0.23 SDs depending on condition. For acupuncture compared with non-acupuncture control, there was a larger effect size of around 0.5 SDs, although there was some variation in the effect size for individual RCTs, a variation that could be partly explained in terms of the type of control used. For example, acupuncture had a smaller benefit in patients who received a programme of ancillary care involving physiotherapist-led exercise[135] than in patients who continued to be treated with usual care. Nevertheless, the average effect, as expressed in the meta-analytic estimate of approximately 0.5 SDs, was of clear clinical relevance either considered as a standardised difference or when converted back to a pain scale.

This meta-analysis has not been compromised by study quality or sample size, on the basis that only high-quality studies were included and the total sample size was large. In addition, we saw no evidence that publication bias, or failure to identify published eligible studies, might affect our conclusions. With regard to limitations, the comparisons between acupuncture and non-acupuncture controls cannot be blinded and therefore both performance and response bias are possible. Although we considered the risk of bias from unblinding to be low in most studies comparing acupuncture and sham acupuncture, health-care providers could not but be aware of the treatment that they provided and therefore a certain degree of bias of our effect estimate for specific effects is possible. However, it should be kept in mind that this problem applies to almost all studies on non-pharmacological interventions. We would argue that the risk of bias in the comparison between acupuncture and sham acupuncture was low compared with the risk of bias for other non-drug treatments for chronic pain, such as cognitive therapies, exercise or manipulation, which are rarely subject to placebo control. The meta-analyses combined different end points, such as pain and function, measured at different times. However, when we restricted the analysis to pain end points measured at a specific follow-up time, 2–3 months after randomisation, our results did not change.

Many previous systematic reviews of acupuncture for chronic pain have included RCTs of low methodological quality because of liberal eligibility criteria. As a result, the authors have come to the circular conclusion that weaknesses in the data did not allow conclusions to be drawn.[57,140] Because of variation in the study end points, other reviews have not included meta-analyses.[110,141] Both limitations have been avoided by including only high-quality RCTs and obtaining raw data for IPD meta-analysis. Some more recent systematic reviews have published meta-analyses[48,64,65,142] and reported findings that are broadly comparable with ours, with clear differences between acupuncture and non-acupuncture controls and smaller differences between true and sham acupuncture. Our findings have greater precision: all previous reviews have analysed summary data, an approach of reduced statistical precision compared with IPD meta-analysis.[105,143] In particular, we have demonstrated a robust difference between acupuncture and sham acupuncture that can be distinguished from bias. This is a novel finding that moves the evidence base beyond the existing literature.

These findings are important both clinically and scientifically. The total effects of acupuncture, as experienced by the patient in routine clinical practice, are clinically relevant as a result of our determination of an effect size of around 0.5. Our finding that acupuncture also has statistically significant effects over and above those of sham acupuncture, with an effect size of around 0.2, is therefore of major importance for clinical practice. We have found the effect size of sham acupuncture compared with non-acupuncture controls to be around 0.3. Accordingly, there appears to be no evidence to support the claim that it makes little difference whether one receives real or sham acupuncture. The size of this effect may be associated with sham acupuncture's potentially potent placebo or context effects,[144–146] or with the additional physiological effects associated with needle penetration even when at the wrong acupuncture points.[147] If one accepts that, on average, the differences in effect sizes between acupuncture and sham

acupuncture are small, the clinical decision made by physicians and patients is not between true and sham acupuncture but between a referral to an acupuncturist or avoiding such a referral. The total effects of acupuncture, as experienced by the patient in routine practice, include the specific effects associated with correct needle insertion according to acupuncture theory, non-specific physiological effects of needling and non-specific psychological (placebo) effects related to the patient's belief that treatment will be effective.

Further research will include the ongoing addition of new trials to the database subject to eligibility. This will require further literature searches and then updating of the main meta-analysis. Although six substudies were planned as part of the original programme of work, we report in detail on only two here.[45,123]

### Discussion for substudy 1: influence of control group on effect size

This substudy has extended the primary findings of the main study, confirming that acupuncture is significantly superior to either sham acupuncture or non-sham controls by also showing that this result holds irrespective of which type of sham or non-sham control is used. This substudy has produced robust data on the differences in effect sizes between trials with different control conditions. We found that the acupuncture in trials with sham controls involving penetrating needles had smaller effect sizes than the acupuncture in trials that did not use a needle control or in which the needles in the control group did not penetrate the skin. An important implication is that the central estimates from our primary meta-analysis may have underestimated the effects of acupuncture compared with sham acupuncture. Although differences did not reach statistical significance, we have also found evidence that the effect size of acupuncture when compared with protocolled care is smaller than when compared with the less intensive routine care.

With regard to the impact of placebo controls for non-pharmacological therapies, one approach has been to investigate trials that included both a placebo arm and a no-treatment arm and then compare outcomes between these two, and in this way explore variations in the impact of the different types of placebo. An example of this is a Cochrane review of placebo controls covering a wide range of trials for different conditions, including some acupuncture trials.[44] In a subgroup analysis the authors found that trials using 'physical placebos' (including sham acupuncture) were associated with greater placebo effects than trials using pharmacological placebos. This finding is consistent with the results of a trial that was specifically designed to compare a sham device (sham acupuncture) with an inert pill, the sham device being associated with a greater reduction in self-reported pain.[144] Our finding that different types of sham control led to different estimates of treatment effects is consistent with these findings.

By combining patient data from 29 high-quality trials in a single database, we have had, for the first time, sufficient power to explore the role of controls in trials of acupuncture for chronic pain. This is because the power of metaregression is strongly influenced by the number of trials and their variation. However, even with this large data set we are not able to obtain a full understanding of the different physiological and psychological effects of sham acupuncture. One limitation within the field generally is that the mechanisms for a persistent effect of acupuncture on chronic pain are incompletely understood and therefore we have no clear idea of whether or not a sham control inadvertently activates these mechanisms. This lack of understanding about the physiological mechanisms of acupuncture limits any firm conclusions that we can draw regarding the extent that any of the sham controls discussed above can be considered as a true 'placebo'. Moreover, when implementing sham acupuncture trials the outcome may also be influenced by factors not included in our analysis. These include practical implementation issues such as how carefully the ring that comes with the Streitberger needle[108] is taped in place, the believability of the control, patients' prior knowledge of acupuncture, whether or not the true acupuncture group is treated identically and the extent to which the acupuncturists are able to maintain equipoise.

Although sham acupuncture involving penetrating needles may well have a place when addressing questions of point specificity in explanatory trials, our results provide support to the contention that needle penetration should be avoided as a sham technique when controlling for non-specific effects associated with

acupuncture for chronic pain. We are more cautious with regard to recommending the use of non-penetrating needles. Many forms of Japanese acupuncture use shallow insertion or non-insertion (the toya hari method).[148] Using non-penetrating needles in controlled trials is not without its challenges: although apparently less active than other types of sham control, we cannot assume that non-penetrating needles have complete physiological inactivity. Furthermore, there are practical and generic questions regarding the use of sham acupuncture, for example whether or not to enrol only acupuncture-naive patients and whether or not practitioners can maintain equipoise in large trials over reasonable periods of time.

The choice of control needs to be driven by the research question. For instance, in the UK NHS trial of acupuncture for chronic headache, the study question of Vickers et al.[60] was related to the effects of making acupuncture more widely available in primary care, a pragmatic comparison of 'use acupuncture' and 'avoid acupuncture'. On the other hand, Foster et al.[135] were interested in the impact of acupuncture when added to an existing rehabilitation programme. Our findings have clear implications for sample size calculations, with larger sample sizes needed in trials in which care in the control arm is carefully specified. Further discussion has been reported separately.[45]

### Discussion for substudy 2: characteristics of acupuncture

Our results from high-quality trials of acupuncture for chronic pain provide robust data on treatment effect modifiers related to treatment characteristics. Across the many characteristics that might have been commonly expected to modify outcomes, such as style of acupuncture, use of electrical stimulation, addition of moxibustion, experience of the acupuncturist and the frequency and duration of sessions, we found no evidence of a modifying effect on pain outcomes in trial-level analyses. We also found in patient-level analyses, involving a subset of only five trials, that more treatment sessions were associated with better pain outcomes in acupuncture treatment groups compared with non-acupuncture controls ($p < 0.001$). It appears that the 'dose' of acupuncture has an impact on treatment outcomes. We need to be careful when interpreting these results in the context of testing multiple hypotheses, which increases the risk of falsely rejecting at least one null hypothesis. We did not feel that formal statistical correction to account for multiple testing was justified given the largely null results.

Our observations on the minimal impact of style of acupuncture practice on outcomes contrasts with findings from qualitative studies asserting the importance of the theoretical affiliation and institutional context of the acupuncturists.[149,150] For example, theoretical affiliation has been linked to impact on 'almost all aspects of treatment', with 'demonstrable implications for the practice and research of acupuncture'.[149] Similarly, others have argued that without a theoretical approach to acupuncture that is 'holistic', with an emphasis that includes the process of care, there are likely to be reduced or absent outcomes.[150]

In a study that pooled data from four German-based trials of acupuncture for chronic pain,[151] all of which are included in our study, we see some concordance with our results, as might be expected. The only physician characteristic that the authors found had a significant influence on outcome in these four trials was that internists performed slightly better than an average physician and 'orthopaedists' slightly worse.

Given the results of the primary analyses in this study, which showed small differences between real and sham acupuncture, it is perhaps unsurprising that this substudy showed little evidence of substantial differences between alternative approaches to, or characteristics of, acupuncture. It is likely that we had insufficient power to determine the extent, if any, of the contribution of each individual component of acupuncture to outcome, even with the large data set at our disposal.

As for any meta-analysis, the main limitation was data availability. The total number of trials was relatively modest and analyses with IPD included no more than five trials. Consequently, many of our analyses had relatively low power, with wide CIs around central estimates. Furthermore, heterogeneity of treatment characteristics was relatively limited. For example, nearly 75% of trials involved between six and 15 treatments, and in no trial was acupuncture administered more than twice a week. It is not unusual in China for acupuncture to be given four or five times a week. A feature that characterises acupuncture

when practised according to the principles of traditional Chinese medicine involves syndrome differentiation, but we had no data on which to assess the resulting impact on outcome.

The quality of the current evidence supports the case that contemporary acupuncture trials do not systematically underestimate treatment effects. Although acupuncturists have long been concerned about what constitutes 'correct' practice,[152,153] it can be argued that the consensus methods that are often used to determine acupuncture characteristics – number of treatment sessions, duration of sessions, needle prescriptions and training and experience of acupuncturists – are appropriate. This is because the variations in outcome associated with these factors are likely to be small. This study would suggest that the most useful characteristics to test would be the number of needles and number of treatment sessions.

Few characteristics of acupuncturists were reported sufficiently consistently by triallists, namely age, sex and minimum experience as an acupuncture practitioner, and consequently our results on related effects on outcome were compromised. We know that some practitioners have better results than others, yet we do not have the data from this substudy to understand why, a concern also raised in relation to other therapist-led interventions.[154,155] It is likely that more sophisticated measures are needed, such as a measure of the patients' perception of a practitioner's empathy, which has been shown to correlate with enablement and in turn with outcome.[156]

### Recommendations for future research

Four additional substudies are planned by the ATC, as well as an update to the database by adding in a set of more recently published trials. The first additional substudy will explore the time course of acupuncture, establishing for how long the effects of acupuncture last beyond the end of treatment. Another will explore the relationship between characteristics of the patients and variations in outcome, for example whether or not there any baseline characteristics of patients, such as age, psychological distress or baseline severity, that might influence the effects of acupuncture. In the third substudy we will determine if there is a certain type of patient who could be classed as a super-responder. Finally, a substudy will explore whether or not the effects of acupuncture are significantly influenced by the acupuncture practitioner.

An important recommendation for future research is that trials need to be designed with sufficient power to detect the small differences in outcome that might be associated with differences between acupuncture and sham acupuncture. Given our analyses involving acupuncture trials for chronic pain, an effect size of 0.2 requires a sample size in a two-arm trial of > 1000 patients, a number rarely reached in almost all sham-controlled trials. To determine differences between two different styles of acupuncture, for example when one might hypothesise that the difference in outcome between two individual acupuncture characteristics of treatment is associated with a smaller effect size of say 0.10 or 0.05, would require > 2000 or > 8000 patients, respectively, assuming 90% power and a 0.05 significance level.

To explore contextual factors, further research into empathy and enablement as well as other measures such as the therapeutic alliance or success of patient–practitioner interactions may be useful lines of enquiry. Interestingly, a widely accepted principle underlying traditional Chinese medicine is that it is not the techniques and methods used but the cultivation of the practitioner that is the key to effective practice.[157] A qualitative study might provide a useful way of exploring the key factors which explain the observation that some practitioners consistently have better patient outcomes than others.

## Conclusion

Acupuncture is superior to both non-acupuncture controls and sham acupuncture for the treatment of chronic pain and is therefore a reasonable referral option for patients with chronic pain. The data indicate that acupuncture is more than a placebo; however, the differences between true and sham acupuncture are relatively modest. Other factors, in addition to the specific effects of needling, are important contributors to therapeutic effects. Given that the results are from IPD meta-analyses of nearly 18,000

randomised patients in high-quality RCTs, it can be concluded that they provide the most robust evidence to date on acupuncture for chronic pain.

Moreover, acupuncture is significantly superior to control, irrespective of the subtype of control. With regard to sham acupuncture, non-penetrating needles appear to be a more effective sham control. In addition, the effect size of acupuncture is greater when compared with routine care than when compared with protocol-guided care. Our findings can help inform study design in acupuncture, particularly with respect to sample size, in the context that the choice of control should be driven by the research question.

We found little evidence that different characteristics of acupuncture or acupuncturists were effect modifiers. There was modest evidence that more needles and more sessions were associated with better outcomes when comparing acupuncture with non-sham controls, suggesting that the dose of acupuncture is important. Trials designed to evaluate the potentially small differences in outcome associated with different acupuncture or acupuncturist characteristics are likely to require large sample sizes.

# Chapter 3 Comparison of acupuncture with other physical treatments for pain caused by osteoarthritis of the knee: a network meta-analysis

## Background

Osteoarthritis is a degenerative condition involving the progressive wearing down of (joint) bone and cartilage, normally resulting in pain, stiffness and functional disability. These symptoms usually worsen according to how much the affected joint is used. In adults aged ≥ 45 years, the knee represents the most common site of peripheral joint pain and the prevalence of painful, disabling knee osteoarthritis in people aged > 55 years is 10%.[158] Risk factors for knee osteoarthritis include age, sex, obesity, bone density, genetic factors and injury.

Diagnosis is usually made using clinical features of knee osteoarthritis, by radiological assessment of the knee or by a combination of the two. Radiographic features – the severity of which are commonly summarised using the Kellgren and Lawrence score[159] – have been significantly associated with knee pain.[160]

The WOMAC index is a self-administered disability status measure for knee (or hip) osteoarthritis.[115] Its individual components assess pain, stiffness and function, with the summed scores producing an overall measure of disability (WOMAC index). As a standardised and comprehensive assessment of disability and its components, the WOMAC index increases transparency and comparability within clinical research.

The treatment of knee osteoarthritis should be tailored according to knee risk factors (obesity, adverse mechanical factors, physical activity), general risk factors (age, comorbidity, polypharmacy), level of pain intensity and disability, sign of inflammation, and location and degree of structural damage.[161] The main objective of a GP treating a patient with knee osteoarthritis is normally alleviation of pain; failure to control pain may result in reduced mobility and daily activities, leading to a reduction in quality of life.[161] The more sedentary lifestyle that might follow may, in turn, exacerbate the symptoms of knee osteoarthritis through lack of exercise and joint movement, and weight gain.

In clinical practice, treatment often begins with analgesia [paracetamol and/or topical non-steroidal anti-inflammatory drugs (NSAIDs)] and, when these are ineffective, a cyclooxygenase-2 inhibitor is recommended. GP advice about exercise and weight loss, which NICE guidelines[82] recommend as part of core therapy, is often given in addition to (rather than instead of) analgesic drugs. The regular and long-term use of pharmacological agents such as NSAIDs for pain may be associated with side effects, such as gastrointestinal bleeding, without necessarily resulting in worthwhile pain reduction.[162] A UK review of qualitative studies of medicine taking[163] revealed considerable reluctance to take drugs and a preference to take as little as possible; many knee osteoarthritis patients want non-pharmacological treatments for pain relief.[164] The use of physical (i.e. non-pharmacological) treatments such as acupuncture is therefore likely to be attractive for patients seeking alternatives, particularly for a condition such as osteoarthritis of the knee for which there is currently no cure.

In patients for whom insufficient pain relief has been provided by the core interventions mentioned above (as recommended by NICE), coupled with paracetamol and/or topical NSAIDs, GPs may consider a range of physical treatments as the next step in the treatment pathway. The NICE guidelines[82] list muscle-strengthening and aerobic exercise, manual therapy, TENS, braces and insoles, weight loss, and heat and cooling treatments as being among such alternatives, but acupuncture was not recommended.

Other non-pharmacological interventions used for osteoarthritis of the knee, but which would not be considered as alternatives to acupuncture, include surgery, an intervention that would be considered at a later stage in the treatment pathway. Similarly, structured psychosocial/educational interventions are generally considered for a different group of patients, that is, when pain-reducing therapies have failed and the emphasis is on a need for pain-coping skills, rather than pain reduction.[165]

Many reviews have been undertaken of the varying types of physical therapies for osteoarthritis of the knee, but evaluation of a single therapy for a single condition provides only a limited basis for decision-making. Few randomised trials have directly compared physical therapies and no review has attempted to address the question of how effective such treatments are relative to each other using statistical methods. The focus of interest within our study was on acupuncture, as this review was funded as part of this programme of projects on acupuncture and chronic pain, and because of the uncertainty within the NICE decision-making process with regard to the level of evidence on acupuncture for osteoarthritis relative to other physical treatments.[84] The purpose of this systematic review, therefore, was to comprehensively synthesise both the direct and the indirect evidence – using mixed-treatment comparison methods in a network meta-analysis – to compare the effectiveness of different physical therapies used for the alleviation of knee pain caused by osteoarthritis.

In a separate substudy within this project we summarise the reporting methods of the WOMAC pain subscale and the WOMAC index from the trials identified for the main study, and make recommendations to improve reporting in future studies.

## Methods

Using the methods recommended by the Centre for Reviews and Dissemination (CRD)[166] and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement,[167] the systematic review and network meta-analysis was first conducted in 2010. A report based on this study is available on the CRD website.[168] This chapter reports an update of this systematic review and network meta-analysis conducted in 2013 and published in *Osteoarthritis and Cartilage*.[169]

### Literature search
We searched 17 electronic databases from inception to June 2013. A combination of relevant free-text terms, synonyms and subject headings relating to osteoarthritis of the knee and named physical treatments were included in the strategy. A search filter was used to limit retrieval of studies to RCTs. No language or date restrictions were applied.

The base search strategy developed in MEDLINE was translated to run on the databases listed in *Appendix 3*. Adaptations to the search strategy were necessary for certain databases: Manual, Alternative and Natural Therapy Index System (MANTIS), PASCAL (database of the Institut de l'Information Scientifique et Téchnique), Inside Conferences, Physiotherapy Evidence Database (PEDro), CAMbase (Complementary and Alternative Medicine), Latin American and Caribbean Health Sciences Literature (LILACS) and ClinicalTrials.gov. Supplementary internet searches of websites relating to osteoarthritis were undertaken to locate any additional studies not found from the database searches. The bibliographies of all relevant reviews and guidelines were checked for further potentially relevant studies. The base MEDLINE search strategy can be found in *Appendix 3*.

### Study selection and definitions of interventions and outcome
All abstracts were screened by two reviewers independently, followed by all relevant full papers. Disagreements were resolved by discussion or, when necessary, by a third reviewer. We included RCTs in adults with osteoarthritis of the knee (in which the mean age of the population was ≥ 55 years) that assessed pain as an outcome. Studies with mixed populations (e.g. including both patients with osteoarthritis of the knee and those with osteoarthritis of the hip) that presented results by site of

osteoarthritis were eligible for inclusion. Trials of acute knee pain or trials in which the mean age of the population was < 55 years were excluded.

We included treatment with the following: acupuncture, balneotherapy, braces, aerobic exercise, muscle-strengthening exercise, heat treatment, ice/cooling treatment, insoles, interferential therapy, laser/light therapy, manual therapy, neuromuscular electrical stimulation (NMES), pulsed electrical stimulation (PES), pulsed electromagnetic fields (PEMFs), static magnets, t'ai chi, TENS and weight loss. We aimed not to be restrictive with regard to selecting the types of intervention within these categories. However, exercise interventions that were predominantly home based and unsupervised were excluded as being too similar to standard care. Trials evaluating surgery or medication were also excluded, as were studies evaluating the combination of two or more physical treatments and studies comparing only different regimens/durations/modalities of the same type of intervention.

When considering the electrotherapy interventions, we classed studies using 'pulsed short-wave' interventions as being PES. Although interferential therapy works in a similar way to TENS, we classed it as a distinct intervention. Similarly, NMES was considered separately from TENS as it is commonly used to elicit muscle contraction, as opposed to TENS, which stimulates nerves with the aim of blocking pain inputs to the brain. PEMFs (in which an electric current is generated in the treated area by means of a magnetic field) was also classed separately.

We classified adjunctive components to the main intervention into five categories, based on what was reported in the trials: (1) treatment as usual, (2) treatment as usual plus specified home exercise or education, (3) treatment as usual plus specified (trial-specific) analgesics, (4) no medication and (5) no medication plus specified home exercise or education. Using this coding we explored the impact of different adjunctive components on the main interventions and on variations in standard care.

Eligible comparators included any form of standard or usual care, including waiting-list control (which could incorporate one or more of analgesics, education and exercise advice), all of which we classed as being 'standard care'. Placebo interventions, no intervention and sham acupuncture were also eligible. Because of evidence suggesting that sham acupuncture is more active than an inert 'placebo', it was treated as a separate comparator.[145] Commonly, sufficient details about standard care or usual treatment were not reported.

### Data extraction

Using a standardised data extraction form created using EPPI-Reviewer software (version 4.0; Evidence for Policy and Practice Information and Co-ordinating Centre, University of London, London, UK), data were extracted on population characteristics [population type, method of diagnosis, age, sex, weight, body mass index (BMI) and Kellgren and Lawrence score], intervention parameters and study quality. Data on pain assessment at baseline, at the end of treatment and at all subsequent time points were extracted onto a Microsoft Excel® 2010 spreadsheet (Microsoft Corporation, Redmond, WA, USA). Data extraction was performed by one reviewer and independently checked by a second reviewer. Any disagreements were resolved by discussion or by a third reviewer when necessary. Data from non-English-language papers were extracted by one reviewer together with a native speaker. Multiple publications of the same study were extracted as one study, using all of the information available.

### Assessment of trial quality

Trial quality was assessed using 14 questions adapted from a checklist used in a previous review by researchers at the CRD.[170] Based on the number of criteria satisfied, studies were then graded as excellent, good, satisfactory or poor. To be of satisfactory quality studies had to report the number of randomised participants; have groups with comparable baseline characteristics for important variables, such as pain; adequately report eligibility criteria; clearly report on losses to follow-up; report data for the intention-to-treat population; and use an appropriate placebo (if relevant). Poor-quality trials were those that failed to satisfy one or more of the criteria required for satisfactory study quality. Beyond questions relating to the

above grading, other questions in the assessment covered methods of randomisation and allocation concealment, level of blinding, use of a power calculation and level of losses to follow-up. A further quality assessment was conducted using the Cochrane risk-of-bias tool.[171] Quality assessments were performed by one reviewer and independently checked by a second reviewer. Any disagreements were resolved by discussion or by a third reviewer when necessary.

### Outcomes and data transformations

It was anticipated that pain – our primary outcome – would be measured using a variety of measures, for example a VAS, Likert scale, WOMAC pain subscale and Arthritis Impact Measurement Scale (AIMS), with all scales accepted.

The WOMAC is a widely used, self-administered health status measure that assesses the dimensions of pain, stiffness and physical function in patients with osteoarthritis of the hip or knee. It is available in five-point Likert, 11-point numerical rating and 100-mm VAS formats. Under each dimension there are a number of questions designed to assess the clinical severity of the disease (five questions for pain, two questions for stiffness and 17 questions for physical function). The patient's response to each question produces a score, with the scores summed to derive an aggregate score for each dimension. There are three subscale scores (pain, stiffness and physical function) and a total score (WOMAC index), which reflects disability overall.

The WOMAC pain score range has been reported in various ways: a VAS 0–10 scale (commonly reported across a 0–50 range), a VAS 0–100 scale (commonly reported as a 0–500 range) or a Likert scale (commonly reported as a 0–20 range). The overall WOMAC score (index) is determined by summing the scores across the three dimensions and the score range includes the following: a VAS 0–10 scale (commonly reported as a 0–240 range), a VAS 0–100 scale (commonly reported as a 0–2400 range) and a 0–4 Likert scale (commonly reported as a 0–96 range). A number of transformations and modifications are reported in the literature.

The preferred measure of pain was the WOMAC pain scale (using either a VAS or a Likert scale). Another pain scale was included in the analysis when a trial did not measure the WOMAC pain scale, with prioritisation of pain scales made on a clinical, or prevalence, basis. The secondary outcome was the WOMAC index. Studies that did not report a pain outcome were excluded from the review. Outcome data were extracted for different time points: baseline, end of treatment and any follow-up time point.

As a variety of pain scales were used, Hedges' *g* SMDs between treatment groups were calculated for the meta-analyses (studies reporting medians could not be analysed). Different doses/regimens of the same type of treatment within a study were pooled. Final values were used in the analysis to maximise the evidence available and to avoid the need to make assumptions about within-patient correlation between baseline and final values, which the use of change from baseline data would have necessitated. For trials reporting change from baseline but not final values, we calculated final values provided baseline data were reported along with variance estimates (e.g. SDs). When the number of patients included in a trial's analysis was not reported, but the number of patients randomised was, we estimated the number of analysed patients by multiplying the number of patients randomised by the average proportion of patients included in an analysis across trials. SEs or 95% CIs were used to derive SDs when they were not reported. When this was not possible, trials that used the same or a similar scale as that used in the trial with missing SDs were identified and their SDs pooled, with this imputed estimate being used. We present results as SMDs, as well as SMDs converted to the WOMAC VAS 0–100 pain scale, to provide more clinically meaningful results.

### Evidence synthesis

A network meta-analysis draws on both direct (treatments compared in the same trial) and indirect evidence (different treatments studied in separate trials, but compared when they share the use of a common comparator treatment). The summary treatment effect from each study is utilised, so the benefit

of randomisation in each study is retained. To conduct a meta-analysis of trials, study characteristics must be similar within a comparison. For indirect and direct evidence to be consistent, study characteristics must be similar across comparisons.[85,86,172–175]

We planned analyses for three different time points to assess both the immediacy and the durability of effects: (1) end of treatment, which was our primary time point, as defined in the individual studies; (2) 3 months from the start of treatment, which was the time point closest to 3 months from the start of treatment (excluding outcomes recorded at < 4 weeks from the start of treatment); and (3) 3 months after the end of treatment, which was the time point closest to 3 months, but between 8 and 16 weeks, from the end of treatment. However, there was a paucity of long- and medium-term data across the trials, and for the 3 months after the end-of-treatment time point no connected network incorporating acupuncture existed and this time point was evaluated by only 21 trials. Data for the 3 months from the start of treatment analysis were very similar to data for the end-of-treatment analysis because for around two-thirds of trials the two time points were the same. Furthermore, in most studies, the primary time point specified by investigators was the end of treatment and this time point produced the largest network, incorporating more interventions, and studies, than the other time points. We therefore report the results for the end-of-treatment time point.

To evaluate the impact of study quality on the results, two sets of analyses were performed: one included all studies regardless of quality (labelled 'any quality') and one was a primary sensitivity analysis including only studies of satisfactory, or better, quality (labelled 'better quality'). Studies with atypical populations, interventions or results were excluded in a second sensitivity analysis. These studies were identified from pairwise meta-analyses conducted (in RevMan 5.0; The Cochrane Collaboration, The Nordic Cochrane Centre, Copenhagen, Denmark) using outcomes recorded at the end of treatment only. These studies were not intended as a comprehensive stand-alone synthesis, but as a means of informing and complementing the network meta-analysis. In particular, they were used to investigate the within-intervention clinical and statistical heterogeneity. We assessed for possible publication bias using a funnel plot when enough studies within individual treatments were available. This was deemed to be appropriate only for muscle-strengthening exercise and the funnel plot provided no evidence to suggest publication bias.[168]

Analyses were conducted using WinBUGS software (version 1.4; MRC Biostatistics Unit, Cambridge, UK), which uses Markov chain Monte Carlo (MCMC) simulation to estimate model parameters and follows a Bayesian approach in which prior probabilities are specified for parameters (these were specified to be vague throughout the analysis). The treatment difference was assumed to be normally distributed and a random-effects network meta-analysis model was selected as clinical and methodological heterogeneity within the treatment definitions appeared likely.[176] A common between-study variance was modelled to allow a between-study variance to apply for comparisons with few data points.

Convergence of the MCMC chains was assessed by observing the history of the traces of the starting values for selected priors, the Brooks–Gelman–Rubin statistic and posterior distributions.[177] The first 10,000 iterations were discarded and then a further 50,000 iterations were conducted on which parameter estimates were based. The model fit was evaluated using the residual deviance, with this being approximately equal to the number of data points if the fit was good.[174,178–180]

Inconsistency in the treatment effect estimates derived separately from direct and indirect evidence was assessed for many of the comparisons distributed across the networks using the node-splitting method where a $p$-value is $2 \times \min(\text{prob}, 1 - \text{prob})$, where 'prob' is the probability that the direct estimate is higher than the indirect estimate.[172,179]

Uncertainty in all estimates is presented using the upper and lower limits of the 95% credible intervals (CrIs) of these estimates. These credible limits describe the boundaries within which it is believed that there is a 95% chance that the true value lies. The median rank of each intervention and the 95% CrIs of the rank are presented to summarise the uncertainty across all of the treatment effect estimates.[181]

To present more clinically meaningful network meta-analysis results, we present both SMDs and the SMDs converted to the WOMAC VAS 0–100 pain scale (although it is acknowledged that back-transformation can be of limited value in heterogeneous populations).[182] A pooled SD for the WOMAC VAS 0–100 pain scale was calculated from all of the arms of the six trials in the analysis that utilised this scale. The SMDs were then multiplied by this pooled SD (16.49) to produce a difference in WOMAC VAS 0–100 score.

In the substudy within this study that explored reporting of the WOMAC pain subscale and WOMAC index, further details were extracted for those trials that utilised the WOMAC: scale used (Likert/VAS 0–10, VAS 0–100/NRS); whether the WOMAC pain subscale or the WOMAC index was used; and whether any modifications were reported. In the light of inconsistencies and lack of clarity identified during the review, the WOMAC outcome details were re-examined by a third reviewer (NFW) and further information was extracted as necessary to address the following four questions:

1. Was it clear that all assessments had been conducted?
2. Was the score range clear?
3. Were details reported on how the final score had been calculated (sum, average or transformation to 0–100 scale)?
4. Were baseline scores reported (and approximate baseline score)?

In addition, the scale used and the score range that could be deduced from the information provided in each paper was recorded and the ease of identification of these was categorised as clearly stated in the paper (stated), required assumptions to be made (assumed) or unclear. All information that could support any assumptions, including baseline score (or, when not reported, follow-up scores), was also recorded. Further details of this substudy's methods are reported elsewhere.[183]

## Results

In total, 3820 references were retrieved from searches, of which 156 trials (detailed in *Appendix 3*) including 18 distinct interventions and four comparators met the inclusion criteria. Four of 10 foreign-language papers that appeared eligible based on their English abstracts could not be translated and so had to be excluded. Thirty-eight trials reported data in ways that meant they could not be incorporated in the network meta-analyses. One study was found to have been retracted and was subsequently removed from all analyses. A study flow diagram is presented in *Figure 4*.

### Study characteristics

An overview of all eligible studies is presented in *Table 15*, regardless of whether or not they reported data suitable for network meta-analysis. The mean treatment duration (and timing of the end-of-treatment assessment) varied widely, from just a single session (TENS) to 69 weeks (weight-loss interventions), although the majority of interventions were administered over a 2- to 6-week period. Most studies were classified as having recruited a general knee osteoarthritis population, although weight-loss trials (as expected) recruited only overweight or obese participants. The mean BMI in some studies recruiting a general population fell into the overweight or obese classification, although most studies did not report BMI.

Around three-quarters of the studies (110/152) were classed as 'poor-quality' studies. Of the remainder, 33 studies were classed as 'satisfactory' and nine studies were classed as 'good', which together were classed as 'better quality'. Only 12 trials were considered to be at low risk of bias in the network meta-analysis. Trial quality was commonly compromised by a lack of adequate blinding and small sample sizes, which limited the effectiveness of randomisation, resulting in baseline imbalances. Quality assessment data are presented in *Appendix 3*. Study quality did vary by intervention, making the evidence base more robust in some areas than in others. No evidence was found for publication bias (only assessable for
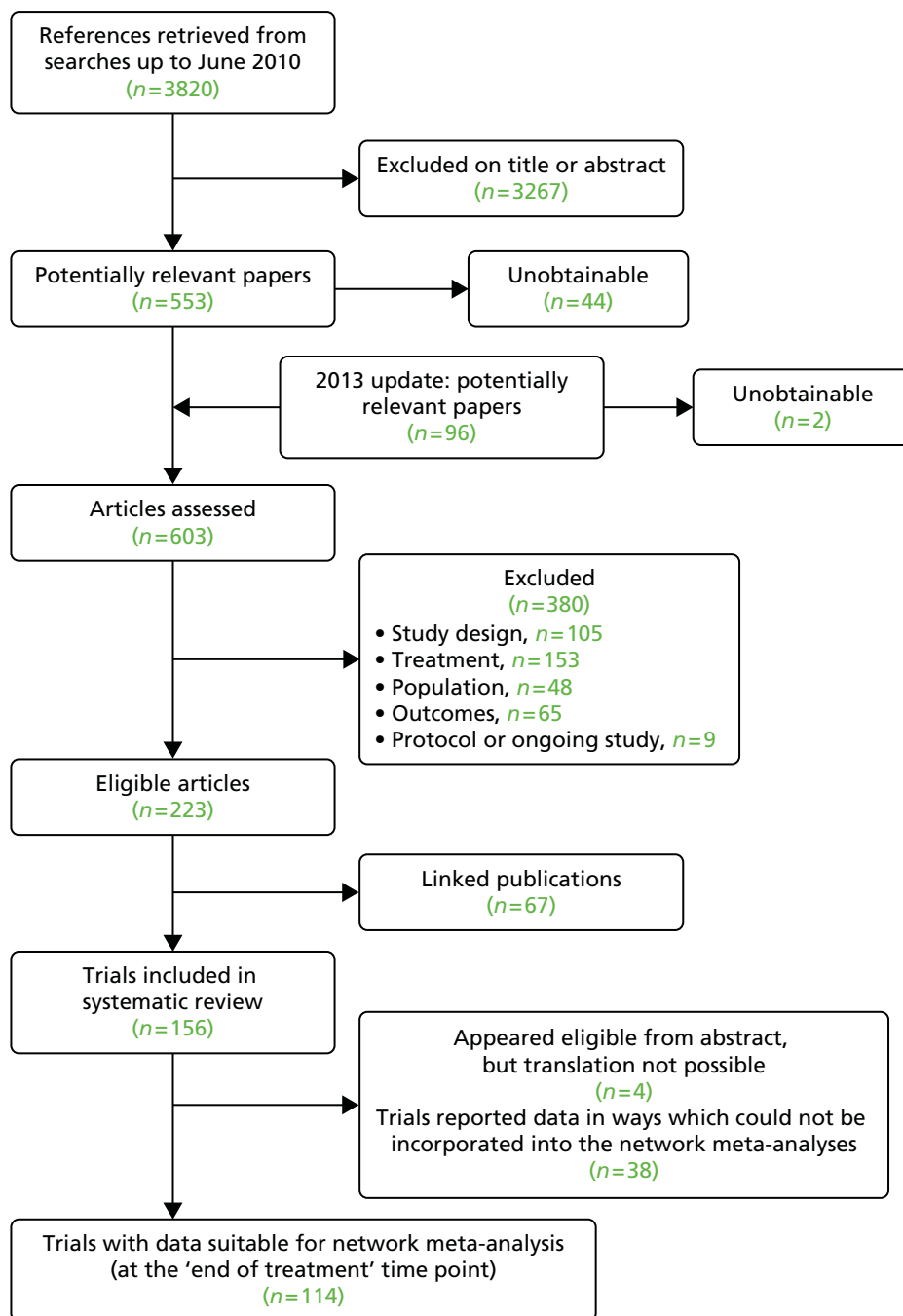
**FIGURE 4** Study flow chart.

muscle-strengthening exercise). Individual study characteristics of all studies included in the systematic review can be found in *Appendix 3*.

### Network meta-analysis

Suitable data for the end-of-treatment analyses were reported in 114 trials (9709 patients) (detailed in *Appendix 3*). This includes data from the 22 new studies identified from the search update conducted in 2013 and nine studies that had been excluded from the original review analyses but which were now included by calculating final values from the change from baseline data. In our original analyses (based on searches up to 2010) there was no indication that the majority of the adjunctive components of the experimental interventions were associated with a treatment effect difference.[169] The one exception was that standard care incorporating active analgesia was more effective than standard care with 'treatment as

**TABLE 15** Summary of characteristics of eligible trials[a] included in the review

| Intervention | Number of trials eligible for the review (number of patients[b]) | Type of population recruited[c] (number of studies) | Mean age (years), range | Female (%), range | Comparators (number of treatment arms[d]) |
|---|---|---|---|---|---|
| Acupuncture | 25 (2794) | General (23), both knees affected (1), awaiting surgery (1) | 58–85 | 50–96 | Sham acupuncture (15), standard care (13), TENS (3), muscle-strengthening exercise (1), ice/cooling (1) |
| Balneotherapy | 14 (1008) | General (12), both knees affected (2) | 54–70[e] | 47–100 | Placebo (8), standard care (6), heat treatment (1) |
| Braces | 1 (24) | General (1) | 59.5 | 63 | Insoles (1) |
| Aerobic exercise | 13 (1136) | General (9), both knees affected (2), overweight or obese (2) | 54–75[e] | 50–100 | Standard care (13), muscle-strengthening exercise (2), weight loss (1) |
| Muscle-strengthening exercise | 34 (3013) | General (26), both knees affected (5), awaiting surgery (2) | 53–77[e] | 31–100 | Standard care (22), placebo (4), no treatment (2), aerobic exercise (2), heat treatment (1), TENS (1), acupuncture (1), PES (1), manual therapy (1), NMES (2) |
| Heat treatment | 7 (412) | General (7) | 61–74 | 63–100 | Placebo (4), standard care (1), TENS (1), muscle-strengthening exercise (1), balneotherapy (1), ice/cooling (1) |
| Ice/cooling treatment | 4 (211) | General (4) | 56–61 | 48–91 | TENS (2), acupuncture (1), standard care (1), heat treatment (1), placebo (1), no treatment (1) |
| Insoles | 6 (893) | General (6) | 58–68 | 54–100 | Placebo (5), braces (1) |
| Interferential therapy | 5 (240) | General (5) | 59–67 | 67–80 | Placebo (3), TENS (1), no treatment (1) |
| Laser/light therapy | 9 (379) | General (6), both knees affected (3) | 58–74 | 68–90 | Placebo (8), standard care (1) |
| Manual therapy | 6 (486) | General (6) | 56–68 | 63–78 | Standard care (4), placebo (2), muscle-strengthening exercise (1) |
| NMES | 3 (78) | General (3) | 60–71 | 42–79 | Standard care (2), muscle-strengthening exercise (2) |

| Intervention | Number of trials eligible for the review (number of patients[b]) | Type of population recruited[c] (number of studies) | Mean age (years), range | Female (%), range | Comparators (number of treatment arms[d]) |
|---|---|---|---|---|---|
| PES | 8 (392) | General (8) | 55–70 | 46–100 | Placebo (7), standard care (1), muscle-strengthening exercise (1), no treatment (1) |
| PEMF | 6 (521) | General (6) | 60–69 | 28–80 | Placebo (6) |
| Static magnets | 3 (131) | General (3) | 63–65 | 60–79 | Placebo (3) |
| T'ai chi | 4 (307) | General (4) | 65–70 | 75–93 | Standard care (4) |
| TENS | 18 (805) | General (17), awaiting surgery (1) | 56–85 | 48–97 | Placebo (12), standard care (3), acupuncture (3), ice/cooling (2), heat treatment (1), interferential (1), no treatment, muscle-strengthening exercise (1) |
| Weight loss (dieting) | 5 (870) | Overweight or obese (5) | 61–70 | 26–89 | Standard care (5), aerobic exercise (1) |

a  Some studies compared two or more different interventions.
b  Number of patients analysed by the primary studies for end-of-treatment pain not always clearly stated.
c  'General' is used as a broad term for studies that did not recruit very specific population types.
d  Different doses of the same treatment in a trial were pooled, counting as one arm.
e  Trial reported mean age by treatment group, and contained a group with a mean age of ≥ 55.

usual' (with or without home exercise/education). However, as analgesic adjuncts were used in only eight trials, and most studies were classified as using the 'treatment as usual' adjunct, with little adjunct detail defined, the focus of this study was on comparing the interventions categorised without adjuncts. The resulting network for any-quality studies, with analysis at the end of treatment and interventions categorised without adjuncts, is illustrated in *Figure 5*.

The interventions drawn from the any-quality trials were compared with standard care and acupuncture (*Tables 16* and *17*, respectively), with caterpillar plots shown in *Figures 6* and *7*, respectively, and interventions ordered by treatment effect. Across all comparisons, inconsistency at a *p*-value of < 0.05 was identified only for the two comparisons involving PES. Eight physical treatments had a statistically significant mean beneficial effect compared with standard care, namely interferential therapy, acupuncture, TENS, PES, balneotherapy, aerobic exercise, sham acupuncture and muscle-strengthening exercise (see *Table 16* and *Figure 6*). When acting as a comparator, acupuncture was statistically significantly better at reducing pain than sham acupuncture, muscle-strengthening exercise, weight loss, PEMF, placebo, insoles, NMES and no intervention (see *Table 17* and *Figure 7*).

Effect sizes for each intervention are presented in terms of both SMDs and the WOMAC VAS 0–100 pain scale. To help evaluate these conversions, one study reported the minimal clinically important change as –15 mm (on a VAS 0–100 scale and derived from a prior Delphi exercise)[184] and the minimal perceptible clinical improvement (MPCI, the smallest change detectable by the patient) as –9.7 mm (on a WOMAC
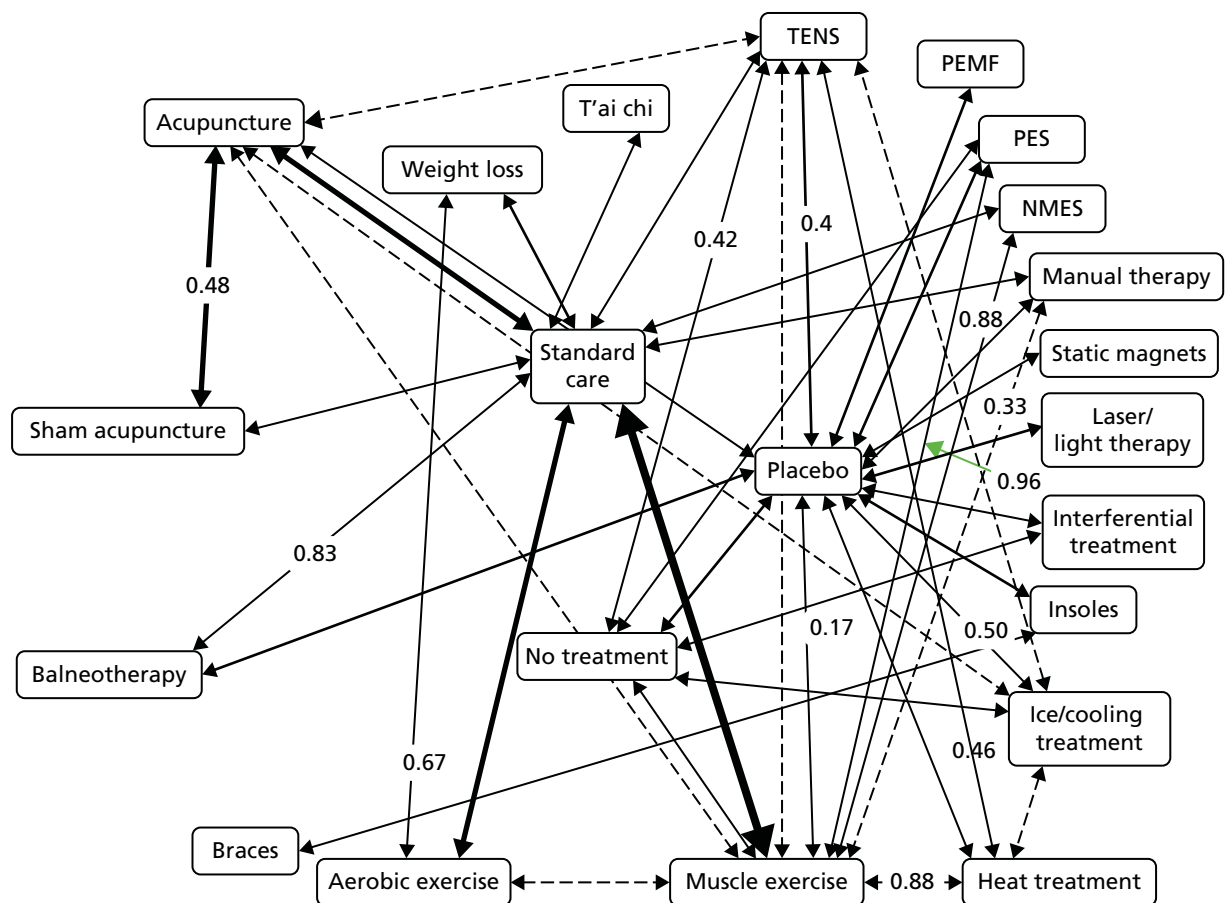


**FIGURE 5** Network diagram for the end-of-treatment analysis of any-quality studies (with interventions categorised without adjuncts). In this figure, each solid arrow indicates that there is a data point for that comparison entered into the analysis. The thickness of the arrow reflects the number of trials. The dotted line reflects an extra comparison in a multiarm trial. The numbers are a measure of inconsistency: 0 is no inconsistency; 1 is complete inconsistency.

**TABLE 16** Results of network meta-analyses for comparisons with standard care: studies of any quality

| Intervention | Number of trials (patients[a]) | SMD (95% CrI) | Difference expressed on a WOMAC VAS 0–100 pain scale (95% CrI) |
|---|---|---|---|
| Standard care (comparator) | | – (–) | – (–) |
| Interferential therapy | 3 (98) | –1.63 (–2.39 to –0.87) | –26.90 (–39.39 to –14.40) |
| Acupuncture | 24 (1219) | –0.89 (–1.18 to –0.59) | –14.69 (–19.52 to –9.80) |
| TENS | 12 (285) | –0.65 (–1.06 to –0.25) | –10.77 (–17.50 to –4.05) |
| PES | 6 (180) | –0.65 (–1.19 to –0.10) | –10.65 (–19.59 to –1.66) |
| Balneotherapy | 9 (275) | –0.60 (–1.04 to –0.15) | –9.87 (–17.15 to –2.48) |
| Aerobic exercise | 11 (428) | –0.55 (–0.89 to –0.21) | –9.02 (–14.68 to –3.51) |
| T'ai chi | 4 (159) | –0.51 (–1.03 to 0.01) | –8.39 (–16.98 to 0.13) |
| Static magnets | 2 (41) | –0.50 (–1.34 to 0.33) | –8.27 (–22.08 to 5.43) |
| Sham acupuncture | 14 (892) | –0.47 (–0.84 to –0.09) | –7.76 (–13.89 to –1.52) |
| Manual therapy | 4 (166) | –0.44 (–0.96 to 0.09) | –7.21 (–15.90 to 1.49) |
| Muscle-strengthening exercise | 28 (1254) | –0.40 (–0.61 to –0.19) | –6.54 (–9.99 to –3.11) |
| Ice/cooling treatment | 3 (51) | –0.35 (–1.03 to 0.33) | –5.81 (–16.94 to 5.44) |
| Heat treatment | 5 (123) | –0.31 (–0.86 to 0.24) | –5.14 (–14.20 to 3.98) |
| Laser therapy | 5 (155) | –0.27 (–0.86 to 0.32) | –4.53 (–14.19 to 5.20) |
| Weight loss | 5 (436) | –0.26 (–0.67 to 0.15) | –4.25 (–10.97 to 2.43) |
| PEMF | 5 (238) | –0.15 (–0.71 to 0.42) | –2.43 (–11.76 to 6.90) |
| Placebo | 42 (1077) | –0.07 (–0.42 to 0.29) | –1.15 (–6.98 to 4.70) |
| Braces | 1 (12) | 0.00 (–1.39 to 1.39) | 0.07 (–22.84 to 22.94) |
| Insoles | 3 (197) | 0.10 (–0.65 to 0.85) | 1.64 (–10.71 to 13.97) |
| NMES | 2 (28) | 0.22 (–0.62 to 1.05) | 3.58 (–10.26 to 17.33) |
| No intervention | 5 (87) | 0.44 (–0.15 to 1.04) | 7.25 (–2.51 to 17.12) |

a Number of patients analysed by the primary studies for end-of-treatment pain not always clearly stated.
Data points: 131; residual deviance: 152.
Between-study SD 0.45 (95% CI 0.36 to 0.55).

**TABLE 17** Results of network meta-analyses for comparisons with acupuncture: studies of any quality

| Intervention | Number of trials (patients[a]) | SMD (95% CrI) | Difference expressed on a WOMAC VAS 0–100 pain scale (95% CrI) |
|---|---|---|---|
| Acupuncture (comparator) | | – (–) | – (–) |
| Interferential therapy | 3 (98) | –0.74 (–1.54 to 0.05) | –12.21 (–25.33 to 0.84) |
| TENS | 12 (285) | 0.24 (–0.22 to 0.70) | 3.92 (–3.70 to 11.50) |
| PES | 6 (180) | 0.25 (–0.35 to 0.84) | 4.04 (–5.78 to 13.87) |
| Balneotherapy | 9 (275) | 0.29 (–0.22 to 0.81) | 4.82 (–3.60 to 13.28) |
| Aerobic exercise | 11 (428) | 0.34 (–0.11 to 0.79) | 5.67 (–1.84 to 13.00) |

**TABLE 17** Results of network meta-analyses for comparisons with acupuncture: studies of any quality (*continued*)

| Intervention | Number of trials (patients[a]) | SMD (95% CrI) | Difference expressed on a WOMAC VAS 0–100 pain scale (95% CrI) |
|---|---|---|---|
| T'ai chi | 4 (159) | 0.38 (–0.22 to 0.98) | 6.30 (–3.58 to 16.12) |
| Static magnets | 2 (41) | 0.39 (–0.48 to 1.25) | 6.41 (–7.86 to 20.61) |
| Sham acupuncture | 14 (892) | 0.42 (0.15 to 0.70) | 6.93 (2.50 to 11.46) |
| Manual therapy | 4 (166) | 0.45 (–0.14 to 1.05) | 7.47 (–2.30 to 17.23) |
| Muscle-strengthening exercise | 28 (1254) | 0.49 (0.15 to 0.84) | 8.14 (2.41 to 13.83) |
| Ice/cooling treatment | 3 (51) | 0.54 (–0.16 to 1.25) | 8.88 (–2.70 to 20.61) |
| Heat treatment | 5 (123) | 0.58 (–0.02 to 1.18) | 9.55 (–0.30 to 19.44) |
| Laser therapy | 5 (155) | 0.62 (–0.02 to 1.25) | 10.16 (–0.28 to 20.61) |
| Weight loss | 5 (436) | 0.63 (0.13 to 1.14) | 10.44 (2.13 to 18.72) |
| PEMF | 5 (238) | 0.74 (0.13 to 1.36) | 12.26 (2.22 to 22.36) |
| Placebo | 42 (1077) | 0.82 (0.40 to 1.25) | 13.53 (6.58 to 20.53) |
| Standard care | 53 (2308) | 0.89 (0.59 to 1.18) | 14.69 (9.80 to 19.52) |
| Braces | 1 (12) | 0.89 (–0.51 to 2.31) | 14.76 (–8.49 to 38.01) |
| Insoles | 3 (197) | 0.99 (0.21 to 1.78) | 16.33 (3.41 to 29.30) |
| NMES | 2 (28) | 1.11 (0.22 to 1.98) | 18.27 (3.57 to 32.72) |
| No intervention | 5 (87) | 1.33 (0.69 to 1.97) | 21.95 (11.30 to 32.52) |

a  Number of patients analysed by the primary studies for end-of-treatment pain not always clearly stated.
Data points: 131; residual deviance: 152.
Between-study SD 0.45 (95% CI 0.36 to 0.55).
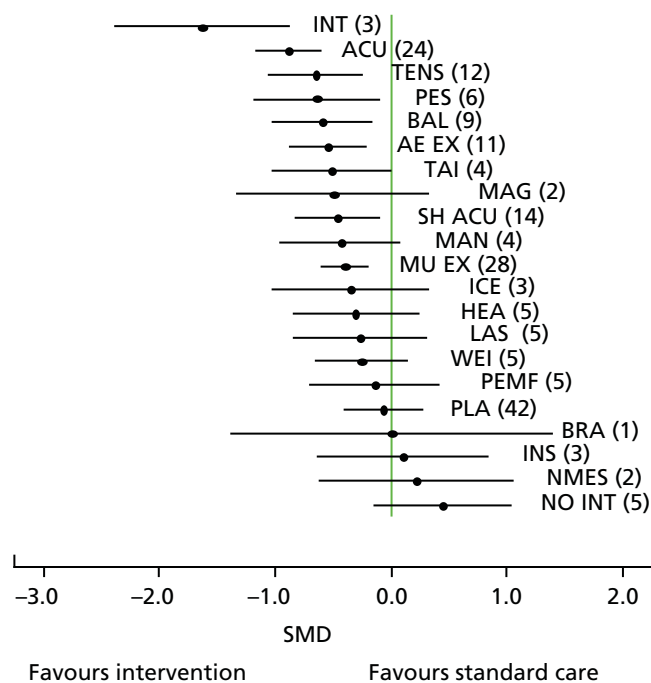


**FIGURE 6** Standardised mean differences of each treatment compared with standard care for the analysis including studies of any quality. Numbers in parentheses indicate the numbers of studies. ACU, acupuncture; AE EX, aerobic exercise; BAL, balneotherapy; BRA, braces; HEA, heat treatment; ICE, ice/cooling treatment; INS, insoles; INT, inferential therapy; LAS, laser therapy; MAG, static magnets; MAN, manual therapy; MU EX, muscle-strengthening exercise; NO INT, no intervention; PLA, placebo; SH ACU, sham acupuncture; TAI, t'ai chi; WEI, weight loss.
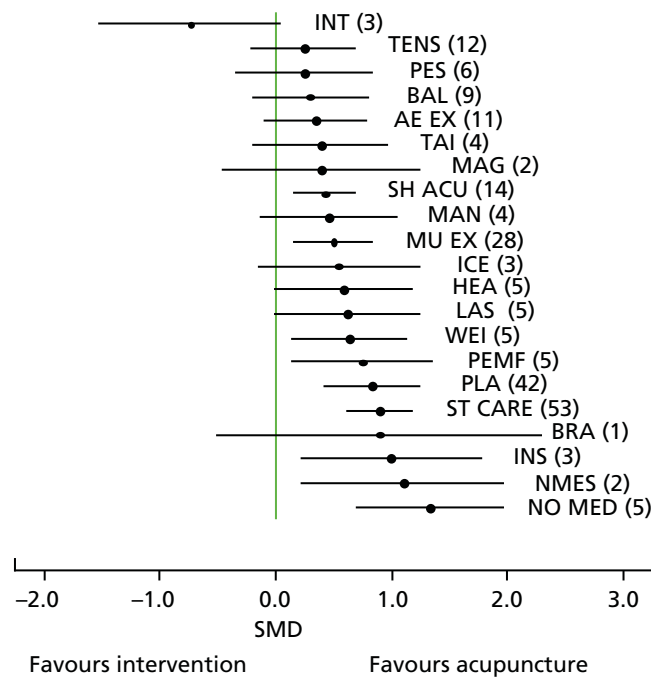
**FIGURE 7** Standardised mean differences of each treatment compared with acupuncture for the analysis including studies of any quality. Numbers in parentheses indicate the numbers of studies. AE EX, aerobic exercise; BAL, balneotherapy; BRA, braces; HEA, heat treatment; ICE, ice/cooling treatment; INS, insoles; INT, inferential therapy; LAS, laser therapy; MAG, static magnets; MAN, manual therapy; MU EX, muscle-strengthening exercise; NO MED, no intervention; PLA, placebo; SH ACU, sham acupuncture; ST CARE, standard care; TAI, t'ai chi; WEI, weight loss.

VAS 0–100 scale).[185] Another study estimated the minimal clinically important improvement (MCII), although only for pain on movement, as –19.9 mm on a VAS 0–100 scale; this figure varied by baseline pain score, with patients with less pain having a smaller MCII (10.8 mm) and patients with severe pain having a larger MCII (36.6 mm).[186]

When analysing only the better-quality studies (see *Appendix 3*) in the primary sensitivity analysis, 35 trials were included, with nine types of intervention and 3499 patients. One study was identified as causing inconsistency in the main analysis (a small study of muscle-strengthening exercise vs. PES) and was therefore excluded.[187] The network is illustrated in *Figure 8*, in which the analysis is at the end of treatment and interventions are categorised without adjuncts. Uncertainty around the true between-study variance increased because of the reduction in the number of studies per comparison, as well as loops in the network. Most studies were of acupuncture ($n = 11$) or muscle-strengthening exercise ($n = 9$), with some interventions represented by few studies.

When compared with standard care, there was a statistically significant reduction in pain for acupuncture, balneotherapy, sham acupuncture and muscle-strengthening exercise (*Table 18* and *Figure 9*). When acupuncture was the comparator, it was statistically significantly better at a 95% level of credibility than sham acupuncture, muscle-strengthening exercise, weight loss, aerobic exercise and no intervention (*Table 19* and *Figure 10*).

In terms of ranking, a probability statistic calculated from the treatment effect distributions showed that acupuncture and balneotherapy were the two interventions with the highest rank (*Table 20*). Because of overlapping CrIs for sham acupuncture, muscle-strengthening exercise and t'ai chi, there is some uncertainty around these rankings.

In a secondary sensitivity analysis several trials were excluded based on population or intervention differences, or on extreme data;[136,188–193] the results were not sensitive to these changes, although the model fit improved, as reported elsewhere.[169] No network link could be made with the placebo-controlled studies
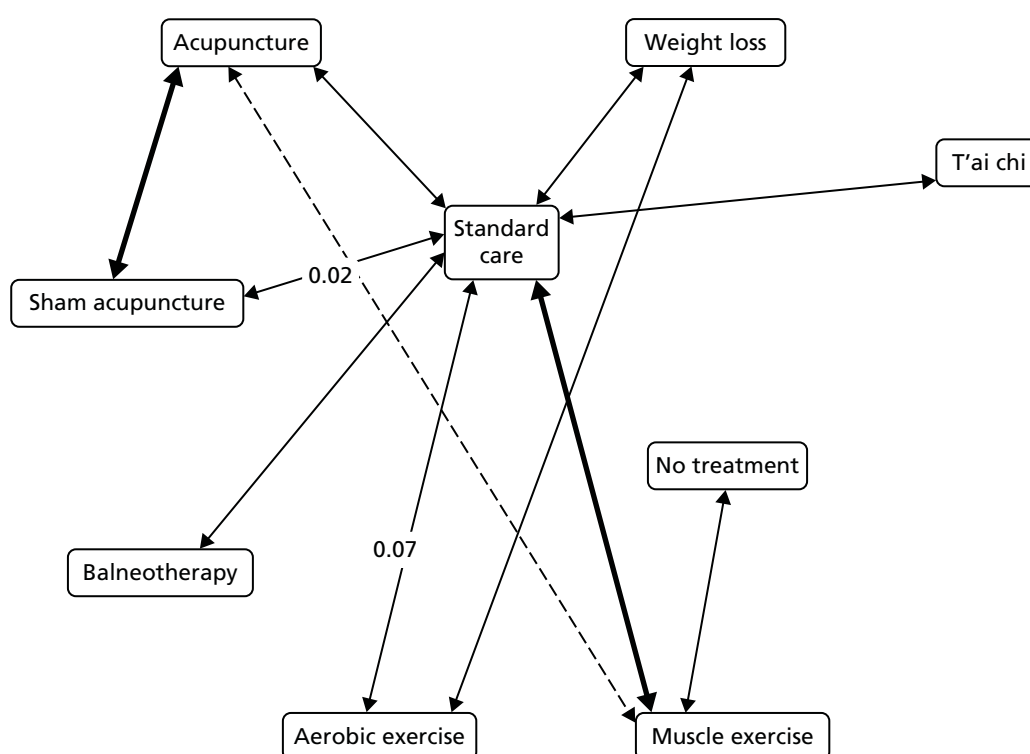
**FIGURE 8** Network diagram for the end-of-treatment analysis of better-quality studies (with interventions categorised without adjuncts).

**TABLE 18** Results of network meta-analyses for comparisons with standard care: studies of better quality

| Intervention | Number of trials (patients[a]) | SMD (95% CrI) | Difference expressed on a WOMAC VAS 0–100 pain scale (95% CrI) |
|---|---|---|---|
| Standard care (comparator) | | – (–) | – (–) |
| Acupuncture | 11 (878) | –1.01 (–1.43 to –0.61) | –16.70 (–23.61 to –10.07) |
| Balneotherapy | 1 (40) | –1.01 (–1.92 to –0.11) | –16.65 (–31.73 to –1.74) |
| Sham acupuncture | 8 (685) | –0.68 (–1.17 to –0.19) | –11.14 (–19.29 to –3.16) |
| Muscle-strengthening exercise | 9 (450) | –0.52 (–0.84 to –0.22) | –8.62 (–13.92 to –3.58) |
| T'ai chi | 2 (51) | –0.26 (–0.96 to 0.44) | –4.29 (–15.87 to 7.23) |
| Weight loss | 3 (357) | –0.08 (–0.55 to 0.39) | –1.34 (–9.10 to 6.41) |
| Aerobic exercise | 1 (80) | 0.07 (–0.69 to 0.84) | 1.23 (–11.30 to 13.78) |
| No intervention | 1 (30) | 0.19 (–0.77 to 1.14) | 3.11 (–12.72 to 18.77) |

a  Number of patients analysed by the primary studies for end-of-treatment pain not always clearly stated.
Data points: 31; residual deviance: 31.4.
Between-study SD 0.39 (95% CI 0.24 to 0.58).
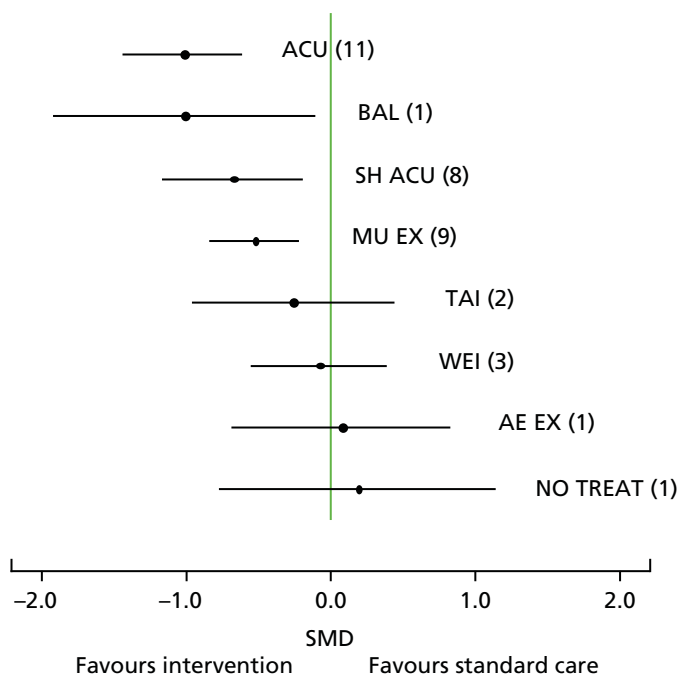
**FIGURE 9** Standardised mean differences of each treatment compared with standard care for the analysis including better-quality studies. Numbers in parentheses indicate the numbers of studies. ACU, acupuncture; AE EX, aerobic exercise; BAL, balneotherapy; MU EX, muscle-strengthening exercise; NO TREAT, no intervention; SH ACU, sham acupuncture; TAI, t'ai chi; WEI, weight loss.

**TABLE 19** Results of network meta-analyses for comparisons with acupuncture: studies of better quality

| Intervention | Number of trials (patients[a]) | SMD (95% CrI) | Difference expressed on a WOMAC VAS 0–100 pain scale (95% CrI) |
|---|---|---|---|
| Acupuncture (comparator) | | – (–) | – (–) |
| Balneotherapy | 1 (40) | 0.00 (–0.99 to 1.01) | 0.05 (–16.36 to 16.62) |
| Sham acupuncture | 8 (685) | 0.34 (0.03 to 0.66) | 5.57 (0.42 to 10.86) |
| Muscle-strengthening exercise | 9 (450) | 0.49 (0.00 to 0.98) | 8.08 (0.02 to 16.21) |
| T'ai chi | 2 (51) | 0.75 (–0.05 to 1.57) | 12.42 (–0.81 to 25.84) |
| Weight loss | 3 (357) | 0.93 (0.31 to 1.57) | 15.36 (5.18 to 25.81) |
| Standard care | 17 (928) | 1.01 (0.61 to 1.43) | 16.70 (10.07 to 23.61) |
| Aerobic exercise | 1 (80) | 1.09 (0.23 to 1.96) | 17.94 (3.82 to 32.27) |
| No intervention | 1 (30) | 1.20 (0.18 to 2.23) | 19.80 (2.94 to 36.81) |

a   Number of patients analysed by the primary studies for end-of-treatment pain not always clearly stated.
Data points: 31; residual deviance: 31.4.
Between-study SD 0.39 (95% CI 0.24 to 0.58).
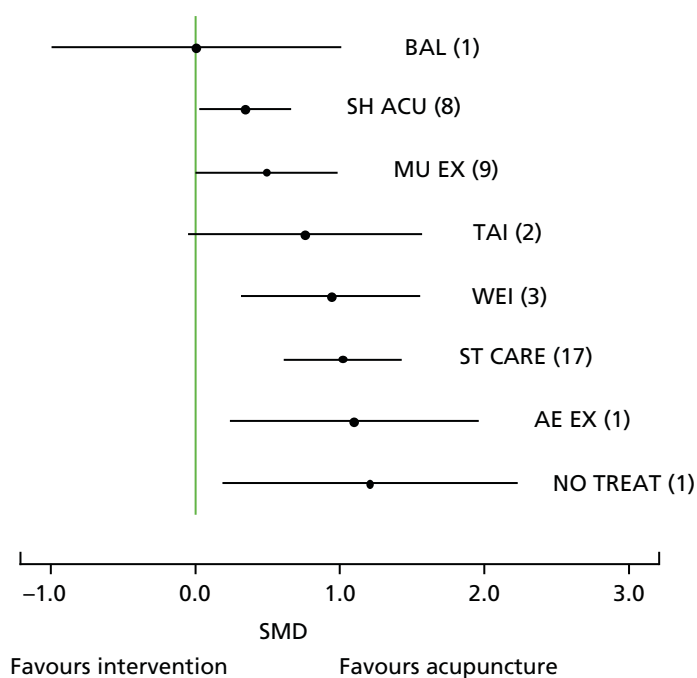
**FIGURE 10** Standardised mean differences of each treatment compared with acupuncture for the analysis including better-quality studies. Numbers in parentheses indicate the numbers of studies. AE EX, aerobic exercise; BAL, balneotherapy; MU EX, muscle-strengthening exercise; NO TREAT, no intervention; SH ACU, sham acupuncture; ST CARE, standard care; TAI, t'ai chi; WEI, weight loss.

**TABLE 20** Ranking of interventions (using only better-quality studies)

| Intervention | Number of trials | Median rank | 95% Crl |
|---|---|---|---|
| Acupuncture | 11 | 2 | 1 to 3 |
| Balneotherapy | 1 | 2 | 1 to 6 |
| Sham acupuncture | 8 | 3 | 2 to 6 |
| Muscle-strengthening exercise | 9 | 4 | 2 to 6 |
| T'ai chi | 2 | 5 | 2 to 9 |
| Weight loss | 3 | 6 | 4 to 9 |
| Standard care | 17 | 7 | 6 to 9 |
| Aerobic exercise | 1 | 8 | 3 to 9 |
| No intervention | 1 | 8 | 3 to 9 |

in the analysis of better-quality studies. We therefore conducted a separate network meta-analysis for these studies. Both interferential therapy and heat treatment were statistically significantly more effective than placebo, but laser therapy, PES and insoles were not; these data are also reported elsewhere.[169]

In the substudy on the reporting of the WOMAC pain subscale and WOMAC index, the former was reported in 60 (45%) trials and the latter in 31 (23%) trials. Reporting of the exact method used in administering the WOMAC pain subscale scoring was poor in many cases. Overall, only 15 (25%) trials reported unambiguously both the scale and the score range for their use of the WOMAC pain subscale. Only four (13%) trials reported unambiguously both the scale and the score range for their use of the WOMAC index. Further details of the results of this substudy are reported elsewhere.[183]

## Discussion

### Principal findings

In the comprehensive network meta-analysis that we report here we compared all physical treatments for osteoarthritis of the knee with each other within a coherent framework. This analysis provides the first estimate of the relative effect of these treatments, which can be viewed as essential for decision-makers when comparing treatment effects. By providing a basis for synthesising all of the available evidence in a consistent framework, a network meta-analysis obviates the need to make decisions based on subjective inferences from disconnected data.

Compared with standard care, eight of the 22 interventions that we evaluated produced a statistically significant reduction in pain: interferential therapy, acupuncture, TENS, PES, balneotherapy, aerobic exercise, sham acupuncture and muscle-strengthening exercise. Of these eight, only two interventions were represented by more than three trials in the sensitivity analysis of better-quality studies: acupuncture (11 trials) and muscle-strengthening exercise (nine trials), with acupuncture having statistically significantly better outcomes. Acupuncture and balneotherapy (only one trial) were the two interventions with the highest rank, although there is some uncertainty around this. For the better-quality placebo-controlled studies, interferential therapy (one trial) showed a strong effect compared with placebo.

### Strengths and limitations

Numerous systematic reviews, some summarised in a review of reviews,[194] have evaluated the interventions (or classes of interventions) included in this review. However, our analysis represents the use of the most practical methods currently available to compare a large number of different types of treatment, enabling a fair comparison to be made of competing physical treatments (including acupuncture) with each other.

A network meta-analysis requires an assumption of exchangeability between the trials in the same way as is required for a standard meta-analysis. With regard to concerns that might arise from within- or between-intervention heterogeneity, we sought to minimise these by using an age restriction as part of our inclusion criteria and by excluding interventions consisting of more than one physical treatment. We found that patient characteristics appeared to be broadly comparable across interventions. Inevitably, there will be some clinical heterogeneity in a wide-ranging study such as this, but as far as it was possible to tell, given the wide variation of scales used, baseline pain did not appear to vary systematically between interventions.

We used a random-effects model to incorporate heterogeneity and we evaluated levels of inconsistency and model fit. We also conducted sensitivity analyses excluding trials causing heterogeneity. Although heterogeneity is accounted for in our results with the CrIs, it is possible that unknown confounding factors may be affecting the results of indirect comparisons. With regard to trials of placebo interventions, the majority used electrical or electromagnetic interventions and so it is not unreasonable to assume that the placebo effects were similar (as the interventions were similar). In our review the trials, covering a diverse range of interventions, were all assessed using the same quality assessment tools, which enabled equivalence in the comparisons and better interpretation of the evidence base for each intervention.

Our sensitivity analysis of the better-quality studies resulted in fewer trials per comparison and fewer network loops. This led to greater uncertainty about the true heterogeneity and about the differences between the direct and the indirect evidence. The uncertainty associated with inconsistency may not be fully captured in the results because fewer loops in relation to the size of the network meant that there were fewer data to quantify inconsistency.

We were not able to include all of the studies in our analyses because of the variable reporting of pain results. Moreover, our analyses focused only on the end-of-treatment data and these were available mostly for short-term time periods. Of the trials that investigated effectiveness over medium- or long-term time periods, only a few provided the data required for our analyses. However, a comparison of the maximum

effect of interventions is not without merit, given that the treatments under consideration are not intended as being cures and that any treatment effect might be expected to attenuate over time.

It is important that our results are evaluated in context. Most of the studies in our review were rated as being of poor quality. Many of the better-quality studies were pragmatic trials in which blinding of patients was not possible, that is, most studies are likely to have been subject to some form of bias. For this reason there can be methodological limitations in clinical trials of physical treatments that are often inherent and unavoidable. For the trials in which patients were not blinded and treatments were compared with standard care, the overall treatment effect is likely to incorporate non-specific (placebo) effects. We assumed that such non-specific effects were similar across all interventions, but variation may in fact be present. However, there were also limitations that could have been avoided by triallists using better methodology and reporting practices. For example, in our substudy on WOMAC reporting,[183] we found poor reporting of both the WOMAC pain subscale and the WOMAC index, which in turn resulted in significant uncertainty in the interpretation of the results of individual trials and limited their contribution to our evidence synthesis.

### Comparison with the wider literature
In light of our results, it is worth considering what might be the true (or specific) effect of acupuncture. In a Cochrane review,[142] a statistically significant, clinically relevant, short-term improvement in pain was reported (acupuncture vs. waiting list control: SMD –0.96, 95% CI –1.19 to –0.72), a similar finding to what we have reported. A similar effect to ours was also observed in the comparison of acupuncture with sham acupuncture (SMD –0.35, 95% CI –0.55 to –0.15). It is worth noting that the largest study[70] in this Cochrane analysis, which showed no statistically significant difference between acupuncture and sham acupuncture, was one of two trials that used an intensive sham needling technique, which may have had physiological effects. Also, our analysis included a recent large trial[125] that used what appeared to be a very active sham. Therefore, the inclusion of trials with sham controls that might be more active than an inert placebo control could lead to pooled results that underestimate the short-term effect of acupuncture. It is of interest to note that the effect size of acupuncture compared with sham acupuncture is of the same order as that seen for NSAIDs compared with placebo (SMD 0.32, 95% CI 0.24 to 0.39), a difference that has also been described as being too small to be clinically significant.[195]

An IPD meta-analysis that included an evaluation of acupuncture for patients with knee osteoarthritis was recently reported[196] (see *Chapter 2*). All included studies were deemed be of high quality because the allocation concealment methods were assessed to be unambiguously adequate. This study also found acupuncture to be more effective than sham acupuncture and with a smaller effect size than when acupuncture was compared with no-acupuncture (usual care) controls. These findings indicate that non-specific effects provide a partial contribution to the pain-alleviating effects of acupuncture. Non-specific effects will also be contributing to the effectiveness of other (non-acupuncture) interventions in our network meta-analysis. When interventions were not controlled by a placebo or a relevant sham, commonly when blinding was not possible, the contribution of non-specific effects to the overall effect cannot be estimated. However, given that there are inherent problems with identifying non-specific effects in interventions involving physical treatments, it is reasonable to assume that fair comparisons between treatments have been made.

There is some evidence to suggest that larger treatment effects are associated with sham acupuncture than with pharmacological or other physical placebos.[145,197] However, one of two contrasting factors may impact on the effect of sham acupuncture in a given trial: either there is inadequate patient blinding because of using unsuitable shams or there is the use of physiologically active shams. The former may lead to an overestimation of the true effect of acupuncture, whereas the latter may lead to an underestimation of the true effect. In the trials that we have reviewed in this study we found that important details about sham acupuncture (e.g. depth of insertion) were sometimes poorly reported or were not reported at all. As with the variations in styles of acupuncture, so too were there variations in the types of sham acupuncture, both contributing to the possibility of clinical heterogeneity.

## Implications for clinical practice

Five guidelines[82,161,198–200] have evaluated treatment effects on key outcomes of knee osteoarthritis (including pain, function and disability). Only the Osteoarthritis Research Society International (OARSI) guideline[200] is unequivocal in its recommendation to offer acupuncture for knee osteoarthritis. The American College of Rheumatology (ACR)[198] conditionally recommended acupuncture but only for patients with moderate to severe pain who are unable or unwilling to undergo total knee arthroplasty. The American Academy of Orthopaedic Surgeons (AAOS)[199] found the acupuncture evidence to be inconclusive and the European League Against Rheumatism (EULAR)[161] and NICE[82] did not recommend the use of acupuncture. Our analyses of the better-quality studies suggest that acupuncture should be considered as one of the short-term physical treatment options for relieving pain caused by osteoarthritis of the knee.

Guidance from all organisations recommended treatment with muscle-strengthening and aerobic exercise, education, weight loss (if required) and, when necessary, paracetamol and/or topical NSAIDs; when these are ineffective, a choice of one or more options from a range of pharmacological and non-pharmacological treatments is sometimes recommended, including TENS, thermal (heat/cooling) treatments, insoles and braces. Some of our results on effectiveness do not concur with existing guidance on the (non-acupuncture) physical treatments: our evidence differs from the EULAR guidelines[161] with regard to insoles, braces and weight loss; from the NICE guidelines[82] with regard to TENS, insoles, braces, weight loss, manual therapy and heat or cooling treatment; from the ACR guidelines[198] with regard to weight loss, insoles, thermal agents and t'ai chi; from the AAOS guidelines[199] with regard to weight loss; and from the OARSI guidelines[200] with regard to insoles, braces, heat or cooling treatment, TENS and weight loss. Our analyses found little evidence (of significant differences from standard care, let alone clinically relevant differences) to support such guidance with respect to treating pain, other than for TENS, for which the evidence was of poor quality and likely to be unreliable. It should be remembered, however, that our review was focused on pain outcomes rather than on function, disability or cost-effectiveness.

The clinical relevance of improvements in knee pain scores has been quantified in several ways. In this context, our better-quality trial results appear to indicate that acupuncture produces both a MPCI[185] and quite possibly a minimum clinically important change,[184,185] but may yield a MCII only for patients with low levels of pain.[186] A MPCI remains a possibility for muscle-strengthening exercise (with evidence from nine trials). Our better-quality results suggest that few physical treatments are likely to have a clinically relevant pain-relieving effect. The exceptions were balneotherapy, interferential therapy and heat treatment for which we found evidence of effectiveness compared with standard care. However, the results for these three interventions were informed by single small studies and so a cautious interpretation is warranted.

When interpreting effectiveness results, other factors to consider beyond effectiveness are acceptability, safety, rapidity and durability of benefit, convenience, cost and likelihood of patient adherence to treatment.[201] Given the diverse range of interventions that we studied, these factors will clearly differ between interventions, as well as in relation to pharmacological and other treatments.

## Recommendations for future research

To comprehensively assess the value of many of these interventions, larger RCTs, with risk of bias reduced and with longer treatment periods, are needed. Given the stronger evidence on acupuncture and muscle-strengthening exercise in the better-quality trials, there is a need in future studies to determine the optimum timing and parameters of treatment. Ideally, trials should examine the effectiveness of retreatment following treatment cessation (to evaluate durability and attenuation effects), which would match the way that these physical treatments are often delivered in practice.

In the substudy on standards of reporting of WOMAC scales,[183] we found that in general the reporting of methods and results in RCTs using the WOMAC assessment tool lacked clarity. Poor reporting of WOMAC scales limits the interpretation of trial results and their useability for evidence synthesis. Given that the various versions of WOMAC available are clearly defined and have all been validated, full descriptions by researchers are needed. Adherence to the standard WOMAC scoring system should be encouraged.

As an absolute minimum, the type of WOMAC used and the score range must be reported. Clear reporting is important and should not be sacrificed to reduce word count.

## Conclusion

The evidence available for our network meta-analyses, in which physical interventions for osteoarthritis of the knee were compared equally with each other within a coherent framework, suggests that the evidence of effectiveness for most interventions is weak. However, when comparing all interventions, whether based on the any-quality or the better-quality trials, acupuncture can be considered as one of the more effective physical treatments for alleviating pain in the short term. Despite the large evidence base found, the methodological limitations associated with many of the trials indicate that high-quality trials of many of the physical treatments are still required.

# Chapter 4　Towards a cost-effectiveness analysis of acupuncture for chronic pain: developing methods in a case study

## Background

### Evidence synthesis in health technology assessment

Cost-effectiveness analyses of health technologies have a number of key requirements.[202] These analyses should entail (1) a clear definition of the decision problem, which should include all relevant comparators; (2) an appropriate time horizon for the analysis; (3) the systematic identification and consideration of all relevant evidence;[203] (4) an appropriate characterisation of all sources of uncertainty; and (5) an assessment of the value of acquiring additional research. It is extremely rare for the evidence base informing a cost-effectiveness analysis to come from a single study.[204] Data (typically summary study level) are derived from multiple sources and are often available in multiple formats, for example using different instruments for measurement and using different measures of effect reported at different time points. Evidence synthesis and decision modelling are used extensively in health technology appraisal to meet the challenge of reflecting these disparate sources of evidence within a coherent framework.

Synthesis tools are increasingly used to obtain pooled estimates of the parameters of interest to inform economic decision models. This is particularly the case for treatment effect estimates when multiple relevant RCTs may be available. In many circumstances the synthesis of treatment effect evidence considers only pairwise comparisons through the use of standard meta-analysis. However, frequently there are more than two treatment choices. Network meta-analysis (also known as mixed-treatment comparisons) is a tool that extends standard pairwise meta-analysis, allowing the estimation of the relative effectiveness of multiple treatments by simultaneously synthesising all relevant evidence. This statistical method is a well-established technique and its methods have been described extensively in the literature.[85,86,178,205–208] As is the case for standard meta-analysis, most published work using network meta-analysis focuses on the synthesis of aggregate data. These data are usually obtained from published literature and consist of an estimate of treatment effectiveness (e.g. mean difference in the case of continuous outcomes) and an appropriate measure of uncertainty (e.g. the variance or SE).

### Network meta-analysis and the use of individual patient data

With the increasing availability of IPD for economic evaluation, together with considerable support for utilising this type of evidence,[209,210] meta-analytic methods have emerged to address the challenges of IPD study synthesis.[211,212] Most progress has been made in the area of the statistical synthesis of clinical effectiveness, whereas little work has been undertaken to address the challenges of synthesising information on other important decision model parameter types.[213] Techniques for pairwise meta-analysis of individual-level evidence exist for most outcome types including binary,[214] continuous[210,215] and time-to-event data.[216,217] Use of IPD to inform decisions creates added value by offering the potential to reduce network heterogeneity, tackle existing evidence inconsistencies[172] and examine subgroup effects in patients in whom interventions might have an effectiveness and cost-effectiveness profile which differs from that of the wider population.[218] Few methodological studies on the synthesis of IPD in network meta-analysis are available in the published literature and even fewer examples of its use within cost-effectiveness analysis exist.[213]

### Objectives and structure

Given the potential benefits of IPD network meta-analysis as a basis for informing cost-effectiveness modelling and the paucity of examples of this approach in the literature, we present a case study of using IPD network meta-analysis to inform a cost-effectiveness analysis of acupuncture for chronic pain.

The objectives of this research were to both develop novel methods for IPD network meta-analysis and demonstrate the application of IPD network meta-analysis for use in economic evaluation.

To our knowledge the current synthesis methods literature does not offer modelling tools for continuous data within an IPD network meta-analysis framework. Using a pairwise meta-analysis framework, Riley *et al.*[219] discuss different approaches to the synthesis of continuous outcome data when IPD are available. Riley *et al.*[219] highlight that modelling the follow-up result, adjusted for the baseline value, commonly called ANCOVA,[219,220] is the preferred approach. The availability of IPD is crucial for such models. If IPD are not available, the use of ANCOVA would require all original study authors to have reported appropriate treatment effect estimates, ideally at the same follow-up time. These requirements, in most circumstances, make this option unfeasible. Flexibility is introduced with the availability of IPD as the analyst can apply the same modelling approach across trials and derive consistent outputs. This report describes a novel, methodological framework for IPD network meta-analysis of continuous data within the Bayesian framework, which builds on the work described in Riley *et al.*[219]

Two approaches to synthesising data on heterogeneous continuous outcomes are explored. The first involves standardising outcomes by dividing primary outcome scores by study-specific SDs. This creates a dimensionless measure of treatment effect usually termed the SMD.[221–223] Although commonly used, this approach does not produce results that can directly feed into cost-effectiveness analysis models, as absolute treatment effect estimates are required. Furthermore, health-care policy-makers require a common health outcome measure to be able to make decisions across different conditions and clinical areas. In many jurisdictions, including England and Wales,[224] this measure is the QALY.[225] The QALY is a composite measure and provides an estimate of an individual's remaining life expectancy weighted by a preference-based measure of HRQoL. The most popular HRQoL measure for generating quality-of-life weights is the EQ-5D.[226] These considerations motivate the second synthesis approach used, which involves translating (or 'mapping') the available HRQoL data from the trials to EQ-5D values and synthesising the resulting data.

After describing the motivating data set in *Motivating case study: the cost-effectiveness of acupuncture for chronic pain in primary care*, the core of *Methods* outlines the novel statistical models for the IPD network meta-analysis. A variety of modelling approaches are described and discussed. This section also describes how comparable end points suitable for synthesis were obtained, the estimation of costs and the cost-effectiveness modelling methods. The section *Application* provides the results of the IPD network meta-analysis and cost-effectiveness analysis. The discussion section offers concluding remarks and discusses relevant issues, including extensions to the current work.

## Motivating case study: the cost-effectiveness of acupuncture for chronic pain in primary care

### Background to acupuncture and acupuncture guidance
There is currently a lack of agreement about the effectiveness of acupuncture as a treatment for chronic pain, as reflected in debates about recent UK guidance surrounding its value.[83,84,227] Acupuncture received a positive recommendation from NICE for its use in back pain[81] and headache/migraine,[80] whereas a negative recommendation was given for its use in osteoarthritis in 2008[228] and 2014.[82] The methods in this chapter were developed as part of a project to improve evidence regarding the clinical effectiveness and cost-effectiveness of acupuncture for chronic non-specific pain to inform decision-making in the UK NHS.

### Data description and network of evidence
Data for the current study were made available by the ATC. To address the lack of good-quality evidence in acupuncture, the ATC undertook a systematic review in which relevant high-quality trials were identified and, for a large proportion, IPD were obtained.[196] From 31 eligible RCTs from the ATC database, IPD were obtained from 29. However, data from Cherkin *et al.*[131] were not available to us because of sharing

restrictions. The data set analysed here included 28 high-quality RCTs[60,61,66–77,117–122,130,132–138] that assessed the effectiveness of acupuncture for three pain conditions: osteoarthritis of the knee (seven trials), headache [including tension-type headache (TTH)] and migraine (six trials), and musculoskeletal pain, encompassing lower back, shoulder and neck pain (15 trials), totalling approximately 17,500 patients from the USA, UK, Germany, Spain and Sweden. These studies are summarised in *Table 21*.

Nine of these studies were three-arm trials, assessing the three treatments simultaneously, 11 evaluated acupuncture and sham acupuncture only, and eight considered acupuncture and usual care only. Thus, the comparison of acupuncture with usual care is informed by 17 studies, acupuncture with sham acupuncture by 20 studies and sham acupuncture with usual care by nine studies. The resulting evidence network is presented in *Figure 11*.

Resource use information in the data set was limited to five[60,61,66,69,71] of the 28 studies. Of these, three are specific to the German health-care system[66,69,71] and, given the jurisdiction-specific nature of health-care resource use data, non-UK studies are of limited value to inform decision-making in the UK.[229] The remaining two studies that provided resource use evidence[60,61] were carried out in the UK, although only one of these[60] collected resource use for time points that matched the 3-month time frame of the effectiveness assessment. The study by Thomas *et al.*[61] recorded resource use at 12 and 24 months only. The study by Vickers *et al.*[60] focused on one of the three clinical areas of interest (headache/migraine). It was decided to seek additional external data sets outside the ATC data set instead of assuming that health-care resource use data from headache/migraine trials could be generalised to musculoskeletal conditions and osteoarthritis of the knee. Therefore, IPD were obtained from the UK Back pain Exercise And Manipulation (BEAM) study[230] for musculoskeletal pain and the UK Topical or Oral IBuprofen (TOIB) study[231] for knee osteoarthritis pain.

## Methods

This section describes how individual-level comparable values for the two end points of interest were generated, that is, EQ-5D index values and standardised pain scores. The section then goes on to describe the Bayesian IPD network meta-analysis synthesis modelling framework for both end points. Extensions to the modelling approach are then considered. Following this, methods used to analyse resource use and costs, and to generate cost-effectiveness results, are described.

### *Overview of the analysis*

The ATC data set includes trials comparing acupuncture with sham acupuncture or usual care or comparing all three comparators. All trials were included in the synthesis to maximise use of the available data. In the context of clinical or other health-care decision-making, a sham comparator is not a clinically meaningful intervention as it would not be prescribed; the focus of the cost-effectiveness analysis was therefore on acupuncture compared with usual care. The range of possible treatment options for chronic non-specific pain in primary care is much wider than those considered here. The cost-effectiveness results presented in this report should not therefore be interpreted as providing a definitive answer to the question of whether or not acupuncture is cost-effective for the treatment of chronic non-specific pain. Instead, the cost-effectiveness analysis provides an illustration of how IPD network meta-analysis can be used.

Two outcome measures were used to value benefit in the present analysis: pain and EQ-5D index values. Given the spectrum of pain conditions considered, the availability of multiple instruments with which to measure pain, the lack of agreement about the preferred outcomes with which to measure pain and variable quality in reporting, pain measurement was highly heterogeneous across trials. SMDs were used to synthesise the pain outcomes. The HRQoL data were also obtained using a variety of instruments – some generic [e.g. Short Form questionnaire-12 items (SF-12) and SF-36] and others disease specific (e.g. the WOMAC index). These data were therefore mapped to EQ-5D values using a series of statistical mapping algorithms, which are described below.

**TABLE 21** Data set main characteristics and study outcomes used for analysis

| ID | Study | Location | Pain group (type) | Age (years), mean (SD) | Follow-up/time point used (months) | Treatment | Observations | HRQoL outcome mapped | Pain outcome standardised |
|----|-------|----------|-------------------|------------------------|-----------------------------------|-----------|--------------|----------------------|---------------------------|
| 1 | Diener 2006[72] | Germany | Headache (migraine) | 37.62 (10.4) | 6/3 | Usual care | 328 | SF-12 | Migraine days |
| | | | | | | Sham acupuncture | 202 | | |
| | | | | | | Acupuncture | 305 | | |
| 2 | Endres 2007[73] | Germany | Headache (TTH) | 38.44 (11.77) | 6/3 | Sham acupuncture | 200 | SF-12 | TTH days |
| | | | | | | Acupuncture | 209 | | |
| 3 | Jena 2008[76] | Germany | Headache (headache) | 43.66 (12.69) | 6/3 | Usual care | 1613 | SF-36 | Headache days |
| | | | | | | Acupuncture | 1569 | | |
| 4 | Linde 2005[68] | Germany | Headache (migraine) | 42.55 (11.35) | 6/3 | Usual care | 76 | SF-36 | Days of moderate to severe pain |
| | | | | | | Sham acupuncture | 81 | | |
| | | | | | | Acupuncture | 145 | | |
| 5 | Melchart 2005[67] | Germany | Headache (TTH) | 42.68 (13.18) | 6/3 | Usual care | 75 | SF-36 | Headache days |
| | | | | | | Sham acupuncture | 62 | | |
| | | | | | | Acupuncture | 132 | | |
| 6 | Vickers 2004[60] | UK | Headache (headache) | 46.34 (10.39) | 12/3 | Usual care | 161 | SF-36 | Severity score |
| | | | | | | Acupuncture | 140 | | |
| 7 | Brinkhaus 2006[66] | Germany | Musculoskeletal (lower back) | 58.81 (9.13) | 12/2 | Usual care | 79 | SF-36 | VAS pain score |
| | | | | | | Sham acupuncture | 73 | | |
| | | | | | | Acupuncture | 146 | | |
| 8 | Carlsson 2001[119] | Sweden | Musculoskeletal (lower back) | 49.84 (15.4) | 6/3 | Sham acupuncture | 16 | VAS pain | VAS pain score |
| | | | | | | Acupuncture | 34 | | |
| 9 | Guerra de Hoyas 2004[138] | Spain | Musculoskeletal (shoulder) | 59.19 (11.37) | 6/3 | Sham acupuncture | 65 | VAS pain | VAS pain score |
| | | | | | | Acupuncture | 65 | | |

| ID | Study | Location | Pain group (type) | Age (years), mean (SD) | Follow-up/time point used (months) | Treatment | Observations | HRQoL outcome mapped | Pain outcome standardised |
|---|---|---|---|---|---|---|---|---|---|
| 10 | Haake 2007[71] | Germany | Musculoskeletal (lower back) | 50.15 (14.68) | 6/3 | Usual care | 388 | SF-12 | von Korff pain intensity score |
|  |  |  |  |  |  | Sham acupuncture | 387 |  |  |
|  |  |  |  |  |  | Acupuncture | 387 |  |  |
| 11 | Irnich 2001[118] | Germany | Musculoskeletal (neck) | N/A | 3/3 | Sham acupuncture | 61 | VAS pain | VAS pain score |
|  |  |  |  |  |  | Acupuncture | 56 |  |  |
| 12 | Kennedy 2008[130] | Northern Ireland | Musculoskeletal (lower back) | 45.58 (11.1) | 3/3 | Sham acupuncture | 24 | VAS pain | Roland Morris disability score |
|  |  |  |  |  |  | Acupuncture | 24 |  |  |
| 13 | Kerr 2003[117] | Northern Ireland | Musculoskeletal (lower back) | N/A | 6/1 | Sham acupuncture | 20 | VAS pain | VAS pain score |
|  |  |  |  |  |  | Acupuncture | 26 |  |  |
| 14 | Kleinhenz 1999[137] | Germany | Musculoskeletal (shoulder) | N/A | 3/1 | Sham acupuncture | 27 | CMS and predicted VAS pain | CMS |
|  |  |  |  |  |  | Acupuncture | 25 |  |  |
| 15 | Salter 2006[133] | UK | Musculoskeletal (neck) | 47.71 (16.51) | 3/3 | Usual care | 14 | No mapping – EQ-5D available | Northwick Park pain score |
|  |  |  |  |  |  | Acupuncture | 10 |  |  |
| 16 | Thomas 2006[61] | UK | Musculoskeletal (neck) | 42.62 (10.71) | 24/3 | Usual care | 80 | No mapping – EQ-5D available | SF-36 bodily pain score |
|  |  |  |  |  |  | Acupuncture | 159 |  |  |
| 17 | Vas 2006[121] | Spain | Musculoskeletal (neck) | 46.73 (13.2) | 6/1 | Sham acupuncture | 62 | SF-36 | VAS pain score |
|  |  |  |  |  |  | Acupuncture | 61 |  |  |
| 18 | Vas 2008[122] | Spain | Musculoskeletal (shoulder) | 55.68 (11.37) | 12/3 | Sham acupuncture | 220 | VAS pain | CMS |
|  |  |  |  |  |  | Acupuncture | 205 |  |  |
| 19 | White 2004[132] | UK | Musculoskeletal (neck) | 53.36 (15.61) | 12/3 | Sham acupuncture | 65 | VAS pain | VAS pain score |
|  |  |  |  |  |  | Acupuncture | 70 |  |  |
| 20 | Witt 2006[75] | Germany | Musculoskeletal (neck) | 50.57 (12.93) | 6/3 | Usual care | 1698 | SF-36 | Neck pain and disability score |
|  |  |  |  |  |  | Acupuncture | 1753 |  |  |

**TABLE 21** Data set main characteristics and study outcomes used for analysis *(continued)*

| ID | Study | Location | Pain group (type) | Age (years), mean (SD) | Follow-up/time point used (months) | Treatment | Observations | HRQoL outcome mapped | Pain outcome standardised |
|---|---|---|---|---|---|---|---|---|---|
| 21 | Witt 2006[74] | Germany | Musculoskeletal (lower back) | 52.83 (13.33) | 6/3 | Usual care | 1390 | SF-36 | Hanover functional ability score |
| | | | | | | Acupuncture | 1451 | | |
| 22 | Foster 2007[135] | UK | Osteoarthritis of the knee | 63.23 (8.81) | 12/1 | Usual care | 116 | VAS pain | WOMAC pain score |
| | | | | | | Sham acupuncture | 119 | | |
| | | | | | | Acupuncture | 117 | | |
| 23 | Berman 2004[134] | USA | Osteoarthritis of the knee | 65.46 (8.62) | 6/2 | Usual care | 189 | No mapping – EQ-5D available | WOMAC pain score |
| | | | | | | Sham acupuncture | 191 | | |
| | | | | | | Acupuncture | 190 | | |
| 24 | Scharf 2006[70] | Germany | Osteoarthritis of the knee | 62.81 (10.07) | 6/3 | Usual care | 316 | SF-12 | WOMAC total score |
| | | | | | | Sham acupuncture | 365 | | |
| | | | | | | Acupuncture | 326 | | |
| 25 | Vas 2004[120] | Spain | Osteoarthritis of the knee | 67.04 (10.09) | 3/3 | Sham acupuncture | 49 | WOMAC total | WOMAC total score |
| | | | | | | Acupuncture | 48 | | |
| 26 | Williamson 2007[136] | UK | Osteoarthritis of the knee | 70.67 (8.94) | 3/3 | Usual care | 61 | WOMAC total | Oxford Knee Score |
| | | | | | | Acupuncture | 60 | | |
| 27 | Witt 2005[69] | Germany | Osteoarthritis of the knee | 64.01 (6.49) | 12/2 | Usual care | 70 | SF-36 | WOMAC total score |
| | | | | | | Sham acupuncture | 75 | | |
| | | | | | | Acupuncture | 149 | | |
| 28 | Witt 2006[77] | Germany | Osteoarthritis of the knee | 61.2 (10.39) | 6/3 | Usual care | 310 | SF-36 | WOMAC total score |
| | | | | | | Acupuncture | 322 | | |

CMS, Constant–Murley Score; N/A, not applicable; SF-12, Short Form questionnaire-12 items.
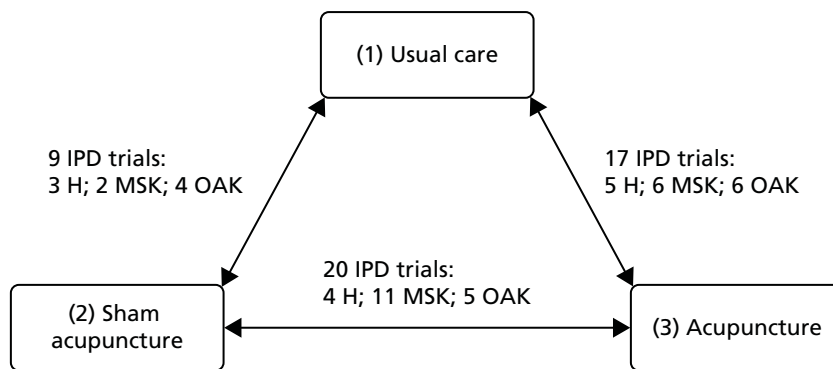
**FIGURE 11** Network of RCTs. In the network, a unique treatment category is indicated by a box. Arrows between boxes indicate that these treatments had been compared in a trial. Pain groups: H, headache/migraine; MSK, musculoskeletal; OAK, osteoarthritis of the knee (as set out in *Table 21*).

End points were measured at a variety of follow-up times across trials. To consistently assess the effect of the treatments of interest across trials, the analysis focused on the time point closest to 3 months from the start of treatment. The 3-month time point was used as this was typically the measurement taken after the end of an acupuncture treatment course in the trials forming the evidence base and was reported for the majority (21/28) of the trials (see *Appendix 1*).

### Generating homogeneous health-related quality-of-life scores and pain outcomes

The EQ-5D index score was the preferred end point for our analysis because of its importance for cost-effectiveness analysis. The conventional three-level version of the EQ-5D questionnaire includes five domains (mobility, self-care, usual activities, pain/discomfort and anxiety/depression), each of which can be at one of three severity levels (no problems, some or moderate problems, or extreme problems), to generate a health status descriptor of one of 243 ($3^5$) health states (245 states in total when also considering the 'unconscious' and 'dead' states). The descriptor is quality adjusted using a score derived from analysing the preferences of approximately 3400 members of the UK public.[232] Bounded by full health and by the worst imaginable health state, the score ranges from 1 to –0.594. The distribution of EQ-5D health-state utility data is commonly non-normal. This, among other features, makes statistical modelling of the EQ-5D particularly challenging.[233] Only a small number of trials in the data set (see *Table 21*) provided EQ-5D data.

When EQ-5D data were not available they were predicted using other generic and disease-specific measures (see *Table 21*) through published mapping algorithms. Mapping algorithms were identified using the University of Oxford's Health Economics Research Centre (HERC) database of studies mapping from HRQoL or clinical measures to the EQ-5D.[234,235] When using this tool and when multiple mapping algorithms were available for a given instrument, the preferred algorithm was selected on the basis of the sample size, adequacy of statistical modelling and relevance of study population. The selection of the outcome to be mapped was not at random. Preference was given to generic health status-based instruments (i.e. SF-12 and SF-36) and, in their absence, to condition-specific instruments [i.e. WOMAC, VAS pain and Constant–Murley Score (CMS)], conditional on the existence of a valid and published algorithm. The WOMAC was used in preference to VAS pain and CMS as it covers a broader definition of HRQoL. In 50% of the trials (*n* = 14) (see *Appendix 4* for trial details), well-established published algorithms were used to map from SF-36 dimensions and SF-12 summary scores to the EQ-5D[61,236,237] [a random-effects generalised least-squares algorithm considering dimensions, dimensions squared and interactions from Rowen *et al.*[236] was used (model $R^2$ = 0.71); a multinomial logit using physical component summary (PCS) and mental component summary (MCS) scores, summary scores squared and interaction terms (mean square error = 0.021) from Gray *et al.*[237] was used to map the SF-12 to the EQ-5D. In 10 of the 28 trials, published algorithms that map VAS pain scores [an ordinary least squares (OLS) regression including VAS pain and VAS pain squared as covariates from Maund *et al.*[238] was used ($R^2$ = 0.101)] and WOMAC scores [an OLS regression including total WOMAC score, total WOMAC score squared, age and sex as covariates from Barton *et al.*[239] ($R^2$ = 0.313) was used] to the EQ-5D were used. For one trial,[137] a double mapping was

necessary as, to our knowledge, no direct mapping algorithm exists to obtain EQ-5D values from the CMS. Thus, an in-house unpublished mapping algorithm[240] was used to derive VAS pain estimates from the CMS, which were used to obtain individual-level EQ-5D predictions using the algorithm mentioned above (available on request from Kamran Khan – K.A.Khan@warwick.ac.uk). For further details on mapping to the EQ-5D, see *Chapter 5* (*Health-related quality of life for cost-effectiveness analysis*).

A high level of unexplained variation was found in the majority of the mapping algorithms used, that is, the proportion of total variation of the outcome(s) explained in these models (quantified by the coefficient of determination, $R^2$, in most cases) was low. To account for this source of uncertainty in the mapping process, an additional variance component was included in the EQ-5D predictions.[241] A mapping process involves additional sources of uncertainty – the uncertainty in the mapping function regression coefficients and the structure of the mapping model. These additional sources of uncertainty are not accounted for in this analysis. This was achieved by drawing from a normal distribution with a mean of zero and variance equal to the study-specific residual variance. The residual variance was calculated as the difference between the total variance (calculated by dividing the variance of the mapped data by the $R^2$ for the mapping algorithm) and the mapped outcome variance. Each random draw was then added to each individual-level EQ-5D prediction.

The second outcome measure assessed was standardised pain. Across the 28 trials, the primary outcome of each study was used to generate patient-level standardised pain estimates. Pain measures varied from days with headache in the headache/migraine pain condition to VAS pain in the musculoskeletal group or WOMAC pain in the osteoarthritis of the knee group, as reported in *Table 21*. Individual-level standardised pain estimates were obtained for each trial by dividing the primary outcome scores by the study-specific SD. Note that, although these estimates were used as inputs in the synthesis models, the outputs of the synthesis are in the SMD format, as differences between treatments were estimated within the modelling [considering $s_{tx}^t$ as the standardised value of the pain measurement $p$ made at the time point $t$ in patients under treatment $tx$, it can be demonstrated that $(s_{tx1}^{t1} - s_{tx1}^{t0}) - (s_{tx0}^{t1} - s_{tx0}^{t0}) = (s_{tx1}^{t1} - s_{tx0}^{t1}) - (s_{tx1}^{t0} - s_{tx0}^{t0}) = \Delta SMD$].

Health-related quality of life and standardised pain estimates were obtained at baseline and at the follow-up point closest to 3 months following the start of treatment. Changes from baseline were obtained by calculating the difference between values for these two time points.

## *Statistical methods*

This section describes the IPD network meta-analysis models. All analyses were conducted from a Bayesian perspective. Bayesian methods can be considered an alternative to the classical (frequentist) approach to statistical modelling and have been frequently used in the data synthesis and the economic evaluation of heath-care technologies.[202,207] They provide a more appealing, intuitive and flexible modelling framework as both the data and the model parameters are considered as random quantities. The key feature in this framework is the likelihood function, which defines how reasonable the data are given values of those model parameters. A key feature of this approach is that it allows the model to incorporate external information alongside available data in the format of prior distributions. When very little or no information is accessible, or when wanting the data to dominate, the posterior, subjective or 'vague' beliefs are set as priors.[242] This framework also allows the uncertainty in the relative effect estimates to be translated into probabilities of decision uncertainty, that is, the probability of which treatment is best (most efficacious) out of all treatments being compared. This explicit consideration of decision uncertainty leads naturally into a decision theory framework, which usually also considers costs and utilities, typically used in health-care decision-making.[202]

A one-step IPD network meta-analysis modelling approach was preferred as, together with relative treatment effect estimates, the estimation of treatment–covariate interactions for patient-level covariates were of interest.[210,212] In the following model descriptions a random-effects approach was taken because of the expected between-study heterogeneity. Nonetheless, a fixed-effect framework could be attained with straightforward simplifications.[86,243] The models described apply both to the EQ-5D and to the standardised pain outcome.

The main modelling approach considered (model 1) is a variation of the ANCOVA approach, modelling the change score but also adjusting for baseline outcome values.[219,220,243,244] Model 1 was used as changes from baseline more closely approximated a normal distribution than absolute outcomes at 3 months. The model included interaction effects for pain type as it was expected that the impact of acupuncture (and sham acupuncture) may differ across pain types. Interaction effects for each pain type were modelled as exchangeable and related[245] as it was expected that the impact of each pain type on the treatment effect of acupuncture may be related to the impact of each pain type on sham acupuncture effects.

### Individual patient data network meta-analysis considering pain type as a treatment effect modifier

The model considers a set of $J$ studies for which IPD were available. These studies included patients with a specific pain condition, with the pain conditions being headache/migraine, musculoskeletal pain and osteoarthritis of the knee. The set of treatments included in these trials are labelled [A,B,C], where A is the reference treatment, and there are $K (= 3)$ treatments in total. At baseline, patient $i$ in study $j$ allocated to treatment $k$ provides a baseline measurement $Y_{ijk0}$, where 0 indicates time $t$ at baseline). Each patient provides a follow-up measurement (the assessment closest to 3 months) $Y_{ijk3}$. The change from baseline $(Y_{ijk3} - Y_{ijk0})$ is denoted $\Delta Y_{ijk}$.

### Model 1: analysis of covariance variation – change score modelling, adjusted for baseline

This model can be written as:

$$
\begin{aligned}
&\Delta Y_{ijk} \sim N(\theta_{ijk}, V_j) \\
&\theta_{ijk} = \begin{cases} \mu_{jb} + \beta_{0j}Y_{ijk0} & \text{if } k = 0; b, k \in \{A, B, C\} \\ \mu_{jb} + \beta_{0j}Y_{ijk0} + \delta_{jbk} + \beta_{bkp}X_{jp} & \text{if } k > b \end{cases} \\
&\delta_{jbk} \sim N(d_{bk}, \sigma^2) \sim N(d_{Ak} - d_{Ab}, \sigma^2) \\
&\beta_{bkp} = \beta_{Akp} - \beta_{Abp} \\
&\beta_{Akp} \sim N(B_p, \sigma^2_{Bp}) \\
&d_{AA}, \beta_{AAp} = 0,
\end{aligned}
\tag{1}
$$

where $V_j$ represents the study-level variance, the quantity $\mu_{jb}$ represents the outcome for the treatment $b$ in study $j$ for a patient with a baseline utility of 0, the parameter $\beta_{0j}$ represents the impact of the (outcome) baseline on the change outcome for each study $j$, the term $\delta_{jbk}$ represents the study-specific treatment effect for treatment $k$ relative to treatment $b$ and $X_{jp}$ are $p - 1$ dummy variables representing pain type $p$ in the $j$th study. Pain × treatment interaction effects $\beta_{Akp}$ were considered different for each treatment but exchangeable and were assumed to be drawn from a random distribution with a common mean ($B_p$) and between-treatment variance ($\sigma^2_{Bp}$).

Independent prior distributions were defined as follows: $1/V_j \sim Gamma(0.001, 0.001)$; $\mu_{jb} \sim N(0, 10^6)$; $\beta_{0j} \sim N(0, 10^6)$; $d_{Ak} \sim N(0, 10^6)$; $\sigma \sim Unif(0, 2)$; $B_p \sim N(0, 10^6)$; $\sigma_{Bp} \sim Unif(0, 2)$. Correlations in the random effects from trials with three or more arms were accounted for following published methodology.[86,222] In this report, $k > b$ indicates that $k$ is after $b$ in the alphabet.

### Modelling extensions

Model 1 can be extended to consider covariates. Age and BMI were identified as potential treatment effect modifiers, with the clinical expectation being that older age or higher BMI may make patients more difficult to treat (i.e. reduce the effect of treatment). BMI data were, however, rarely reported and were available in only 10 of the 28 studies. This covariate was not therefore adjusted for in the modelling.

Age was assumed to modify outcomes by the same amount across pain types and to modify treatment effects by the same margin for acupuncture and sham acupuncture (i.e. a single interaction term is assumed to apply to all comparisons with usual care). Squared terms were included for main effects and

treatment interaction effects as a non-linear impact of age on outcomes and treatment effects was expected a priori. Age was centred prior to inclusion in the model.

The following model (model 2) extends model 1 by considering the effects of the covariate $Z$:

$$\Delta Y_{ijk} \sim N(\theta_{ijk}, V_j)$$

$$\theta_{ijk} = \begin{cases} \mu_{jb} + \beta_{0j}Y_{ijk0} + \phi_0 Z_{ijk} + \varphi_0 Z_{ijk}^2 & \text{if } k = 0; b, k \in \{A, B, C\} \\ \mu_{jb} + \beta_{0j}Y_{ijk0} + \phi_0 Z_{ijk} + \varphi_0 Z_{ijk}^2 + \delta_{jbk} + \beta_{bkp}X_{jp} & \text{if } k > b \text{ and } b \neq A \\ \mu_{jb} + \beta_{0j}Y_{ijk0} + \phi_0 Z_{ijk} + \varphi_0 Z_{ijk}^2 + \phi Z_{ijk} + \varphi Z_{ijk}^2 + \delta_{jbk} + \beta_{bkp}X_{jp} & \text{if } k > b \text{ and } b = A \end{cases}$$

$$\delta_{jbk} \sim N(d_{bk}, \sigma^2) \sim N(d_{Ak} - d_{Ab}, \sigma^2)$$

$$\beta_{bkp} = \beta_{Akp} - \beta_{Abp}$$

$$\beta_{Akp} \sim N\left(B_p, \sigma_{Bp}^2\right)$$

$$Z_{ijk} \sim N(m, prec)$$

$$d_{AA}, \beta_{AA} = 0.$$

(2)

Coefficients on the main covariate effect and the effect squared are represented by $\phi_0$ and $\varphi_0$, respectively. Coefficients on the treatment–covariate interaction term and the interaction between treatment and the squared covariable term are represented by $\phi$ and $\varphi$, respectively. No interaction term for comparisons of $k$ and $b$ was included when $b \neq A$ because the common regression coefficient cancels out.

Because of the possibility of missing covariate information for some individuals, $Z_{ijk}$ was represented as a normally distributed random variable with mean $m$ and precision $prec$, common across all IPD studies. This represents a multiple imputation technique and assumes that the covariable data were missing at random. Additional priors were required for this model: $\phi_0, \phi, \varphi_0, \varphi, \sim N(0,10^6)$; $m \sim Unif(-50,50)$, $\sqrt{1/prec} \sim Unif(0, 30)$.

### Analysis in the presence of restricted evidence

Although model 1 is the preferred choice, this model would not be feasible in the absence of information at the individual level at the baseline and follow-up time points. Models that do not rely on the availability of IPD were therefore run for comparison purposes. Two options[219] are typically available to the analyst when only aggregate data are available: modelling the change score (model 3) or modelling the final outcome score (model 4), both without baseline adjustment. These models represent simplifications of model 1 in which the baseline outcome variable is omitted.

### Model selection and implementation

Data management was performed in the freely available software package, R version 3.0.0 (The R Foundation for Statistical Computing, Vienna, Austria). The network meta-analysis was undertaken in WinBUGS version 1.4.3,[246] linked to the R software through the packages R2WinBUGS[246] and CodaPkg.[247] Code for the network meta-analysis is provided in *Appendix 4*.

In all models the MCMC Gibbs sampler was initially run for 10,000 iterations and these were discarded as 'burn-in'. Models were run for a further 5000 iterations, on which inferences were based. Chain convergence was checked using autocorrelation and Gelman and Rubin[248] diagnostics. Within the network meta-analysis, goodness of fit was assessed using the deviance information criterion (DIC) and residual deviance.[180] The DIC is a measure that balances fit and complexity, allowing parsimony to be considered in model choice. The DIC is often used for model comparison when smaller DIC models should be preferred. The residual deviance of each data point may be viewed as a measure of the data point's contribution to the total residual deviance (or lack of fit) of the model. A posterior mean for the total residual deviance similar to the number of data points will imply that model predictions fit well to the observed data.

Results are presented as EQ-5D index scores and SMD treatment effect estimates (and associated 95% CrIs), and also the probability of treatment being the 'best' treatment in terms of being the most clinically effective.[207]

### Modelling resource use

Acupuncture was assumed to be administered during 10 sessions with a physiotherapist. Ten sessions of acupuncture have been recommended by NICE in the context of lower back pain[81] and headache/migraine,[80] and it was assumed that this duration of therapy could be generalised to other musculoskeletal conditions and osteoarthritis of the knee. The first session was assumed to last for 40 minutes and subsequent sessions for 30 minutes. All sessions were costed using a unit cost for a physiotherapist (£36 per hour; Schema 9.1, with qualifications[249]).

The NICE recommendations, alongside the above assumptions regarding appointment durations, equate to a total of 5.2 hours of therapist time. A sensitivity analysis using a weighted average of the therapist time observed in the trials was conducted. Data were obtained from the data extractions conducted by Vickers et al.[105] Therapist time was calculated as the duration of sessions multiplied by the number of sessions and included only sessions that occurred within the 3-month time horizon considered for efficacy. The sensitivity analysis used total therapist interaction times of 5.6 hours for headache/migraine, 3.9 hours for musculoskeletal and 4.7 hours for osteoarthritis of the knee chronic pain.

The potential impact of improved health outcomes on resource use was explored using the three data sets described in *Data description and network of evidence*.[60,230,231] EQ-5D predictions (mapped from the available SF-36 physical and mental summary scores) together with the number of primary care (i.e. GP) and secondary care (i.e. specialist) visits from Vickers et al.[60] were used to estimate the relationship between change in HRQoL and change in health resource utilisation for the headache pain group. The relationship estimated from Vickers et al.[60] was assumed to apply for the entire headache group of patients (which includes patients with TTH and migraine pain). A simple OLS analysis was used to regress the change in resource use from 0–3 months to 3–12 months on the change in EQ-5D scores between month 3 and month 12. Primary and secondary care visits were analysed separately. Although not aimed at evaluating acupuncture, the UK BEAM study[230] (with approximately 1300 patients) and the TOIB study[231] (with approximately 280 patients) were used to estimate this relationship for lower back pain and osteoarthritis of the knee patients, respectively, using the same approach. Data from the UK BEAM study were assumed to be applicable to the other patients within the musculoskeletal pain category (i.e. those with neck and shoulder pain).

Resource utilisation at baseline was not collected in these studies (and is generally not collected in clinical trials). Changes in resource use were preferred to absolute resource use estimates as their relationship with EQ-5D changes is less likely to be confounded. To estimate the change in resource use it was therefore necessary to use the change from 0–3 months to 3–12 months. Use of the change from 0–3 months to 3–12 months to infer change in resource use over the 0- to 3-month time horizon of the economic model, however, assumes that a given utility change would drive a given change in resource use regardless of the time frame. Given that this is a strong assumption a secondary analysis was conducted using the absolute resource use in the period 0–3 months and regressing this on the change in EQ-5D score over this period.

The statistical software Stata 13 was used to model resource use for each pain condition.

The average cost of non-intervention resources used for each pain condition was calculated as the product of the EQ-5D estimates derived from the synthesis models, the coefficients on the EQ-5D estimates from the resource use regressions and the relevant unit costs [primary care visits were costed at £46.8 – this represents a weighted average of GP (£45 per consultation) and nurse (£49 per hour) visits with weights taken from the UK BEAM study and unit costs taken from the Curtis[249] – and secondary care visits at £135 – this is the weighted average NHS reference cost for all outpatient procedures taken from Curtis[249]). Costs are reported in UK pounds for the financial year 2012–13. Other treatments and health-care interactions that may form a package of 'usual care' were assumed to have been provided equally to all patients regardless of comparator. These costs were therefore omitted from the analysis.

### Estimation of cost-effectiveness outcomes

Quality-adjusted life-years were estimated assuming that the benefit of acupuncture over usual care estimated from the network meta-analysis of EQ-5D index scores was achieved instantaneously, with benefit maintained from 0 to 3 months, and was then lost instantaneously, illustrated in *Figure 12* by the accrued benefit 2. This is equivalent to assuming that the full benefit was gradually achieved over a specified period and then lost linearly over the same period, which may be viewed as a more realistic scenario (see *Figure 12*, accrued benefit 1). For example, the benefit could be linearly achieved from the start of treatment until 12 weeks and gradually lost over the 12 weeks following treatment completion. Costs and effects beyond the 3-month time horizon were not considered in the current model and given the short time horizon no discounting was applied.

Incremental QALY estimates were compared with incremental cost estimates (intervention costs and non-intervention costs) to calculate incremental cost-effectiveness ratios (ICERs). These can be compared with a threshold value of £20,000–30,000 per QALY as conventionally applied in England and Wales.[203]

Uncertainty in the estimates was quantified through the use of probabilistic analysis. The 5000 posterior samples from the synthesis of effectiveness (extracted from the Convergence Diagnostic and Output Analysis WinBUGS output) were used together with 5000 samples of the cost parameters, generated through Monte Carlo simulation. Uncertainty surrounding the decision to accept/reject acupuncture on the basis of cost-effectiveness was illustrated through cost-effectiveness acceptability curves. The cost-effectiveness modelling was also implemented in R.

## Application

### Results of generating homogeneous health-related quality-of-life scores and pain outcomes

*Appendix 4* presents the (mapped) EQ-5D data and standardised pain outcomes. In general, patients' HRQoL increased from baseline to 3 months. Similarly, standardised pain estimates decreased from baseline to 3 months. For both time points, it appears that osteoarthritis of the knee patients had, on average, lower HRQoL (and higher mean values of standardised pain) than patients suffering from headache/migraine or musculoskeletal pain.

For both end points, baseline imbalances between trial arms were observed within trials. For the EQ-5D end point, the biggest within-trial differences at baseline were found in the studies by Carlsson *et al.*[119] and Salter *et al.*[133] For the SMD end point, the largest differences were found in the same two trials and
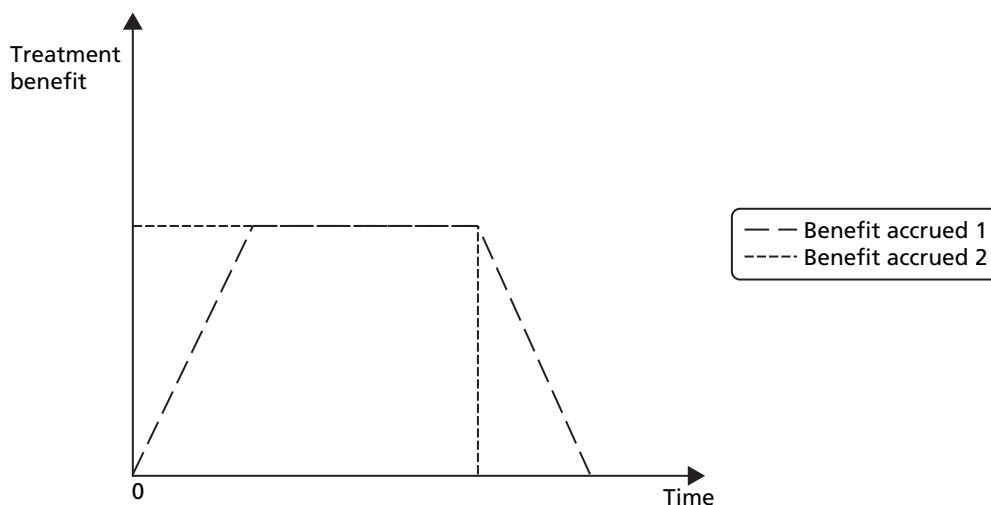


**FIGURE 12** Illustrative diagram of the treatment benefits over the period of assessment.

also in the studies by Kleinhenz *et al.*[137] and White *et al.*[132] These large imbalances are not surprising as most of these trials included only a small number of patients (around 50 or fewer). These observations supported the use of a modelling framework that allows for baseline adjustment,[219] involving the use of either model 1 or 2 as the appropriate tool to synthesise this evidence.

## *Analysis of covariance (model 1)*

*Table 22* shows the parameter estimates obtained from model 1 applied to the EQ-5D and the standardised pain outcome data. For each parameter estimate the median of the MCMC posterior sample and 95% CrI are shown. Relative treatment effect estimates are shown, adjusted for baseline and treatment–pain interaction effects, together with measures of model fit (total residual deviance and DIC). The osteoarthritis of the knee pain group is the reference category for the pain interaction effects.

For both end points, model 1 indicates that acupuncture treatment increases the HRQoL of patients and/or reduces pain more than usual care and sham acupuncture treatments, irrespective of the pain group they belong to. For the EQ-5D end point, the median treatment effect of acupuncture compared with usual care in the osteoarthritis of the knee population is 0.079 (95% CrI 0.042 to 0.114); for headache/migraine and musculoskeletal pain patients the comparable median treatment effects are 0.056 (95% CrI 0.021 to 0.092) and 0.082 (95% CrI 0.047 to 0.116), respectively. The results also favour acupuncture over sham acupuncture, although with a greater degree of uncertainty, as reflected by the fact that the CrIs include zero for all pain types (osteoarthritis of the knee 0.022, 95% CrI –0.014 to 0.060; headache/migraine 0.004, 95% CrI –0.035 to 0.042; musculoskeletal pain 0.023, 95% CrI –0.007 to 0.053). The probability that acupuncture is the best treatment at improving HRQoL is 0.89 for osteoarthritis of the knee, 0.64 for headache/migraine and 0.95 for musculoskeletal pain.

**TABLE 22** Parameter estimates from fitting the novel network meta-analysis ANCOVA synthesis model (model 1) to the EQ-5D preference score and standardised pain end points

| IPD NMA results | Comparison | | Model 1 (ANCOVA extension, change in outcome score, adjusted for baseline), median MCMC posterior sample (95% CrI) | |
| --- | --- | --- | --- | --- |
| | | | Change EQ-5D | Change standardised pain |
| Relative treatment effects for OAK | SHAM vs. UC | | 0.057 (0.013 to 0.095) | 0.271 (–0.007 to 0.537) |
| | ACU vs. UC | | 0.079 (0.042 to 0.114) | 0.703 (0.399 to 0.984) |
| | ACU vs. SHAM | | 0.022 (–0.014 to 0.060) | 0.438 (0.121 to 0.715) |
| Pain exchangeable interactions (vs. OAK[a]) | Headache[b,c] | SHAM (vs. UC) | –0.005 (–0.060 to 0.054) | 0.057 (–0.351 to 0.485) |
| | | ACU (vs. UC) | –0.023 (–0.071 to 0.029) | –0.121 (–0.467 to 0.254) |
| | Musculoskeletal[b,c] | SHAM (vs. UC) | 0.002 (–0.052 to 0.062) | –0.199 (–0.595 to 0.218) |
| | | ACU (vs. UC) | 0.003 (–0.046 to 0.054) | –0.108 (–0.465 to 0.307) |
| Between-study variance | | | 0.001 (0 to 0.003) | 0.090 (0.049 to 0.170) |
| Total residual deviance[d] | | | 15,850 (15,480 to 16,230) | 17,060 (16,660 to 17,450) |
| DIC[e] | | | –6420.4 | 37,394.2 |

ACU, acupuncture; NMA, network meta-analysis; OAK, osteoarthritis of the knee; SHAM, sham acupuncture; UC, usual care.
a The reference pain group for the interaction terms is the OAK group.
b The reference treatment is usual care.
c Headache group = headache, migraine and TTH; musculoskeletal group = neck, shoulder and lower back pain.
d For the EQ-5D end point, models used approximately 14,800 observations; for the SMD end point, models used approximately 15,900 observations. Models should be preferred when the total residual deviance mean posterior is close to the actual number of data points.
e The DIC is a statistical measure of model fit and model comparison. Models with a smaller DIC should be preferred.

For the SMD end point the median treatment effect of acupuncture compared with usual care in the osteoarthritis of the knee population is 0.703 (95% CrI 0.399 to 0.984); for headache/migraine and musculoskeletal pain patients the comparable median treatment effects are 0.588 (95% CrI 0.311 to 0.869) and 0.588 (95% CrI 0.334 to 0.863), respectively. The results also favour acupuncture over sham acupuncture. In contrast to the EQ-5D analysis, the CrIs do not include zero in the standardised pain analysis for osteoarthritis of the knee (0.438, 95% CrI 0.121 to 0.715) and musculoskeletal pain (0.527, 95% CrI 0.323 to 0.735), although the CrI for headache/migraine does (0.256, 95% CrI –0.073 to 0.560). The probability that acupuncture is the best treatment at improving standardised pain is 0.96–1.00, depending on pain type. These results are presented as a forest plot in *Figure 13*.
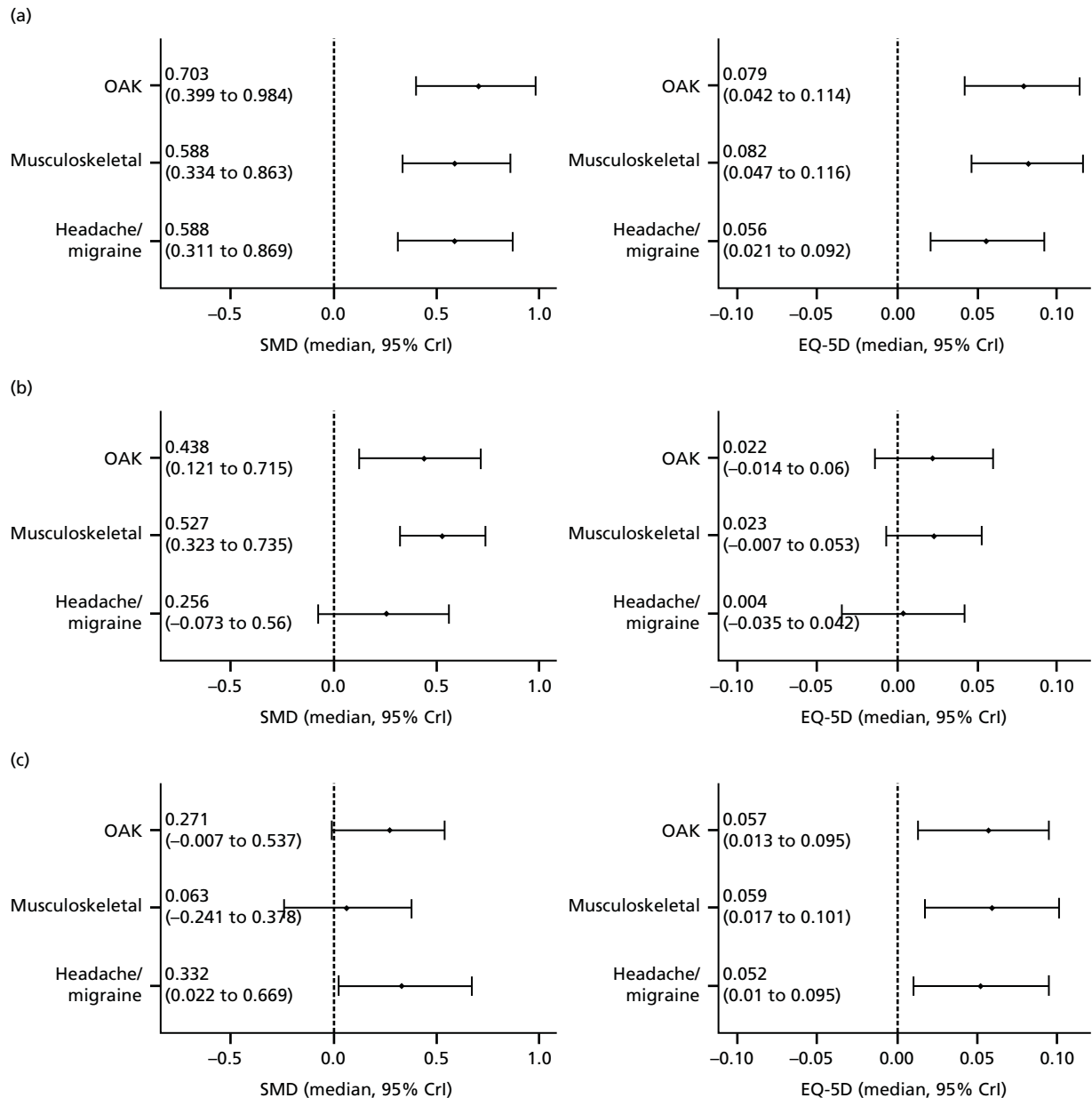


**FIGURE 13** Forest plot showing the network meta-analysis results for the standardised pain and EQ-5D outcomes: (a) acupuncture vs. usual care; (b) acupuncture vs. sham acupuncture; and (c) sham acupuncture vs. usual care. OAK, osteoarthritis of the knee.

The expectations were that some level of heterogeneity existed between trials. Possibly as a consequence of the mapping work performed, this expectation was not fulfilled for the EQ-5D end point (the between-study variance estimate is 0.001). For the standardised pain end point, the between-study variance was also small relative to the magnitude of the treatment effects (the between-study variance estimate is 0.09). The total residual deviance suggests that the models provide an adequate fit to the data (see *Table 22*).

### Controlling for patient-level characteristics

*Table 21* provides information on age of participants for each of the trials included in the data set. On average, age was lower in the headache/migraine pain group than in the musculoskeletal and osteoarthritis of the knee groups.

Using the change in EQ-5D as the outcome for synthesis, *Table 23* presents the results of applying model 2 (an extension of model 1) to include patient-level information on age, age being a potential treatment effect modifier. The model fit statistics show that this adjusted by age model is marginally better than model 1, providing lower DIC statistics and reduced posterior residual deviance. The results of this model are very similar to those of model 1 and do not suggest that age is a strong effect modifier or that non-linear effects of age on the effect of treatments exist.

**TABLE 23** Parameter estimates from fitting the synthesis model including age as a covariate

| IPD NMA results | Comparison | | Model 2 (with adjustment for baseline age and treatment by age interactions), median MCMC posterior sample (95% CrI) |
|---|---|---|---|
| Relative treatment effects | SHAM vs. UC | | 0.040 (–0.006 to 0.084) |
| | ACU vs. UC | | 0.066 (0.025 to 0.105) |
| | ACU vs. SHAM | | 0.026 (–0.012 to 0.066) |
| Main effects | Age | | –0.002 (–0.002 to –0.001) |
| | Age$^2$ | | 0.000 (0.000 to 0.000) |
| Pain exchangeable interactions (vs. OAK[a]) | Headache[b,c] | SHAM (vs. UC) | 0.016 (–0.044 to 0.079) |
| | | ACU (vs. UC) | –0.006 (–0.059 to 0.047) |
| | Musculoskeletal[b,c] | SHAM (vs. UC) | 0.006 (–0.057 to 0.070) |
| | | ACU (vs. UC) | 0.008 (–0.045 to 0.062) |
| Age common interactions | Age | | 0.000 (0.000 to 0.001) |
| | Age$^2$ | | 0.000 (0.000 to 0.000) |
| Between-study variance | | | 0.001 (0.000 to 0.003) |
| Total residual deviance[d] | | | 15,590 (15,210 to 15,970) |
| DIC[e] | | | –6462.0 |

ACU, acupuncture; NMA, network meta-analysis; OAK, osteoarthritis of the knee; SHAM, sham acupuncture; UC, usual care.
a  The reference pain group for the interaction terms is the OAK group.
b  The reference treatment is usual care.
c  Headache group = headache, migraine and TTH; musculoskeletal group = neck, shoulder and lower back pain.
d  Compared with approximately 14,800 observations.
e  The DIC is a statistical measure of model fit and model comparison. Models with a smaller DIC should be preferred.

### *Analysis with restricted evidence (models 3 and 4)*

Results for models 3 and 4 are presented in *Table 24*, together with the model 1 results for comparison. Generally, all three models convey the same message in relation to which treatment provides higher increases in patients' HRQoL, that is, acupuncture is found to be better than sham acupuncture and usual care treatments. Nevertheless, given the presence of baseline imbalance, models 3 and 4 (but model 3 in particular) provide very different and potentially inappropriate summary results of treatment effects when compared with model 1. These two models show also a fit to the data that is worse than model 1 (higher DIC of –6420 in model 1 compared with –69 and –3824 in models 3 and 4, respectively). In the absence of baseline outcome data, if the choice was between modelling change (model 3) or modelling follow-up scores (model 4), results from the latter model indicate that this would be a better option as the relative treatment effect estimates and pain interaction effects are closer to those in model 1.

**TABLE 24** Parameter estimates from fitting three different network meta-analysis models to the EQ-5D end point

| IPD NMA results | Comparison | | Model | | |
|---|---|---|---|---|---|
| | | | 1 (ANCOVA extension, change in EQ-5D scores, adjusted for baseline), median MCMC posterior sample (95% CrI) | 3 (change in EQ-5D scores without baseline adjustment), median MCMC posterior sample (95% CrI) | 4 (follow-up EQ-5D scores without baseline adjustment), median MCMC posterior sample (95% CrI) |
| Relative treatment effects | SHAM vs. UC | | 0.057 (0.013 to 0.095) | 0.077 (0.033 to 0.118) | 0.051 (0.008 to 0.094) |
| | ACU vs. UC | | 0.079 (0.042 to 0.114) | 0.093 (0.054 to 0.129) | 0.074 (0.035 to 0.113) |
| | ACU vs. SHAM | | 0.022 (–0.014 to 0.060) | 0.016 (–0.022 to 0.054) | 0.023 (–0.014 to 0.065) |
| Pain exchangeable interactions (vs. OAK[a]) | Headache[b,c] | SHAM (vs. UC) | –0.005 (–0.060 to 0.054) | –0.032 (–0.089 to 0.029) | 0.001 (–0.059 to 0.064) |
| | | ACU (vs. UC) | –0.023 (–0.071 to 0.029) | –0.035 (–0.082 to 0.014) | –0.021 (–0.074 to 0.032) |
| | Musculoskeletal[b,c] | SHAM (vs. UC) | 0.002 (–0.052 to 0.062) | –0.016 (–0.074 to 0.045) | 0.003 (–0.056 to 0.064) |
| | | ACU (vs. UC) | 0.003 (–0.046 to 0.054) | –0.009 (–0.059 to 0.042) | 0.005 (–0.048 to 0.058) |
| Between-study variance | | | 0.001 (0 to 0.003) | 0.001 (0 to 0.003) | 0.001 (0 to 0.003) |
| Total residual deviance[d] | | | 15,850 (15,480 to 16,230) | 16,990 (16,570 to 17,420) | 15,370 (15,010 to 15,730) |
| DIC[e] | | | –6420.4 | –69.9 | –3823.7 |

ACU, acupuncture; NMA, network meta-analysis; OAK, osteoarthritis of the knee; SHAM, sham acupuncture; UC, usual care.
a The reference pain group for the interaction terms is the OAK group.
b The reference treatment is usual care.
c Headache group = headache, migraine and TTH; musculoskeletal group = neck, shoulder and lower back pain.
d Compared with approximately 14,800 observations.
e The DIC is a statistical measure of model fit and model comparison. Models with a smaller DIC should be preferred.

## Results of analysing resource use

*Table 25* shows the results from regressing change in primary and secondary care health resources on change in EQ-5D index score for each study. Generally, an increase in EQ-5D score over time implies a reduction in health-care resource use. The analysis of secondary care resource use for osteoarthritis of the knee was an exception, with improvements in EQ-5D score being associated with increased secondary care attendances; however, this result was not statistically significant.

## Results of the illustrative cost-effectiveness analysis

Illustrative cost-effectiveness results are presented in *Table 26*. The ICERs in each indication are well below the threshold of £20,000–30,000 per QALY generally considered acceptable in the UK. Results using the 0- to 3-month data for the resource use regressions were very similar to the results in the base case and are therefore not shown here. Acupuncture has close to a 100% probability of being cost-effective in patients with osteoarthritis of the knee and musculoskeletal pain types, and an 86% probability of being cost-effective for the headache/migraine indication, assuming a threshold of £20,000 per QALY. The sensitivity analysis using trial data with a weighted average of the therapist time observed in the trials provided fairly similar results, with musculoskeletal pain now obtaining the lowest estimated ICER compared with the other two pain groups, as shown in *Table 26*. The results of the probabilistic sensitivity analysis are presented as cost-effectiveness acceptability curves in *Figure 14*.

# Discussion

## Principal findings

Policy-makers faced with difficult resource allocation decisions require estimates of the costs and effects of alternative treatment options. These estimates should reflect all relevant data and compare treatments using a metric that can be used across clinical areas – in the UK the QALY is typically used. Synthesising all relevant evidence to produce comparable estimates of costs and effects generates a series of challenges as the available evidence base rarely captures all costs and effects of treatment (because of the nature of data collection or the duration of follow-up), and often requires evidence to be generalised from different populations. The available trial evidence may compare different sets of treatments and in many instances the HRQoL data required to estimate QALYs directly are not available.

The National Institute for Health and Care Excellence has recommended acupuncture for the treatment of chronic headache and musculoskeletal pain but not in the context of chronic pain associated with osteoarthritis of the knee and headache/migraine.[80–82] This decision in part reflected concerns regarding the available evidence. The current study was commissioned as part of a programme intended to improve

**TABLE 25** Resource use regression results

| Study | Pain group (type) | Analysis | | | |
| | | Main 3–12 months of EQ-5D on type of resource use, mean (95% CI) | | Secondary 0–3 months of EQ-5D on resource use, mean (95% CI) | |
| | | Primary care[a] | Secondary care[b] | Primary care[a] | Secondary care[b] |
|---|---|---|---|---|---|
| Vickers 2004[60] | Headache (headache) | −1.247 (−2.409 to −0.085) | −0.17 (−0.63 to 0.289) | −0.719 (−1.849 to 0.412) | −0.057 (−0.503 to 0.390) |
| UK BEAM Trial Team[230] | Musculoskeletal (lower back) | −0.296 (−0.988 to 0.393) | −0.193 (−0.58 to 0.193) | −0.414 (−0.925 to 0.098) | −0.19 (−0.863 to 0.483) |
| Underwood 2008 (TOIB)[231] | Osteoarthritis of the knee | −0.885 (−1.93 to 0.16) | 0.47 (−0.638 to 1.579) | −0.294 (−0.966 to 0.378) | 0.069 (−0.421 to 0.559) |

a Primary care resources include GP and practice nurse visits.
b Secondary care resources include NHS hospital outpatient visits.

**TABLE 26** Cost-effectiveness results (probabilistic analysis) for each pain condition

| Analysis | Pain group | Treatment | Incremental QALYs, mean (95% CrI)[a] | Incremental costs, mean (95% CrI) (£)[a] | ICER (£ per QALY) | Probability cost-effective at £20,000 per QALY |
|---|---|---|---|---|---|---|
| Main analysis (NICE guidance treatment regimen) | Osteoarthritis of the knee | Usual care | – | – | – | 0.02 |
| | | Acupuncture | 0.0196 (0.0101 to 0.0287) | 189 (176 to 202) | 9673 | 0.98 |
| | Headache/ migraine | Usual care | – | – | – | 0.14 |
| | | Acupuncture | 0.0140 (0.0053 to 0.0231) | 183 (176 to 188) | 13,076 | 0.86 |
| | Musculoskeletal | Usual care | – | – | – | 0.01 |
| | | Acupuncture | 0.0205 (0.0118 to 0.0291) | 184 (179 to 189) | 8997 | 0.99 |
| Sensitivity analysis (trial-based treatment regimen) | Osteoarthritis of the knee | Usual care | – | – | – | 0.01 |
| | | Acupuncture | 0.0196 (0.0101 to 0.0287) | 169 (156 to 182) | 8651 | 0.99 |
| | Headache/ migraine | Usual care | – | – | – | 0.18 |
| | | Acupuncture | 0.0140 (0.0053 to 0.0231) | 197 (191 to 202) | 14,110 | 0.82 |
| | Musculoskeletal | Usual care | – | – | – | 0.00 |
| | | Acupuncture | 0.0205 (0.0118 to 0.0291) | 138 (132 to 143) | 6745 | 1.00 |

a Data are incremental QALYs (95% CrIs) and incremental costs (95% CrIs).
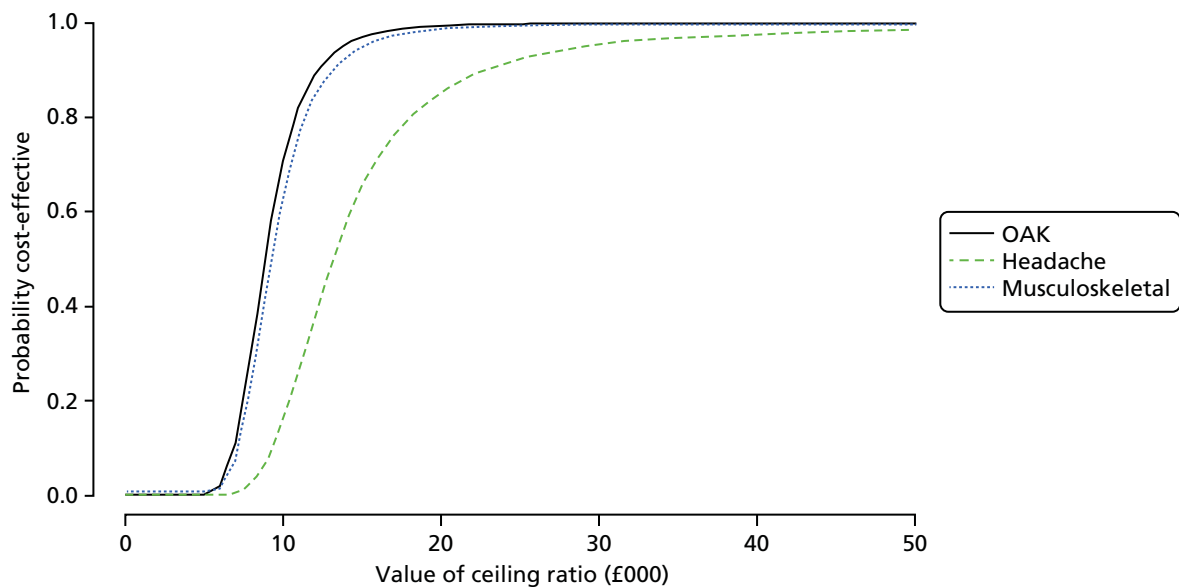


**FIGURE 14** Cost-effectiveness acceptability curves for each pain condition. The curves indicate the probability that acupuncture is cost-effective at different values of the ceiling ratio. Separate curves are shown for osteoarthritis of the knee (OAK), headache/migraine and musculoskeletal chronic pain types.

evidence around the costs and effects of acupuncture. This study synthesised IPD from RCTs of acupuncture in headache/migraine, musculoskeletal and osteoarthritis of the knee chronic pain. Trials compared acupuncture with usual care, sham acupuncture or both control interventions. Bayesian network meta-analysis synthesis modelling was therefore used in this study to leverage all available evidence to inform estimates of relative treatment effects. The studies reported heterogeneous and distinct outcome sets. Methods to homogenise outcomes for synthesis were therefore used. The availability of IPD for all studies expanded the set of feasible analyses and allowed development of de novo methods to fully exploit the benefits of access to these data.

Novel methods for network meta-analysis of IPD on continuous outcomes were developed, building on previous work on ANCOVA models for pairwise meta-analysis.[219] Analysis of the pain outcome required development of methods for conducting SMD analysis with IPD. Analysis of the EQ-5D data required an extensive mapping exercise whereby separate mapping functions were applied to each study, with choice of mapping dependent on the available outcome data. Access to IPD allowed ANCOVA models to be applied, thus improving precision and adjusting for baseline imbalance. Access to IPD also avoided the use of any assumptions regarding the distribution of HRQoL instrument scores, thus allowing the observed distributions to be adequately reflected in the mapped utilities. Finally, access to IPD provided the opportunity to adjust for covariates based on within- and across-trial information. Given the demonstrable benefits of access to IPD, more effort should be made to share and develop repositories for data. A recent survey indicated a high level of support from reviewers affiliated with Cochrane Collaboration's IPD Meta-analysis Methods Group for the development of a central repository for storing IPD.[250]

Analyses were conducted to explore the importance of modelling change scores in the presence of non-normally distributed outcome data and to explore the implications of using non-ANCOVA models, as would be necessary in the absence of IPD. The results showed that modelling final scores or change scores without baseline adjustment produced estimates of treatment effect that differed by up to 26% compared with the baseline adjusted model, emphasising the importance of baseline adjustment and therefore of having access to IPD.

The results of the network meta-analysis show acupuncture to be more effective than usual care with respect to reducing pain and improving HRQoL. There remains uncertainty regarding whether or not the benefit of acupuncture varies across the pain types analysed. The analysis of EQ-5D preference scores suggests that patients with the headache/migraine pain type may benefit from acupuncture, but less so than patients with osteoarthritis of the knee or musculoskeletal pain, although interaction effects are relatively uncertain. A reduced benefit in patients with headache/migraine-related chronic pain could be caused by ceiling effects as individuals with chronic headache/migraine pain had higher baseline EQ-5D index values. Results for the standardised pain analysis were more consistent across indications. Differences between acupuncture and sham acupuncture were relatively small. The large effect of the sham acupuncture intervention compared with usual care may reflect the potency of the sham comparators in the higher-quality trials included in the ATC systematic review. In contrast to the NICE guidelines, our results suggest that if anything the evidence base for acupuncture is stronger in the osteoarthritis of the knee and musculoskeletal conditions (for which acupuncture is recommended only for lower back pain) than in the headache/migraine pain group (for which acupuncture is recommended). The recommendations in the NICE osteoarthritis guidelines were heavily driven by comparisons with sham acupuncture. The network meta-analysis found strong evidence of an effect of acupuncture when compared with sham acupuncture in osteoarthritis of the knee for the standardised pain outcome but not the EQ-5D outcome (for which the CrI contained zero).

Considerable commonalities exist between the methodologies and the results presented in *Chapter 2* and this chapter; however, there are some differences. Across pain types, the two chapters report minor differences in effect between acupuncture and usual care, and acupuncture and sham acupuncture. Nevertheless, the results were broadly consistent across the two chapters. The exact magnitude of the treatment effects and their precision inevitably varied given that there are differences in the data and

methods being used. The current analysis used 28 trials (rather than 29) and consistently used the 3-month end point rather than the primary end point as in *Chapter 2*. For example, in *Chapter 2* the two headache trials used the primary end point, which was at 6 months. Additionally, a different methodology was used. *Chapter 2* used IPD pairwise meta-analysis based on a frequentist approach. In contrast, in this chapter, IPD network meta-analysis was implemented using a Bayesian random-effects framework. The synthesis model implemented in the current analysis considered all evidence and all available treatments of interest in a single analysis, simultaneously deriving relative treatment effects for all comparisons. Finally, *Chapter 2* focused on the standardised pain outcome whereas this chapter analysed standardised pain and HRQoL (EQ-5D) estimates.

The cost-effectiveness results suggest that, compared with usual care alone, acupuncture is cost-effective with ICERs ranging from £9000 to £13,000 per QALY. These values fall within both the NICE plausible threshold range (i.e. between £20,000 and £30,000 per QALY gained) and a more recent empirical threshold estimate of £13,000 per additional QALY obtained.[251] These values are comparable to those in other studies in the UK comparing acupuncture with usual care for the same pain indications, which have estimated ICERs of £4000–17,000.[62,63,227] These ICERs were derived from individual studies, whereas the ICERs presented here reflect the synthesis of a large number of studies.

### *Limitations*
The study has a series of limitations. First, synthesis of heterogeneous outcomes relied on imperfect standardisation processes (which assume that any differences in within-trial outcome variability result from the use of different instruments) and mappings, which are typically able to explain only a minority of variation in EQ-5D scores. Clearly, the use of any mapping tool is considered a second best approach to directly eliciting relevant preference-based measures from study participants. The magnitude of bias introduced by using standardisation processes and mapping functions (and different mapping functions across trials) is unknowable. The availability of key outcomes across trials would have reduced these concerns, as would the collection of generic preference-based measures of HRQoL in all trials. A 'core outcome set' for osteoarthritis is available, along with the recommendation that future Phase III trials of knee, hip and hand osteoarthritis should evaluate the following domains: pain, physical function, patient global assessment and, for studies of ≥ 1 year, joint imaging.[252] Other recommendations have tended to focus on domains rather than specific instruments. Recommendations that go beyond Phase III regulatory trials, and which define the instruments that should be used to measure outcomes in these domains, are warranted.

Second, outcome data closest to 3 months were selected for synthesis. The synthesis therefore requires the assumption that, in the minority of trials not reporting at 3 months, the available data are reflective of the 3-month time point. Some trials reported outcomes at months 1 and 2. If the effect of acupuncture is gradual, these effects may underestimate 3-month outcomes. For the cost-effectiveness analysis, the HRQoL effects observed at 3 months were applied from 0 to 3 months to generate QALYs. Other quality-of-life trajectories may, however, be more plausible. For example, quality of life may increase gradually during treatment and reduce gradually following treatment completion. Moreover, there is some evidence of benefits increasing for some time after the first 3 months when treatment was provided, for example at 12 months for headache/migraine[60] and at 24 months for lower back pain.[61] Depending on the nature and magnitude of these effects, the incremental benefit of sham acupuncture and acupuncture could be larger or smaller than presented here. Further work analysing repeated outcome measurements in a network meta-analysis could be used to evaluate the importance of these effects.

All sham interventions were assumed to be equivalent in the analysis, as were the usual care controls. Evidence from work recently conducted by the ATC suggests that the effect of sham acupuncture may vary depending whether penetrating or non-penetrating needles are used and that the effect of usual care may depend on whether or not a treatment protocol for usual care is specified.[45] Exploration of a network including more refined comparator definitions may, therefore, be of value.

The impact of each pain condition on treatment effects was assumed to be exchangeable;[173] this assumption could be explored further by comparing the fit of models assuming a common pain–treatment effect interaction and models assuming completely separate pain–treatment effect interactions.

The studies analysed here are from a range of countries, which may differ in terms of the method and intensity with which acupuncture is administered. For instance, following NICE recommendations for lower back pain, we assumed that acupuncture treatments are fixed at 10 sessions, irrespective of the pain condition. This assumption might be questionable as the optimum number of treatment sessions may vary according to setting and pain type. Also, acupuncture sessions were costed using a unit cost for a physiotherapist of £36 per hour. This is also an assumption of the current work as unit costs will depend on how the NHS will provide the service. In addition, differences in the nature of health care for chronic pain more generally could have impacted on outcomes.

The analysis of non-intervention resource use assumed that only primary care and specialist visits are impacted on by changes in outcomes following acupuncture, and that the impact of treatment on resource use can be captured through changes in the EQ-5D. It is possible, however, that this did not capture the full impact of treatment on resource use.

Our analysis of standardised pain included the primary end point for each study and, therefore, the outcomes on which we would expect the trials to have been powered. The outcomes included in the analysis ranged from pure pain measures to wider measures of HRQoL (e.g. total WOMAC score). Both pain and functioning outcomes have been highlighted in previous NICE Guidance Development Groups to be of critical importance to decision-making.[80–82] Our analysis suggests that, based on the standardised pain outcome, acupuncture is better than usual care and sham acupuncture for all indications, although CrIs include zero for the headache group when acupuncture is compared with sham acupuncture.

### Recommendations for future research

First, a key limitation of this work is the use of imperfect standardisation processes to adequately combine available heterogeneous evidence. Thus, we consider it a research priority to identify key outcomes for the conditions considered here and improve reporting so that consistency exists across the body of evidence. Second, in the impossibility of achieving complete homogeneity of outcomes across the relevant evidence resulting in mapping tools being required, a worthwhile methodological extension of the current work would be to develop a model that would map the existing evidence to the desired outcome and simultaneously synthesise it together with other relevant evidence. Finally, it was highlighted in this work how important it is to have access to, and analyse, evidence at the individual level. It showed that IPD has clear value over summary data for both synthesis and decision modelling aspects of the analysis. Thus, continuing efforts to share this data type across the research community is highly commended.

Although results from this analysis provide robust estimates of the incremental costs and effects of acupuncture compared with usual care, they are unlikely to provide a suitable basis for decision-making. There is a wide range of alternative treatments for chronic pain and the relative value of these alternatives should be appraised alongside the costs and effects of acupuncture and usual care to reliably inform decision-making. In the context of osteoarthritis of the knee, an evaluation of a broader set of treatment options has been conducted and is presented in the following chapter.

## Conclusions

This study presents methods for conducting IPD network meta-analysis of continuous outcomes when the instruments used to measure outcomes differ between trials. Using the example of acupuncture for the treatment of chronic pain, our novel methods show how heterogeneous outcomes can be analysed using standardisation and mapping approaches, and how the resulting outcomes can be translated into cost-effectiveness results to inform resource allocation decisions.

The methods developed allowed all available trials to inform the synthesis. Availability of IPD allowed the true distribution of outcome measures to be reflected in the mapping to EQ-5D and avoided the use of non-baseline-adjusted models, which produced quite different results. Use of baseline-adjusted change score models produced better results than non-adjusted models, suggesting the superiority of the ANCOVA framework in the context of treatment effect estimation.

The analysis found acupuncture to be more effective than usual care with respect to reducing pain and improving EQ-5D preference scores in patients with chronic pain of osteoarthritis of the knee, musculoskeletal and headache/migraine origin. The benefits of acupuncture over sham acupuncture are smaller than when compared with usual care. The probability that acupuncture is associated with better pain outcomes than sham acupuncture and usual care is high (> 0.96) across indications. The probability that acupuncture is associated with higher EQ-5D preference scores than sham acupuncture and usual care is high in osteoarthritis of the knee (0.89) and musculoskeletal chronic pain (0.95). For headache/migraine this probability is 0.64, reflecting the smaller benefit of acupuncture compared with sham acupuncture for this indication. The methods used provide outputs in a format that can be used to directly inform cost-effectiveness considerations once the full set of relevant comparators is considered.

# Chapter 5 Cost-effectiveness of non-pharmacological adjunct treatments for patients with osteoarthritis of the knee

## Introduction

Health policy-makers worldwide are under increasing pressure to provide the best and most affordable care to their fellow citizens to maximise population health given existing constraints (e.g. budgetary, ethical, structural). Cost-effectiveness analysis is now being used in many jurisdictions to support 'value for money' appraisals as part of the health technology assessment of competing interventions.[253] Examples of national agencies that use cost-effectiveness analysis for health technology assessment to inform their deliberation process include NICE in England and Wales, the Pharmaceutical Benefits Advisory Committee in Australia and the Common Drugs Review in Canada.

Many other similar agencies exist around the world[253] and, although the methods they use may vary slightly across jurisdictions, there are a number of essential information requirements, common to all, that must be met for these decision-makers to be able to formulate their funding recommendations. These requirements include the systematic consideration and quantification of the clinical effectiveness, quality of life and health-care cost implications associated with each competing treatment strategy relevant to the decision problem in the jurisdiction of interest.

Unfortunately, the above information is often either not available at all or not available in the format required by the decision-maker. For instance, the evidence base may (1) lack (or be informed by a limited set of) studies comparing head to head all of the relevant treatment strategies, (2) present a fragmented picture with different studies reporting different sets of outcomes and summary statistics, (3) include clinical studies with too short a follow-up duration to directly inform questions about the long-term (cost) effectiveness of the technologies, (4) reflect large variations in clinical practice (between and within jurisdictions) and (5) provide little or no health-care resource utilisation data relevant to the jurisdiction of interest.[204,254]

In all of these cases, to use the existing evidence base to inform health policy inevitably requires the application of statistical evidence synthesis and decision-analytic (cost-effectiveness) models. These types of models facilitate the organisation and synthesis of the available information within a coherent mathematical framework developed to evaluate the outcomes of interest (e.g. long-term costs and effects), identify and quantify their key drivers, and appropriately reflect all sources of uncertainty surrounding the decision problem.[255]

In many clinical areas and for many interventions, particularly those not subject to strict regulation, the evidence base may include a significant proportion of poor-quality studies, making decision-making challenging. For example, not all interventions that come under the category of complementary and alternative medicine (CAM) therapies[256] are subject to rigorously conducted RCTs. However, there is growing public interest in (and demand for) the use of these therapies, with patients often making direct contact with practitioners, with or without their primary care physician's referral.[257,258] Health policy-makers are therefore responsible for assessing the role and position of CAM therapies within the management of patients with certain conditions and health-care staff need to be familiar with the various treatment options, their possible benefits and risks, and potential interactions with more conventional medical therapies.[259]

The limitations of the evidence base mean that there is considerable uncertainty surrounding the (cost) effectiveness of some of the CAM therapies, so much so that, in a recent review of UK clinical guidelines that discussed CAM, 62% did not reach a conclusive recommendation for or against treatment.[260]

In its guideline on the care and management of osteoarthritis, NICE[228] acknowledges the limitations in the published evidence base and advises people with osteoarthritis of the knee to use a range of *core treatments* (access to appropriate information, exercise and weight loss for people who are overweight or obese) and a range of pharmacological and non-pharmacological (e.g. manipulation and stretching, electrotherapy) *adjunct treatments*. The guideline stated that acupuncture should not be offered for the management of osteoarthritis. This recommendation spurred reactions from patient groups and practitioners,[84] particularly because it is in contrast to the fact that, of the various forms of CAM, acupuncture is one of the most popular referrals (based on a review of three different surveys) and approximately 4 million sessions are provided annually in the UK each year.[1,261,262]

In this chapter we address some of the concerns about the limitations in the published evidence base on CAM, particularly with regard to the role of acupuncture for osteoarthritis of the knee. The study featured in this chapter builds on the IPD meta-analysis of *Chapter 2* in which we found that acupuncture was superior to both usual care and sham acupuncture for the treatment of back and neck pain, osteoarthritis of the knee, and chronic headaches and migraine. In *Chapter 3* we conducted a comprehensive evidence synthesis (i.e. systematic literature review and network meta-analysis) of physical therapies for patients with pain related to osteoarthritis of the knee in which we found acupuncture to be one of the more effective therapies for alleviating knee pain. However, the study did not attempt to quantify the wider quality-of-life benefits of the included interventions or their value for money. In *Chapter 4*, we reanalysed the IPD of *Chapter 2* in which pain outcome measures and measures of HRQoL mapped to EQ-5D preference weights were analysed and a number of methodological challenges were addressed.

This chapter tackles the methods issues identified in *Chapter 4* and builds on the work of *Chapter 3* with its focus on osteoarthritis of the knee. We report on novel methods and results of a network meta-analysis of multiple sources of aggregate data and IPD, and a cost-effectiveness (or 'value for money') assessment of non-pharmacological adjunct interventions in patients with knee pain caused by osteoarthritis in the UK NHS primary care setting. The study compares the costs and effects of 13 therapies that could be (and in some cases already are being) used within the UK NHS as adjunct treatments for knee osteoarthritis. *The evidence base and economic decision problem associated with osteoarthritis of the knee* describes the decision problem and the evidence base. The methods section describes the network meta-analysis models developed for the simultaneous synthesis of continuous aggregate data and IPD, and the cost-effectiveness model methods. The results section presents the results of the application of the methods to the decision problem and data at hand. The discussion section discusses the study findings, its strengths and limitations, and recommendations for future applied and methodological research. This is followed by the final conclusions.

## The evidence base and economic decision problem associated with osteoarthritis of the knee

### The decision problem

Osteoarthritis is most commonly located in the knee and is a major cause of pain, activity limitation and health-care utilisation, especially among older people.[228] There appears to be an unmet need for treatment for this condition.[263,264] Although acupuncture has been advocated as a potentially effective therapy to manage osteoarthritis-related pain,[265] NICE does not recommend its use because of questions over its (cost) effectiveness.[228] Other international guidelines have varied in their recommendations. EULAR[161] and AAOS[266] do not recommend the use of acupuncture, the ACR[198] recommends acupuncture for those with moderate to severe pain who are unwilling, unable or ineligible to undergo total knee arthroplasty and OARSI[267] makes an 'uncertain' recommendation regarding acupuncture. The Scottish Intercollegiate Guidelines

Network[268] recommendations regarding the management of chronic pain recommend acupuncture for short-term pain relief in osteoarthritis and guidance on the management of pain in older people has deemed acupuncture worthy of further investigation.[269] In addition, a range of other adjunct treatments have been recommended and/or are available that could be used to manage osteoarthritis-related pain.

This evaluation aimed to assess the cost-effectiveness of alternative non-pharmacological adjunct interventions in patients experiencing pain attributable to osteoarthritis of the knee. The setting of the study was UK primary care and the costing perspective was that of the NHS and Personal Social Services. The outcome measure used was the QALY. The comparators included in the economic evaluation are documented in *Table 27*. The exact delivery of comparators reflects the delivery (and heterogeneity in delivery) of the underlying trials (see *Appendix 3*).

### The evidence base

To inform estimates of the effectiveness of the alternative interventions, we conducted an extensive systematic literature review, presented in *Chapter 3*, synthesising the network of available RCT data using network meta-analysis methods, to assess the effectiveness of acupuncture and other relevant physical treatments for alleviating osteoarthritis-related knee pain. Of the 22 main interventions in the studies forming this evidence base, we found that muscle-strengthening exercise, acupuncture, TENS and balneotherapy were the interventions most commonly investigated. Studies typically recruited from general populations with osteoarthritis of the knee, although weight-loss trials (as expected) recruited only overweight or obese participants. Mean ages in the studies ranged from 53 to 85 years and the proportion of female patients ranged from 26% to 100%. Usual care and placebo were the most frequently studied comparators, with 'no intervention' being used rarely. There was considerable variation in the average treatment duration across the interventions, although a majority of interventions were administered over a 2- to 6-week period. For five of the studies included in this review, IPD were available from the repository prepared by the ATC for the IPD meta-analysis study presented in *Chapter 2*.

**TABLE 27** Adjunct non-pharmacological interventions for knee osteoarthritis

| Intervention | Intervention subtype |
| --- | --- |
| Acupuncture | |
| Appliances | Braces |
| | Insoles |
| Electrotherapy | Interferential therapy |
| | Laser/light therapy |
| | NMES |
| | PES |
| | PEMFs |
| | TENS |
| Manual therapy | |
| Static magnets | |
| Heat treatment | |
| Usual care | |

### The evidence base in the context of the economic decision problem

The NICE clinical pathway for the management of osteoarthritis distinguishes between core treatments (i.e. information, exercise and weight loss) and a number of non-core treatments that may be given to patients as needed, depending on preferences, needs and risk factors.[228] These additional treatments are classified by NICE as pharmacological adjunct, non-pharmacological adjunct and possible joint surgery following referral. The 22 interventions included in the systematic literature review and network meta-analysis presented in *Chapter 3* included core and non-pharmacological adjunct treatments. Furthermore, NICE will undertake a full review of evidence on the pharmacological management of osteoarthritis, which will be carried out after a review of the safety of over-the-counter analgesics is completed by the Medicines and Healthcare products Regulatory Agency. Therefore, the decision problem addressed in this chapter is, 'which is the most cost-effective non-pharmacological adjunct treatment for individuals in England and Wales with pain of the knee caused by osteoarthritis'? It follows that information provision, activity and exercise and interventions to achieve weight loss are excluded from the cost-effectiveness comparisons on the grounds that they are core treatments and the decision about which to use is expected to be independent of the choice of adjuvant therapy. Activity and exercise interventions were retained in the network meta-analysis as they provide indirect evidence regarding the relative efficacy of the interventions of interest in the cost-effectiveness analysis. Trials of weight-loss interventions were excluded from the network meta-analysis as these interventions were trialled only in overweight osteoarthritis of the knee patients and inclusion of these data was expected to increase heterogeneity in the network.

Usual care and no intervention were pooled in the current analysis as information available from the trials did not allow these comparators to be clearly distinguished. Furthermore, balneotherapy was included in the network meta-analysis but not the cost-effectiveness analysis as it seemed unlikely that the NHS in England and Wales would invest in provisions for mineral bathing. Ice/cooling treatment was included in the network meta-analysis but not the cost-effectiveness analysis as use of local cold application (packs, massage) is widely used as part of self-management at no (or minimal) cost and no known risk.[228]

Comparator interventions were classified as in *Chapter 3*. This included classifying sham acupuncture separately from other placebo interventions but otherwise considering all sham interventions in a single placebo category. The potential for this to increase heterogeneity in the network is addressed in the discussion.

## Data included in the synthesis model

All studies from the systematic review in *Chapter 3* that provided data suitable for the network meta-analysis were initially considered for inclusion in the synthesis (see *Appendix 5* for a description of the data required for the network meta-analysis). The characteristics of the studies included in this study have been reported in detail in *Appendix 3* as well as in a previous report[168] and are briefly summarised in *Appendix 5*. The network meta-analysis included 88 studies (out of a possible 152) and 7507 patients. The remaining 64 studies were not included because of limitations in the collection and reporting of data. IPD were available for five of the 88 trials,[69,77,120,134,136] including 1329 patients. As IPD were made available by the ATC, IPD were available only for acupuncture trials. The systematic review in *Chapter 3* identified 25 trials including acupuncture as a comparator; of these, 16 provided data suitable for inclusion in the current analysis and five were included by the ATC and therefore contributed IPD.

The studies were generally small, with only 15 of the 88 studies including > 50 patients per trial arm. It was felt that, because of the heterogeneity in follow-up assessment and duration of treatment, the analysis should focus only on data reported while patients were on treatment (or within 2 weeks of treatment discontinuation). Based on feedback from clinical experts (a GP and a physiotherapist), patients in the NHS are typically offered 6–10 weeks of treatment to alleviate pain related to osteoarthritis of the knee. The analysis therefore included data reported closest to the 8-week time point. This time point was also used in the review of acupuncture conducted as part of the development of the NICE osteoarthritis guideline.[228] The time points available for analysis ranged from < 1 day to 1 year, as determined by the nature of the intervention, treatment duration, trial-specific design and planned follow-up assessment. *Figure 15* provides a visual representation of the network of all included trial data.
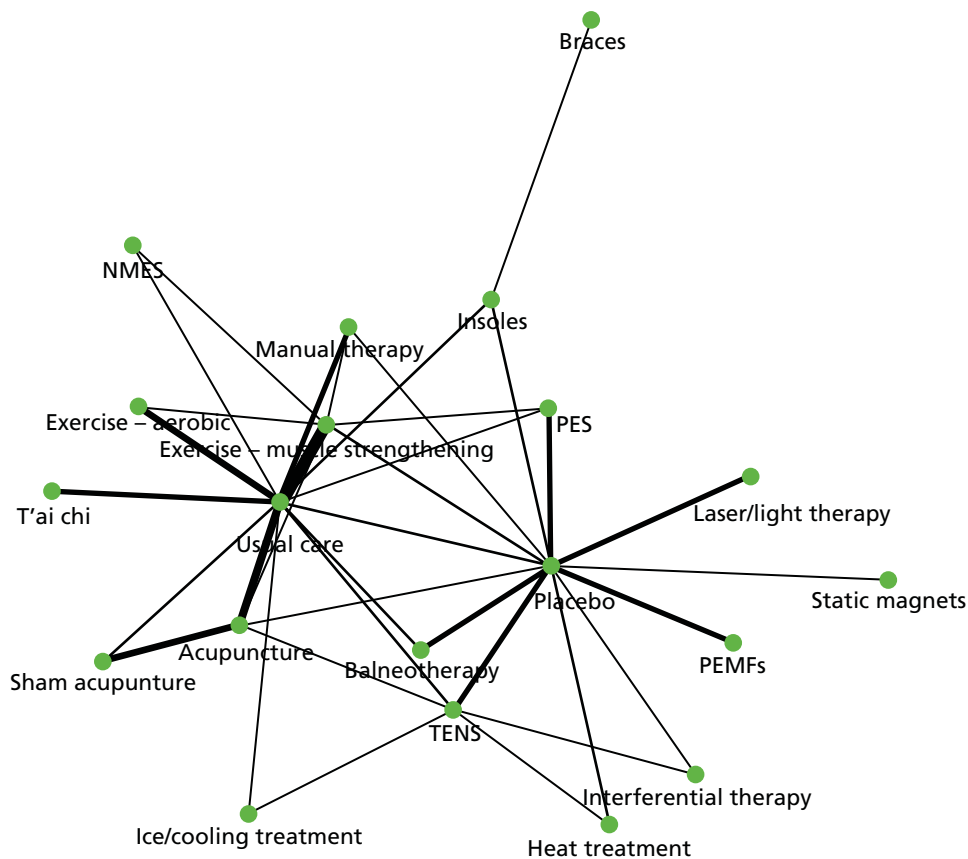
**FIGURE 15** Network of evidence for interventions in osteoarthritis of the knee.

Here, points represent competing treatment strategies and solid lines describe treatment comparisons for which direct trial evidence exists. Each line's thickness is proportional to the number of studies informing a comparison between two strategies, thus, the thicker the line the larger the number of trials available for that comparison. The largest numbers of trials were available for the comparisons between muscle strengthening exercise and usual care ($n = 14$), acupuncture and sham acupuncture ($n = 8$), aerobic exercise and usual care ($n = 7$), and acupuncture and usual care ($n = 7$).

Given the variable and often poor quality of the underlying evidence base, the appropriate data set to inform decision-making is uncertain. Three different networks of evidence were therefore used in the primary analysis: first, a network meta-analysis in which all trials were included as described above; second, a subset of trials that restricted the network meta-analysis to studies with a low risk of bias for allocation concealment (39 trials); and, third, a network meta-analysis that used this criterion as well as further restricting the data set to those studies that reported outcomes between 3 and 13 weeks (31 trials). *Figures 16* and *17* show the network of evidence available for the second and third scenarios, respectively. *Appendix 5* documents which studies were included within each network.

## Health-related quality of life

Instruments designed to measure patient HRQoL can be classified according to whether they are generic or condition specific and whether they are preference based or not preference based (depending on whether or not the values used to score them have been derived using methods consistent with economic theory).[270,271] Examples of generic preference-based HRQoL instruments include the EQ-5D,[272] Health Utilities Index-3[273] and Short Form questionnaire-6 Dimensions,[274] whereas examples of generic non-preference-based instruments include the SF-36[275] and the SF-12.[276] Similarly, the WOMAC[115] and the Health Assessment Questionnaire (HAQ)[277] are examples of condition-specific non-preference-based HRQoL instruments in osteoarthritis, whereas an example of a condition-specific preference-based HRQoL instrument is the HAQ preference-based measure.[278]
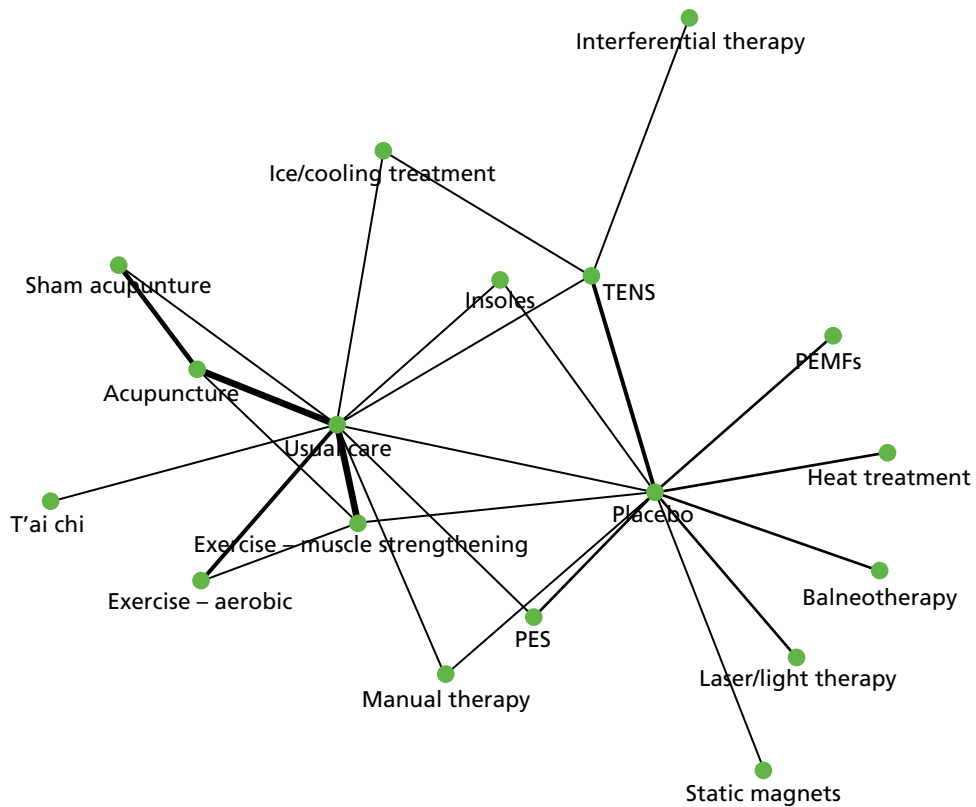
FIGURE 16 Network of evidence: trials with a low risk of bias for allocation concealment.
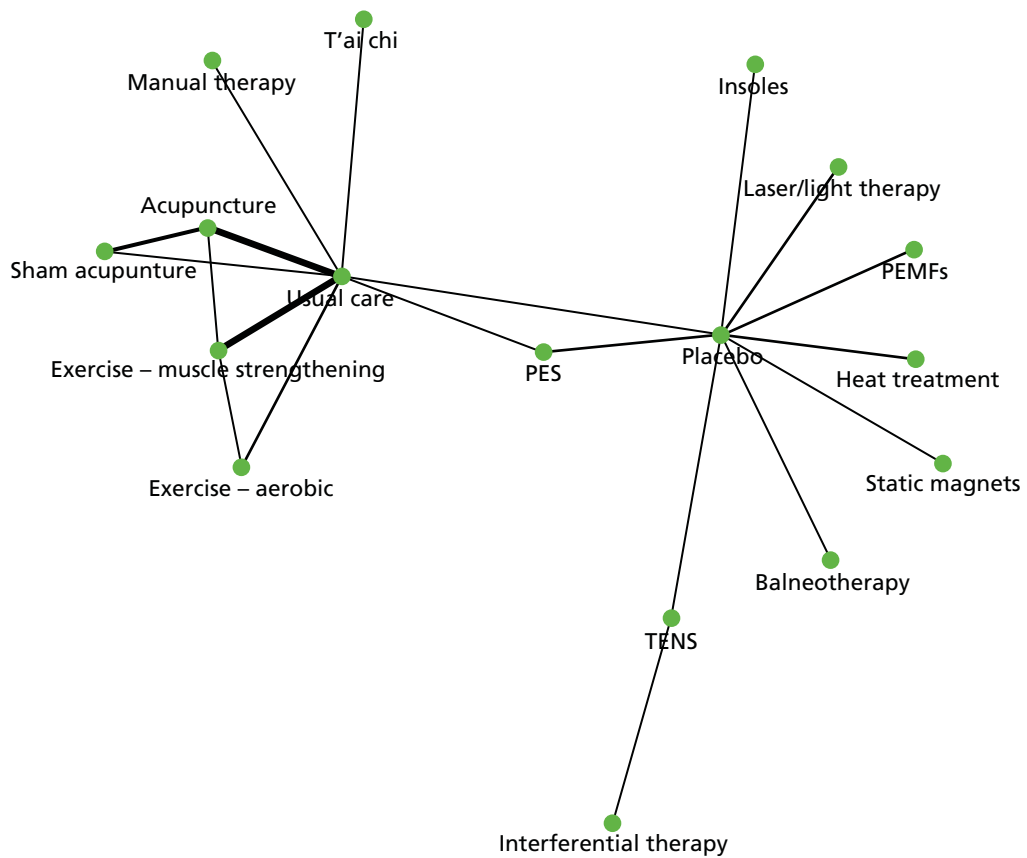


FIGURE 17 Network of evidence: trials with a low risk of bias for allocation concealment and reporting at 3–13 weeks.

The HRQoL instrument(s) reported in each study varied considerably across the 88 trials and many trials reported more than one instrument. The EQ-5D was our preferred end point for the economic assessment given the preferences of UK decision-makers, such as NICE, for this HRQoL measure.[224] We therefore focused on HRQoL instruments for which a mapping algorithm to the EQ-5D was available. The instruments for which a mapping algorithm to the EQ-5D was available, and the mapping algorithms themselves, were identified using the University of Oxford's HERC database of studies mapping from HRQoL or clinical measures to EQ-5D as reported by Dakin.[235] When multiple mapping algorithms were available for a given instrument, the preferred algorithm was selected on the basis of the sample size, adequacy of statistical modelling and relevance of the study population.

In a number of studies data for multiple HRQoL instruments for which mappings were available were reported. The preferred instrument was selected from these studies based on the extent to which the instrument was expected to reflect all dimensions of the EQ-5D. This resulted in the following hierarchy: EQ-5D preference values; SF-36 dimension scores; SF-36 MCS and PCS scores; SF-12 MCS and PCS scores; WOMAC total score; VAS measures of pain; and NRS measures of pain.

The HRQoL instrument used for each study is presented in *Appendix 5*. Generic HRQoL instruments were available for 19 studies (EQ-5D, $n = 3$; SF-36 dimensions, $n = 6$; SF-36 MCS and PCS, $n = 9$; SF-12 MCS and PCS, $n = 1$). The remaining 69 studies provided WOMAC ($n = 33$); pain VAS ($n = 33$) or pain NRS ($n = 3$) data. The distribution of instruments used according to treatment comparison and study size is shown in *Figure 18*. Those studies that included generic HRQoL data tended to be larger and of higher quality. Generic HRQoL instruments were used in studies including muscle-strengthening or aerobic exercise, acupuncture, PES, t'ai chi and NMES comparators.

### Health-care resource use

Only one study in the IPD data set provided resource use data.[69] This study was conducted in Germany and the only resource use item provided was the number of acupuncture sessions. It was, thus, deemed appropriate to bring in external study data to inform resource use. Resource use associated with administering each intervention was obtained using information from the clinical trials regarding the intensity of treatment and information from clinical experts (a GP and a physiotherapist) regarding the typical method of administering the treatment and equipment required.

It was anticipated that the frequency of primary care and outpatient specialist visits might be impacted on by interventions that reduce individuals' chronic pain. Resource use data from the TOIB RCT[231] was therefore used to relate changes in health outcomes (measured by the EQ-5D) to changes in these resource use items.

## Methods

This section describes the methods developed to carry out the synthesis of IPD and aggregate data, the rationale and methods for mapping generic and condition-specific HRQoL into EQ-5D scores, the approach used to derive health-care resource use and cost estimates for each treatment strategy being compared and the methods used to carry out the cost-effectiveness analysis. Given the complexity of the evidence base at hand and the range of methods used to address the challenges described in the previous section, *Figure 19* provides a visual representation of the relationships between data, models and outputs to facilitate exposition in this section.

### *Health-related quality of life for cost-effectiveness analysis*

As mentioned in the previous section, the 88 studies included in the base-case analysis used a range of different HRQoL instruments, which were also reported in a fragmented way (see *Appendix 5* for more details). To overcome this problem, in *Chapter 3* we synthesised study-level standardised mean (pain) differences as a measure of effectiveness. This is one of the possible solutions when combining treatment
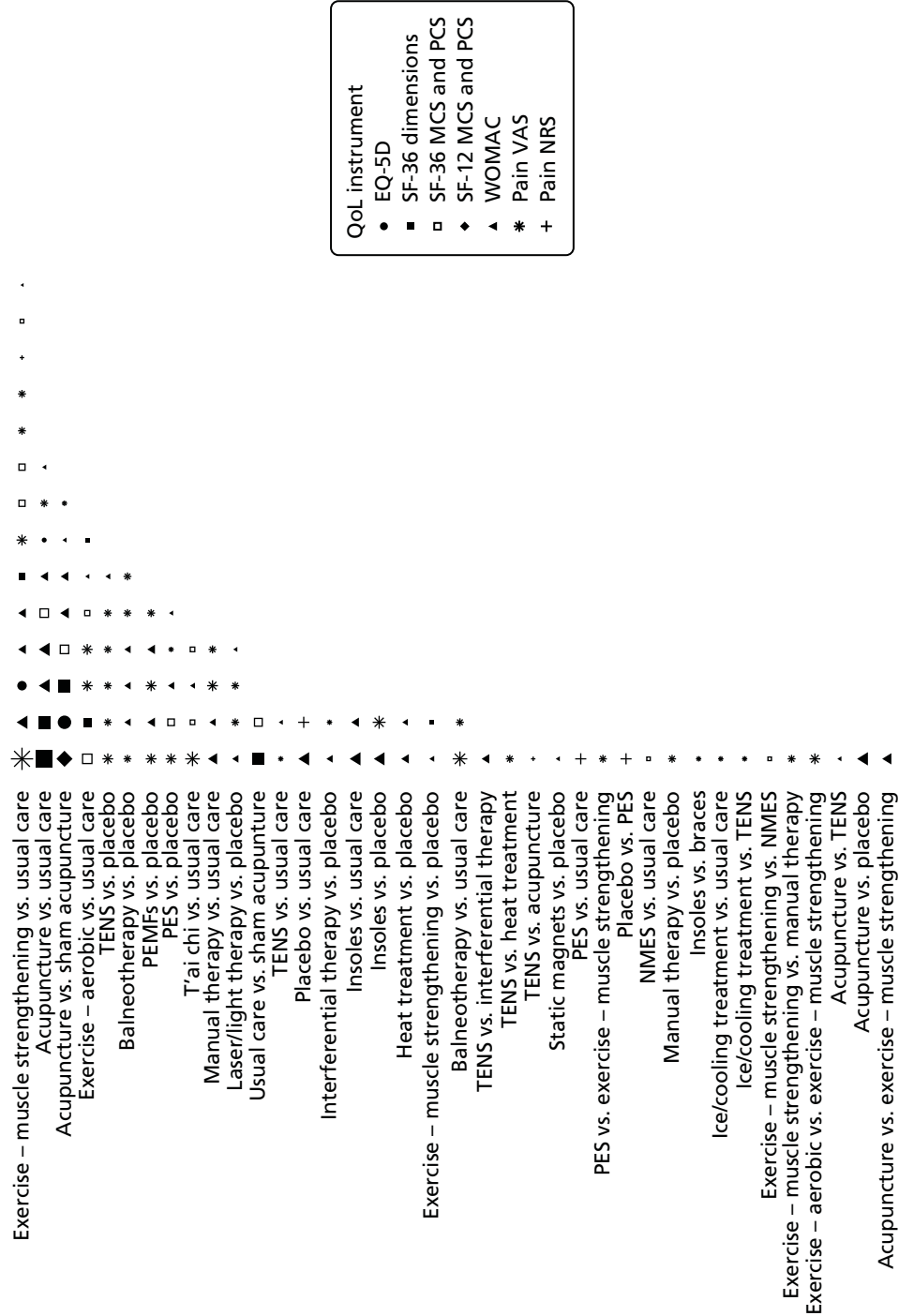
**FIGURE 18** Best available quality-of-life data by comparison. Note: shape size is proportional to study size. Shapes are used to represent different HRQoL instruments. Studies informing multiple comparisons (because of the presence of three or more trial arms) appear for each comparison.
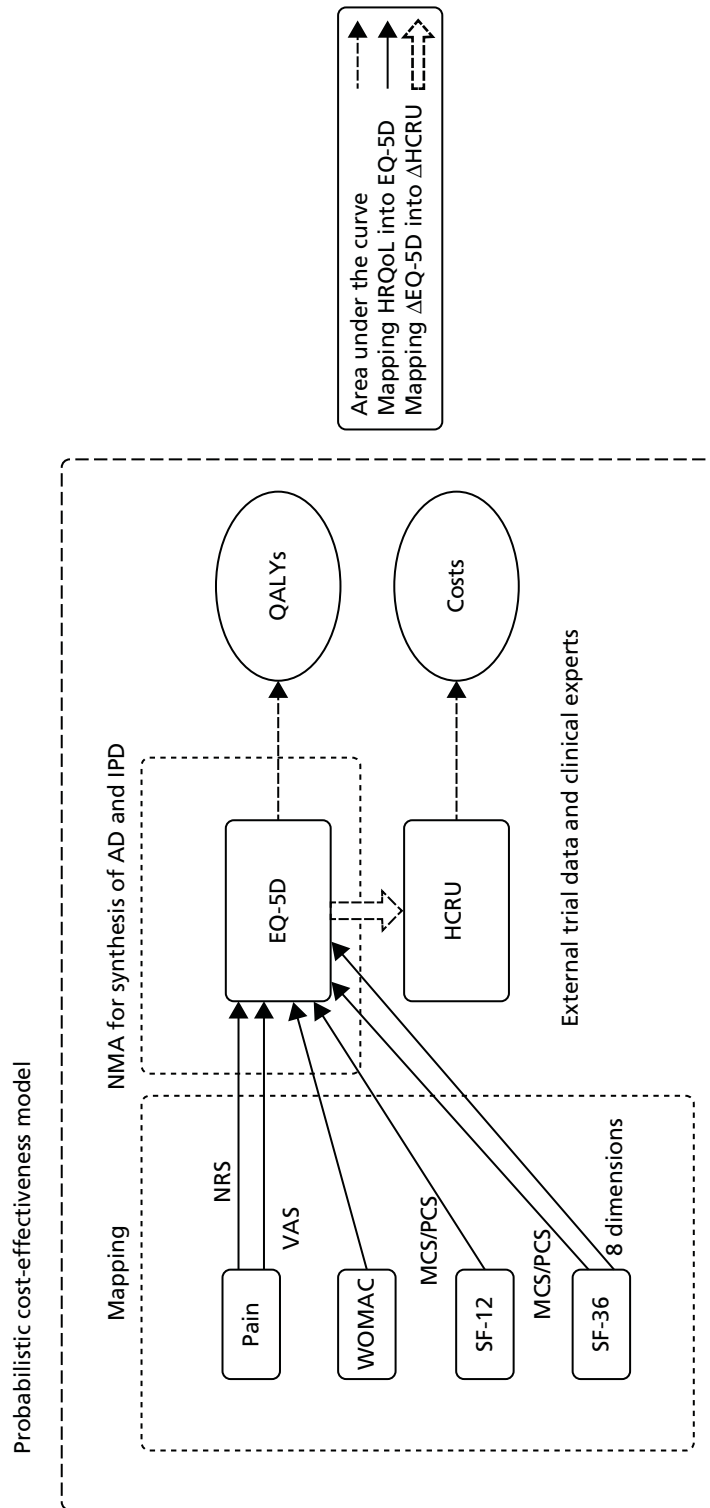
**FIGURE 19** Visual map of the model integrating synthesis, mapping and economic analysis. AD, aggregate data; HCRU, health-care resource use; NMA, network meta-analysis.

effect estimates from studies that measured the same outcome using a variety of different instruments.[279,280] However, the method has limitations and these are important when the SMD outcome is to be used to inform cost-effectiveness considerations. This is because health-care policy-makers require a common health outcome measure to use as a yardstick to be able to make decisions across different conditions and clinical areas. In many jurisdictions, including England and Wales,[224] this yardstick is the QALY.[225] The QALY is a composite measure that combines mortality and morbidity into a single numeraire, thus providing an estimate of an individual's remaining life expectancy weighted by some measure of HRQoL. In England and Wales, HRQoL is typically measured using a preference-based instrument such as the EQ-5D.[272]

The EQ-5D has been developed by the EuroQol Group [see www.euroqol.org (accessed 20 July 2016)] as a standardised instrument for describing and valuing HRQoL. It describes an individual's health state on five domains – mobility, self-care, usual activity, pain/discomfort, anxiety and depression – each of which has three levels of severity (no problem, some problems, extreme problems), giving rise to 243 possible health states. Studies in the general population have been used to derive societal preference values for each of the EQ-5D health states in several countries, including the UK.[232] The UK survey estimated the EQ-5D index score using the time trade-off method to elicit preference values for 42 of the 243 health states in the EQ-5D using a representative sample ($n = 3395$) of the UK population, and regression methods to predict the values of the remaining health states. The EQ-5D index value (also referred to as 'tariff' or 'social tariff') for the UK ranges from –0.594 (health state 33333, i.e. the worst possible state) to 1 (health state 11111, i.e. full health).

As the metric of interest for quantifying health benefits in cost-effectiveness analysis is the (absolute) difference in mean QALYs between treatment groups, a series of network meta-analysis models were developed to synthesise the EQ-5D data. These models were designed to include data available at both the study and the patient level. Unfortunately, most studies had not reported EQ-5D data. Therefore, the synthesis was preceded by a series of 'mapping' exercises used to predict EQ-5D scores from observed generic or condition-specific HRQoL data. Details of the mapping are provided in the next section.

## Synthesis of multiple heterogeneous outcomes

As described in the previous section, a range of mapping algorithms was used to predict EQ-5D estimates from the available HRQoL data. This generated a series of challenges. First, the mapping algorithms used were all non-linear and frequently used multiple correlated input dimensions. When IPD are available, these non-linearities and correlations between dimensions are appropriately reflected simply by applying the algorithm to the data reported for each patient. However, when only summary statistics for the HRQoL dimension scores are available, methods were required to appropriately reflect the non-linear nature of the mapping algorithms and the correlation between the dimension scores. This was addressed by generating simulated data sets for each aggregate data study's HRQoL instrument dimension scores, applying the relevant mapping algorithm and then estimating the statistics required for the synthesis from the resulting simulated EQ-5D data set. The statistics required for synthesis were baseline EQ-5D score and follow-up EQ-5D score as well as their variances. As the distribution of HRQoL dimensions was unknown, individual dimensions from the HRQoL instrument were assumed to follow a multivariate normal distribution. This required an estimate of the variance–covariance matrix for the HRQoL instrument dimensions at baseline and follow-up. Variances were obtained directly from the aggregate data studies. Covariances were estimated using these variances combined with correlations estimated from the IPD studies using standard formulae for the variance–covariance matrix. Correlations between EQ-5D index values at baseline and follow-up were estimated from one study.[134] Correlations between SF-36 dimension scores (and SF-36 PCS and MCS scores) at baseline and follow-up were estimated from three studies.[69,77,134] Correlations between SF-12 PCS and MCS scores were obtained from a study not included in the systematic review reported in *Chapter 3*, but made available by the ATC.[70] Correlations between baseline and follow-up WOMAC scores were obtained from four studies.[69,77,120,136] Correlations between baseline and follow-up VAS scores were estimated from four studies[120,134,136] and an additional trial obtained from the ATC.[135] No IPD were available for the pain NRS end point; correlations between baseline and follow-up NRS data were therefore assumed to be the same as for the VAS data. Samples of HRQoL measures that fell outside of the feasible

range were truncated to the minimum or maximum possible values for each HRQoL instrument. Mapped utilities generated by both the IPD and the aggregate data were truncated using the minimum and maximum EQ-5D tariff values.

A second challenge faced in all analyses of mapped HRQoL data was that simply applying a mapping algorithm does not adequately reflect the uncertainty in the mapping process. This is because of both the presence of unmeasured predictors (reflected in the residual error of the mapping algorithm) and the fact that the coefficients of the mapping algorithm are also random variables. Recent methodological developments[241] were therefore applied to capture residual error in the mapping algorithms. As recommended in Chan et al.,[241] variances for the mapped aggregate data were inflated by the inverse of the $R^2$ statistic from the corresponding mapping algorithm. For the IPD studies the observed variances were inflated in the same way. As data entered the model at the individual patient level, the additional variance was incorporated by adding a random deviate to the observed outcomes. This deviate was drawn from a normal distribution with a mean of zero and variance equal to the difference between the original and the inflated variances. Uncertainty in the regression coefficients from the mapping algorithm and in the valuation of the EQ-5D preference scores was not captured.

As reported in *Appendix 5*, the HRQoL measures reported in the evidence base of knee pain related to osteoarthritis included several different instruments, including the SF-36 (sometimes reported in terms of its eight dimensions and sometimes reported in terms of its PCS and MCS scores), SF-12 (MCS and PCS scores), WOMAC and pain (pain VAS or pain NRS). The following published mapping algorithms were applied to data collected using these instruments to derive EQ-5D estimates.

### SF-36 to EQ-5D

Those studies that used the SF-36 questionnaire reported the results either in terms of the score for the eight dimensions of the instrument or in terms of its PCS and MCS scores. The model used the mapping algorithm published by Rowen et al.[236] for those studies that reported results for the eight dimensions of the SF-36 questionnaire and the mapping algorithm published by Maund et al.[238] for those studies that reported the SF-36 PCS and MCS scores. Rowen et al.[236] mapping coefficients were obtained from a generalised least-squares regression of individual patient-level EQ-5D scores against the values of the eight dimensions of the SF-36, their squares and their interactions, as this was identified as the preferred model by the authors. Rowen et al.[236] analysed data from a wide range of inpatients and outpatients at Cardiff and Vale NHS Hospitals Trust. Maund et al.[238] mapping coefficients were obtained from an OLS regression of individual patient-level EQ-5D scores against the values of the SF-36 PCS and MCS scores, using data from patients with rotator cuff disease in primary care recruited in the SAPPHIRE (Stenting and Angioplasty With Protection In Patients at High Risk for Endarterectomy) trial.[281] The authors estimated five models: three OLS regressions (one with main effects for PCS and MCS scores only, another adding squared terms and a third adding both squared and interaction terms), one Tobit regression and one censored least absolute deviations (CLAD) model (both the Tobit and CLAD models included main effects, squared terms and interaction terms). All models performed similarly (with mean absolute errors of 0.18–0.19). Given this, for simplicity OLS models were preferred and the model including the main effects, their squares and their interaction was used on the basis of marginal improvements in explanatory power and model fit. The analysis of 1-, 3- and 12-month data was used as these analyses had slightly improved explanatory power and model fit compared with the 3-month analysis.

### SF-12 to EQ-5D

Those studies that used the SF-12 questionnaire reported its results in terms of its PCS and MCS scores. The model used the mapping algorithm published by Gray et al.,[237] which was obtained from a multinomial logistic regression model of individual patient-level EQ-5D scores against the values of the two summary scores of the SF-12, their squares and interaction terms. This model is based on an analysis of data from the Medical Expenditure Panel Survey [see https://meps.ahrq.gov/mepsweb/ (accessed 13 October 2016)], which reflects the HRQoL outcomes of non-institutionalised US civilians.

## WOMAC to EQ-5D

Barton et al.[239] developed a series of algorithms to map the WOMAC instrument on to EQ-5D scores. Using IPD from patients who participated in the Lifestyle Interventions for Knee Pain study,[282] the authors conducted a series of OLS regressions, relating the EQ-5D index value to various possible ways in which the WOMAC questionnaire may be reported. Five models were estimated in total: the first included total WOMAC score as the only explanatory variable; the second used the WOMAC pain, stiffness and functioning subscales; the third used the total WOMAC score and total WOMAC score squared; the fourth included pain, stiffness and functioning, their interactions and their squares; and the final model was the best fitting of the previous four plus age and sex. Access to the five different models proved useful as the WOMAC had been reported in different ways in the literature forming the evidence base. The model preferred by Barton et al.[239] included total WOMAC score, total WOMAC score squared, age, age squared and sex. This model was therefore applied to the IPD. Because of variable reporting of age and sex across studies, for aggregate data studies the model including only total WOMAC score and total WOMAC score squared was used.

## Pain to EQ-5D

A final set of studies included in the analysis reported pain, measured using either a NRS or a VAS. The report by Maund et al.[238] that provided the mapping algorithm for the SF-36 MCS and PCS scores also provided an algorithm to map pain VAS to EQ-5D scores. The authors estimated four models: two OLS regressions (one with the pain VAS and another with the pain VAS and its squared term), one Tobit regression and one CLAD model, using the pain VAS and its squared term as explanatory variables. All models performed similarly (with mean absolute errors of 0.18–0.20). Given this, for simplicity OLS models were preferred and the model including the main effects and their squares was used as there were no observed differences in model fit or explanatory power between the model with and without the squared term, and it seemed plausible that the relationship between pain VAS and EQ-5D scores was non-linear. The analysis of 1-, 3- and 12-month data was used as these analyses had slightly improved explanatory power and model fit compared with the 3-month analysis.

A mapping algorithm from the 11-point Pain Intensity Numerical Rating Scale (PI-NRS-11), ranging from 0 ('no pain') to 10 ('pain as bad as you can imagine'), to the EQ-5D was available from Gu et al.[283] To estimate two mapping algorithms, the authors used survey data from a US sample of patients who had at least 3 months of neuropathic pain, either painful diabetic peripheral neuropathy or post-herpetic neuralgia, and were receiving medications treating neuropathic pain. The first related the EQ-5D index score to a set of pain NRS dummy variables using OLS and the second used an ordered logistic regression model to predict the response levels (i.e. 1, 2, 3) for each of the EQ-5D dimensions using the same explanatory variables. Models were run with and without patient age, sex and pain duration as independent variables. The reduced models, which excluded age, sex and disease duration, were used given variable reporting of these variables across studies. The OLS model was used in the current analysis as it had a better fit than the ordered logistic model.

## Synthesis of individual-level and aggregate reporting of continuous outcomes

The second challenge that the analysis reported here had to address is common to the synthesis of all continuous end points when only summary statistics are available. ANCOVA is the preferred method for analysing continuous outcome data from RCTs.[219] In an ANCOVA analysis the post-baseline outcome of interest is regressed on both a treatment indicator and the baseline value of the outcome of interest. ANCOVA is preferred as it offers improved precision and lower bias than other available methods when there is imbalance in baseline outcomes. However, ANCOVA analyses are rarely reported in clinical publications, particularly for secondary end points such as HRQoL. Meta-analyses therefore typically analyse either post-baseline (or 'final') values or change from baseline values. An analysis of final scores would be expected to be biased in favour of the treatment with a higher baseline value. In the current data set, patients with higher EQ-5D values experienced lower changes in values; a change score analysis would therefore be expected to bias against the treatment with higher baseline values. Existing proposals to handle this have significant limitations. Fu et al.[284] recommend sensitivity analyses using final scores and change

scores. It is not clear how the results of this analysis would be interpreted, as there is no guarantee that, at the (network) meta-analysis level, the true treatment effects will be bounded by the results of these scenarios. Riley *et al.*[219] propose dropping studies with significant imbalance; however, this results in a loss of information and requires a demarcation of what constitutes 'significant imbalance'. Methods were therefore developed to address this concern, given the significant imbalances observed in this data set.

Differences in study designs, study populations and the implementation of individual interventions seemed likely to generate heterogeneity in the underlying true effects. We therefore developed random-effects models as in *Chapter 3*. This model extends the IPD model presented in *Chapter 4* to the more commonly faced situation in which IPD are available for only a subset of studies, with only aggregate data reporting available for the remainder.

The outcome variable, absolute EQ-5D score at follow-up, was assumed to be normally distributed. Although this assumption is unlikely to characterise EQ-5D scores,[233] simulation results have shown treatment effect estimates from ANCOVA to be very close to the true treatment effects for a range of non-normal outcome distributions.[285]

Individual patient data were included in the model using an ANCOVA network meta-analysis model:

$$Y_{ikjt} \sim N\left(\theta_{ikj}, \sigma_j^2\right)$$
$$\theta_{ikj} = \begin{cases} \alpha_{bj} + \beta_{j0}Y_{ikj0} & \text{if } k = b \\ \alpha_{bj} + \beta_{j0}Y_{ikj0} + \delta_{kbj} & \text{if } k > b, \end{cases} \tag{3}$$

where $Y_{ijkt}$ and $Y_{ijk0}$ are the values of the (continuous) outcome at time point $t$ and baseline ($t = 0$), respectively, for participant $i$ in treatment arm $k$ of study $j$; $\sigma_j^2$ represents the study-level variance; the quantity $\alpha_{bj}$ represents the outcome for the baseline treatment $k = b$ in study $j$ for a patient with an EQ-5D score $Y_{ijk0}$ of zero at time 0 (this is study specific to respect within-study randomisation); the parameter $\beta_{j0}$ represents a study-specific estimate of the impact of baseline EQ-5D score on the final outcome; and $\delta_{kbj}$ represents the treatment effect for treatment $k$ relative to treatment $b$ in trial $j$.

The aggregate data are included in a final outcomes model that has been modified to adjust for the potential bias caused by differences in outcomes between treatments at baseline ($Y_{kbj0} = Y_{kb0} - Y_{bj0}$):

$$Y_{kjt} \sim N\left(\theta_{kj}, \sigma_{kj}^2\right)$$
$$\theta_{jk} = \begin{cases} \alpha_{bj} & \text{if } k = b \\ \alpha_{bj} + \beta_{pred}Y_{kbj0} + \delta_{kbj} & \text{if } k > b, \end{cases} \tag{4}$$

where $Y_{kjt}$ and $\sigma_{kj}^2$ are the mean and variance for the final outcome for treatment arm $k$ of study $j$ at time $t$, respectively.

This model adjusts the treatment effects to emulate the ANCOVA model results. This is achieved by considering the reported unadjusted results as being subject to omitted variable bias, where the omitted variable is baseline EQ-5D score. Omitted variable bias can be estimated as the product of the coefficient on the baseline EQ-5D score from the 'correctly' specified model (the ANCOVA model) multiplied by the coefficient from a regression of the omitted variable (baseline EQ-5D score) on the included variable (treatment).[286] The coefficient on the baseline EQ-5D score from the ANCOVA model ($\beta_{j0}$) is obviously unknown for the aggregate data studies. This coefficient is assumed to be exchangeable with the coefficients obtained from the studies for which IPD were available and can therefore be estimated as:

$$\beta_{pred} \sim N(\beta_0, \sigma_{\beta0}). \tag{5}$$

$\beta_0$ and $\sigma_{\beta0}$ are estimated by assuming that the coefficient on the baseline outcome measure for each IPD trial ($\beta_{j0}$) is drawn from a normal distribution with mean $\beta_0$ and variance $\sigma_{\beta0}$.

The covariate from the regression of baseline EQ-5D score on treatment is simply the difference in baseline EQ-5D score between the treatment and the control for the comparison of interest (i.e. $Y_{kbj0}$).

The estimated study-specific treatment effects ($\delta_{kbj}$) from the IPD and aggregate data inform a common random-effects consistency model:

$$\delta_{kbj} \sim N(d_{kb}, \sigma^2)$$
$$d_{kb} = d_{k1} - d_{1b},$$

(6)

where $d_{kb}$ represents the treatment effects for the comparison of treatment $k$ with treatment $b$; these are assumed to be derived from a set of basic parameters ($d_{k1}$), which estimate the treatment effect of treatment $k$ relative to the reference treatment ($k = 1$), where $d_{11} = 0$.

Priors were defined to be vague and were specified as follows: $\alpha_{bj} \sim N(0,10^4)$; $d_{k1} \sim N(0,10^3)$; $\sigma k \sim Unif(0,2)$; $\beta_0 \sim N(0,10^6)$; $\sigma_{\beta0} \sim Unif(0, 2)$; $\sigma \sim Unif(0, 2)$.

If outcome data were missing for patients within the IPD, these cases were dropped from the analysis (this occurred in 21% of all cases). If baseline data were missing for the aggregate data studies, the difference at baseline ($Y_{kbj0}$) was assumed to be zero.

Correlation in the random effects from trials with three or more arms were accounted for using previously published methods.[222] When SEs were not reported and could not be derived from the reported data, they were imputed using the methods described by Dakin et al.,[234] which allow for the uncertainty in the imputation process.

Model fit was assessed using the posterior residual deviance and the DIC. In the IPD each patient (data point) should contribute approximately 1 to the residual deviance, whereas for the aggregate data each study arm (data point) should contribute approximately 1 to the residual deviance. Residual deviance contributions from each patient in each IPD study arm were therefore summed and divided by the total number of patients in the study arm to allow an aggregate measure of mean posterior residual deviance to be estimated and to allow deviance contributions to be compared across studies.

Inconsistency was assessed using the methods set out in Dias et al.,[287] whereby the fit of the consistency model is compared with the fit of an inconsistency model. As recommended in Dias et al.,[287] changes in the model fit associated with using the inconsistency model were assessed using 'omnibus' diagnostics (DIC; posterior mean residual deviance). Residual deviance contributions for each data point were also reviewed for both consistency and inconsistency models.[180]

Convergence was assessed by inspecting trace plots, density plots and Brooks–Gelman–Rubin plots as well as autocorrelation plots.[177] Efficiency was assessed by comparing the Monte Carlo error to the posterior SD. All models were run for 100,000 'burn-in' iterations and a further 100,000 samples.

Data manipulation and analysis was carried out using the software packages R version 3.0.2 and WinBUGS[246] version 1.4.3. The network meta-analyses were undertaken in WinBUGS, linked to the R software through the package R2WinBUGS.[247,288] The statistical software Stata 12.0 was used to analyse the relationship between changes in resource use and changes in EQ-5D score. The annotated WinBUGS code is provided as *Appendix 5*.

## Study heterogeneity

Three sources of heterogeneity were anticipated to be potential treatment effect modifiers based on a review of the trial data and clinical opinion: bias associated with trial quality, variable reporting times across trials and differences in patients across trials.

Study quality has been found to be a study-level treatment effect modifier in the context of osteoarthritis trials.[289] A meta-epidemiological study by Nüesch et al.[289] of 16 meta-analyses comparing active with control interventions (or placebo) in patients with hip or knee osteoarthritis found effect sizes to be less

beneficial in trials with adequate allocation concealment than in those with inadequate or unclear allocation concealment according to the Cochrane risk-of-bias tool.[171] The Cochrane risk-of-bias tool assesses allocation concealment based on whether or not the method used to conceal the allocation sequence is sufficient to avoid intervention allocations being foreseen before or during enrolment. The risk-of-bias score (low, unclear or high) is assigned based on the risk of selection bias because of inadequate concealment of allocations prior to assignment.

The difference in effect size was most pronounced when the trial data set was restricted to trials of complementary medicine interventions, which included acupuncture, chondroitin, glucosamine, PEMFs and static magnets. The impact of patient blinding was also explored; however, observed differences in treatment effects were less consistent and disappeared after accounting for allocation concealment. We therefore ran analyses restricting the study set to those with adequate allocation concealment according to the Cochrane risk-of-bias tool.[171]

It was expected that trials reporting at very short or very long time points may have both different characteristics (e.g. very short trials may be of a more experimental nature) and reflect different points on a patients' HRQoL change trajectory. An additional analysis was therefore conducted further restricting the analysis of trials with adequate allocation concealment to only those trials reporting within the 3- to 13-week period (note that originally an analysis restricted to trials reporting in the 4- to 12-week range was planned; however, this resulted in the network becoming disconnected).

Increased age and BMI were potentially expected to reduce the efficacy of the interventions considered in this synthesis. Given the low number of studies available for many of the contrasts we used the common interaction effect model outlined in the study by Dias $et\ al.$[245] A common interaction term $\beta_{xk1}$ was assumed to apply to all treatment effects relative to treatment 1 (usual care). This required modifications to $Equations\ 3$ and $4$ as shown in $Equations\ 7$ and $8$, where $X_{ikj}$ represents the centred age (BMI) of patient $i$ in treatment arm $k$ of study $j$ in the IPD studies, and $X_j$ represents the centred mean age (BMI) of study $j$ in the aggregate data studies. The coefficient $\beta_{xj}$ represents the study-specific impact of centred age (BMI) on the outcome measure. The logical operator I($u$) takes the value 1 if the condition $u$ is met and 0 otherwise.

$$Y_{ikjt} \sim N\left(\theta_{ikj}, \sigma_j^2\right)$$
$$\theta_{ijk} = \begin{cases} \alpha_{bj} + \beta_{j0}Y_{ikj0} + \beta_{xj}X_{ikj} & \text{if } k = b \\ \alpha_{bj} + \beta_{j0}Y_{ikj0} + \delta_{kbj} + \beta_{xk1}X_{ikj}I(b=1) & \text{if } k > b. \end{cases} \tag{7}$$

$$Y_{kjt} \sim N\left(\theta_{kj}, \sigma_{kj}^2\right)$$
$$\theta_{jk} = \begin{cases} \alpha_{bj} & \text{if } k = b \\ \alpha_{bj} + \beta_{pred}Y_{kbj0} + \delta_{kbj} + \beta_{xk1}X_jI(b=1) & \text{if } k > b. \end{cases} \tag{8}$$

The following additional priors were required for the covariate modelling: $\beta_{xj} \sim N(0,10^4)$; $\beta_{xk1} \sim N(0,10^4)$. Missing covariate values were imputed assuming an ignorable missing covariate mechanism using the methods described in Lunn.[290] Missing data for BMI and age were assumed to be drawn from independent normal distributions, and separate models were assumed for missing IPD and aggregate data covariates.

### Methods for estimating the total cost
The total cost included the costs of providing the interventions, primary care visits and outpatient hospital visits. The cost year was 2012–13.

### Intervention costs
The main purpose of the cost-effectiveness analysis was to estimate the costs and QALYs associated with each treatment and identify which strategy provided 'value for money' in the NHS for England and Wales. Unfortunately, most of the studies in the data set either did not collect health-care resource use data to

inform a costing exercise relevant to the NHS for England and Wales or did not collect resource use data at all. Health-care resource use information is typically considered not transferable from country to country as it is a function of country-specific health-care system factors (e.g. the incentive structure faced by providers and patients, relative prices of available health-care resource use, the treatment strategies available in the country), which are difficult to control for even when the analyst has access to IPD.

To estimate the resource use associated with delivering each intervention, clinical experts were consulted, as well as the published literature (including the published trial data) and UK NHS data, such as NHS foundation trust information available on the internet. The resulting intervention resource use and cost data are summarised in *Table 28*.

For interventions assumed to be administered via regular physiotherapy attendance (acupuncture, electrotherapy excluding TENS, thermotherapy and manual therapy), weekly attendance durations were estimated from the clinical trial data. Acupuncture is administered in the UK by a range of practitioners. Recent survey data suggest that, of acupuncture administered within the NHS, 59% is administered by physiotherapists with the remainder administered by medical doctors (29%), nurses (4%), other health-care professionals (5%) or independent practitioners (4%).[1] For the purpose of estimating costs, we therefore assume that acupuncture is carried out by a physiotherapist. Unless a marked difference in outcomes was expected between delivery by a physiotherapist and delivery by a medical doctor, it seems unlikely that provision by medical doctors would be cost-effective given the large difference in hourly costs.[249]

Weekly attendance was estimated for each trial as the total number of sessions multiplied by the number of minutes per session divided by the number of weeks of treatment. A weighted average for weekly attendance was then estimated from the trials included in the network meta-analysis. This was carried out separately for the analysis of all trials and the analysis restricted to trials with adequate allocation concealment to ensure that in each cost-effectiveness analysis scenario the resource use was aligned with the efficacy data. Equipment administered by physiotherapists was not included in the costing, as the associated per-patient costs were expected to be small given the high throughput of individuals experiencing osteoarthritis of the knee and chronic pain of other aetiologies.

Insoles, braces and static magnets would be typically used for longer durations, beyond the 8-week cost-effectiveness model time horizon. Therefore, it was assumed that for these therapies, the (upfront) costs could be spread across their useable lifetimes and that the effects observed within the network meta-analysis would be maintained throughout their useable lifetimes. For TENS we assumed that the patient used the TENS machine at home for 8 weeks and then returned it to the physiotherapy unit for use by other patients.

Unit costs were obtained from published national sources when possible and are reported in *Table 29*. The financial year 2012–13 was used to value resource use relevant to each intervention. Costs reported for previous financial years were inflated using the Hospital and Community Health Services prices index.[249] Total costs were calculated by multiplying resource use by the unit cost. Other treatments and health-care interactions that may form a package of 'usual care' were assumed to have been provided equally to all patients regardless of comparator. These costs were therefore omitted from the analysis.

## Other health-care costs

Non-treatment-specific health-care resource utilisation was assumed to be a function of any change in EQ-5D over the 8 weeks of the treatment period and was derived from the TOIB trial.[231,292] Briefly, the TOIB trial collected EQ-5D and health-care resource utilisation data at baseline and follow-up; having gained access to the TOIB trial IPD, it provided an opportunity to estimate a series of simple models that regressed changes in health-care resource utilisation (i.e. GP visits, nurse practitioner visits, outpatient hospital specialist visits) against changes in EQ-5D score. A simple OLS analysis was used to regress the change in resource use from 0–3 months to 3–12 months on the change in EQ-5D score between month 3 and month 12. Primary and secondary care visits were analysed separately. The estimated regression

**TABLE 28** Interventions for the management of chronic pain in NHS patients with osteoarthritis of the knee in primary care

| Interventions | | Resource use | Trials (weekly attendance in minutes), mean (range) | |
| --- | --- | --- | --- | --- |
| | | | All | With adequate allocation concealment[a] |
| Acupuncture | | Administered weekly by a physiotherapist over 8 weeks | 37 (18–80) | 40 (20–50) |
| Biomedical appliances | Braces | Initial 40-minute visit to a physiotherapist to prescribe brace and a subsequent 30-minute visit to fit brace. Assume brace lasts for 6 months | N/A | N/A |
| | Insoles | Initial 40-minute visit to a podiatrist to prescribe insole and a 30-minute follow-up visit. Assume use of a ready-made insole, which lasts for 1 year | N/A | N/A |
| Electrotherapy | Interferential therapy | Administered weekly by a physiotherapist over 8 weeks | 159 (40–245) | 245 (N/A[b]) |
| | Laser/light therapy | | 105 (25–210) | 60 (N/A[b]) |
| | NMES | | 100 (N/A[b]) | N/A[c] |
| | PES | | 82 (57–114) | 85 (57–114) |
| | PEMFs | | 303 (80–600) | 120 (N/A[b]) |
| | TENS | Initial 40-minute visit to a physiotherapist to provide intervention, which is subsequently self-administered. Assume TENS machine lasts for 1 year | N/A | N/A |
| Thermotherapy | Heat treatment | Short-wave diathermy administered weekly by a physiotherapist over 8 weeks. In one trial, heat treatment consisted of a heat-retaining knee sleeve. This is assumed to be prescribed by a physiotherapist (40-minute consultation) and then used as required for 6 months. Total cost for heat treatment is a weighted average of diathermy and heat-retaining sleeve costs, with weights of 73%/27% in line with the number of patients enrolled in the heat treatment trial arms | 84 (60–143) | 60 (N/A[b]) |
| Manual therapy | | Administered weekly by a physiotherapist over 8 weeks | 63 (30–90) | 57 (30–90) |
| Static magnets | | Initial 40-minute visit to a physiotherapist to prescribe static magnet. Assume static magnets last for 2 years and 50% of patients require one replacement wrist strap | N/A | N/A |

N/A, not applicable.
a Data were identical for the analysis restricted to 3–13 weeks.
b Ranges not available as they are not meaningful when only one trial informed an analysis.
c Intervention does not provide data to inform network.

**TABLE 29** Unit costs of interventions (UK £, 2012–13)

| Health-care professional/appliance | Unit cost[a] (£) | Unit | Source |
|---|---|---|---|
| Hospital physiotherapist[b] | 36 | Per hour | Schema 13.1[291] |
| Community podiatrist | 30 | Per hour | Schema 9.4[291] |
| GP visit | 45 | Per 11.7-minute consultation | Schema 10.8b[291] |
| Secondary care specialist visit | 135 | Per consultation | Schema 7.1[291] |
| Brace | 88 | Brace | Bauerfeind GenuTrain Knee Support [www.healthandcare.co.uk (accessed 9 May 2014)] |
| Insole | 50 | Insole | Ready-made lateral wedge foot insole [www.rcht.nhs.uk/DocumentsLibrary/PeninsulaCommunityHealth/OperationsAndServices/Podiatry/PodiatryDischargeLetter.pdf (accessed 9 June 2014)] |
| Heat-retaining sleeve | 11 | Sleeve | www.amazon.co.uk/TITANIUM-ADJUSTABLE-KNEE-HEATING-STRAP/dp/B007EBTXDS/ref=sr_1_3?ie=UTF8&qid=1401871047&sr=8–3&keywords=heat+knee (accessed 9 June 2014) |
| Static magnets | 35 | Per static magnet | www.magnetsforall.com/category/40397320 (accessed 8 July 2014) |
| | 15 | Per additional strap | |
| TENS | 35 | Per TENS machine | www.boots.com/en/Boots-TENS-Digital-Pain-Relief-Unit_1405593/ (accessed 22 May 2014) |

a  Labour costs inclusive of qualifications; equipment costs exclusive of value-added tax.
b  Note that the cost of a community physiotherapist is very similar (£34 per hour including qualifications).

coefficients were then used (in combination with the unit costs from *Table 29*) to derive changes in cost components and, following aggregation, change in total costs as a function of changes in EQ-5D score over the 8-week time period of the base-case analysis. The regression coefficients estimated that a 1-unit increase in the EQ-5D resulted in a 0.89 (SE 0.53) decrease in primary care visits and a 0.47 (SE 0.56) increase in specialist visits.

## Cost-effectiveness analysis

Linking predicted changes in EQ-5D score from the network meta-analysis for each of the competing strategies, and changes in total costs (treatment and non-treatment related) based on the analysis described in the previous section, facilitated the value for money assessment of competing adjunct non-pharmacological interventions for pain of the knee in individuals with osteoarthritis in the NHS. Direct HRQoL benefits were the focus of the evaluation as the interventions under appraisal are expected to impact on pain and functioning but not on disease progression. QALYs were therefore estimated from the EQ-5D estimates using an area-under-the-curve (AUC) approach.

The time horizon for the analysis was 8 weeks, in line with the preferred time point for the synthesis. There is limited evidence regarding whether or not the effects of the therapies appraised continue beyond the treatment period. For example, an earlier report by Corbett *et al.*[168] found that only 23% of studies reporting data suitable for synthesis reported data 8–16 weeks from the end of treatment. The possibility that interventions may offer longer-term benefits was explored in scenario analyses. Some studies did, however, report results for multiple on-treatment time points. This could have provided information on interim on-treatment outcomes. However, it was not expected that there would have been sufficient data to provide robust treatment-specific estimates of the HRQoL profile from baseline to 8 weeks.

Quality-adjusted life-years were therefore estimated assuming that the benefit of treatment was achieved instantaneously, was maintained for 8 weeks and was then lost instantaneously. This is equivalent to assuming that the full benefit was gradually achieved over a specified period and then lost linearly over the same period, which may be viewed as a more realistic scenario for some interventions. For example, the benefit could be linearly achieved from the start of treatment until 8 weeks and gradually lost over the 8 weeks following treatment completion. The model is therefore expected to reasonably approximate a range of alternative HRQoL profiles (*Figure 20*). For example, expert opinion suggested that electrotherapy, static magnets and heat treatment may be more likely to follow an instant gain/instant loss-type profile, whereas acupuncture, manual therapy and the appliances may be associated with a gradual gain/gradual loss-type profile. These issues are explored further in the sensitivity analyses.

Placebo and sham acupuncture were not included as comparators in the cost-effectiveness analysis as it is not expected that the NHS would be willing to prescribe these as treatments. The analysis therefore controls for regression to the mean effects (as all improvements are relative to usual care) but not patient expectation effects.[293] The underlying assumption made here is therefore that the patient expectation effects observed in the trials would also be observed in clinical practice.

The model was run probabilistically by assigning the posterior distribution from the network meta-analysis to the effect sizes and reflecting the uncertainty in the relationship between the EQ-5D and primary care/outpatient resource use using a normal distribution (parameterised using data described in *Other health-care costs*).

After obtaining changes in expected costs and QALYs for each strategy, these were ranked by mean change in costs, starting from the least costly. ICERs were then calculated by dividing incremental (change in) costs by incremental (change in) QALYs. Treatment strategies that were dominated and those subject to extended dominance were subsequently excluded and ICERs recalculated if necessary. A treatment is dominated if it generates worse health outcomes and has equal or higher costs when compared with an alternative treatment. Extended dominance occurs when a treatment is less effective and has a higher ICER than an alternative treatment.

Derived posterior distributions for the incremental costs and incremental QALYs were used to calculate the probability that each strategy is cost-effective at the conventional cost-effectiveness threshold levels used by NICE (i.e. between £20,000 and £30,000 per QALY gained[224]). Given the large number of treatment strategies, it was considered more appropriate to plot the study results as a cost-effectiveness acceptability frontier,[294] which provides a direct indication of both the optimal strategy and the probability of that strategy being cost-effective for a given threshold value. Given that the time horizon was < 1 year, costs and QALYs were not discounted.
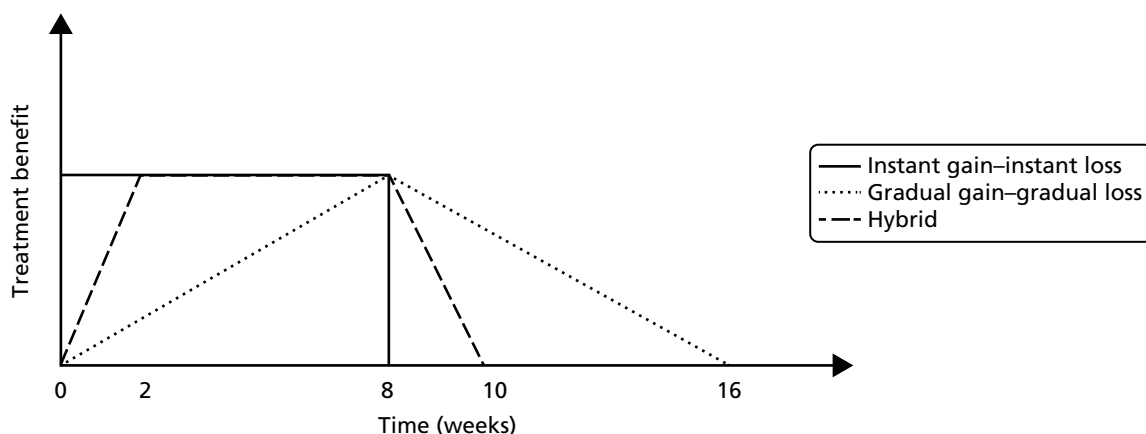


**FIGURE 20** Health-related quality-of-life profiles compatible with base-case assumptions.

The cost-effectiveness model was analysed using outputs of the synthesis of all trials; of trials that had a low risk of bias for adequate allocation concealment according to the Cochrane risk-of-bias tool;[171] and of trials with a low risk of bias for allocation concealment and reporting end-point data between weeks 3 and 13 post baseline.

For each scenario a series of sensitivity analyses were run. For treatments delivered via regular physiotherapy attendance, sensitivity analyses were run assuming that each session lasted as long as the shortest and the longest duration observed in the trial data for each intervention (as shown in *Table 28*). In the case of therapies for which the cost was driven by both labour and equipment costs (braces, insoles, static magnets and TENS), the costs were varied by ±50%. We also explored the impact of assuming that acupuncture was delivered by a private practitioner at the higher cost of £47.50 for the first session and £37.50 for subsequent sessions (for data sources see *Chapter 6, Methods for substudy 4: cost-effectiveness analysis*).

The intensity of therapy (hours per week) varied considerably across trials, with many therapies being administered much more intensively than would typically be expected in the context of a UK musculoskeletal outpatient physiotherapy service. Given the small number of studies for many of the interventions it was not possible to explore the impact of treatment intensity on treatment efficacy within the synthesis. We therefore conducted sensitivity analyses using appointment times more typical in the NHS for the purposes of costing [based on clinical opinion: 40-minute consultation followed by 20-minute treatment sessions (or 30 minutes for acupuncture and manual therapy)] and exploring different assumptions with respect to how the efficacy of treatment might vary when the session duration was reduced. Four assumptions with respect to the dose–response relationship of session duration and EQ-5D outcome were explored: (1) outcomes increase linearly to achieve a maximum benefit with the mean session duration, as listed in *Table 28*; (2) outcomes increase linearly up to the maximum value for 1 hour of session time per week, beyond which no further benefit is observed; (3) 75% of outcomes are achieved using 30-minute sessions with the remaining 25% of benefit achieved by moving from 30- to 60-minute sessions; and (4) the full benefit of treatment observed is obtained with 20-minute sessions (or 30 minutes for acupuncture), that is, no adjustment to outcomes. Moving from scenario (1) through to scenario (4), the benefits of increased session duration are assumed to be reduced. Scenarios (1) and (4) probably represent quite extreme scenarios. As an example, *Figure 21* shows the method for estimating outcomes under each scenario assuming that the weighted average session duration for the treatment is 80 minutes, which we wish to extrapolate to a 20-minute session duration.

Sensitivity analyses using the upper and lower values of the 95% CrIs from the network meta-analysis were also conducted. The threshold value treatment effect at which the decision would change, as well as the probability of observing a value at least this extreme, were calculated. These estimates provide an indication of the likelihood of observing an outcome sufficiently extreme for the conclusions of the analysis to alter. However, they are univariate analyses and do not account for other sources of uncertainty or the correlation between treatment effects estimated from the network meta-analysis.

In the base case the benefit of treatment is assumed to be instantaneous at time zero and to disappear immediately at the end of week 8. Two sensitivity analyses were conducted: the first was a threshold analysis estimating the duration of benefit beyond the 8-week period of treatment that would be required to alter the decision, assuming that all interventions delivered benefit over this period; the second extended the benefit of each intervention individually by 50% (equivalent to 4 weeks at full benefit or 8 weeks linear decline).

Given the large number of scenarios and comparators, the results are summarised by stating the intervention that is cost-effective at a cost-effectiveness threshold of £20,000 per QALY under each scenario.

The expected value of perfect information (EVPI) was also calculated for the decision problem. The EVPI provides the upper bound on the value of reducing uncertainties in the decision problem. The EVPI associated with each instance of a decision was calculated using established methods.[202] It is important that the EVPI is calculated for the total population eligible to benefit from additional information.
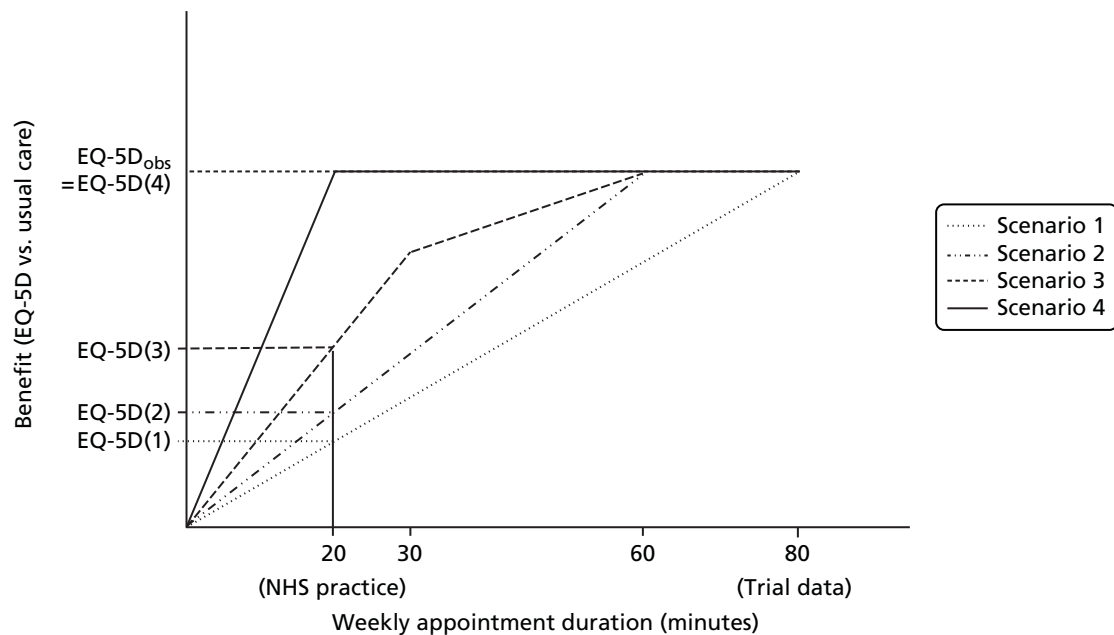
**FIGURE 21** Methods for treatment appointment duration sensitivity analysis.

This requires an assessment of the period over which information will be useful and the estimated incidence of the decision during this period.

Studies of new GP referrals for knee pain and of incident radiographic osteoarthritis suggest that among older patients (typically aged > 45 years or > 50 years) incidence is 2–4% per year.[158,295] The most recent data suggest that the incidence of consultations for knee pain in those aged > 50 years is 4% per year.[295] We estimate that 56% of consultations for knee pain are given a diagnosis of osteoarthritis.[296] Applying these data to UK population data on the number of people aged ≥ 50 years in the UK (22.773 million[297]) yields an estimate of 490,721 new consultations per year. Some data are available in the literature regarding the proportion of individuals with osteoarthritis of the knee (or with knee pain) who receive non-pharmacological treatments. Estimates of the proportion of patients receiving physiotherapy range from 13% to 41%.[296,298] One study reported the use of complementary therapy (21%)[296] and another reported the use of TENS (13%), acupuncture (8%), wedged insoles (6%) and appliances (5%).[298] Based on this information we estimate that approximately 50% of patients will be considered for adjunct non-pharmacological interventions. We assume that patients are considered for non-pharmacological interventions only once during their entire treatment for osteoarthritis of the knee; this may therefore underestimate the number of instances in which the decision of interest is faced.

We used the value of information analysis to compare three possible policy options that are available for cost-effective interventions:[299]

1. adopt the cost-effective intervention (adopt)
2. adopt the cost-effective intervention and commission research or 'adopt with research' (AWR)
3. delay adoption until new research is available or 'only in research' (OIR).

Given the nature of the uncertainties in the model we anticipate that any further research would take the form of a clinical study, which would take approximately 5 years to be planned, conducted and reported. We also assume that current practice is usual care. In reality there is likely to be some use of the appraised interventions in current practice. This assumption will not impact on the difference between the adopt and the AWR policies but will increase (reduce) the benefit of OIR relative to these policies if the net health benefit of current usage is positive (negative) relative to usual care.

We therefore consider the net health benefits associated with adoption (the population net health benefits generated by the cost-effective technology, which we assume would be used for 10 years by which point new interventions may be available and the management pathway for osteoarthritis of the knee may have changed dramatically), the health benefits associated with AWR (the health benefits generated by adoption for 5 years followed by a revised decision based on perfect information from year 6 to year 10) and the health benefits associated with OIR (the health benefits generated by no adoption from year 1 to year 5 and a decision based on perfect information from year 5 to year 10). The AWR and OIR estimates should be considered upper bounds as they are based on an assumption that research will resolve all uncertainties, which will not be the case for real research designs that reduce but do not remove uncertainty.

Adopt with research will always offer the highest health benefits if net health benefits are constant over time. There are two reasons why AWR may not, however, be the preferred policy choice. First, it may not be feasible, for instance if adoption removes incentives for health-care professionals to conduct research or patients to participate in research. Second, if there are high upfront costs OIR may offer higher health benefits as under this policy such upfront (or 'sunk') costs are incurred only if the intervention turns out to be cost-effective.

The main source of sunk costs is likely to be the cost of training physiotherapists to deliver acupuncture. Our estimates suggest that 245,360 patients would be eligible for treatment each year. Assuming that these patients receive eight sessions of treatment and that physiotherapists who deliver acupuncture typically deliver 208 sessions per year,[1] then 9437 physiotherapists would require training. We assume that training consists of an 80-hour course (40 hours taught) costing £495.[300] The opportunity cost of this course is unclear as many accredited courses run during weekends and require out-of-hours study. We therefore conservatively assume an opportunity cost of 40 hours of time (at £36 per hour) or £1440. The total cost of training would therefore be £18.3M, which we assume would be incurred in the first year of adoption. Spreading this across all treated patients over 10 years (or more if training provides benefits over a career) would have a minimal impact on the cost-effectiveness results. However, if this cost is incurred in the first year, this may increase the benefit of OIR over adoption or AWR, and is therefore included in our comparison of the three policies.

## Results

### Mapping and synthesis

*Appendix 5* presents the results of the mapping from individual HRQoL measures to the EQ-5D. *Figure 22* displays the results of the network meta-analysis of (1) all trials, (2) trials that were assessed to be at low risk of bias with respect to the adequacy of allocation concealment and (3) trials that were assessed to be at low risk of bias with respect to the adequacy of allocation concealment and which reported in the restricted 3- to 13-week range.

In the all-trials analysis, interferential therapy followed by acupuncture, TENS, PES and t'ai chi offered the largest benefit in terms of the EQ-5D, based on point estimates of the effect size alone. However, the CrI crossed zero for t'ai chi. The least uncertainty, as reflected by the narrowest CrIs, was associated with acupuncture and muscle-strengthening exercises. In the analysis of trials with a low risk of bias for allocation concealment, t'ai chi followed by acupuncture, interferential therapy, manual therapy and sham acupuncture offered the largest benefits, again based on point estimates of effect. However, the CrIs touched or crossed zero for t'ai chi, interferential therapy and manual therapy. When this analysis was restricted to trials reporting between week 3 and week 13, t'ai chi followed by manual therapy, acupuncture, interferential therapy and sham acupuncture offered the largest benefit. In this case the CrIs touched or crossed zero for t'ai chi, manual therapy and interferential therapy. The magnitude of the effect sizes differed between analyses. The effects of interferential therapy and TENS diminished when the analysis was restricted to trials with a low risk of bias with respect to allocation concealment. The effects of t'ai chi and sham acupuncture were stronger in the analyses restricted to trials with a low risk of bias with

(a)
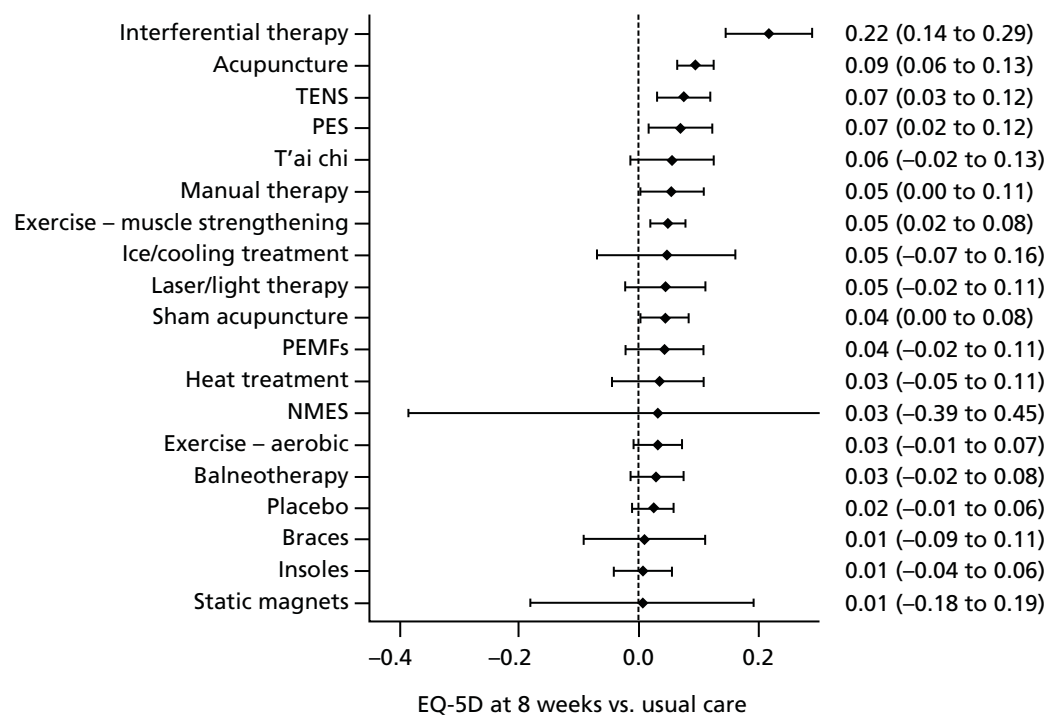


(b)
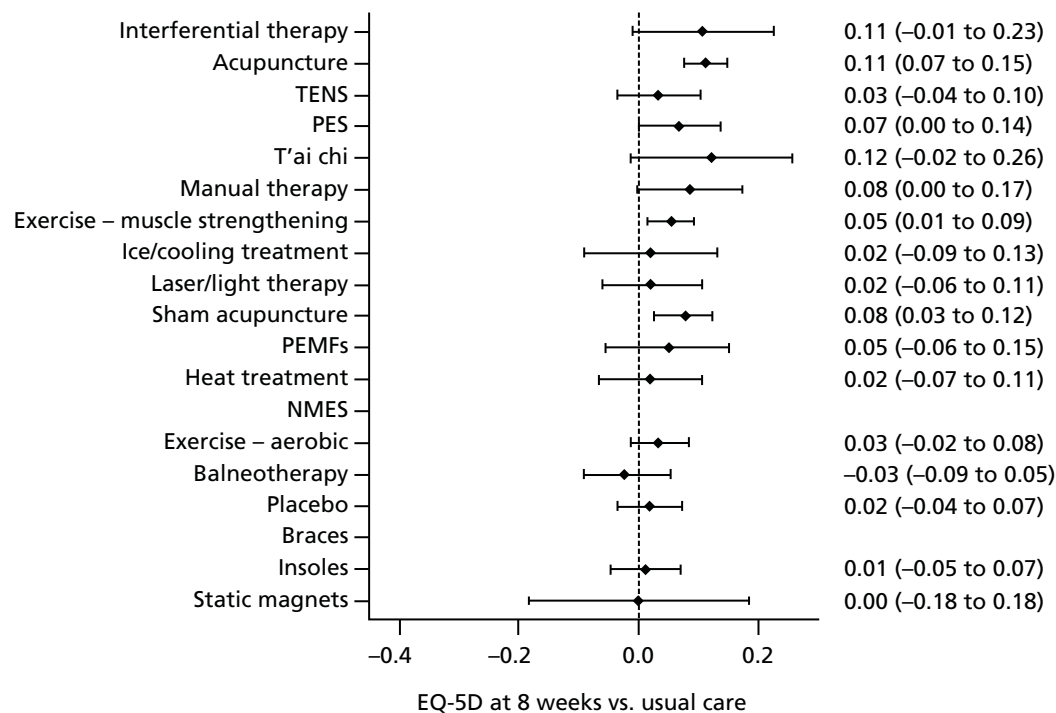


**FIGURE 22** Results of the network meta-analysis with 95% CrIs (results presented as values corresponding to 2.5%, 50% and 97.5% of the posterior distribution): (a) all trials; (b) trials with adequate allocation concealment; and (c) trials with adequate allocation concealment and reporting in the range 3–13 weeks. (*continued*)

(c)
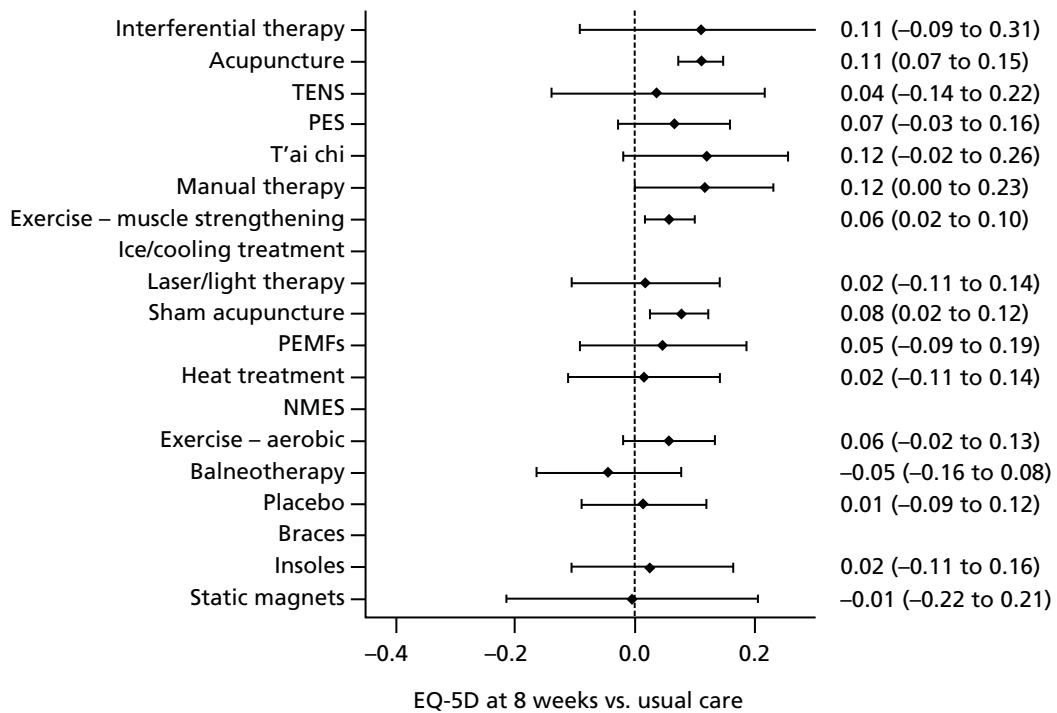


| | |
|---|---|
| Interferential therapy | 0.11 (−0.09 to 0.31) |
| Acupuncture | 0.11 (0.07 to 0.15) |
| TENS | 0.04 (−0.14 to 0.22) |
| PES | 0.07 (−0.03 to 0.16) |
| T'ai chi | 0.12 (−0.02 to 0.26) |
| Manual therapy | 0.12 (0.00 to 0.23) |
| Exercise – muscle strengthening | 0.06 (0.02 to 0.10) |
| Ice/cooling treatment | |
| Laser/light therapy | 0.02 (−0.11 to 0.14) |
| Sham acupuncture | 0.08 (0.02 to 0.12) |
| PEMFs | 0.05 (−0.09 to 0.19) |
| Heat treatment | 0.02 (−0.11 to 0.14) |
| NMES | |
| Exercise – aerobic | 0.06 (−0.02 to 0.13) |
| Balneotherapy | −0.05 (−0.16 to 0.08) |
| Placebo | 0.01 (−0.09 to 0.12) |
| Braces | |
| Insoles | 0.02 (−0.11 to 0.16) |
| Static magnets | −0.01 (−0.22 to 0.21) |

EQ-5D at 8 weeks vs. usual care

**FIGURE 22** Results of the network meta-analysis with 95% CrIs (results presented as values corresponding to 2.5%, 50% and 97.5% of the posterior distribution): (a) all trials; (b) trials with adequate allocation concealment; and (c) trials with adequate allocation concealment and reporting in the range 3–13 weeks.

respect to allocation concealment, as was the effect of manual therapy, particularly when the analysis was restricted to trials reporting in between 3 and 13 weeks.

The results are associated with a high degree of uncertainty for most but not all of the interventions as indicated by the wide CrIs around many of the treatment effect estimates. Muscle-strengthening exercise and acupuncture are associated with the least uncertainty, as they have the narrowest CrIs across analyses, and static magnets and NMES are associated with the most uncertainty. The treatment effects associated with TENS and interferential therapy are considerably more uncertain when the evidence is restricted to trials with adequate allocation concealment reporting in between 3 and 13 weeks.

Forest plots are unable to represent the full set of information regarding treatment effect uncertainty as they do not reflect the correlation in treatment effects. *Figure 23* therefore presents a 'rankogram'[181] showing the probability that each treatment ranks as the first, second, third etc. best of the treatments considered in each analysis. The majority of the rankograms are fairly flat, indicating high uncertainty with respect to where the treatment would rank. Exceptions to this are acupuncture, for which there is a peak at ranks 2–3 (depending on analysis) with the majority of the density covering ranks 1–5, and interferential therapy, which peaks at rank 1 in all analyses but which is subject to considerable uncertainty in the analyses restricted to trials with adequate allocation concealment. Static magnets is likely to be one of the worst treatments across analyses, with balneotherapy likely to be one of the worst treatments in the analyses restricted to trials with adequate allocation concealment. T'ai chi peaks at rank 1 in the analyses restricted to trials with adequate allocation concealment. For NMES, the rankogram displays a 'bucket' shape as the posterior density is spread over a very wide range and at the lowest and highest extremes it competes with few treatments to be the 'best' and 'worst' of those considered.

FIGURE 23 Results of the network meta-analysis (results presented as the probability that each treatment is ranked first, second, third etc.): (a) acupuncture; (b) balneotherapy; (c) braces; (d) insoles; (e) NMES; (f) placebo; (g) TENS; (h) usual care; (i) sham acupuncture; (j) muscle-strengthening exercise; (k) aerobic exercise; (l) heat treatment; (m) ice/cooling treatment; (n) inferential therapy; (o) laser/light therapy; (p) manual therapy; (q) PES; (r) PEMF; (s) static magnets; and (t) t'ai chi. The solid line represents the all-trials analysis; the small dots the analysis of trials with adequate allocation concealment and the long dashes the analysis of trials with adequate allocation concealment reporting at 3–13 weeks. (*continued*)

**FIGURE 23** Results of the network meta-analysis (results presented as the probability that each treatment is ranked first, second, third etc.): (a) acupuncture; (b) balneotherapy; (c) braces; (d) insoles; (e) NMES; (f) placebo; (g) TENS; (h) usual care; (i) sham acupuncture; (j) muscle-strengthening exercise; (k) aerobic exercise; (l) heat treatment; (m) ice/cooling treatment; (n) inferential therapy; (o) laser/light therapy; (p) manual therapy; (q) PES; (r) PEMF; (s) static magnets; and (t) t'ai chi. The solid line represents the all-trials analysis; the small dots the analysis of trials with adequate allocation concealment and the long dashes the analysis of trials with adequate allocation concealment reporting at 3–13 weeks.

The level of uncertainty regarding the effect of some comparators was very high, particularly for NMES and static magnets. The 95% CrIs cross zero for many interventions, with the exception of muscle-strengthening exercise and acupuncture in all three analyses; interferential therapy, PES and TENS in the all-trials analysis only; and sham acupuncture in both analyses involving a low risk of bias for allocation concealment.

Inclusion of age as a potential treatment effect modifier did not improve the model fit according to the residual deviance or DIC measures and had a very minimal impact on results. The results of this analysis are therefore not shown here. The model analysing the role of BMI as a treatment effect modifier did not converge, which was likely because of the high volume of missing data for BMI. This was not addressed through more advanced imputation approaches as the available data were unlikely to provide a useful basis for imputation (weight was rarely reported in the studies that did not report BMI).

Model fit statistics are provided in *Table 30*. The mean residual deviance suggests an adequate fit of each model to the data. Comparison of mean residual deviance and DIC statistics between the inconsistency and the consistency models for each analysis suggests no global evidence of inconsistency. However, inspection of individual data points' posterior mean deviance contributions suggests that the model is a poor fit for some data points. In the all-trials analysis, four data points had very high deviance contributions (3.4–6.9) in both the consistency and the inconsistency models. Two of the points were from trials comparing interferential therapy with placebo.[301,302] These studies are clearly inconsistent with each other. The difference in EQ-5D scores for interferential therapy compared with placebo is 0.07 (no variance measure reported) in the study by Adedoyin *et al.*[301] and 0.30–0.33 (95% CrI 0.19 to 0.42), depending on interferential therapy arm, in the study by Gundog *et al.*[302] These effects are also quite different from the indirect data on interferential therapy from Burch *et al.*[303] (interferential therapy vs. TENS 0.07, 95% CrI –0.01 to 0.16); thus, although all studies suggest that interferential therapy offers benefit, the magnitude of this benefit is inconsistent across studies and therefore more uncertain than the analysis of all trials suggests. No clear source of the uncertainty was identified, although all studies were assigned a high or unclear risk-of-bias score. One potential source of heterogeneity is the treatment intensity: patients in the Adedoyin *et al.*[301] study were delivered the equivalent of a 40-minute weekly dose, whereas patients in the Gundog *et al.*[302] study were delivered 100 minutes of treatment and patients in the Burch *et al.*[303] study were delivered 245 minutes of treatment (some of this period involved patterned muscle stimulation rather than interferential therapy). The two other points contributing high residual deviances were from a study comparing aerobic exercise with usual care.[189] This study is considered to be at high risk of bias and has been identified as an outlier in previous analyses.[168] The benefit from treatment estimated in this study may also be viewed as implausibly large (the difference in EQ-5D scores for aerobic exercise vs. usual care is 0.37, 95% CI 0.20 to 0.54). As aerobic exercise does not form part of a closed loop, any impact of this study will be limited to the aerobic exercise–usual care contrast. Residual deviance contributions for all other data points in the all-trials consistency model were < 2.3. For the other models, residual deviance contributions for all data points were < 2.5.

## Cost-effectiveness analysis

The results of the cost-effectiveness analysis are presented in *Table 31*. Intervention costs represented the vast majority of costs. Expected outcomes calculated using the posterior distributions of costs and effects are presented; uncertainty around these results is discussed in the remainder of the results section. In the analysis including all trials, TENS is the cost-effective intervention assuming a willingness-to-pay threshold in the range £20,000–30,000 per QALY gained, with an ICER of £2690 per QALY compared with usual care. In the analysis restricted to trials with adequate allocation concealment, acupuncture is the

**TABLE 30** Model diagnostics

| Statistic | Trials | | | | | |
| | **All** | | **With adequate allocation concealment** | | **With adequate allocation concealment, end point reporting 3–13 weeks** | |
| | **Consistency model** | **Inconsistency model** | **Consistency model** | **Inconsistency model** | **Consistency model** | **Inconsistency model** |
|---|---|---|---|---|---|---|
| Mean residual deviance (data points) | 206 (210) | 214 (210) | 84 (92) | 87 (92) | 69 (73) | 68 (73) |
| pD | 173 | 182 | 87 | 92 | 77 | 77 |
| DIC[a] | 12,872 | 12,891 | 12,814 | 12,821 | 12,787 | 12,787 |
| Random effect SD (95% CrI) | 0.03 (0.02 to 0.05) | 0.03 (0.01 to 0.05) | 0.02 (0.00 to 0.04) | 0.02 (0.00 to 0.04) | 0.02 (0.00 to 0.05) | 0.01 (0.00 to 0.04) |

pD, number of effective parameters.
a   Lower values indicate improved model fit.

**TABLE 31** Cost-effectiveness results

| Intervention | Trials | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | All | | | With adequate allocation concealment | | | With adequate allocation concealment, end point reporting 3–13 weeks | | |
| | Cost (£) | QALYs | ICER (£ per QALY) | Cost (£) | QALYs | ICER (£ per QALY) | Cost (£) | QALYs | ICER (£ per QALY) |
| Usual care | 0 | 0.000 | Referent | 0 | 0.000 | Referent | 0 | 0.000 | Referent |
| Static magnets | 5 | 0.001 | ED | 5 | 0.000 | Dominated | 5 | –0.001 | Dominated |
| Insoles | 13 | 0.001 | ED | 13 | 0.002 | ED | 14 | 0.004 | 3540 |
| TENS | 31 | 0.011 | 2690 | 30 | 0.005 | 6142 | 30 | 0.006 | 9750 |
| Braces | 40 | 0.001 | Dominated | NA | NA | NA | NA | NA | NA |
| Acupuncture | 179 | 0.014 | ED | 192 | 0.017 | 13,502 | 192 | 0.017 | 14,275 |
| Heat treatment | 297 | 0.005 | Dominated | 214 | 0.003 | Dominated | 213 | 0.002 | Dominated |
| Manual therapy | 304 | 0.008 | Dominated | 276 | 0.013 | Dominated | 277 | 0.018 | 86,964 |
| PES | 396 | 0.011 | Dominated | 410 | 0.010 | Dominated | 410 | 0.010 | Dominated |
| NMES | 481 | 0.005 | Dominated | NA | NA | NA | NA | NA | NA |
| Laser light therapy | 503 | 0.007 | Dominated | 288 | 0.003 | Dominated | 288 | 0.003 | Dominated |
| Interferential therapy | 770 | 0.033 | 33,866 | 1179 | 0.016 | Dominated | 1179 | 0.017 | Dominated |
| PEMF | 1453 | 0.007 | Dominated | 577 | 0.008 | Dominated | 577 | 0.007 | Dominated |

ED, extendedly dominated; NA, not available because of intervention not providing information to network.
**Note**
Shading differentiates the three groups of trials.

cost-effective intervention, with an ICER of £13,502 per QALY compared with TENS. When the analysis is restricted to trials with adequate allocation concealment and which reported in between 3 and 13 weeks, acupuncture is the cost-effective intervention, with an ICER of £14,275 per QALY compared with TENS.

Cost-effectiveness acceptability frontiers for the three primary analyses are presented in *Figure 24*. The probability that each intervention would be cost-effective at decision thresholds of £20,000 and £30,000 per QALY gained is presented in *Table 32*. These summaries provide the probability that each intervention is cost-effective in the context of a fully incremental analysis comparing all interventions. The probability is calculated as the proportion of simulations in which the intervention has the highest ICER (under a given threshold) or, equivalently, the highest incremental net benefit. At a cost-effectiveness threshold of £20,000 per QALY, the probability that TENS is cost-effective in the all-trials analysis is 49%. In the analysis of trials with adequate allocation concealment the probability that acupuncture is cost-effective is 47%. In the analysis of trials with adequate allocation concealment reporting between 3 and 13 weeks the probability that acupuncture is cost-effective is 25%.

## Sensitivity analyses
A summary of all sensitivity analyses that altered the optimal choice of intervention at a cost-effectiveness threshold of £20,000 per QALY is provided as *Table 33*. Detailed results of the sensitivity analyses are provided in *Appendix 5*.
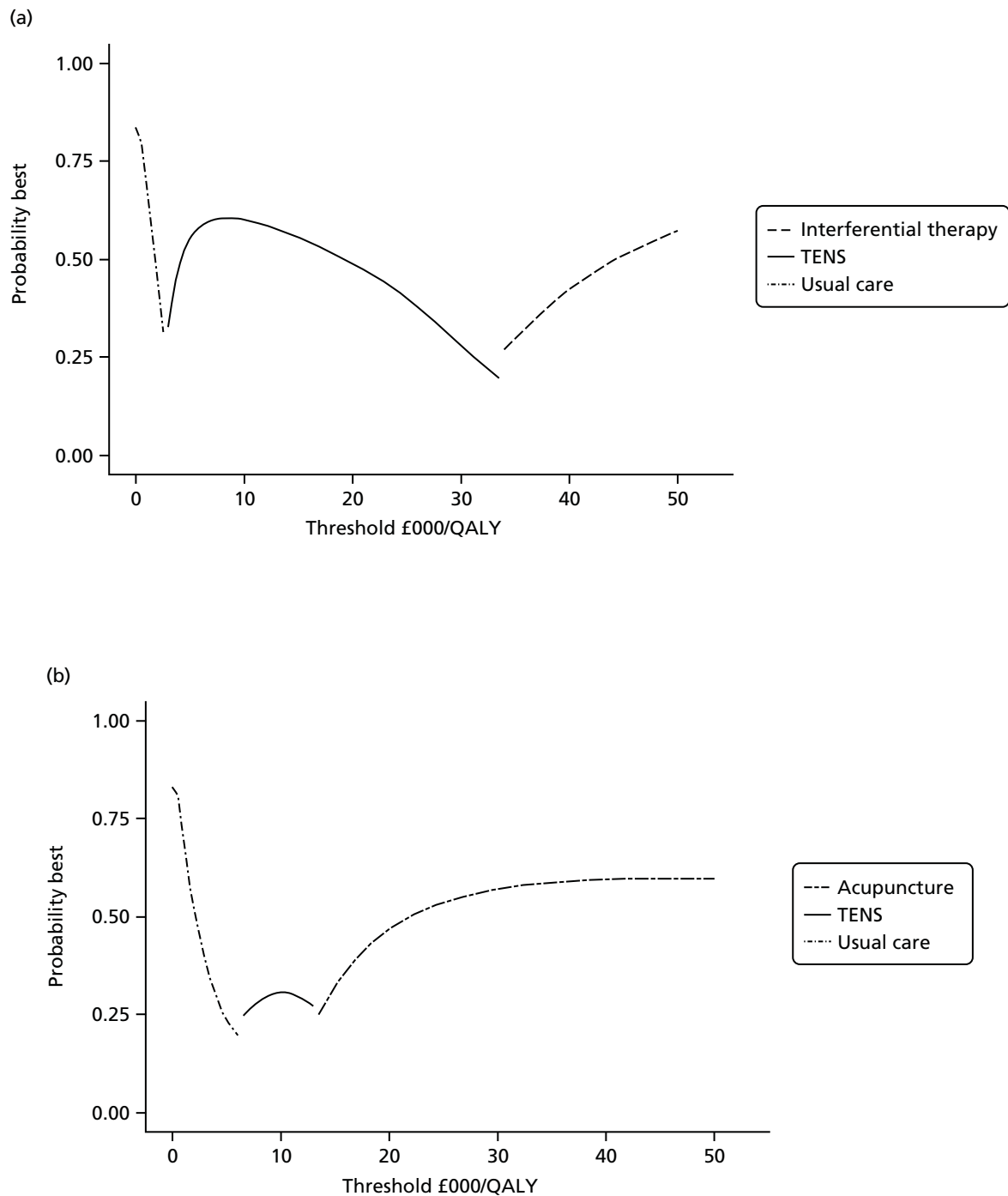
**FIGURE 24** Cost-effectiveness acceptability frontiers: (a) all trials; (b) trials with adequate allocation concealment; and (c) trials with adequate allocation concealment and reporting in the range 3–13 weeks. (*continued*)
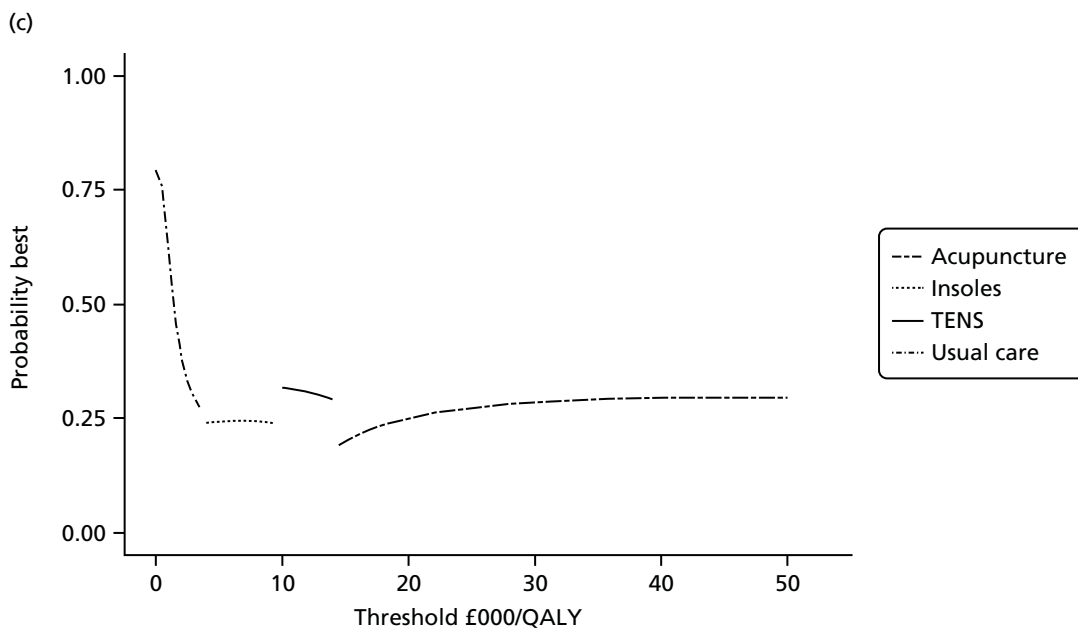
**FIGURE 24** Cost-effectiveness acceptability frontiers: (a) all trials; (b) trials with adequate allocation concealment; and (c) trials with adequate allocation concealment and reporting in the range 3–13 weeks.

**TABLE 32** Probability intervention is cost-effective at cost-effectiveness thresholds of £20,000 and £30,000 per QALY

| | Trials | | | | | |
| | All | | With adequate allocation concealment | | With adequate allocation concealment 3–13 weeks | |
| Intervention | £20,000 | £30,000 | £20,000 | £30,000 | £20,000 | £30,000 |
|---|---|---|---|---|---|---|
| Usual care | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Static magnets | 0.22 | 0.18 | 0.26 | 0.20 | 0.17 | 0.13 |
| Insoles | 0.00 | 0.00 | 0.04 | 0.02 | 0.13 | 0.09 |
| TENS | 0.49 | 0.28 | 0.15 | 0.08 | 0.25 | 0.20 |
| Braces | 0.06 | 0.05 | NA | NA | NA | NA |
| Acupuncture | 0.06 | 0.12 | 0.47 | 0.57 | 0.25 | 0.28 |
| Heat treatment | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.01 |
| Manual therapy | 0.00 | 0.00 | 0.07 | 0.12 | 0.20 | 0.28 |
| PES | 0.00 | 0.00 | 0. 00 | 0.00 | 0.00 | 0.00 |
| NMES | 0.16 | 0.21 | NA | NA | NA | NA |
| Laser light therapy | 0.00 | 0.00 | 0. 00 | 0.00 | 0.00 | 0.00 |
| Interferential therapy | 0.00 | 0.15 | 0.00 | 0.00 | 0.00 | 0.00 |
| PEMF | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

NA, not available because of intervention not providing information to network.
**Note**
Shading differentiates the three groups of trials.

**TABLE 33** Results of sensitivity analyses that altered the optimal choice of intervention

| Data set | Scenario | Cost-effective intervention at threshold of £20,000 per QALY[a] |
|---|---|---|
| All trials | Base case | TENS |
| | Shortest session duration used for acupuncture costing | Acupuncture |
| | Shortest session duration used for interferential therapy costing | Interferential therapy |
| | Shortened sessions – 75% of benefit in first 30 minutes, remainder by 1 hour | Interferential therapy |
| | Shortened sessions – all benefit achieved within 20–30 minutes | Interferential therapy |
| | Increase in duration of benefit for all interventions by 6 weeks | Interferential therapy |
| | Increase in duration of benefit of acupuncture by 31% | Acupuncture |
| | Increase in duration of benefit of interferential therapy by 45% | Interferential therapy |
| | Upper 95% CrI from NMA for acupuncture | Acupuncture (0.05) |
| | Upper 95% CrI from NMA for braces | Braces (0.08) |
| | Upper 95% CrI from NMA for NMES | NMES (0.18) |
| | Upper 95% CrI from NMA for static magnets | Static magnets (0.25) |
| | Lower 95% CrI from NMA for TENS | Acupuncture (0.15) |
| Trials with low risk of bias for allocation concealment | Base case | Acupuncture |
| | Use of private cost for acupuncture sessions | TENS |
| | Shortened sessions – all benefit achieved within 20–30 minutes | Interferential therapy |
| | Lower 95% CrI from NMA for acupuncture | TENS (0.05) |
| | Upper 95% CrI from NMA for insoles | Insoles (0.09) |
| | Upper 95% CrI from NMA for manual therapy | Manual therapy (0.10) |
| | Upper 95% CrI from NMA for static magnets | Static magnets (0.29) |
| | Upper 95% CrI from NMA for TENS | TENS (0.21) |
| Trials with low risk of bias for allocation concealment, end point reporting 3–13 weeks | Base case | Acupuncture |
| | Use of private cost for acupuncture sessions | Manual therapy |
| | Shortest session duration used for manual therapy costing | Manual therapy |
| | Shortened sessions – all benefit achieved within 20–30 minutes | Interferential therapy |
| | Increase in duration of benefit for all interventions of 27 weeks | Manual therapy |
| | Increase in duration of benefit of manual therapy by 18% | Manual therapy |
| | Lower 95% CrI from NMA for acupuncture | TENS (0.14) |
| | Upper 95% CrI from NMA for heat treatment | Heat treatment (0.35) |
| | Upper 95% CrI from NMA for insoles | Insoles (0.35) |
| | Upper 95% CrI from NMA for manual therapy | Manual therapy (0.35) |
| | Upper 95% CrI from NMA for static magnets | Static magnets (0.30) |
| | Upper 95% CrI from NMA for TENS | TENS (0.40) |

NMA, network meta-analysis.
a Values in parentheses represent the probability that the intervention is cost-effective at a cost-effectiveness threshold of £20,000 per QALY.

## Costs

Varying the costs of treatment as specified in *Cost-effectiveness analysis* altered the optimal decision at a decision threshold of £20,000 per QALY gained in only three of the analyses (out of the 70 conducted; see *Appendix 5* for full results). In the analysis of all trials, using the shortest duration of session time for acupuncture resulted in this intervention becoming cost-effective (ICER £19,033 per QALY gained) and using the shortest duration for interferential therapy resulted in this intervention becoming cost-effective (ICER £7626 per QALY gained). In the remainder of the all-trials analyses, TENS remained the cost-effective treatment, with all ICERs < £4000 per QALY gained. In the analysis restricted to trials with adequate allocation concealment, acupuncture remained cost-effective across all analyses, with ICERs ranging from £6000 to £18,000 per QALY gained. In the analysis restricted to trials with adequate allocation concealment and reporting within 3–13 weeks, using the shortest duration for manual therapy resulted in this intervention becoming cost-effective (ICER £9500 per QALY gained). In the other analyses acupuncture was cost-effective, with ICERs ranging from £7000 to £19,000 per QALY gained.

Assuming that acupuncture was delivered by private practitioners resulted in acupuncture no longer being cost-effective in the analysis restricted to trials with adequate allocation concealment and the analysis restricted to trials with adequate allocation concealment reporting within 3–13 weeks. In these analyses TENS and manual therapy became cost-effective, respectively.

As described in *Cost-effectiveness analysis* a series of scenarios explored shorter treatment sessions and a range of assumptions about the response of intervention effects to session duration. In the all-trials analysis, interferential therapy became cost-effective if we assumed that 75% of the benefit of treatment occurs in the first 30 minutes and the remainder of the benefit occurs between 30 minutes and 1 hour. In both analyses restricted to trials with adequate allocation concealment, the preferred treatment choice switched only under the extreme scenario whereby all of the benefit observed within the trials is achieved within the first 20 minutes (or 30 minutes in the case of acupuncture or manual therapy).

## Effectiveness

The optimal choice of intervention at a cost-effectiveness threshold of £20,000 per QALY gained using 95% CrI values was altered in our sensitivity analyses for the following interventions: acupuncture, braces, heat treatment, insoles, manual therapy, NMES, static magnets and TENS. The values required to alter decisions were generally at the extremes of the posterior distributions (see *Table 33*), although there were exceptions to this.

In the analysis of all trials the probability that the treatment effect for TENS would be sufficiently small for acupuncture to become cost-effective was 0.15. In the analyses of trials with adequate allocation concealment, the probabilities that the treatment effect for TENS would be sufficiently large for TENS to become cost-effective were 0.21 and 0.40 (analysis restricted to trials reporting between 3 and 13 weeks). Conversely, in the analysis of trials with adequate allocation concealment reporting in the 3- to 13-week period the probability that the treatment effect for acupuncture would be sufficiently small for TENS to become cost-effective was 0.14. In the analyses of trials with adequate allocation concealment, the probabilities that the treatment effect for manual therapy would be sufficiently large for manual therapy to become cost-effective were 0.10 and 0.35 (analysis restricted to trials reporting between 3 and 13 weeks). In the analysis of trials with adequate allocation concealment reporting in the 3- to 13-week period, insoles and heat treatment had a probability of 0.35 of being sufficiently effective to become the cost-effective treatment. Across analyses, the probability of observing a treatment effect for static magnets sufficiently large for it to become cost-effective was 0.25–0.30, whereas for NMES this probability was 0.18. In both cases this reflected the high levels of uncertainty around the efficacy of these comparators.

In the analysis of all trials the extension of benefit would have to be 6 weeks for interferential therapy to become cost-effective. In the analysis restricted to trials with adequate allocation concealment no extension of the duration of benefit impacted on the decision. In the analysis restricted to trials with adequate allocation concealment and which reported between 3 and 13 weeks, the extension of benefit

would have to be 27 weeks for manual therapy to become cost-effective. Extending the duration of benefit of each intervention individually (by up to 50%) altered the decision in only three cases. In the all-trials analysis, increasing the benefit of acupuncture and interferential therapy by 50% resulted in these interventions becoming cost-effective. In the analysis of trials with adequate allocation concealment reporting in the 3- to 13-week period, increasing the benefit of manual therapy resulted in this intervention becoming cost-effective. Further threshold analyses indicated that the extension of benefit required to alter these decisions was 31%, 45% and 18%, respectively.

### Value of further information

Assuming a decision threshold of £20,000 per QALY, the per-decision EVPI for the all-trials scenario is £113, for the analysis of trials with adequate allocation concealment is £74 and for the trials with adequate allocation concealment reporting in the 3- to 13-week period is £164. The EVPI for the total population eligible to benefit from improved decision-making is estimated as £239M, £156M and £346M, respectively, over a 10-year period.

*Figure 25* shows the cumulative net health benefits (in QALYs) generated under the policies of adoption, AWR and OIR. This shows that AWR is always the preferred policy. If AWR was not available, adoption is the preferred option in the all-trials analysis and the analysis of trials with adequate allocation concealment. In the analysis restricted to trials with adequate allocation concealment reporting between 3 and 13 weeks, there is higher uncertainty and the adoption and OIR policies deliver very similar health benefits. Given that this analysis reflects the maximum value of further research and that the real value will be lower (as research will not deliver perfect information and is associated with research costs), adoption is the preferred policy.

## Discussion

### Principal findings

Health-care decision-makers, aiming to maximise population health, require evidence of the costs and effects of competing treatment strategies. In many cases the available information is not in a format that readily facilitates comparisons of costs and effects. RCTs rarely compare all of the treatments of interest and rarely provide sufficiently long or detailed follow-up for all costs and outcomes of treatment to be observed. In addition, evidence may need to be generalised across different settings (e.g. countries, disease subpopulations). In the case of complementary therapies and other non-pharmaceutical interventions for which there are more limited regulatory processes, there is an increased potential for heterogeneous conduct of trials and data availability, especially when combined with a focus on subjective outcome measures. However, regardless of the limitations and uncertainties in the evidence base, policy-makers must make decisions.

The National Institute for Health and Care Excellence has recommended acupuncture for the treatment of chronic headache and migraine,[80] and for musculoskeletal pain,[81] but not in the context of chronic pain associated with osteoarthritis of the knee.[82] This decision in part reflected concerns regarding the available evidence. The current study was commissioned as part of a programme intended to improve evidence around the costs and effects of acupuncture. The work presented in this chapter builds on work presented in *Chapter 3* to systematically identify and synthesise outcome data on a wide range of interventions for osteoarthritis of the knee. The work extends the synthesis in *Chapter 3* to the EQ-5D end point, allowing calculation of QALYs and a cost-effectiveness analysis. The available trials were of varying quality, reporting distinct outcomes at different time points. IPD were available for a minority of studies with the rest providing only summary statistics. Novel Bayesian network meta-analysis models were developed to synthesise continuous outcome data reported at both the individual patient level and the aggregate level. Mapping of the available HRQoL data to EQ-5D preference weights was conducted to produce a common statistic for synthesis and one that could inform the cost-effectiveness analysis. The simulation outputs from the network meta-analysis informed a decision-analytic cost-effectiveness model that compared the costs and QALYs of the competing interventions, and was used to reflect the degree of uncertainty

(a)



(b)



(c)
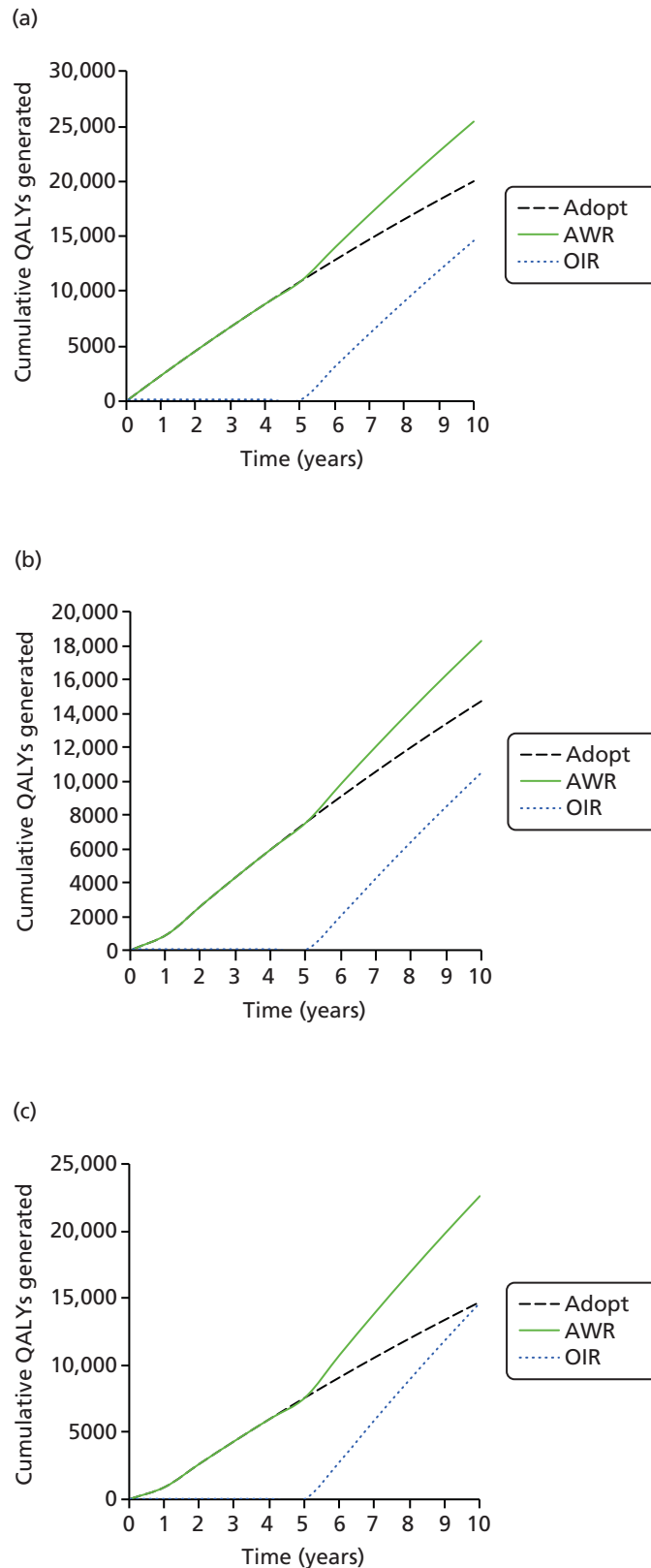


**FIGURE 25** Cumulative health generated under the adoption, AWR and OIR policies: (a) all trials; (b) trials with adequate allocation concealment; and (c) trials with adequate allocation concealment reporting at 3–13 weeks.
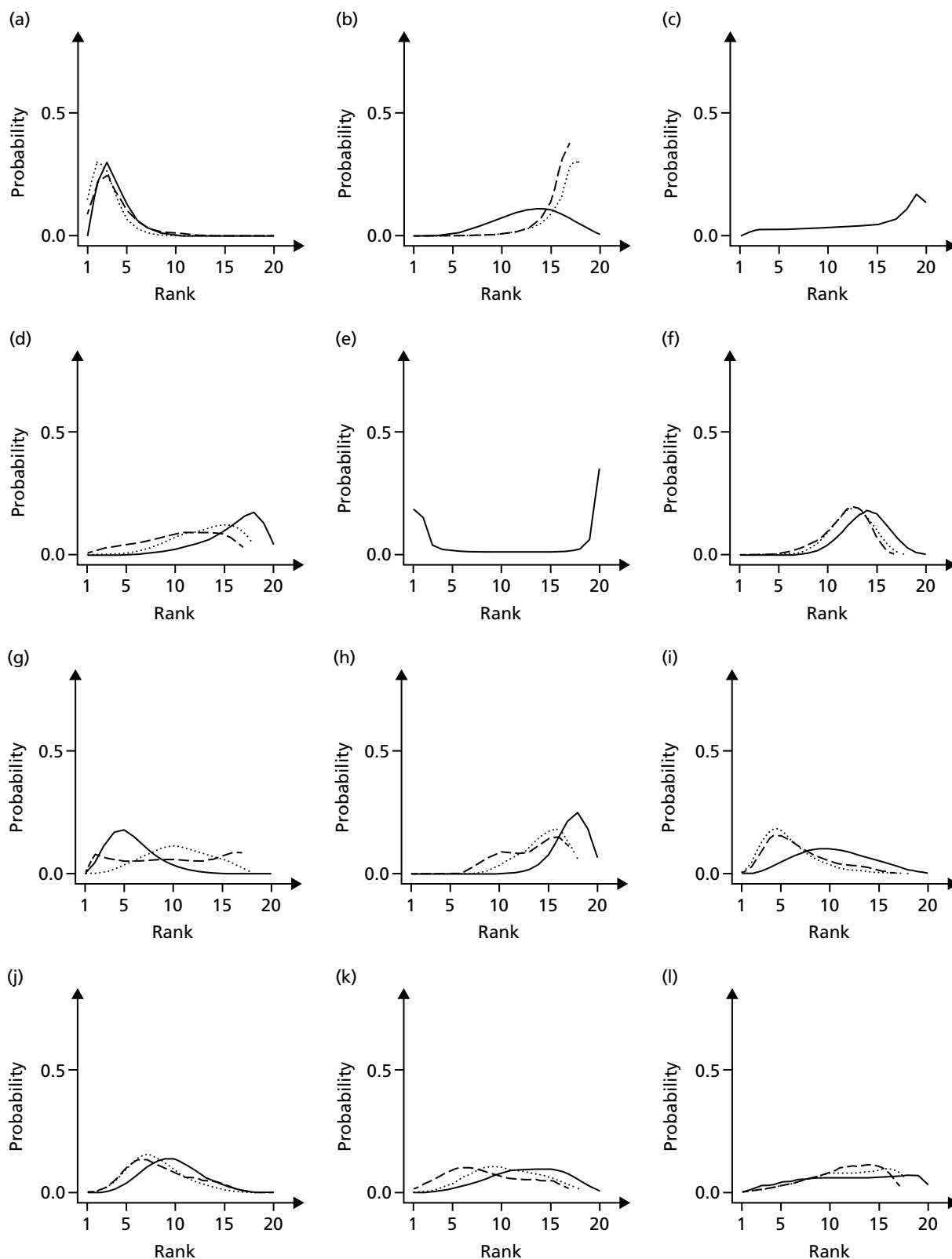
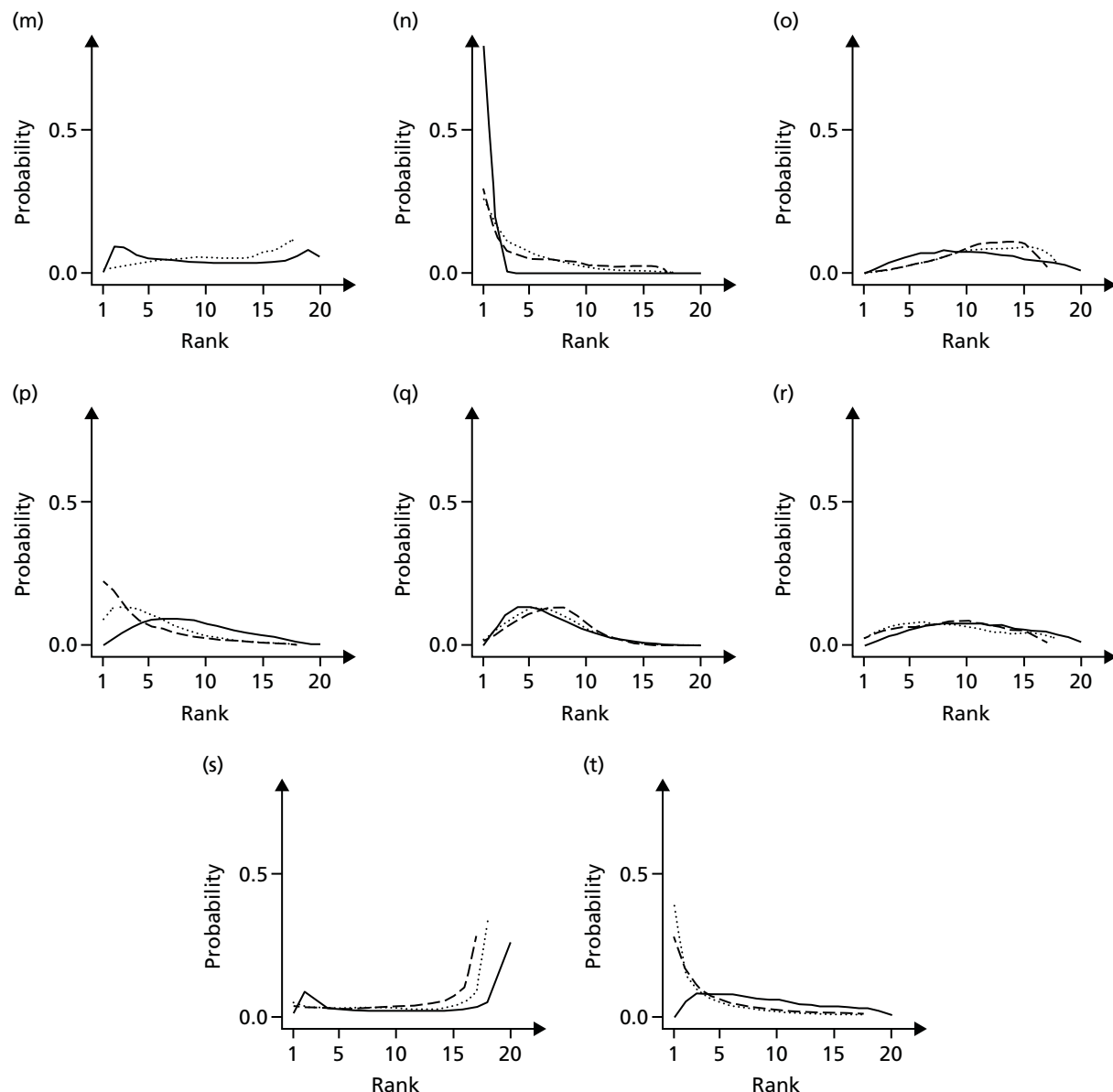surrounding the optimal treatment decision under different assumptions concerning quality of the evidence, treatment effectiveness, duration of treatment effectiveness and costs.

Network meta-analyses were conducted for three data sets: (1) all trials, (2) trials with a low risk of bias for allocation concealment and (3) trials with a low risk of bias for allocation concealment and reporting within the 3- to 13-week window. In the analysis of all trials, interferential therapy was the most effective treatment based on the point estimate of effect, followed by acupuncture, TENS, PES and t'ai chi. These results are broadly consistent with the work conducted in *Chapter 3* on the standardised pain outcome, which found that the least uncertainty, as reflected by the narrowest CrIs, is associated with acupuncture and muscle-strengthening exercises. Although this analysis of all trials suggested a high probability that interferential therapy was the most effective treatment, analyses of model fit suggested that the model estimates were inconsistent with the observed trial-level treatment effect estimates because of large differences in the effects observed in different studies. No obvious sources of heterogeneity were observed, although differences in treatment intensity may have played a role. Results from this analysis should therefore be interpreted with caution.

The analysis including trials with a low risk of bias for allocation concealment found that t'ai chi was the most effective treatment based on the point estimate of effect, followed by acupuncture, interferential therapy, manual therapy and sham acupuncture. The analysis including trials with a low risk of bias for allocation concealment and reporting within the 3- to 13-week window found that t'ai chi was the most effective treatment based on the point estimate of effect, followed by manual therapy, acupuncture, inferential therapy and sham acupuncture. The analyses suggested an incremental effect of acupuncture over usual care that ranged from 0.09 (95% CrI 0.06 to 0.13) to 0.11 (95% CrI 0.07 to 0.15), depending on the analysis. These results for acupuncture are similar to those reported for osteoarthritis of the knee in *Chapter 4* (0.08, 95% CrI 0.05 to 0.12). The level of uncertainty regarding the treatment effect of some comparators was very high. By contrast, in both the analyses of higher-quality trials, acupuncture and muscle-strengthening exercises stand out as being associated with the narrowest CrIs and therefore the least uncertainty. The 95% CrIs cross zero for all interventions with the exception of muscle-strengthening exercise and acupuncture in all three analyses; for interferential therapy, PES and TENS in the all-trials analysis only; and for sham acupuncture in the two analyses involving trials with a low risk of bias for allocation concealment.

The cost-effectiveness results varied according to the data set considered relevant for decision-making. The analysis of all trials found that TENS was cost-effective (ICER £2690 per QALY vs. usual care) whereas analyses (2) and (3) found that acupuncture was cost-effective (ICER £13,502–14,275 per QALY vs. TENS) as an adjunct treatment for knee osteoarthritis in the UK NHS. The difference between these results is attributable to the reduced effect of TENS in analyses (2) and (3), in which trials with a high risk of bias were excluded.

The sensitivity analyses conducted suggested that the optimal intervention with respect to cost-effectiveness was uncertain. In analysis (1), plausible reductions in the cost of acupuncture, reductions in the efficacy of TENS and increases in the duration of benefit of acupuncture resulted in acupuncture becoming the cost-effective treatment. Using the shortest duration of interferential therapy sessions observed in the studies included in the synthesis resulted in this intervention becoming cost-effective, as was the case in analyses assuming that the duration of interferential therapy sessions could be reduced without proportionate reductions in efficacy. In analyses (2) and (3) plausible increases in the efficacy of TENS resulted in TENS becoming the cost-effective treatment and plausible increases in the efficacy of manual therapy resulted in this comparator becoming cost-effective. In analysis (3) using the shortest duration of manual therapy sessions observed in the studies included in the synthesis resulted in this comparator becoming cost-effective, as did carrying out plausible increases to the duration of benefit of manual therapy. Analysis (3) was particularly uncertain and use of the upper and lower CrIs from the network meta-analysis altered the cost-effective intervention in six instances.

There is considerable uncertainty in terms of both the effects of many of the interventions and the probability that each intervention is cost-effective. The latter reflects both uncertainty in the effect of

treatment and the large number of interventions appraised. Despite these uncertainties the value of information analysis suggested that adoption (of TENS or acupuncture, depending on the analysis) will generate more health to the NHS than delaying adoption until further research is conducted. There may be value to conducting further research alongside adoption if this is feasible and possible avenues for this research are discussed later in this section.

The conclusions of the current work are based on a cost-effectiveness threshold of £20,000–30,000 per QALY as this has been used historically by NICE. A cost-effectiveness threshold is used to assess if the health benefits offered by an intervention are greater than the health likely to be lost because the additional resources required are not available to fund other effective treatments in the NHS. Research conducted during the course of this study suggests that cost-effectiveness thresholds of £20,000–30,000 per QALY may be too high as it was estimated that £13,000 of NHS resources adds 1 QALY to NHS patients.[251] In the context of this study, using this lower estimate of the cost-effectiveness threshold would result in TENS being the cost-effective choice in the base-case analyses.

### Strengths and limitations

This study has a number of strengths and limitations. The study methods and the presentation of the results are consistent with the requirements of funding agencies, which need to ensure efficient allocation of resources across a wide range of programmes and diseases. The analysis was based on evidence identified by the previous systematic literature review presented in *Chapter 3*, all decisions with respect to inclusion/exclusion of studies and evidence are made transparent and the assumptions required to make comparisons across therapies are made explicit. The analysis was able to include a large number of studies and interventions despite a lack of head-to-head trial data, access to IPD and many studies not collecting (or reporting) the outcome of interest (EQ-5D preference weights).

The synthesis of heterogeneous outcomes relied on imperfect mappings, which are typically able to explain only a minority of variation in EQ-5D scores. The magnitude of bias introduced by using mapping functions (and different mapping functions across trials) is unknowable. The availability of key outcomes across trials would have reduced these concerns, as would the collection of generic preference-based measures of HRQoL in all trials. Mapping is always a second-best approach compared with directly collecting data on preference-based measures, such as the EQ-5D, as it leads to increased uncertainty and error around HRQoL estimates.[304] A 'core outcome set' for osteoarthritis is available. This recommended that future Phase III trials of knee, hip and hand osteoarthritis should evaluate the following domains: pain, physical function, patient global assessment and, for studies of ≥ 1 year, joint imaging.[252] Recommendations that go beyond Phase III regulatory trials and which define the instruments that should be used to measure outcomes in these domains are warranted to foster consistency.

Our analysis accounted for the limited explanatory power of the mapping algorithms; however, it did not explicitly account for measurement error around the HRQoL measures or the EQ-5D. None of the mapping algorithms used reflected measurement error and methods to reflect measurement error in this context have only recently been piloted.[305] Future research should explore how measurement error can be adequately reflected in mapping algorithms and synthesis of multiple outcomes (some work on this has begun – see, for example, Lu *et al.*[306]).

Individual patient data were available only for a minority of studies. Means and variances for the EQ-5D data mapped from aggregate data on HRQoL were required for the synthesis. The HRQoL instrument dimension scores were therefore assumed to be distributed according to a multivariate normal distribution. This may not have been a good approximation in some cases, thus increasing uncertainty regarding the resulting mean or variance estimates. In addition, samples from the HRQoL instruments that were outside the feasible range for each instrument were truncated at the minimum or maximum value. This will have altered the means and variances from those reported in the publications. Checks were performed to identify any trials for which the direction of treatment effect on utility was not as would have been expected using the mean values of the HRQoL instrument. This found discrepancies for five of the 88 trials.

The relationship between baseline and final outcomes observed in the IPD studies was assumed to generalise to the aggregate data studies. This allowed potential baseline imbalances in studies for which only summary data were available to be adjusted for. This assumption may not have been appropriate in all cases, particularly when follow-up time points in the aggregate data studies were particularly short or long. Missing data from the IPD studies were excluded from the analysis, which may have biased the results. Development of methods to account for missing data in the context of IPD and aggregate data network meta-analysis is warranted.

Outcome data closest to 8 weeks' follow-up were selected for synthesis. The synthesis therefore required the assumption that, in the many trials not reporting outcomes at 8 weeks, the available data are reflective of the 8-week time point. Sensitivity analyses around this assumption suggested that the results were not particularly sensitive to the inclusion of studies with very short or very long follow-up periods, with the exception of manual therapy. For the cost-effectiveness analysis, the HRQoL effects observed at 8 weeks were applied from 0 to 8 weeks to generate QALY estimates. Sensitivity analyses were conducted to assess the impact of extending the duration of benefit of the interventions. These analyses found that the model was sensitive to the duration of benefit of acupuncture, interferential therapy and manual therapy. Further work to understand the long-term effects of therapies would be of value.

All usual care interventions were assumed to be equivalent in the analysis, as were all sham acupuncture interventions and all other placebo interventions. Evidence from work recently conducted by the ATC suggests that the effect of sham acupuncture may depend on whether it involves the use of penetrating or non-penetrating needles and that the effect of usual care may depend on whether or not usual care is delivered according to a pre-specified protocol.[45] It seems plausible that other non-acupuncture sham procedures may also vary in their effects. Exploration of a network including more refined comparator definitions may therefore be of value.

There was substantial variation in the duration of therapy course, frequency of sessions and duration of sessions administered. Further research should explore the impact of these attributes of intervention delivery on outcomes and cost-effectiveness.

The studies analysed here are from a range of countries, which may differ in terms of the method by which and intensity with which interventions are administered. In addition, differences in the nature of health care for chronic pain more generally could have impacted on outcomes.

The analysis of non-intervention resource use assumed that only primary care and specialist visits are impacted on by changes in outcomes following interventions and that the impact of treatment on resource use can be captured through changes in EQ-5D scores. It is possible that this did not capture the full impact of treatment on resource use. Intervention costs were derived using a combination of expert opinion on required resource use and trial data on the intensity with which interventions were administered. Empirical resource use data for each intervention would have increased the robustness of the intervention costs. In addition, the costing involved a number of strong assumptions. Insoles, braces and static magnets were assumed to deliver the benefits estimated from the network meta-analysis for their usable lifespan. Costs of equipment used by physiotherapists were not included in the analysis as per-use costs were expected to be small. Inclusion of these costs is unlikely to change the conclusions of the analysis as equipment costs for acupuncture and manual therapy are likely to be small and many musculoskeletal outpatient physiotherapy departments are likely to already have interferential therapy devices.[307] We assumed that acupuncture would be administered by physiotherapists. Survey data suggest that a non-negligible proportion of acupuncture sessions administered within the NHS are administered by medical doctors. We did not evaluate this as, unless a marked difference in outcomes was expected between delivery by a physiotherapist and delivery by a medical doctor, it seemed unlikely that provision by medical doctors would be cost-effective given the large difference in hourly costs. Survey data also suggest that independent acupuncturists currently deliver the majority of acupuncture sessions per annum in the UK, primarily for musculoskeletal pain.[1] However, a large majority of this is delivered outside of the NHS.

The analysis presented here evaluated adjunct treatments assuming that their benefit was independent of any concomitant core treatments. Further work could explore interactions between core and adjunct therapies. The analysis also assumes that health-care professionals are faced with a single opportunity to select treatment for an individual, whereas in reality individuals with chronic pain are likely to be treated with a series of interventions. Further work could explore the costs and effects of different sequences of interventions.

### Recommendations for future research

The EVPI in this area is relatively high, suggesting that additional research may be cost-effective. Further analysis is required to identify the most cost-effective and clinically appropriate specifications for further research. Such analyses could explore the cost-effectiveness of research to address the following key areas of uncertainty:

- The impact of the exact attributes of the treatment approach on HRQoL outcomes. This could include exploring variation in attributes such as the treatment protocol, overall duration of the course of therapy and frequency and duration of sessions.
- The time profile of HRQoL benefits of interventions both within the treatment period and beyond the treatment period.

The value of new clinical research to inform these areas of uncertainty should ideally be assessed by extending the existing model to include data (and when necessary expert opinion) that quantify our current understanding of these uncertain quantities. This would allow the uncertainties that are most important in determining cost-effectiveness to be identified. Clinical research to reduce these key uncertainties could take a number of forms including prospective trials or observational studies.

More generally, there would be value in developing a set of recommendations defining which HRQoL instruments should be used when evaluating therapies aimed at improving the HRQoL of knee osteoarthritis patients. Development of these recommendations would help to ensure that appropriate instruments are used in future trials and foster consistency in outcomes collected across trials. This in turn would improve the reliability of future meta-analyses and network meta-analyses.

## Conclusion

When all trials are included in the network meta-analysis irrespective of quality, TENS is the cost-effective non-pharmacological adjunct treatment for patients with chronic pain related to osteoarthritis of the knee. However, the effect of TENS may be exaggerated because of biases associated with poor trial conduct. When the network meta-analysis is restricted to trials with adequate allocation concealment, an attribute known to be associated with treatment effect magnitude in osteoarthritis, the effect of TENS diminishes and is associated with wide CRIs that include zero. In this analysis acupuncture becomes cost-effective.

# Chapter 6 Acupuncture, Counselling or Usual Care for Depression (ACUDep): a randomised controlled trial

## Introduction

A substantial proportion of the global disease burden involves depression,[308] which is a major cause of suicide. Depression involves more than just everyday mood fluctuations; it also involves feelings of severe sadness, anxiety, hopelessness and worthlessness. Those affected individuals lose interest in the activities that they used to enjoy and often have physical symptoms such as chronic pain, fatigue and insomnia.

The front-line treatment for depression in primary care usually involves antidepressants; however, they do not work well for more than half of patients.[95] Moreover, the effectiveness of newer antidepressants has recently come into question for patients with mild to moderate depression.[309] Non-pharmacological treatment options for depression are of interest to many patients, in part because of concerns about dependency on antidepressants.[97]

Among non-pharmacological treatments, acupuncture and counselling show some promise in the treatment of depression, yet further evidence is needed. A Cochrane review of acupuncture for depression was inconclusive on whether or not acupuncture is an effective intervention for depression, in part because of the high risk of bias in the majority of studies.[310] Counselling is a widely used intervention for patients with depression and is provided in approximately half of the 9000 primary care practices in England.[311] A Cochrane review of counselling for mental health and psychosocial problems in primary care found short-term but not long-term benefits from counselling.[312] Both Cochrane reviews recommended extending the evidence base to include comparisons not just with usual care but also with other interventions, as this would have the potential to increase patient choice.

The aims of this study were to determine the clinical effectiveness and cost-effectiveness of short courses of either acupuncture or counselling compared with usual care for patients with depression. We also planned to compare acupuncture and counselling on the basis that there would be structural equivalence regarding time and attention. Additional exploratory studies nested within the trial focused on (1) the experience of treatment from the patient perspective; (2) the impact of comorbid pain on depression outcomes; (3) the approaches that practitioners use to enhance longer-term benefits; and (4) a health-economic analysis to investigate HRQoL and the costs of these treatments, and understand whether or not they should be considered a good use of limited health resources.

## Methods

### Design

Patients were randomised by the York Trials Unit to one of three arms in a pragmatic RCT (registration number ISRCTN63787732), as detailed in a published protocol.[313] An allocation ratio of 2 : 2 : 1 was used for the randomised groups: acupuncture plus usual care, counselling plus usual care and usual care alone, respectively. The York Trials Unit ensured that the allocation was securely concealed from the researchers who subsequently informed patients of their allocation. The unit recorded patient details prior to using Structured Query Language software for computer-generated block randomisation, with block sizes of five and 10. Randomisation was conducted by an investigator with no clinical involvement in the trial. The York NHS Research Ethics Committee (reference number 09/H1311/75) provided ethical approval on 21 September 2009.

### Population

Patients were recruited from a primary care population who had consulted with depression within the past 5 years, who were continuing to experience depression and who were aged ≥ 18 years. We identified potential participants from general medical practice databases. Signed consent forms and baseline questionnaires were returned by patients. Patients were not eligible if they were receiving acupuncture or counselling at the time, had a terminal illness, significant learning disabilities, haemophilia, hepatitis or human immunodeficiency virus infection, were pregnant or had confounding psychiatric conditions (bipolar disorder, post-partum depression, adjustment disorder, psychosis, dementia or personality disorder). Patients who had suffered a close personal bereavement or given birth during the previous 12 months were also excluded. Spoken English was a requirement. Patients were eligible if they scored ≥ 20 on the Beck Depression Inventory-II (BDI-11),[314] which is classified within this scale as 'moderate' or 'severe' depression.

### Interventions

Up to 12 sessions were offered, usually on a weekly basis, to patients allocated to the acupuncture or counselling groups. Practitioners providing acupuncture were registered with the BAcC, with at least 3 years of post-qualification experience. In consultation with participating acupuncturists, an acupuncture treatment protocol was developed, which allowed for customised treatments within a standardised theory-driven framework.[315] Practitioners providing counselling were members of the British Association for Counselling and Psychotherapy with accreditation or who were eligible for accreditation having completed 400 supervised hours post qualification. Counselling competences using a humanistic approach drawn from those independently developed for Skills for Health[316] were incorporated into a manualised protocol. All practitioners recorded in logbooks the number and length of sessions, treatment provided and adverse events. Further details of the two interventions are presented in *Appendix 6*. In all three patient groups, usual care, both NHS and private, was available according to need and was monitored for all for the purposes of comparison.

### Outcome measures

The Patient Health Questionnaire-9 items (PHQ-9)[317] at 3 months was the primary outcome measure. We also evaluated the overall impact over the 12-month period. PHQ-9 scores range from 0 to 27, with depression considered mild (5–9), moderate (10–14), moderately severe (15–19) or severe (≥ 20). As a preference-based measure of health outcome, we used the EQ-5D.[226] Medication use was ascertained by asking patients if they had taken any prescribed medication for depression or any prescribed analgesics/painkillers. Patients reported health service use, including the number of times that they had consulted a health professional because of their depression. Patients also reported out-of-pocket spending on acupuncture, counselling or psychotherapy, including cognitive–behavioural therapy. Data were collected at baseline and by postal questionnaire at 3, 6, 9 and 12 months. Because these data were self-reported by patients independently of the research team, we minimised potential bias associated with unblinded researchers or clinicians measuring outcomes. We also collected demographic data and patients' prior preferences and expectations of the interventions at baseline and BDI-II data at baseline and 12 months. We measured patients' perceptions of their practitioners' empathy at 3 months using the Consultational and Relational Empathy (CARE) measure[14] in both the acupuncture group and the counselling group. A payment of £5 as reimbursement was enclosed with the final questionnaire to enhance the response rate.[15]

### Sample size

We sought an effect size of 0.39 on the PHQ-9 when comparing either acupuncture with usual care alone or counselling with usual care alone, an effect size that was taken to be both moderate and realistic. We used the allocation ratio of 2 : 2 : 1 to increase the power to detect statistically significant differences between acupuncture and counselling. A smaller effect size of 0.32 was sought when comparing acupuncture and counselling, although the anticipated difference between these two treatments was expected to be small and not clinically meaningful. With a two-sided significance level of 5% and 90% power the required group sizes were 204, 204 and 102 in the acupuncture, counselling and usual care

alone arms, respectively. The total sample size required was 640 (i.e. groups of 256, 256 and 128 respectively) when allowing for 20% attrition.

### Analysis

Primary comparisons were between acupuncture plus usual care and usual care alone, and between counselling plus usual care and usual care alone. A secondary comparison was between acupuncture plus usual care and counselling plus usual care. The PHQ-9 was the primary outcome measure and 3 months was the primary end point, and we used ANCOVA, with baseline PHQ-9 score as covariate. For missing data we used multiple imputation by chained regression using treatment group, baseline measures (PHQ-9, BDI-II, SF-36, EQ-5D anxiety/depression) and demographics (age and sex). We used imputed rather than raw data for the primary analysis to take account of the profile of non-responders.

In a secondary analysis we assessed the overall clinical impact over 12 months, using the AUC method for the PHQ-9 by linear regression, predicting the average AUC while controlling for baseline PHQ-9. We also explored in more detail the PHQ-9 outcomes across all time points using random intercept linear mixed models with fixed effects for treatment arm, time and arm–time interaction for each treatment comparison. The models nested time points within patients and controlled for baseline PHQ-9 score and potential mediators including patients' prior expectations and preferences regarding the treatments. Potential mediators were identified by univariate regressions ($p < 0.1$) of the PHQ-9 at 3 months for the whole patient sample, controlling for baseline PHQ-9 score. We controlled for treatment time (combined length of sessions) and quality of attention (CARE score) when comparing acupuncture and counselling. We used ANCOVA to evaluate treatment differences between BDI-II depression scores at 12 months, while controlling for BDI-II baseline scores and covariates for depression as identified above. We used multiple imputation for missing data as for the PHQ-9.

Using an intention-to-treat basis, all analyses were carried out in Stata 12.1. Statistical tests were at a two-sided 0.05 significance level. To assess model assumptions, analysis of residuals was undertaken for all regression models.

### Patient and public involvement

In preparation for conducting this research into acupuncture for depression, we conducted a feasibility study that included a prospective case series of 10 patients who received acupuncture treatment for their depression and a focus group with six patients with experience of depression. We used a focus group because as a research team we wanted to learn from patients about their depression and experiences of treatment. We also needed guidance on how best to address the question of whether or not acupuncture is a suitable treatment option for people with depression.

A local patient-centred mental health group, York & District Mind, agreed to work with us on this research and became a long-term partner in its support of research into acupuncture for depression. As a first step we drew on this support in forming a focus group, which met once for 2 hours. Participants in the group were two researchers and six volunteers, identified through York & District Mind, who had experienced depression and were either staff or users of mental health services. A topic guide provided structure for the meeting, covering personal experiences of depression, experiences of both conventional and alternative treatments for depression, and feedback on our research plans. It is the last of these that is most relevant to the patient and public impact on our research.

With regard to feedback on our research plans, participants were concerned about recruitment through GPs on the basis that people who do not consult their GP could not be part of the study. It was agreed that reaching out to those not consulting their GP was a good idea; however, the methods of recruitment would be challenging as there was no way of identifying this hard-to-reach group in sufficient numbers for a RCT.

Another issue raised was to do with the definition of depression amid concerns that there might be some cut-off in severity, for example having a cut-off point so that only people with mild to moderate depression could be recruited. There was a strong feeling among participants that people with severe depression, particularly those who had exhausted all other treatment options, would want to be included in the offer to participate in a trial. The feelings regarding discrimination against mental health users were clearly resonating in this discussion; nobody wanted to be excluded.

This point about not including more severely depressed patients in the trial had a major impact on our trial design when finalising the eligibility criteria. The resulting design included a cut-off to exclude mild depression; however, there was no cut-off for the more severely depressed. This was operationalised with the BDI-II, which we used to recruit patients with both moderate (score of 20–28) and severe (score 29–63) depression. This eligibility requirement led to 62% of patients recruited being categorised as having severe depression and the remaining 38% of patients being categorised as having moderate depression.

York & District Mind continued to be involved with our research into acupuncture for depression. Two lay members and their chief executive joined our independent steering group, which met roughly every 6 months throughout the duration of the trial. Beyond the steering group, their contributions also involved editing the wording in documentation related to the questionnaires that were to be filled in by patients. The chief executive took responsibility for consulting members of his organisation, including users of mental health services, and then aggregated the feedback, which led to improvements in the wording of the questionnaires. Once we had agreed the documentation, and received ethics approval, there was less scope for the York & District Mind participants to contribute, as the trial became primarily a logistic task for the next 3 years. In 2011 York & District Mind merged with another charity called Our Celebration to become York Mind. This change also brought a new chief executive, a transition that took some time to be completed. The new chief executive attended one independent steering committee meeting in June 2011. The combination of this change of personnel and the associated participation gap led to a reduced involvement by patients and the public in the trial from this point on.

### Methods for substudy 1: experience of treatment from the patient perspective

The aim of this substudy was to explore patients' experiences of depression, both with and without comorbid pain, and the perceived changes as a result of receiving the treatments of acupuncture or counselling and usual care. A secondary aim was to report the aspects of treatment that patients reported might have had a positive influence on any long-term change.

In terms of methods, this was a nested qualitative substudy within the Acupuncture, Counselling or Usual Care for Depression (ACUDep) trial. To recruit patients for interview, we selected only from those who had already consented to interview. We used a purposive sampling frame and recruited in the same 2 : 2 : 1 proportions as in the main trial from the treatment groups of acupuncture, counselling and usual care, respectively. We balanced the proportions of men and women and whether or not in pain at baseline. On receipt of their final 12-month postal questionnaire, and based on a sampling frame, a sample of participants who had previously consented to be contacted for an interview was invited to engage in a telephone interview of approximately 30 minutes' duration. Altogether, 52 people consented to a one-to-one, audio-recorded, semistructured telephone interview for this study.

A researcher (AH) interviewed all participants and all interviews were conducted between February and May 2012. The researcher was unknown to the participants prior to the interview and therefore the interview opened with an introduction designed to set the participant at ease, reveal the context for their depression and draw out their account of treatment received as part of the trial. Prompts from a prepared topic guide were used to elicit the participants' experiences of depression and treatment. On average, the interviews lasted approximately 25 minutes (range 11–46 minutes). To encourage participants to relax, participants were asked to introduce themselves by speaking about things they like to do or hoped to do and then how depression had entered their lives, before moving on to the research-related questions within the topic guide. Interviews were audio taped, transcribed verbatim and checked for accuracy. All recordings

were of sufficient clarity and content that no repeat interviews were necessary. Each transcription was checked to remove any names and was then assigned a participant identification number.

Across the data set of 52 transcribed interviews we used an inductive thematic analysis.[318] With a constructivist approach to grounded theory,[319] data were analysed initially by AH. Coding was performed sequentially on each transcript by working systematically through the entire data set. Coding and extractions were checked by JE to verify that the participants' experiences were reflected and summarised accurately. Coding was conducted within each interview and across interviews resulting in themes associated with the experience of depression and further themes associated with the experience of the treatment. Further details of the methods are provided in a separate publication.[320]

### Methods for substudy 2: impact of comorbid pain on depression outcomes

The aim of this substudy was to find out whether people with depression who were also in pain had better or worse depression outcomes than those without pain. We documented the prevalence of pain in this trial population, the demographic profile of patients in pain and the relationship between pain and depression. We determined depression and pain outcomes following treatment with acupuncture or counselling compared with usual care alone. A full report has been published elsewhere.[321]

In terms of methods, the participants were first divided into two groups according to their response to the EQ-5D pain statements at baseline. People who reported either 'moderate' or 'extreme pain' were considered together as the 'pain group'; the remainder were assigned to a 'no-pain comparator group'.

Analysis of variance (ANOVA) was used to compare the mean baseline PHQ-9 scores between the pain group and the no-pain comparator group. For the pain group alone, a Kendall's tau correlation was used to test the association between the baseline scores of the PHQ-9 and SF-36 bodily pain score. A series of regression models was applied to determine the influence of baseline pain and other demographic variables on the PHQ-9 depression outcome at 3 months, the primary end point of the trial. To take into account the additional impact of other demographic predictors of PHQ-9 depression at 3 months, a systematic univariate analysis of demographic variables was conducted on the BDI-II depression items and the five EQ-5D items while controlling for PHQ-9 scores at baseline. For this scoping exercise, the level of significance was set at $p < 0.1$ to maintain consistency with the analysis within the main ACUDep trial. The variables identified by univariate analysis were then included in a linear regression model ($p < 0.05$); the odds ratios and 95% CIs of the variables identified by the regression models are presented within the results section.

Using any remaining significant predictors and controlling for baseline PHQ-9 depression scores, ANCOVA was used to test if baseline pain affected treatment outcomes measured by the PHQ-9 at 3 months. An interaction term between treatment and pain group in this model was used to establish whether or not patients in pain responded differently to the treatments with regard to their depression from patients reporting no pain. Also, using any remaining significant predictors and controlling for baseline PHQ-9 depression scores, ANCOVA was used to test whether or not baseline pain affected treatment outcome measured by the number of depression-free days at 3 months; an approximate summary measure was derived from PHQ-9 cut-off scores averaged over the period between measurements.[322]

An ANCOVA model controlling for baseline depression and baseline SF-36 bodily pain was used to assess the impact of treatment for depression on the bodily pain scores at 3 months. Descriptive analysis of the depression and pain scores was conducted over the 12-month follow-up period. Finally, a comparison of the adverse events reported between baseline and 12 months' follow-up was conducted using proportions and odds ratios.

### Methods for substudy 3: approaches that practitioners used to enhance longer-term benefits

The aim of this qualitative substudy was to explore practitioners' experiences of providing treatment within the trial and to specifically identify the strategies that practitioners reported using to promote longer-term benefits for their patients.

Nested within the trial, this substudy involved in-depth interviews and focus groups with acupuncturists and counsellors who provided treatments to at least two patients within the trial. Our aim was to recruit up to 30 counsellors and acupuncturists (approximately 15 in each group). All participants provided written consent. The interviews and focus groups were informed by topic guides, which were developed by the research team to address the substudy aims. Along with questions, the topic guides also included prompts.

The interviews and focus group were conducted by two experienced health services researchers from the research team who were not involved in the conduct of the trial itself (Liz Newbronner and Ruth Chamberlain). The interviews commonly lasted around 45 minutes (range 40–90 minutes) and the focus group 90 minutes. They were all audio-recorded (with participants' permission) and transcribed verbatim.

The transcripts were analysed thematically.[318] A framework approach facilitated the thematic content analysis, which involved analysing the data across the two groups of practitioners contributing to the study.[323] An inductive process was used to identify themes, drawing on shared experiences or points of difference identified from the data. A full report of the methods has been published separately.[324]

### Methods for substudy 4: cost-effectiveness analysis

The main outcome of the cost-effectiveness analysis was the cost per QALY, which takes into account the treatment differences in HRQoL and mortality. HRQoL was measured using the EQ-5D[325] at baseline and at months 3, 6, 9 and 12. The cost perspective of the NHS was used, although out-of-pocket expenses were also identified. Using patient questionnaires in the trial, resource use data were collected at 3, 6, 9 and 12 months. The sum of the resource use collected during each 3-month period was used to determine the total annual resource use. The total annual cost was calculated by multiplying the total annual resource use by the unit costs using publicly available 2012 national unit costs (see *Appendix 6*). We used the costs of acupuncture as estimated previously[326] and the average of the ranges reported by NHS Choices:[327] £47.50 for an initial session and £37.50 for subsequent sessions. The costs of counselling were those currently used in the NHS: £65 per hour of client contact.[291]

Multiple imputation methods were used to manage the uncertainty caused by the missing data. Chained imputation using predictive mean matching was undertaken using resource use data, PHQ-9 and BDI scores, QALYs and patient characteristics such as age, sex and education. EQ-5D data were analysed using ordered logit models on each of the five dimensions of the instrument. Analysis at 3 months controlled for the baseline response and analysis over 12 months used random-effects models and controlled for the baseline response and the timing of each response (i.e. the day from randomisation). HRQoL weights on the 0 (equivalent to death) to 1 (full health) scale, including negative values (health states worse than death), were calculated using an independent pre-defined algorithm obtained by the elicitation of societal preferences for the health dimensions in a random population sample through a time trade-off technique.[328] QALYs were calculated by applying an individual's HRQoL weights and the time between EQ-5D measures using the AUC approach.[329] For all cost-effectiveness analyses, seemingly unrelated regressions were used to account for the correlation between costs and QALYs.[330] QALYs were regressed on baseline HRQoL and treatment arm, and costs were regressed on the treatment arm only. We estimated ICERs using fully incremental analysis.

Using total costs we calculated base-case results taking into account the uncertainty from the multiple imputation and the seemingly unrelated regression. Probabilistic sensitivity analysis was used to reflect uncertainty in mean total costs and QALYs, and we estimated the probability of cost-effectiveness conditional on alternative cost-effectiveness thresholds. Further exploratory scenario analyses were undertaken to understand the influence on cost-effectiveness of (1) the differential cost of the acupuncture

and counselling interventions, (2) depression-related resource use, (3) the complete case and (4) a population for whom acupuncture is not appropriate or unavailable. All analyses considered the published NICE cost-effectiveness thresholds of £20,000 and £30,000 per QALY gained.[224]

## Results

### Participants

We exceed our recruitment target of 640, recruiting 755 patients in total. Patients registered with 27 general medical practices were recruited between December 2009 and April 2011 in Yorkshire and north-east England. These practices recruited an average of 28 patients (range 0–122 patients). The baseline characteristics of patients were balanced between the trial arms and are presented in *Table 34*. Notably, baseline expectations of treatment effectiveness were lowest for acupuncture yet, by contrast, a majority (58%) expressed a baseline preference to be allocated to acupuncture. These baseline variations were taken into account in the linear mixed-model analysis. Comparison of baseline data between patients with and without missing data at 3 months is presented in *Appendix 6*. *Figure 26* presents the patient flow in the trial.

### Interventions

Of the patients allocated to acupuncture, 266 (88.1%) received one or more treatment sessions (mean 10.3, SD 3.14 treatment sessions) with one of 23 acupuncturists. The average number of sessions received at 3 months was 8.7 (SD 3.34), with 133 patients (50.0%) having completed all of their sessions. Of the patients allocated to counselling, 231 (76.5%) received one or more treatment sessions (mean 9.0, SD 3.74 treatment sessions) with one of 37 therapists. The average number of sessions received at 3 months was 7.5 (SD 3.60), with 114 patients (49.4%) having completed all of their sessions. The mean time from randomisation to last treatment was 117 days in both treatment arms (SD 47.0 and 51.2 days, respectively), a period that included time to first appointment. An average of 13 patients were allocated to each acupuncturist (range 2–45 patients) and an average of eight patients were allocated to each counsellor (range 1–27 patients).

We collected data on acupuncturists' and counsellors' self-reports of intervention protocol violations in logbooks. There were reports of four cases of violation of the acupuncture protocol (which incidentally did not involve counselling), one of which was deemed to be a true case involving prescription of a herbal lotion application to reduce swelling and pain. There were seven reports of violation of the counselling

**TABLE 34** Demographics and variables of interest at baseline

| Characteristic | Intervention | | | |
| --- | --- | --- | --- | --- |
| | Acupuncture + usual care (*N* = 302) | Counselling + usual care (*N* = 302) | Usual care (*N* = 151) | Total (*N* = 755) |
| Age (years) | | | | |
| Mean (SD) | 43.4 (13.24) | 43.5 (13.26) | 43.5 (13.93) | 43.5 (13.37) |
| Median (min., max.) | 43 (18, 86) | 43 (18, 93) | 42 (18, 89) | 43 (18, 93) |
| Interquartile range | 34–52 | 33–52 | 32–54 | 33–53 |
| Missing, *n* (%) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Sex, *n* (%) | | | | |
| Male | 88 (29.1) | 69 (22.8) | 44 (29.1) | 201 (26.6) |
| Female | 214 (70.9) | 233 (77.2) | 107 (70.9) | 554 (73.4) |
| Missing | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |

**TABLE 34** Demographics and variables of interest at baseline (*continued*)

| Characteristic | Intervention | | | Total (*N* = 755) |
| | Acupuncture + usual care (*N* = 302) | Counselling + usual care (*N* = 302) | Usual care (*N* = 151) | |
| --- | --- | --- | --- | --- |
| Age left education (years) | | | | |
| Mean (SD) | 18.0 (4.69) | 18.0 (3.92) | 18.1 (4.62) | 18.0 (4.37) |
| Median (min., max.) | 16 (13, 54) | 16 (14, 48) | 16 (14, 54) | 16 (13, 54) |
| Interquartile range | 16–18 | 16–19 | 16–19 | 16–19 |
| Missing, *n* (%) | 17 (5.6) | 7 (2.3) | 7 (4.6) | 31 (4.1) |
| Employment, *n* (%) | | | | |
| Full-time education | 13 (4.4) | 5 (1.7) | 5 (3.3) | 23 (3.1) |
| Working full-time | 112 (38.0) | 107 (36.4) | 62 (41.3) | 281 (38.0) |
| Working part-time | 57 (19.3) | 59 (20.1) | 28 (18.7) | 144 (19.5) |
| Unable to work | 38 (12.9) | 42 (14.3) | 15 (10.0) | 95 (12.9) |
| Looking after home | 37 (12.5) | 32 (10.9) | 14 (9.3) | 83 (11.2) |
| Retired | 23 (7.8) | 30 (10.2) | 12 (8.0) | 65 (8.8) |
| Other | 15 (5.1) | 19 (6.5) | 14 (9.3) | 48 (6.5) |
| Missing | 7 (2.3) | 8 (2.6) | 1 (0.7) | 16 (2.1) |
| Depression, *n* (%) | | | | |
| In last 2 weeks | 224 (75.7) | 235 (78.6) | 115 (77.7) | 574 (77.3) |
| Missing | 6 (2.0) | 3 (1.0) | 3 (2.0) | 12 (1.6) |
| Not first major episode | 196 (89.5) | 217 (93.5) | 100 (87.7) | 513 (90.8) |
| Missing | 5 (2.2) | 3 (1.3) | 1 (0.7) | 9 (1.6) |
| Four or more previous episodes | 143 (73.0) | 165 (76.7) | 81 (82.7) | 389 (76.4) |
| Missing | 0 (0.0) | 2 (0.9) | 2 (2.0) | 4 (0.8) |
| Age at first major depressive episode (years) | | | | |
| Mean (SD) | 25.8 (12.69) | 24.9 (11.73) | 24.4 (12.55) | 25.2 (12.28) |
| Median (min., max.) | 23 (3, 79) | 22 (6, 71) | 20 (0, 78) | 22 (0, 79) |
| Interquartile range | 16–33 | 16–31 | 16–30 | 16–31 |
| Missing, *n* (%) | 9 (3.0) | 3 (1.0) | 4 (2.6) | 16 (2.1) |
| Medication, *n* (%) | | | | |
| Depression medication in last 3 months | 189 (62.6) | 220 (72.8) | 110 (72.8) | 519 (68.7) |
| Missing | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Analgesic medication in last 3 months | 147 (48.8) | 126 (42.3) | 86 (57.3) | 359 (47.9) |
| Missing | 1 (0.3) | 4 (1.3) | 1 (0.7) | 6 (0.8) |
| EQ-5D anxiety/depression, *n* (%) | | | | |
| Not anxious/depressed | 8 (2.7) | 8 (2.6) | 5 (3.3) | 21 (2.8) |
| Moderately anxious/depressed | 219 (73.0) | 221 (73.2) | 114 (75.5) | 554 (73.6) |
| Extremely anxious/depressed | 73 (24.3) | 73 (24.2) | 32 (21.2) | 178 (23.6) |
| Missing | 2 (0.7) | 0 (0.0) | 0 (0.0) | 2 (0.3) |

**TABLE 34** Demographics and variables of interest at baseline (*continued*)

| Characteristic | Intervention | | | |
| --- | --- | --- | --- | --- |
| | Acupuncture + usual care (*N* = 302) | Counselling + usual care (*N* = 302) | Usual care (*N* = 151) | Total (*N* = 755) |
| PHQ-9 score | | | | |
| Mean (SD) | 15.3 (5.33) | 16.6 (5.27) | 16.2 (5.09) | 16.0 (5.29) |
| Median (min., max.) | 15 (3, 27) | 17 (4, 27) | 16 (5, 27) | 16 (3, 27) |
| Interquartile range | 11–19 | 13–21 | 13–20 | 12–20 |
| Missing, *n* (%) | 1 (0.33) | 0 (0.0) | 0 (0.0) | 1 (0.1) |
| PHQ-9 group, *n* (%) | | | | |
| None (0–4) | 4 (1.3) | 2 (0.7) | 0 (0.0) | 6 (0.8) |
| Mild (5–9) | 44 (14.6) | 29 (9.6) | 14 (9.3) | 87 (11.5) |
| Moderate (10–14) | 97 (32.2) | 74 (24.5) | 46 (30.5) | 217 (28.7) |
| Moderately severe (15–19) | 88 (29.2) | 96 (31.8) | 47 (31.1) | 231 (30.6) |
| Severe (20–27) | 68 (22.6) | 101 (33.4) | 44 (29.1) | 213 (28.2) |
| Missing | 1 (0.3) | 0 (0.0) | 0 (0.0) | 1 (0.1) |
| BDI-II score | | | | |
| Mean (SD) | 32.0 (8.54) | 33.3 (9.11) | 31.8 (8.17) | 32.5 (8.72) |
| Median (min., max.) | 31 (20, 57) | 32 (20, 60) | 30 (20, 56) | 31 (20, 60) |
| Interquartile range | 25–37 | 26–39 | 25–37 | 26–38 |
| Missing, *n* (%) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| BDI-II group, *n* (%) | | | | |
| Moderate (20–28) | 124 (41.1) | 104 (34.4) | 56 (37.1) | 284 (37.6) |
| Severe (29–63) | 178 (58.9) | 198 (65.6) | 95 (62.9) | 471 (62.4) |
| SF-36 bodily pain | | | | |
| Mean (SD) | 58.8 (27.99) | 58.0 (29.17) | 54.4 (27.83) | 57.6 (28.44) |
| Median (min., max.) | 62 (0, 100) | 62 (0, 100) | 51 (0, 100) | 52 (0, 100) |
| Interquartile range | 41–84 | 31–84 | 31–74 | 32–84 |
| Missing, *n* (%) | 1 (0.3) | 3 (1.0) | 0 (0.0) | 4 (0.5) |
| Expectation acupuncture, *n* (%) | | | | |
| Very ineffective | 10 (3.3) | 6 (2.0) | 1 (0.7) | 17 (2.3) |
| Fairly ineffective | 9 (3.0) | 11 (3.7) | 6 (4.0) | 26 (3.5) |
| Can't decide | 187 (61.9) | 204 (68.0) | 103 (68.2) | 494 (65.6) |
| Fairly effective | 66 (21.9) | 44 (14.7) | 29 (19.2) | 139 (18.5) |
| Very effective | 30 (9.9) | 35 (11.7) | 12 (7.9) | 77 (10.2) |
| Missing | 0 (0.0) | 2 (0.7) | 0 (0.0) | 2 (0.3) |
| Expectation counselling, *n* (%) | | | | |
| Very ineffective | 23 (7.7) | 18 (6.0) | 15 (9.9) | 56 (7.5) |
| Fairly ineffective | 50 (16.7) | 43 (14.3) | 21 (13.9) | 114 (15.2) |
| Can't decide | 101 (33.8) | 100 (33.2) | 39 (25.8) | 240 (32.0) |

**TABLE 34** Demographics and variables of interest at baseline (*continued*)

| Characteristic | Intervention | | | |
| | Acupuncture + usual care (*N* = 302) | Counselling + usual care (*N* = 302) | Usual care (*N* = 151) | Total (*N* = 755) |
| --- | --- | --- | --- | --- |
| Fairly effective | 95 (31.8) | 97 (32.2) | 65 (43.0) | 257 (34.2) |
| Very effective | 30 (10.0) | 43 (14.3) | 11 (7.3) | 84 (11.2) |
| Missing | 3 (1.0) | 1 (0.3) | 0 (0.0) | 4 (0.5) |
| Expectation of usual care, *n* (%) | | | | |
| Very ineffective | 29 (9.6) | 34 (11.3) | 6 (4.1) | 69 (9.2) |
| Fairly ineffective | 60 (19.9) | 62 (20.7) | 44 (29.7) | 166 (22.1) |
| Can't decide | 108 (35.8) | 92 (30.7) | 50 (33.8) | 250 (33.3) |
| Fairly effective | 95 (31.5) | 97 (32.3) | 46 (31.1) | 238 (31.7) |
| Very effective | 10 (3.3) | 15 (5.0) | 2 (1.4) | 27 (3.6) |
| Missing | 0 (0.0) | 2 (0.7) | 3 (2.0) | 5 (0.7) |
| Expectation of actual treatment randomised to, *n* (%) | | | | |
| Very ineffective | 10 (3.3) | 18 (6.0) | 6 (4.1) | 34 (4.5) |
| Fairly ineffective | 9 (3.0) | 43 (14.3) | 44 (29.7) | 96 (12.8) |
| Can't decide | 187 (61.9) | 100 (33.2) | 50 (33.8) | 337 (44.9) |
| Fairly effective | 66 (21.9) | 97 (32.2) | 46 (31.1) | 209 (27.8) |
| Very effective | 30 (9.9) | 43 (14.3) | 2 (1.4) | 75 (10.0) |
| Missing | 0 (0.0) | 1 (0.3) | 3 (2.0) | 4 (0.5) |
| Treatment preference, *n* (%) | | | | |
| Acupuncture | 177 (58.8) | 171 (57.6) | 82 (54.7) | 430 (57.5) |
| Counselling | 55 (18.3) | 75 (25.3) | 34 (22.7) | 164 (21.9) |
| Usual care | 2 (0.7) | 7 (2.4) | 1 (0.7) | 10 (1.3) |
| No preference | 67 (22.3) | 44 (14.8) | 33 (22.0) | 144 (19.3) |
| Missing | 1 (0.3) | 5 (1.7) | 1 (0.7) | 7 (0.9) |
| Treatment concordance, *n* (%) | | | | |
| Randomised to preferred treatment | 177 (58.8) | 75 (25.3) | 1 (0.7) | 253 (33.8) |
| Randomised to non-preferred treatment | 57 (18.9) | 178 (59.9) | 116 (77.3) | 351 (46.9) |
| No preference | 67 (22.3) | 44 (14.8) | 33 (22.0) | 144 (19.3) |
| Missing | 1 (0.3) | 5 (1.7) | 1 (0.7) | 7 (0.9) |

Max., maximum; min., minimum.

protocol, of which two were deemed to be true cases; one case involved goal setting and in another the counsellor reported being analytical and interpretative beyond the scope of humanistic counselling.

Details of all three interventions are provided in *Appendix 6*, with usual care documented in the following categories: patients seeing health professionals; patients attending hospital accident and emergency departments; patients admitted to hospital; patients paying for private health care. Medication use is detailed later in this chapter. A more complete report on acupuncture provision has been published[331] as well as a report reflecting on experiences of providing the counselling.[332]
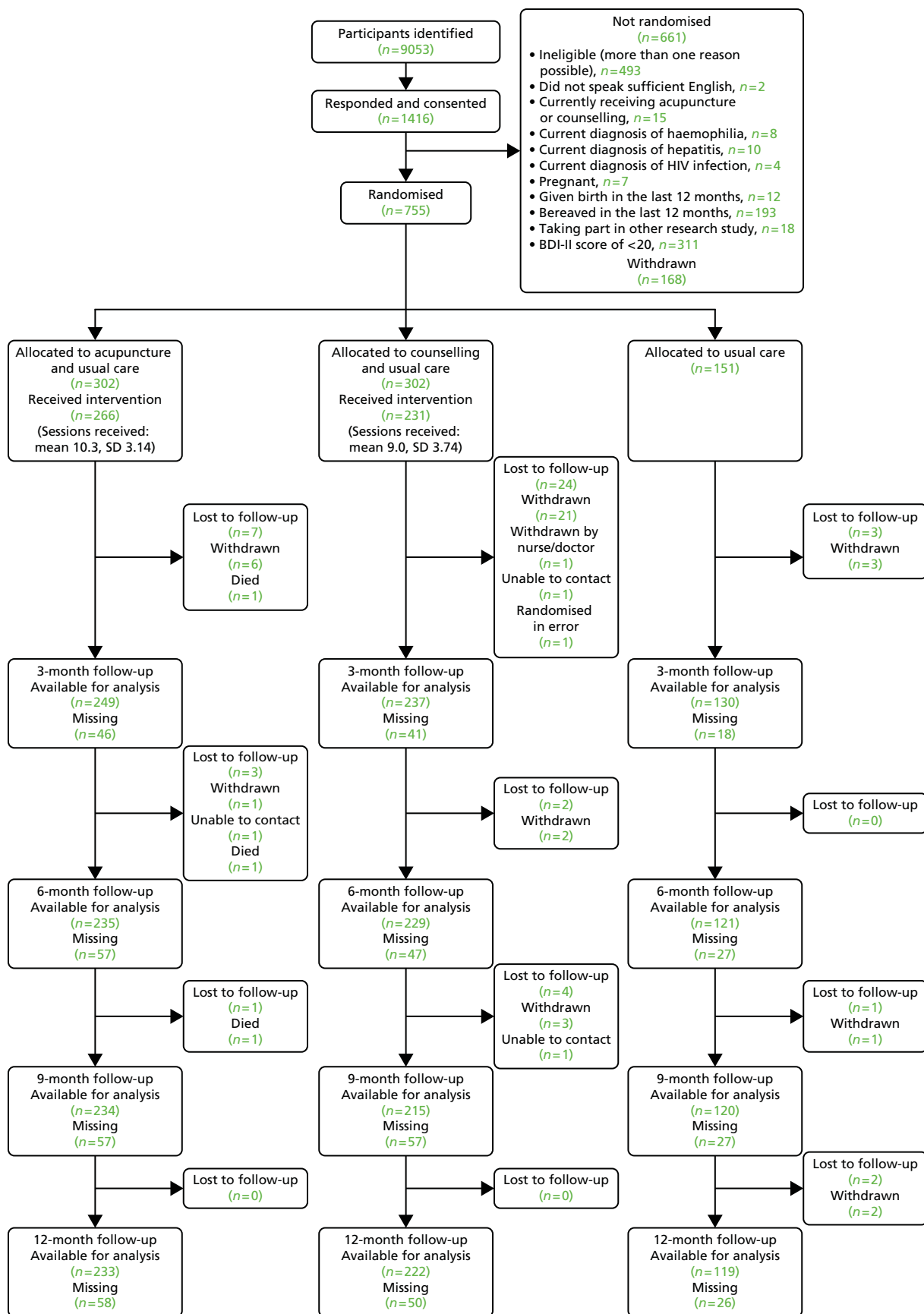
**FIGURE 26** Consolidated Standards of Reporting Trials (CONSORT) flow diagram. HIV, human immunodeficiency virus.

## Clinical outcomes

For the PHQ-9, unadjusted mean scores at all time points are presented in *Table 35* and *Figure 27*. *Table 36* provides the primary outcome results for between-group differences in PHQ-9 depression at 3 months using ANCOVA. With regard to missing data at 3 months, patients for whom no data were available tended to be younger with higher levels of baseline depression. Imputed data were used to take their profile into account. Compared with usual care, patients in the acupuncture arm experienced an average additional reduction in depression of –2.46 points on the PHQ-9 (95% CI –3.72 to –1.21 points; $p < 0.001$), an observed effect size equivalent to a Cohen's *d* of –0.39 (95% CI –0.58 to –0.19). Compared with usual care, patients allocated to the counselling arm experienced an average additional reduction in depression of –1.73 points (95% CI –3.00 to –0.45 points; $p = 0.008$), equivalent to a Cohen's *d* of –0.27 (95% CI –0.47 to –0.07). There was no statistically significant difference between acupuncture and counselling (–0.76 points on the PHQ-9, 95% CI –1.77 to 0.25 points; $p = 0.41$). These data are conservative as non-imputed data showed slightly larger treatment effects (see *Appendix 6*).

**TABLE 35** Unadjusted depression outcomes

| Outcome measure | Intervention | | | | | | | |
| | Acupuncture + usual care | | Counselling + usual care | | Usual care | | Total | |
| | *n* | Mean (SD) | *n* | Mean (SD) | *n* | Mean (SD) | *n* | Mean (SD) |
| **PHQ-9** | | | | | | | | |
| Baseline | 301 | 15.3 (5.33) | 302 | 16.6 (5.27) | 151 | 16.2 (5.09) | 754 | 16.0 (5.29) |
| 3 months | 249 | 9.4 (6.33) | 237 | 10.9 (6.45) | 128 | 12.7 (6.47) | 614 | 10.7 (6.51) |
| 6 months | 235 | 9.1 (6.51) | 228 | 10.1 (6.87) | 120 | 12.0 (6.85) | 583 | 10.1 (6.80) |
| 9 months | 234 | 9.7 (6.90) | 215 | 10.1 (7.03) | 120 | 11.9 (7.04) | 569 | 10.3 (7.02) |
| 12 months | 233 | 9.3 (6.68) | 220 | 10.1 (6.86) | 119 | 11.5 (6.98) | 572 | 10.1 (6.85) |
| **BDI-II** | | | | | | | | |
| Baseline | 302 | 32.0 (8.54) | 302 | 33.3 (9.11) | 151 | 31.8 (8.17) | 755 | 32.5 (8.72) |
| 12 months | 226 | 20.4 (13.19) | 211 | 21.4 (13.64) | 151 | 23.8 (12.63) | 588 | 21.4 (13.29) |
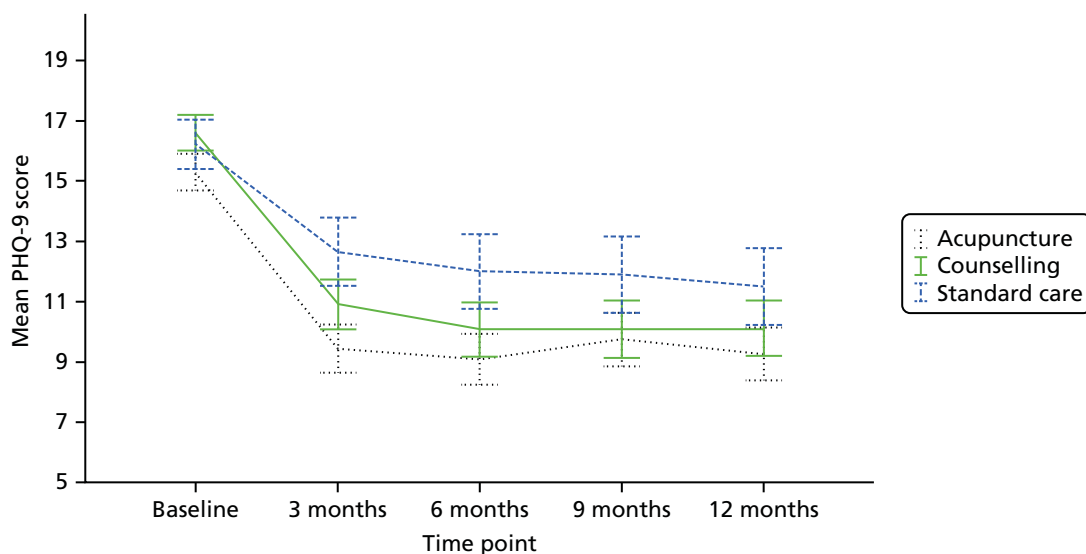


**FIGURE 27** Unadjusted PHQ-9 mean scores from baseline to 12 months.

**TABLE 36** Effect of trial arm on PHQ-9 depression

| Analysis | Month | n | Group 1 Mean | SE | Group 2 Mean | SE | Group difference Mean | SE | 95% CI | p-value |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | **Acupuncture** | | **Usual care** | | | | | |
| ANCOVA[a] | 3 | 452 | 9.8 | 0.41 | 12.3 | 0.58 | −2.46 | 0.636 | −3.72 to −1.21 | < 0.001 |
| Mixed[b] | 3 | 372 | 9.8 | 0.27 | 12.1 | 0.38 | −2.29 | 0.475 | −3.22 to −1.36 | – |
| | 6 | 350 | 9.6 | 0.33 | 11.5 | 0.46 | −1.90 | 0.569 | −3.02 to −0.79 | – |
| | 9 | 348 | 10.2 | 0.39 | 11.1 | 0.54 | −0.83 | 0.671 | −2.15 to 0.49 | – |
| | 12 | 347 | 9.7 | 0.45 | 10.7 | 0.64 | −0.99 | 0.785 | −2.53 to 0.55 | – |
| AUC[c] | 3–12 | 407 | 10.9 | 0.25 | 12.5 | 0.35 | −1.55 | 0.435 | −2.41 to −0.70 | – |
| | | | **Counselling** | | **Usual care** | | | | | |
| ANCOVA[a] | 3 | 453 | 11.1 | 0.40 | 12.8 | 0.58 | −1.73 | 0.648 | −3.00 to −0.45 | 0.008 |
| Mixed[b] | 3 | 362 | 10.9 | 0.28 | 12.8 | 0.38 | −1.83 | 0.477 | −2.76 to −0.90 | – |
| | 6 | 345 | 10.4 | 0.33 | 12.2 | 0.47 | −1.78 | 0.576 | −2.91 to −0.65 | – |
| | 9 | 332 | 10.5 | 0.40 | 11.8 | 0.55 | −1.26 | 0.688 | −2.61 to 0.08 | – |
| | 12 | 336 | 10.4 | 0.47 | 11.4 | 0.65 | −1.00 | 0.805 | −2.58 to 0.57 | – |
| AUC[c] | 3–12 | 402 | 11.6 | 0.28 | 13.1 | 0.38 | −1.50 | 0.470 | −2.43 to −0.58 | – |
| | | | **Acupuncture** | | **Counselling** | | | | | |
| ANCOVA[a] | 3 | 603 | 10.0 | 0.41 | 10.8 | 0.40 | −0.76 | 0.514 | −1.77 to 0.25 | 0.140 |
| Mixed[b] | 3 | 402 | 9.5 | 0.29 | 9.4 | 0.32 | 0.11 | 0.439 | −0.75 to 0.97 | – |
| | 6 | 371 | 9.6 | 0.35 | 9.1 | 0.38 | 0.45 | 0.527 | −0.58 to 1.49 | – |
| | 9 | 360 | 10.0 | 0.41 | 9.0 | 0.45 | 0.97 | 0.621 | −0.25 to 2.19 | – |
| | 12 | 361 | 9.6 | 0.48 | 9.0 | 0.53 | 0.59 | 0.721 | −0.82 to 2.01 | – |
| AUC[c] | 3–12 | 531 | 11.1 | 0.27 | 11.2 | 0.27 | −0.06 | 0.378 | −0.81 to 0.68 | – |

a  ANCOVA: effect of trial arm on imputed PHQ-9 scores at 3 months (adjusting for baseline PHQ-9 score).
b  Mixed: effect of trial arm on PHQ-9 scores over 12 months (mixed-effects model, including time, trial arm–time interaction, baseline PHQ-9 score, baseline expectation of counselling and baseline expectation of allocated treatment; for the comparison between acupuncture and counselling, treatment time completed within 3 months and practitioner's empathy were additionally included).
c  AUC: effect of trial arm on average PHQ-9 depression over 12 months (adjusting for baseline PHQ-9 score).

In the AUC analysis over the 12-month period, the statistically significant benefit of acupuncture and counselling over usual care alone in terms of PHQ-9 score reduction seen at 3 months remained: acupuncture reduced the PHQ-9 score by −1.55 (95% CI −2.41 to −0.70) and counselling by −1.50 (95% CI −2.43 to −0.58) (see *Table 36*).

In a secondary analysis exploring potential mediators, PHQ-9 scores at 3 months were found to be associated with two factors: expectations of counselling ($p = 0.064$) and expectations about the treatment that patients were allocated to ($p = 0.015$). These factors were included as potential mediators of the effect of trial arm in further analyses. Related to the comparison of the acupuncture and counselling arms, additional significant factors were total session time at 3 months ($p < 0.001$) and perceived empathy of practitioners ($p < 0.001$), which were added as further covariates.

From the linear mixed modelling, PHQ-9 depression scores in the acupuncture and counselling groups were reduced compared with usual care at 3 and 6 months (see *Table 36*). The scores in the usual care group continued to reduce over time, such that differences were no longer statistically significant at 9 and 12 months. There was no evidence of significant differences between acupuncture and counselling throughout.

Using ANCOVA, depression scores on the BDI-II were found to be reduced in the acupuncture arm (–2.88, 95% CI –5.68 to –0.08) and the counselling arm (–2.74, 95% CI –5.50 to 0.02) to a greater extent than by usual care alone at 12 months, but with no statistically significant difference between acupuncture and counselling (*Table 37*).

### Prescribed medication

Among all patients, the majority (68.7%) were taking antidepressants at baseline (*Table 38*) and prescribed antidepressant utilisation decreased steadily by an average of 12% over the 12-month study period, a rate

**TABLE 37** Effect of trial arm on BDI-II depression

| Analysis | Month | n | Group | | | | | Group difference | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 1 | | 2 | | | | | |
| | | | Mean | SE | Mean | SE | | Mean | SE | 95% CI |
| | | | *Acupuncture* | | *Usual care* | | | | | |
| ANCOVA[a] | 12 | 445 | 22.8 | 1.34 | 25.7 | 1.82 | | –2.88 | 1.419 | –5.68 to –0.08 |
| | | | *Counselling* | | *Usual care* | | | | | |
| ANCOVA[a] | 12 | 449 | 22.7 | 1.47 | 25.4 | 1.74 | | –2.74 | 1.399 | –5.50 to 0.02 |
| | | | *Acupuncture* | | *Counselling* | | | | | |
| ANCOVA[a] | 12 | 401 | 22.5 | 0.92 | 21.9 | 1.02 | | 0.59 | 1.281 | –1.93 to 3.11 |

a ANCOVA: effect of trial arm on imputed BDI-II scores at 12 months (adjusting for baseline BDI-II score, baseline expectation of counselling and baseline expectation of allocated treatment; for the comparison between acupuncture and counselling, treatment time*3 months and practitioner's empathy were additionally included).

**TABLE 38** Numbers of patients taking prescribed medication in the preceding 3 months

| Prescribed medication | Intervention | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Acupuncture + usual care | | | Counselling + usual care | | | Usual care | | | Total | | |
| | n | %[a] | Change from baseline (%) | n | %[a] | Change from baseline (%) | n | %[a] | Change from baseline (%) | n | %[a] | Change from baseline (%) |
| *Antidepressants* | | | | | | | | | | | | |
| Baseline | 189 | 62.6 | – | 220 | 72.9 | – | 110 | 72.9 | – | 519 | 68.7 | – |
| 3 months | 147 | 60.7 | –0.9 | 155 | 69.2 | –3.7 | 80 | 66.7 | –6.2 | 382 | 65.2 | –3.5 |
| 6 months | 126 | 54.6 | –8.0 | 145 | 65.3 | –7.6 | 75 | 63.6 | –9.3 | 346 | 60.6 | –8.1 |
| 9 months | 123 | 54.0 | –8.6 | 124 | 60.8 | –12.1 | 74 | 66.1 | –6.8 | 321 | 59.0 | –9.7 |
| 12 months | 121 | 52.4 | –10.2 | 124 | 58.2 | –14.7 | 69 | 61.6 | –11.3 | 314 | 56.5 | –12.2 |
| *Analgesics* | | | | | | | | | | | | |
| Baseline | 147 | 48.8 | – | 126 | 42.3 | – | 86 | 57.3 | – | 359 | 47.9 | – |
| 3 months | 73 | 30.8 | –18.0 | 93 | 42.5 | 0.2 | 57 | 48.3 | –9.0 | 223 | 38.9 | –9.0 |
| 6 months | 99 | 43.4 | –5.4 | 84 | 38.5 | –3.8 | 52 | 46.4 | –10.9 | 235 | 42.1 | –5.8 |
| 9 months | 82 | 36.8 | –12.0 | 71 | 34.8 | –7.5 | 55 | 50.0 | –7.3 | 208 | 38.7 | –9.2 |
| 12 months | 84 | 36.8 | –12.0 | 87 | 41.0 | –1.3 | 54 | 49.5 | –7.8 | 225 | 41.0 | –6.9 |

a All percentages given out of total number of patients with valid responses.

comparable between trial arms. Around half of patients (47.9%) were taking analgesics at baseline, which decreased on average to 41.0% over 12 months. Patients in the acupuncture arm showed a marked decrease in the use of analgesics in the first 3 months.

### Adverse events

Over the 12 months, the number of patients experiencing a serious adverse event, as judged by a clinician (IW), was 16 out of 302 (5.3%) in the acupuncture group, 26 out of 302 (8.6%) in the counselling group and nine out of 151 (6.0%) in the usual care group, of whom nine had more than one serious adverse event (range 2–4). None of the serious adverse events, including three deaths, was known to be related to treatment. Over the 12 months, the number of patients experiencing a non-serious adverse event was 56 (18.5%), 47 (15.6%) and 40 (26.5%), respectively, of whom 17 had more than one non-serious adverse event (range 2–4).

### Results for substudy 1: experience of treatment from the patient perspective

Of the 61 participants invited, four declined participation and three did not respond; in addition, two recordings failed for technical reasons. In total, 52 participants, 24 men and 28 women, with an age range of 22–89 years (mean 46 years, SD 13.8 years), were interviewed. At baseline, 26 of these participants had reported having moderate or extreme pain or discomfort on the EQ-5D questionnaire; these people formed the pain group, with the remainder forming the no-pain comparator group. As part of the ACUDep trial, 22 of the 52 had been randomised to receive acupuncture, 20 had been randomised to receive counselling and 10 had been randomised to receive usual care alone. A summary is presented in the sampling frame in *Appendix 6*. On average, those allocated to acupuncture attended 11 sessions (range 4–12 sessions) and those allocated to counselling attended 10 sessions (range 6–12 sessions).

Participants' experiences of depression varied considerably between those with comorbid pain and those with depression alone. Those participants with depression and comorbid pain commonly experienced a number of other physical symptoms concurrently such as fatigue, low energy and sleep problems. For some this meant withdrawing from social and day-to-day activities. This reduced ability to engage in social activity was one of the factors that led to this group having less in the way of internal and external resources available to manage their depression effectively. The majority in the no-pain comparator group were in full- or part-time employment, or were relatively affluent retired professional people. For many, their experience of depression concerned feelings of low self-esteem brought about by high expectations of themselves within their working life or hectic social schedules.

Those with physical symptoms who were receiving acupuncture commonly reported that these broader symptoms were usually addressed as part of the treatment. As treatment progressed, many participants reported that their acupuncturist began guiding them to make changes to their lifestyle to engender beneficial long-term outcomes. For most people with pain, fear of pain and potential injury posed a barrier to engaging in physical activity. Nevertheless, the majority of the pain group reported being encouraged to take up gentle exercise for their overall health. The advice given to the no-pain comparator group more often focused on dietary change and a reduction in caffeine and alcohol intake, as well as encouragement of some form of relaxation. Overall, longer-term improvement in depression was developed through the participants' active engagement in health-promoting behaviours, supported by a positive therapeutic relationship.

For those receiving counselling there was less emphasis on physical symptoms and more on help with gaining an understanding of themselves and their situation. Despite many having low expectations of counselling based on their previous experiences, the majority of participants engaged with the counselling process. Most reported feeling relieved to have someone to talk to in confidence. For both the pain and the no-pain groups, the process of change followed a common pathway, beginning with the participants' disclosure of personal information and being listened to and leading to an exploration of their past, which helped to clarify their understanding of themselves and their situation. The final stage in the process of change for many was directed towards enabling them to maintain their progress and recovery independently.

In summary, longer-term change was encouraged by most practitioners over the course of treatment. This was facilitated by strong support to cope with depression and pain independently of treatment, with a focus on relevant lifestyle and behaviour changes. Participants identified as important components of treatment the therapeutic relationship with their practitioner and their active engagement in the treatment process, whether they were receiving acupuncture or counselling, A full report on the experiences of treatment has been published separately.[320]

### Results for substudy 2: impact of comorbid pain on depression outcomes

Patients reporting moderate and extreme pain or discomfort on the EQ-5D questionnaire were merged together to form a single pain group ($n = 384$, 51%) with the remainder forming the no-pain comparator group ($n = 371$, 49%). A summary of the demographic variables in the subgroups allocated to the pain and no-pain groups at baseline is presented in *Appendix 6*. The variables were comparable for most items with the following notable exceptions: (1) the pain group members tended to be older than the no-pain comparator group members (mean 47 years vs. 40 years); and (2) in terms of health and employment, 56% of the pain group (vs. 9% of the no-pain group) reported a painful health condition or illness that predated the onset of depression, for which 64% (vs. 31%) used analgesic medication regularly, and 32% (vs. 9%) were unable to work or were retired.

The baseline PHQ-9 depression scores indicate that the pain group reported higher levels of depression at baseline (mean 17.0, SD 5.2) than the no-pain comparator group (mean 14.9, SD 5.2). Results of the ANOVA confirmed the difference to be highly significant (mean difference 2.02, 95% CI 1.28 to 2.76). For the pain group alone, the correlation between the PHQ-9 scores and SF-36 bodily pain scores was weak but highly significant (Kendall's tau –0.172; $p < 0.001$).

Using the average across all treatment groups, participants in the pain group showed a smaller reduction in depression score at 3 months compared with baseline (mean reduction from 16.70 to 12.06 at 3 months) than the no-pain comparator group (mean reduction from 14.06 to 9.10 at 3 months). A linear regression model found that the presence of moderate or extreme pain at baseline predicted a poorer outcome of depression treatment at 3 months (mean difference –1.72, 95% CI –2.64 to –0.80; $p < 0.001$) when controlling for baseline depression scores. An ANCOVA model controlling for significant predictors and baseline PHQ-9 score revealed that the effect of pain group remained significant, with patients with baseline pain having poorer depression outcomes (mean difference –1.16, 95% CI –2.2 to –0.12; $p = 0.028$).

Regarding PHQ-9 depression scores by pain group and by trial arm at baseline and all follow-up time points, relevant data are shown in *Figure 28*. Controlling for baseline depression and covariates, an ANCOVA model including a pain group–treatment interaction term showed that, in the pain group, participants showed a larger reduction in depression with acupuncture at 3 months (mean reduction in PHQ-9 score from baseline 6.0, 95% CI 5.07 to 7.11), with smaller reductions associated with counselling (mean reduction 4.3, 95% CI 3.3 to 5.4) and usual care (mean reduction 2.7, 95% CI 1.50 to 4.06). In comparison, no notable differences were seen between treatment arms within the no-pain comparator group.

Using the SF-36 bodily pain score at 3 months' follow-up as the end point, and controlling for baseline SF-36 bodily pain and baseline PHQ-9 depression scores, the results of the ANCOVA show that the pain group continued to experience significantly worse pain after treatment for depression compared with the no-pain comparator group (mean difference 14.57, 95% CI 9.73 to 19.40). There was also a significant interaction between pain group and treatment arm ($F_{2,1} = 3.3$; $p = 0.036$), with pain group patients who received acupuncture for depression experiencing a greater reduction in SF-36 bodily pain (represented by an increase in scores) between baseline and 3 months' follow-up (mean reduction 11.2, 95% CI 7.1 to 15.2) than those who received counselling (mean reduction 7.6, 95% CI 3.6 to 11.6) or usual care (mean reduction 7.2, 95% CI 2.3 to 12.1). The reduction in pain at 3 months persisted through to the 12-month follow-up point (*Figure 29*); however, the median score in the pain group after 12 months (median 41, interquartile range 1 = 31, interquartile range 3 = 62) remained below the trial baseline median score of 52 on the SF-36 bodily pain scale. Further details of the results have been published separately.[321]
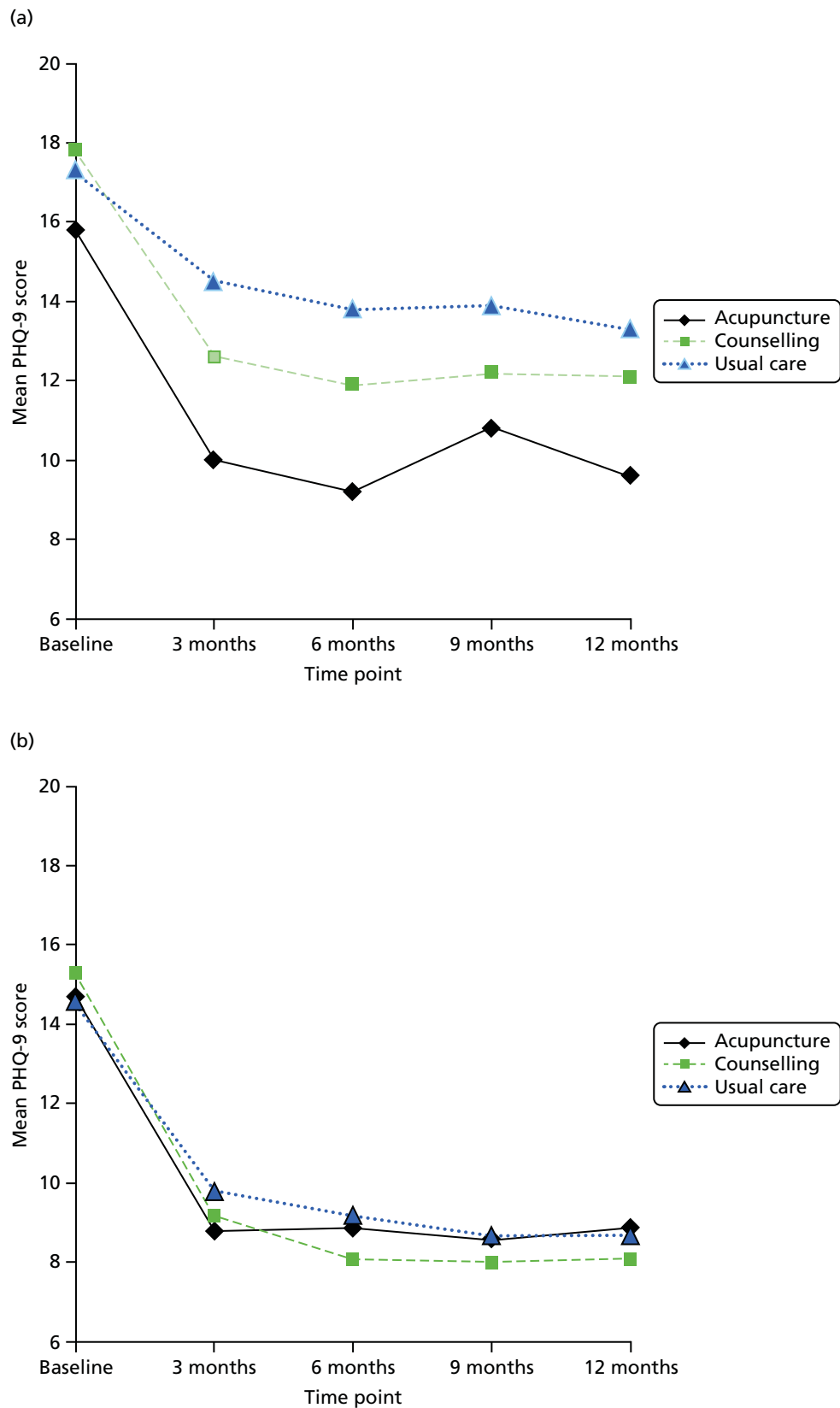
**FIGURE 28** Mean PHQ-9 scores by treatment, comparing the pain and no-pain groups: (a) patients with moderate to extreme pain at baseline (*n* = 384); and (b) patients pain free at baseline (*n* = 371).

(a)



(b)
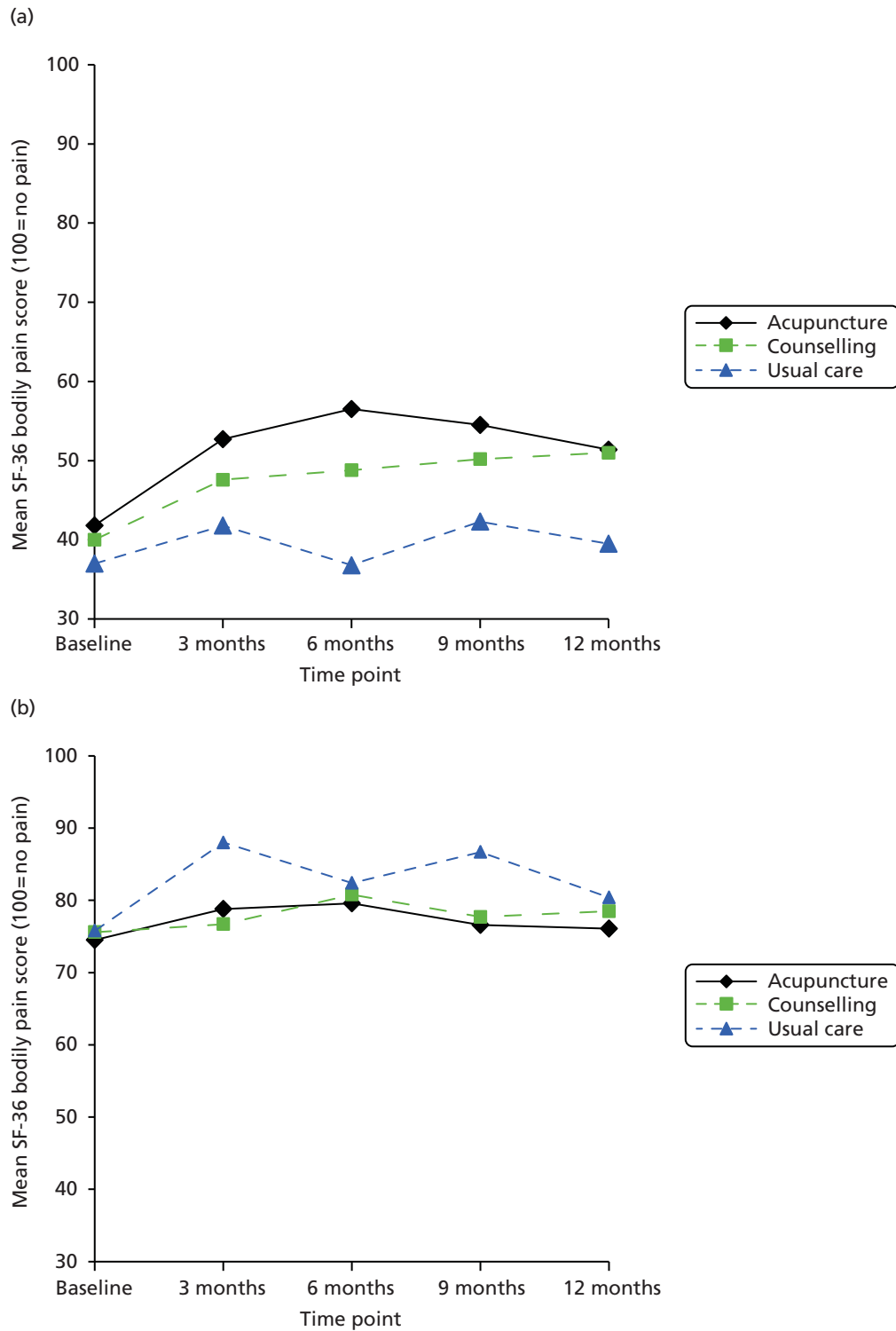


FIGURE 29 Mean SF-36 bodily pain scores by treatment over the course of the trial, comparing the pain and no-pain groups: (a) patients with moderate to extreme pain at baseline (*n* = 384); and (b) patients pain free at baseline (*n* = 371).

## Results for substudy 3: approaches that practitioners used to enhance longer-term benefits

Forty-one therapists consented to be involved in the study but five therapists later withdrew or did not respond when the fieldwork was being set up. The substudy included individual telephone interviews with 15 counsellors and 13 acupuncturists (*n* = 28), and one focus group with four counsellors and four acupuncturists (*n* = 8) (see *Appendix 6*). The sample for the qualitative substudy included over half (56%) of the total number of therapists involved in the main trial. The 17 acupuncturists were predominantly

male (39% were female), with an average duration of practice of 12 years. In the trial they treated on average nine patients, who attended on average for 10 sessions. The 19 counsellors were predominantly female practitioners (79%), with an average duration of practice of 7 years. In the trial they treated on average seven patients, who attended on average for nine sessions.

A cluster of eight themes emerged from the framework analysis. Almost all of the acupuncturists and counsellors stressed the importance that they attached to promoting longer-term benefits.

1. *Importance of a long-term focus*. Almost all of the practitioners mentioned the long-term perspective as inherent to the way they worked. For longer-term impact, both acupuncturists and counsellors encouraged insight into root causes of depression on an individual basis and saw small incremental changes as precursors to sustained benefit.
2. *Identifying root causes*. The commitment to identifying and addressing the root causes of depression within the treatment process was a commonly expressed factor among both acupuncturists and counsellors. Acupuncturists commonly used the theoretical concept of 'root' (ben) and 'branch' (biao). The counsellors were more interested in 'going deeper' and 'further back (in time)' with their clients as a way of getting to the root causes.
3. *Individualisation*. When practised within the theoretical framework of traditional Chinese medicine, acupuncture treatment has been customised to the individual, such that treatment varies not only between patients but also, for the same patient, over time. Likewise for the counsellors in the trial, all of whom were committed to working within the humanistic tradition, a person-centred approach was provided.
4. *Valuing incremental change*. Practitioners reported that they did not expect change to happen all at once. Rather than seeking cathartic experiences, ideally improvement in symptoms would develop slowly with small incremental changes.
5. *Addressing concurrent physical symptoms*. Acupuncturists generally stressed the importance of addressing concurrent physical symptoms, for example helping patients relax or sleep better to be more receptive to change. Counsellors were less focused on the physical symptoms and worked more on the assumption that by treating the underlying causes of depression there should be a knock-on effect in terms of reducing physical symptoms.
6. *Lifestyle changes*. Acupuncturists tended to highlight the importance of giving advice about lifestyle change that was relevant to the Chinese medicine perspective. By contrast, the counsellors had a more non-directive style, consistent with the humanistic tradition.
7. *Therapeutic relationship*. More often than the acupuncturists, the counsellors stressed the importance of the therapeutic relationship as a facilitator of change.
8. *Careful 'pacing' based on 'readiness to change'*. Many of the counsellors also emphasised the need for careful 'pacing' such that the processes and tools that they employed were tailored and timed for each individual, depending on the 'readiness' to change. A number of acupuncturists reflected on the best way of sustaining benefit over the longer term, suggesting that ideally appointments should be spread out towards the end of the course of treatment and beyond.

The above themes capture both similarities and differences in approaches between acupuncture and counselling, with most practitioners of both interventions having a shared sense of the prerequisites for sustained long-term benefit. The impression formed from the interviews is that the various approaches do not operate in isolation. The themes appear to be integrated into a coherent combination that uniquely informs the practice of each practitioner, although the emphasis may vary. A full report of the results has been published separately.[324]

## Results for substudy 4: cost-effectiveness analysis

At 3 months patients treated with acupuncture or counselling were less likely than patients treated with usual care to report that they were moderately or extremely anxious or depressed rather than not anxious or depressed (*Table 39*). The 3-month improvement in anxiety and depression was sustained over the trial period to 12 months. Combining the EQ-5D dimension results with the UK population health state

**TABLE 39** EuroQoL-5 Dimensions data: the proportional odds of being at level 2 or 3 compared with level 1[a] – comparing acupuncture or counselling with usual care

| | At 3 months[b] | | Over 12 months[b] | |
|---|---|---|---|---|
| EQ-5D dimension | Acupuncture OR (95% CI) | Counselling OR (95% CI) | Acupuncture OR (95% CI) | Counselling OR (95% CI) |
| Anxiety and depression | 0.63 (0.40 to 0.98) | 0.66 (0.42 to 1.02) | 0.40 (0.23 to 0.70) | 0.40 (0.23 to 0.70) |
| Pain | 0.77 (0.48 to 1.23) | 0.96 (0.60 to 1.53) | 0.87 (0.49 to 1.54) | 0.88 (0.5 to 1.55) |
| Usual activities | 1.14 (0.48 to 2.71) | 1.05 (0.44 to 2.54) | 0.57 (0.34 to 0.95) | 0.72 (0.43 to 1.21) |
| Self-care | 0.81 (0.52 to 1.27) | 0.85 (0.54 to 1.33) | 0.40 (0.15 to 1.09) | 0.58 (0.22 to 1.53) |
| Mobility | 1.29 (0.64 to 2.61) | 1.19 (0.59 to 2.41) | 0.89 (0.41 to 1.94) | 0.74 (0.35 to 1.59) |

OR, odds ratio.
a Levels 1–3 represent low, moderate and high disability, respectively.
b ORs of < 1 indicate that the treatment was correlated with fewer patients reporting being in the more severe health states compared with patients in the usual care arm (i.e. the odds ratio of 0.63 in column 2 suggests that patients in the acupuncture arm were less likely to report being moderately or extremely anxious or depressed than patients in the usual care arm at 3 months). ORs of > 1 suggest that the treatment is correlated with more patients reporting being in the more severe health states compared with patients in the usual care arm.

preferences resulted in the HRQoL scores over time and by treatment presented in *Figure 30*. For all treatment arms HRQoL increased between baseline and 3 months, with the acupuncture and counselling arms having a higher HRQoL than usual care and remaining higher at 12 months. QALYs were estimated to be 0.604, 0.663 and 0.666 for the usual care, acupuncture and counselling arms, respectively, using imputed data and seemingly unrelated regression and controlling for baseline HRQoL.

Mean NHS resource use is reported in *Table 40*. Total costs and depression-related costs are reported in *Table 41*. Patients reported the amount spent on out-of-pocket acupuncture, counselling or usual care. Patients in the acupuncture arm reported spending a mean of £32 (SD £93) on acupuncture, whereas those in the counselling and usual care arms reported spending £7 (SD £41) and £6 (SD £57), respectively. Patients in the counselling arm reported spending a mean of £42 (SD £173) on counselling, whereas those in the acupuncture and usual care arms reported spending £6 (SD £42) and £5 (SD £32), respectively. Patients spent on average £2 (SD £33), £15 (SD £87) and £3 (SD £23) on psychotherapy in the acupuncture, counselling and usual care arms, respectively. The mean number of days off work over 12 months was
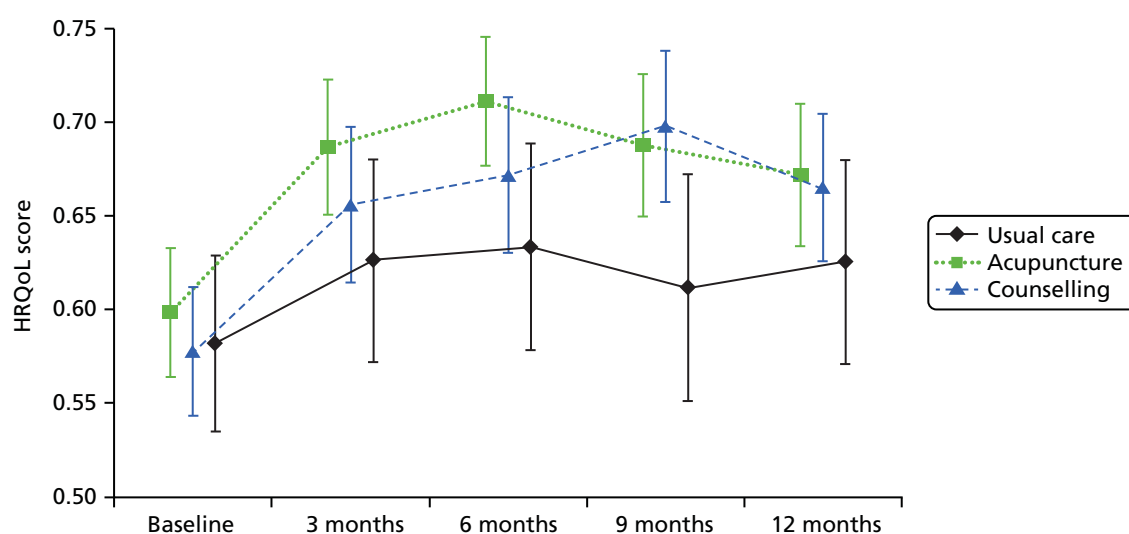


**FIGURE 30** Health-related quality-of-life scores over time and by treatment.

**TABLE 40** Complete case and imputed resource use over 12 months

| Resource | Intervention | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Usual care | | | Acupuncture | | | Counselling | | | | | |
| | n | Complete case mean (95% CI) | Imputed mean (95% CI) | n | Complete case mean (95% CI) | Imputed mean (95% CI) | n | Complete case mean (95% CI) | Imputed mean (95% CI) | | | |
| GP | 69 | 6.48 (5.16 to 7.80) | 6.56 (5.37 to 7.75) | 145 | 5.57 (4.78 to 6.37) | 5.66 (4.88 to 6.43) | 127 | 4.94 (4.19 to 5.7) | 5.06 (4.38 to 5.73) | | | |
| Practice nurse | 60 | 1.40 (0.86 to 1.94) | 1.53 (1.05 to 2.02) | 140 | 1.16 (0.84 to 1.48) | 1.25 (0.95 to 1.54) | 127 | 1.25 (0.95 to 1.55) | 1.36 (1.05 to 1.66) | | | |
| Other health professional | 54 | 1.39 (0.74 to 2.04) | 1.76 (0.93 to 2.6) | 133 | 1.24 (0.7 to 1.79) | 1.37 (0.85 to 1.89) | 116 | 1.4 (0.78 to 2.01) | 1.54 (1 to 2.07) | | | |
| NHS hospital outpatient clinic | 82 | 1.55 (0.99 to 2.1) | 2.01 (1.19 to 2.83) | 175 | 1.4 (0.95 to 1.85) | 1.52 (1.05 to 1.98) | 151 | 1.69 (1.17 to 2.21) | 1.87 (1.35 to 2.39) | | | |
| Hospital ward | 79 | 0.29 (−0.01 to 0.60) | 0.32 (0.04 to 0.6) | 166 | 0.15 (0.02 to 0.28) | 0.21 (0.05 to 0.38) | 138 | 0.27 (0.06 to 0.47) | 0.35 (0.12 to 0.59) | | | |
| Hospital ICU | 76 | – | 0 (0 to 0) | 160 | 0.04 (−0.04 to 0.11) | 0.03 (−0.03 to 0.1) | 136 | – | 0 (0 to 0) | | | |
| Hospital mental health unit | 76 | 0.53 (−0.48 to 1.53) | 0.48 (−0.34 to 1.3) | 160 | 0.04 (−0.04 to 0.13) | 0.12 (−0.18 to 0.43) | 134 | – | 0.42 (−0.13 to 0.97) | | | |
| Other hospital unit | 75 | 0.01 (−0.01 to 0.04) | 0.02 (−0.02 to 0.07) | 152 | 0.05 (−0.04 to 0.14) | 0.04 (−0.02 to 0.09) | 132 | 0.05 (0 to 0.11) | 0.04 (0 to 0.08) | | | |
| Accident and emergency | 84 | 0.37 (0.17 to 0.57) | 0.4 (0.15 to 0.64) | 191 | 0.21 (0.13 to 0.29) | 0.25 (0.13 to 0.37) | 157 | 0.35 (0.11 to 0.59) | 0.35 (0.16 to 0.53) | | | |
| Community mental health nurse | 74 | 0.43 (−0.04 to 0.91) | 0.42 (0.04 to 0.79) | 164 | 0.15 (0.02 to 0.29) | 0.2 (0.04 to 0.37) | 145 | 0.1 (−0.02 to 0.23) | 0.18 (0.01 to 0.35) | | | |
| Psychologist or psychiatrist | 72 | 0.75 (0.04 to 1.46) | 0.73 (0.16 to 1.3) | 164 | 0.21 (0.08 to 0.34) | 0.31 (0.09 to 0.54) | 135 | 0.35 (0.09 to 0.61) | 0.55 (0.22 to 0.88) | | | |
| NHS counsellor not involved in the study | 68 | 0.34 (0.07 to 0.60) | 0.38 (0.07 to 0.68) | 161 | 0.29 (0.08 to 0.49) | 0.28 (0.11 to 0.45) | 142 | 0.17 (0.02 to 0.32) | 0.23 (0.06 to 0.4) | | | |

ICU, intensive care unit.

**TABLE 41** Complete case and imputed costs over 12 months

| Resource | Intervention | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Usual care | | | Acupuncture | | | Counselling | | | | | |
| | n | Complete case mean (95% CI) | Imputed mean (95% CI) | n | Complete case mean (95% CI) | Imputed mean (95% CI) | n | Complete case mean (95% CI) | Imputed mean (95% CI) | | | |
| Total costs (£) | 22 | 621 (365 to 877) | 958 (739 to 1180) | 69 | 1110 (930 to 1291) | 1227 (1103 to 1350) | 59 | 1355 (1082 to 1627) | 1450 (1305 to 1592) | | | |
| Depression-related costs (£) | 18 | 226 (92 to 360) | 496 (288 to 704) | 54 | 769 (644 to 894) | 913 (764 to 1061) | 48 | 962 (759 to 1166) | 1006 (761 to 1251) | | | |

similar across arms: 238 (SD 115), 240 (SD 112) and 231 (SD 113) for the acupuncture, counselling and usual care arms, respectively.

When comparing acupuncture, counselling and usual care, acupuncture was found to be the cost-effective alternative with an ICER of £4560 per additional QALY compared with usual care alone, with probabilities of being cost-effective of 0.68, 0.62 and 0.56 at thresholds of £13,000, £20,000 and £30,000 per QALY, respectively (*Table 42*). Counselling resulted in higher costs and benefits than acupuncture, with an ICER of £71,757 per additional QALY compared with acupuncture.

In a scenario analysis which assumed that each session of acupuncture was the same price as each session of counselling (£65), counselling had higher QALYs and lower costs than acupuncture, that is, acupuncture was dominated (*Table 43*). Restricting the analysis to the complete case data resulted in an ICER for acupuncture of £10,979 per QALY and counselling having higher costs and lower QALYs than acupuncture. For patients in whom acupuncture is inappropriate or unavailable, the incremental cost-effectiveness of counselling compared with usual care was £7935 per additional QALY.

**TABLE 42** Incremental cost-effectiveness of usual care, acupuncture and counselling

| Treatment | QALYs | Total costs (£) | ICER (£ per QALY) | Probability of cost-effectiveness | | |
| | | | | Threshold £13,000 per QALY | Threshold £20,000 per QALY | Threshold £30,000 per QALY |
| --- | --- | --- | --- | --- | --- | --- |
| Usual care | 0.604 | 958 | – | 0.07 | 0.03 | 0.02 |
| Acupuncture | 0.663 | 1227 | 4560 | 0.68 | 0.62 | 0.56 |
| Counselling | 0.666 | 1450 | 71,757 | 0.26 | 0.36 | 0.42 |

**TABLE 43** Incremental cost-effectiveness scenario analyses

| Scenario analysis | QALYs[a] | Total costs (£) | ICER (£ per QALY) | Probability of cost-effectiveness | | |
| | | | | Threshold £13,000 per QALY | Threshold £20,000 per QALY | Threshold £30,000 per QALY |
| --- | --- | --- | --- | --- | --- | --- |
| **1. Assuming acupuncture has the same cost as counselling (£65)** | | | | | | |
| Usual care | 0.558 | 524 | – | 0.15 | 0.06 | 0.03 |
| Counselling | 0.620 | 1050 | 8497 | 0.50 | 0.55 | 0.56 |
| Acupuncture | 0.617 | 1073 | Dominated | 0.35 | 0.39 | 0.42 |
| **2. Using depression-related costs** | | | | | | |
| Usual care | 0.601 | 513 | – | 0.08 | 0.03 | 0.02 |
| Acupuncture | 0.659 | 853 | 5819 | 0.61 | 0.58 | 0.54 |
| Counselling | 0.663 | 1025 | 50,612 | 0.32 | 0.39 | 0.44 |
| **3. Complete case analysis** | | | | | | |
| Usual care | 0.638 | 648 | – | 0.43 | 0.29 | 0.20 |
| Acupuncture | 0.682 | 1121 | 10,979 | 0.57 | 0.70 | 0.79 |
| Counselling | 0.643 | 1378 | Dominated | 0.01 | 0.01 | 0.01 |
| **4. Population for whom acupuncture is not appropriate** | | | | | | |
| Usual care | 0.604 | 958 | – | 0.21 | 0.09 | 0.05 |
| Counselling | 0.666 | 1450 | 7935 | 0.79 | 0.91 | 0.95 |

a Some of the differences between these results and the base-case results (see *Table 42*) are because of the probabilistic nature of the calculations.

## Discussion

### Principal findings

For patients continuing to experience depression in primary care, we found statistically significant benefits at 3 months associated with both acupuncture and counselling interventions when provided as adjuncts to usual care. Prior to recruitment to the trial, these patients experienced recurring bouts of depression (76% having had four or more episodes), with the first episode on average occurring at age 25 years, some 19 years previously, and 69% were on antidepressant medication. We also found statistically significant benefits for patients over the 12-month period as a whole in an AUC analysis. At 12 months the benefits of acupuncture and counselling were no longer significantly better than those of usual care for our primary outcome measure, the PHQ-9, but statistically significant differences remained when depression was measured by the BDI-II. No serious adverse events related to treatment were reported.

### Strengths and limitations

This study design of a pragmatic RCT had a clear and practical research question with an appropriate trial design to model closely what would happen if patient referrals to acupuncturists and counsellors were routine. With an emphasis on external validity, our pragmatic trial was designed to have findings that are generalisable to typical patients and settings. We recruited patients in primary care from among those who had consulted with depression and who continued to be depressed, thereby excluding patients whose symptoms had been alleviated sufficiently by other treatment.

The attrition of patient-reported data between randomisation at baseline and the follow-up time-points was typical of trials that recruit through primary care databases. We used multiple imputation to compensate in part for the limitations related to the loss of follow-up data. Our design controlled for temporal effects, such as the natural history of depression, and other factors across all patients that might have influenced outcomes beyond the treatment itself through the randomisation into groups. The standardised treatment protocols for the acupuncturists and counsellors were designed to reflect routine practice, allowing individualisation to match patient variability while ensuring that all practitioners met appropriate standards of qualification and experience. No attempt was made to standardise usual care; however, we were careful to document the usual care that was provided to patients in all arms of the trial. The majority of patients continued with antidepressants and differences between groups in usual care at all time points were minimal (see *Appendix 6*). For this reason we can assume that the differences in outcomes between arms can largely be ascribed to the treatments provided by the acupuncturists and the counsellors.

Our study was not designed to determine which aspects of the interventions might be most or least beneficial. In contrast to pragmatic trials, explanatory trials are designed to separate out the relative contributions of specific or non-specific components of treatment. Nevertheless, we took into account in our regression model the expectations and preferences reported by all patients in the trial, components that are often considered non-specific effects. Moreover, for two treatment-related components in the acupuncture and counselling groups of our trial that are often considered 'non-specific', namely session time and quality of therapist attention, we found that specific treatment effects remained when these were accounted for. Despite the limitation that a pragmatic approach has in ascribing outcomes to different treatment components, a pragmatic trial design provides a useful estimate of the overall effect, an estimate of most interest to patients, practitioners and providers.

### Relationship to the literature

In a recent Cochrane review of acupuncture for depression,[310] no studies focused on the BDI-II-related categories of moderate to severe depression, unlike patients in our trial who, for eligibility, had to have a score of ≥ 20 on the BDI-II. Moreover, only two trials in the Cochrane review, by the same research team, used a usual care comparator, which was based on a wait list.[333,334] A meta-analysis of these two trials involving 94 patients showed a reduction in the SMD of –0.73 (95% CI –1.18 to –0.29).[310] This is a larger effect size than we found (we found a point estimate of –0.39); however, we need to be cautious in the interpretation of the result, as the patients in these studies were less depressed and the patient numbers were small.

From a Cochrane review of counselling for mental health and psychosocial problems in primary care,[312] there is evidence from six trials of counselling for mild to moderate depression including 772 patients that counselling is more effective than usual care in terms of mental health outcomes over 1–6 months (SMD –0.28, 95% CI –0.43 to –0.13). These advantages were not shown to endure over the longer term from 7 to 12 months (SMD –0.09, 95% CI –0.27 to 0.10). Their number needed to treat for short-term benefit was six, somewhat less than the 10 we found for moderate to severely depressed patients. We found no counselling trial equivalent to the one that we report here, namely one based in primary care that evaluated counselling for moderate to severe depression. In a review of psychological interventions for depression, four cognitive–behavioural therapy studies based in primary care covering 259 patients found a similar short-term effect (SMD –0.33, 95% CI –0.60 to –0.06).[335]

In this report we present the first evidence on acupuncture and counselling for patients in primary care who are representative of those continuing to experience symptoms of moderate to severe depression. At 3 months, which is approximately at the end of the course of 12 treatment sessions, both acupuncture and counselling were shown to be effective treatments compared with usual care alone. There was also evidence that there are benefits over the 12-month period in terms of clinical symptoms.

### Implications for clinical practice

To interpret the results for clinical practice, we found that 33% of acupuncture patients, 29% of counselling patients and 18% of usual care patients achieved a successful treatment outcome when this is defined as improvement from a depressed PHQ-9 score (≥ 10) to a non-depressed score (≤ 9), with an improvement of at least 50%.[336] These percentages apply to the total number of patients with a baseline PHQ-9 score of ≥ 10 and for whom data were available at 3 months. The number needed to treat is also a useful way to interpret results. We found that for one additional treatment success, as defined above, the number needed to treat was 7 for acupuncture (95% CI 4.3 to 17.4) and 10 for counselling (95% CI 5.3 to 47.3). A further illustration of the impact on patient experience can be provided in terms of 'depression-free days',[322] which is a summary measure derived from PHQ-9 cut-off scores averaged over the period between measurements. The mean number of depression-free days over 3 months was 34 (95% CI 31 to 38) for the acupuncture group, 27 (95% CI 24 to 30) for the counselling group and 23 (95% CI 19 to 27) for the usual care group.

We used the BDI-II as a screening tool, with patients having to score ≥ 20 for eligibility, a score that is classified by this measure as moderate or severe depression. Other classification systems, such as the *Diagnostic and Statistical Manual of Mental Disorders* [see www.psychiatry.org/psychiatrists/practice/dsm (accessed 13 October 2016)], have different categories of severity. Our trial did not determine whether acupuncture or counselling performed better or worse for the more mild forms of depression. Our trial also did not determine whether or not patients not receiving antidepressant medication (one-third of our sample) do better or whether fewer sessions would be sufficient for mild to moderate depression or more sessions would improve outcomes for those with severe depression (approximately 60% of our sample). It is not clear from our trial whether or not there was any impact on outcomes because, for those allocated to acupuncture, it was usually their first experience of this intervention, whereas for many randomised to counselling, it was an intervention they had received before.

### Discussion for substudy 1: experience of treatment from the patient perspective

Patients commonly reported that their acupuncturists appeared to have a more physical perspective on treatment, with a focus on directly relieving the symptoms of depression as well as concurrent physical symptoms. This was particularly welcomed by patients who reported comorbid pain at baseline, who appreciated having their physical symptoms treated alongside their depression. In conjunction with the acupuncture sessions, the acupuncturists often helped patients engage in health behaviours that had a positive influence on long-term change. In contrast, patients typically reported that their counsellors helped them identify and confront underlying causes of depression and then helped them find their own way forward. For both treatment modalities, most participants reported that the establishment of a therapeutic relationship and their active engagement as patients helped them develop coping strategies, which in turn helped them be more effective in reducing their depression in the longer term.

These qualitative findings are concordant with, and supplement, the quantitative data on depression and comorbid pain within the trial.[321] Notably, the patients with moderate or extreme pain at baseline (as reported on the EQ-5D) had worse outcomes at 3 months for depression than the no-pain comparator group in all three treatment arms. Our findings extend the findings of the trial's quantitative data in two ways. First, there are identifiable differences in the experiences of depression when comorbid pain is present. Second, acupuncture appears to involve an approach that more directly addresses physical symptoms, including the comorbid pain, whereas counselling more directly addresses the underlying psychological causes of depression, which may or may not be linked to comorbid pain.

Our study has some limitations. Participants may have attributed changes directly to treatment rather than concurrent, coincidental contextual changes. We accept that there is a possibility of recall bias as it has long been known that there is a significant, stable association between depression and memory impairment,[337] which may have altered what was recalled and how it was recalled. The lack of face-to-face contact during the telephone interviews prevented the interviewer gathering non-verbal contextual information, such as social cues, body language, appearance and setting, to supplement the verbal answers of the interviewees. A full report of the discussion related to this substudy has been published separately.[320]

### Discussion for substudy 2: impact of comorbid pain on depression outcomes

Participants with moderate to extreme pain at baseline had worse outcomes for depression than the no-pain comparator group in all three treatment arms after controlling for baseline depression. Participants in the pain group had greater reductions in depression symptoms with acupuncture from baseline to 12 months than those who received counselling or usual care, whereas those who were pain free did relatively well whichever group they were assigned to. Participants in the pain group receiving acupuncture found that their pain reduced markedly in the first 3–6 months compared with those in the other two groups, but by the end of the 12-month trial these differences disappeared. It should be noted that this study, as a substudy of a larger trial, was not powered to detect differences between the subgroups with moderate to extreme pain and no pain.

The estimated prevalence of moderate to extreme pain within our study population of depressed patients was 51%, which is comparable to the 50% identified in previous literature.[92] Evidence from an English longitudinal study of ageing identified pain and mobility disability at baseline as predictors of comorbid pain and depression.[338] In a large European study, a higher number of pain locations, pain of the joints and longer duration of pain (for ≥ 90 days), daily use of pain medication and more severe pain at baseline were found to be associated with a significantly increased risk of still having a depressive or anxiety disorder after 2 years.[338] Together, these factors are known to adversely affect the outcomes of treatment for depression.[338] Consistent with previous research, the majority of painful complaints within the study sample were of musculoskeletal origin and accompanied by poor mobility and a loss of energy. Pain from osteoarthritis is known to determine subsequent depressed mood through its effect on fatigue and disability.[339]

That patients reported reduced pain following acupuncture is not surprising given that 32% of patients had chronic musculoskeletal pain and acupuncturists within the trial were encouraged to work how they normally would, incorporating treatment for pain alongside treatment for the symptoms of depression. Moreover, there is a growing body of evidence supporting the efficacy of acupuncture for chronic pain.[196] Patients who received counselling reported a more gradual reduction in pain over the 12-month follow-up period. This finding is consistent with a Cochrane review[340] which reports that psychological therapies, primarily cognitive–behavioural therapy, can help people with chronic pain reduce negative mood (depression and anxiety), disability and, to a lesser degree, pain over a 6-month period.

Frequently patients with comorbid pain and depression will attribute their condition to one or other of these two symptoms and seek help accordingly.[341] For treatment success, our trial provides some evidence that both pain and depression should be both recognised and treated from the outset. Overall, the evidence emerging in the current study is that both acupuncture and counselling appear to have the potential to reduce symptoms associated with pain and depression when treated concurrently, with the potential to relieve

symptoms of depression and reduce the intensity of pain in both the short and the longer term. Further discussion related to this substudy can found elsewhere.[321]

### Discussion for substudy 3: approaches that practitioners used to enhance longer-term benefits

Encouraging longer-term change was integral to the work of both acupuncturists and counsellors. Although both types of practitioners reported on the need to address the root causes, the approach differed. For the acupuncturists there was a focus drawn from Chinese medicine theory on treating the root cause as well as the manifesting symptoms, with the precise details of the intervention customised to their patients at an individual level. By contrast, the approach of the counsellors was to get below the surface of the clients' problems and to 'go deeper' and 'further back (in time)' as a way of getting a handle on the root cause. For both types of practitioners, this required an individualised approach.

Further differences in approach to facilitating more sustained benefits were noted, in that acupuncturists were more focused on physical symptoms, on whether or not these could be resolved by acupuncture to speed up the improvements in the symptoms of depression and on providing lifestyle advice linked to the Chinese medicine diagnosis. Meanwhile, counsellors were more explicit about the importance of a strong therapeutic relationship accompanied by a careful consideration of what might be a manageable pace of change.

The methods used in this substudy, involving interviews and focus groups with verbatim transcripts and thematic analysis, were consistent with many of the markers of quality in qualitative research.[342] By involving an independent research team, we have helped establish the credibility and dependability of the results. With regard to the transferability of the results, we have provided details of the practitioners as well as the patients who were the focus of this substudy, such that readers can draw conclusions regarding relevance for other areas. In terms of limitations, our data are limited to patients receiving acupuncture as practised by those using the theories of traditional Chinese medicine and to clients receiving counselling as provided in the humanistic and non-directive and person-centred style.

Within the wider literature there is a dearth of evidence on the acupuncture treatment factors that might be associated with longer-term change in the symptoms of depression. The findings from a small study involving interviews with six practitioners in a trial of acupuncture for back pain found that these acupuncturists had a goal of a positive long-term outcome and developed a therapeutic partnership to support the active engagement of patients in their own recovery.[343] Consistent with some of the findings we report here, the authors reported that the key elements were establishing a rapport, using an interactive diagnostic process, matching treatment to the patient and using explanatory models from Chinese medicine to aid a shared understanding and motivate lifestyle changes to reinforce the potential recovery.[343]

There is also limited evidence in the counselling literature from qualitative studies on the factors that improve the long-term effects of counselling. One report contains client interview data drawn from 15 clients who had received counselling from between 1 and 3 years previously.[344] The authors' interpretation of the interview data led them to describe a model of the change process and mechanisms that were perceived as essential to produce a lasting benefit. They identified as key elements of the counselling process the active engagement of the client during and between sessions, and the acquisition of a 'box of skills' to be built on further after the counselling was finished.[344] A full report of the discussion related to this substudy has been published separately.[324]

### Discussion for substudy 4: cost-effectiveness analysis

This economic analysis demonstrated that the HRQoL results are consistent with the clinical results. The cost-effectiveness results, taking into account the uncertainty in the estimates, suggest that acupuncture is the cost-effective option. In the base-case analysis, acupuncture had an ICER of £4560 per additional QALY and was cost-effective with a probability of 0.62 at a cost-effectiveness threshold of £20,000 per QALY.

Currently, acupuncture for depression is not provided by the NHS. To understand the cost-effectiveness of counselling in a population for whom acupuncture is not appropriate (e.g. those who are needle phobic), a scenario analysis of counselling compared with usual care, excluding acupuncture as a comparator, was undertaken. In this population counselling had an estimated ICER of £7935 and a 0.91 probability of being cost-effective.

It is possible that the regulation of acupuncture may increase the per-session costs. A sensitivity analysis was undertaken assuming that each acupuncture session costs £65, the same as the cost of counselling. In this scenario counselling was preferred to acupuncture because not only were the expected benefits higher but the expected costs were lower. This demonstrates that the cost-effectiveness of acupuncture in this study is reliant on it having a lower cost than that of counselling.

### Recommendations for future research

Further research could usefully identify the optimal populations for acupuncture and counselling. It would also be useful to explore the impact of different recruitment methods on the characteristics of those recruited. For example, patients recruited at the point of consultation will be different from those recruited through databases, as we did in this trial, and different again from those with depression who do not consult their GP at all. What was important in the trial that we report here was that all patients had consulted in primary care, all continued to be depressed and all were seeking other interventions that might reduce their depression. Other recruitment methods could be used, for example to assess whether the interventions would be more effective for patients at the time of consulting their GP or for patients who were experiencing their first episode of depression, rather than after many episodes, as was the case in the present study, in which 76% of participants had four or more previous episodes, or for patients who have given up consulting in primary care yet continue to be depressed.

Further research into optimal treatment regimens would also be useful. With regard to acupuncture provision, there is some evidence from the literature that a uniform combination of points can work for depressed patients.[345,346] However, a large-scale head-to-head trial or extensive synthesis in an IPD meta-analysis would be required to determine if a uniform approach for all acupuncture provision might be more beneficial than a style of acupuncture in which treatment is individualised for each patient and with changes over time, as was the case with the trial that we report here.

Although our findings are that both acupuncture and counselling for depression appear to be associated with longer-term benefits, it would be useful to explore the different ways that the interventions work. For example, we found that acupuncturists include in their treatment of patients with depression a focus on physical symptoms, yet it is not clear how important this is for outcomes related to depression. Further research is needed into the patient perspective on the treatment of depression with comorbidities and specifically the value that patients place on their comorbid symptoms being addressed concurrently. Further research is also needed from the perspective of patients on their experiences of treatment from these two modalities. To assist referral, a clearer understanding is needed of which type of person with depression would benefit from acupuncture and which from counselling. When taking into account patient preference, such a typology would provide referring clinicians with valuable guidance on suitability for referral.

Further analyses are needed to explore variations in the time horizon and the related impact on the cost-effectiveness of acupuncture and counselling. In this trial patients were followed up for 12 months and our analysis considered only this 12-month time frame, which assumed no differences between treatment arms beyond 12 months. This is a conservative assumption as there would be no further intervention costs, but our trial results suggest that a continued difference in treatment outcomes after 12 months is plausible, even though these treatment differences seem to be converging (see *Figure 30*). Extrapolating these differences beyond 12 months would result in a lower ICER for acupuncture. Further evaluation of the cost-effectiveness of acupuncture and counselling when compared with other physical and psychological interventions as well as with different levels of usual care will provide a better understand of how to best allocate scarce health-care resources.

## Conclusion

In this report we present what is to our knowledge the first study to rigorously evaluate the clinical and economic impacts of acupuncture and counselling for patients who are representative of those who continue to experience depression in primary care. Our evidence on acupuncture compared with usual care and counselling compared with usual care shows that both treatments are associated with a statistically significant reduction in symptoms of depression in the short to medium term, with no reported serious adverse events related to treatment. Acupuncture is cost-effective compared with counselling or usual care alone, although the ranking of counselling and acupuncture depends on the relative costs of delivering these interventions. For patients in whom acupuncture is unavailable or perhaps inappropriate, counselling has an ICER that is less than most cost-effectiveness thresholds.

# Chapter 7 Conclusions

In this programme of research we have addressed several key questions regarding the evidence base on acupuncture for chronic pain and depression. Our focus has been on assessing the clinical effectiveness of acupuncture compared with usual care as well as the efficacy of acupuncture for chronic pain beyond a placebo. Moreover, we have evaluated acupuncture in terms of cost-effectiveness and value for money. Our questions have led to a range of different methods, some innovative and some not, all of which have been appropriate for the questions being considered. For example, among the more innovative methods, for the first time in acupuncture research we have used an IPD meta-analysis, which has provided more power to explore treatment effects and to explore subgroup variations. We have used network meta-analyses to compare acupuncture with other physical therapies for the first time, leading to new evidence on both clinical impact and cost-effectiveness. We have also used more standard methods, for example in a RCT, in which acupuncture or counselling were compared with usual care for depression. These studies within the programme have provided a comprehensive evidence synthesis on acupuncture for a number of chronic pain conditions (lower back pain and neck pain, osteoarthritis of the knee and headache and migraine), as well as a substantive trial of acupuncture and counselling for depression in primary care.

In the IPD meta-analysis, we determined the effect size of acupuncture for chronic pain based on direct evidence, both when acupuncture is compared with sham acupuncture and when it is compared with a non-acupuncture control. Based on data from 29 of 31 high-quality trials, including a total of 17,922 patients, we found that patients receiving acupuncture had less pain than those receiving sham control treatment, with effect sizes for the different pain conditions being in the order of 0.2 ($p < 0.001$). When comparing acupuncture with non-acupuncture controls, effect sizes were larger for all conditions, with effect sizes in the order of 0.5 ($p < 0.001$). The difference between these two comparisons, which is of the order of 0.3, can be ascribed to 'placebo' effects, sham needle-related effects and other context effects. Given the highly statistically significant effect sizes, these data are relevant to the debate regarding the proportion that 'placebo' effects contribute to the overall effect of acupuncture for chronic pain.

In a network meta-analysis, we addressed the question of how effective physical treatments for osteoarthritis of the knee are, when compared with each other on an equal basis, for relieving pain. When synthesising the data from 114 trials involving 22 treatments and 9709 patients, we found that eight interventions statistically significantly outperformed standard (usual) care, including acupuncture and sham acupuncture. The intervention with the most higher-quality studies was acupuncture, which was also one of the more effective physical treatments for alleviating osteoarthritis knee pain in the short term. The caveat for this study was that much of the evidence on physical therapies was of poor quality, which made it difficult to draw conclusions on the effectiveness of many of them.

In another network meta-analysis, we used innovative methods to analyse IPD, drawn from the study discussed above. HRQoL data are required for a cost-effectiveness analysis. To synthesise heterogeneous outcomes we have made use of standardisation of pain measurements and of published mapping algorithms to convert and compare HRQoL evidence on to the EQ-5D summary index scale. This enabled us to analyse HRQoL data across trials in which these data were not collected. Our analysis used the same evidence base as the IPD pairwise meta-analysis. This included approximately 17,500 patients from 28 trials in which we compared acupuncture, sham acupuncture and usual care with each other. Standardised pain pooled results were broadly similar to the ones obtained from the above-mentioned pairwise meta-analysis. The synthesis of mapped EQ-5D estimates found that acupuncture was effective compared with usual care. However, the EQ-5D benefit of acupuncture over sham acupuncture was found to be smaller and less certain. When combined with resource use and cost estimates, these EQ-5D data can be used to generate estimates of cost-effectiveness. Cost-effectiveness results suggested that acupuncture was cost-effective when compared with usual care alone, with ICERs ranging from £7000 to £14,000 per QALY across pain types. Because this was primarily a methods exercise, not all relevant comparators were included and aggregate data beyond

the IPD from the 28 trials evaluated were not included. These are, therefore, not robust estimates of cost-effectiveness and are not an adequate basis for resource allocation decision-making.

To address the question of the cost-effectiveness of acupuncture for osteoarthritis of the knee, we conducted an economic evaluation of non-pharmacological adjunct interventions. We again used network meta-analysis methods, and IPD when available, this time to synthesise RCT evidence on 17 active interventions and three control interventions for osteoarthritis of the knee. The network meta-analysis included 88 eligible studies and 7507 patients. IPD were available for five of the studies including 1329 patients. Data from HRQoL instruments were mapped to EQ-5D preference weights prior to synthesis using network meta-analysis. We estimated resource use associated with the interventions from trial data, expert opinion, the literature and information obtained from NHS trust websites. When all trials were included in the synthesis, TENS was cost-effective at conventional cost-effectiveness threshold values with an ICER of £2690 per QALY compared with usual care. The effectiveness of TENS may be exaggerated because of biases associated with poor-quality trials. When the analysis was restricted to trials with adequate allocation concealment, acupuncture was cost-effective with an ICER of £13,502 per QALY compared with TENS. The EVPI in this area is relatively high, suggesting that additional research may be cost-effective. Further analysis is required to identify the most cost-effective and clinically appropriate specification of further research. Given the likely 'up-front' training costs associated with expansion of acupuncture services in the NHS, investment in further research may need to precede the widespread availability of acupuncture services.

In a RCT of acupuncture or counselling for patients with depression in primary care, we recruited 755 patients from 27 primary care practices in the north-east of England. This was the largest trial to date evaluating acupuncture for depression. Allocation was to acupuncture ($n = 302$), counselling ($n = 302$) or usual care alone ($n = 151$). A mean of 10 sessions was attended for acupuncture and nine sessions for counselling. We found a statistically significant reduction in mean PHQ-9 depression score at 3 months for both acupuncture and counselling compared with usual care, which was largely sustained at 12 months. When controlling for time and attention, we found no significant differences in clinical outcome between acupuncture and counselling. No serious adverse events were reported that were both unexpected and related to treatment. Acupuncture and counselling were found to have higher mean QALYs and costs than usual care. Acupuncture had an ICER of £4560 per additional QALY in the base-case analysis and was cost-effective with a probability of 0.62 at a cost-effectiveness threshold of £20,000 per QALY.

There are limitations to what we report here for two major reasons. First, in our systematic reviews we were limited by the available literature. The history of acupuncture research, as with much of the literature for physical therapies, has been littered with small trials of generally low quality. The higher the risk of bias, the less certainty we have in drawing conclusions. This programme of research has been possible only because of the number of high-quality acupuncture trials with large sample sizes conducted in the 2000s. Moreover, for all physical therapies, including acupuncture, the difficulty in having available a scientifically acceptable control for non-specific effects means that attempts to delineate the proportion of any effect that is non-specific are inevitably limited. At least in acupuncture trials for chronic pain, in which the research question has been related to efficacy rather than effectiveness, the sham needle has been widely used as a control for 'placebo' effects. Remarkably, across all physical therapies for osteoarthritis, there are more high-quality trials of acupuncture with a sham comparator than of any other physical therapy. Therefore, the limitations in the existing literature on physical therapies for chronic pain are less in acupuncture trials, allowing us to draw conclusions with more certainty than for other physical therapies.

The second broad limitation of the programme is that we have not addressed all of the aspects of the evidence base that continue to be associated with uncertainty. In part, this is because we have yet to complete a number of substudies that were planned as integral to individual projects. For example, the IPD meta-analysis had six substudies planned as part of the original mission, yet we have only managed to report on two at this stage; a further four will be completed and published soon. These are a study of the time course of acupuncture, a study exploring the relationship between patient characteristics and

variations in outcome, a study of practitioner effects and a study to identify whether or not there are 'super-responders'. We have recently published two substudies related to the acceptability, feasibility and validity of using text messaging scores of depression,[347,348] and there remains a further substudy exploring the variation in practitioner outcomes that has yet to be published. In part, this is because some of the much-needed associated research fell outside the scope of the programme at the outset. A case in point here is the absence of relevant comparators when we evaluated the cost-effectiveness of acupuncture for musculoskeletal pain and headache and migraine. There are limited data regarding the long-term effects of many non-pharmacological interventions used to treat osteoarthritis of the knee and sensitivity analyses suggested that the cost-effectiveness model results may be sensitive to the magnitude of these effects. The active and control interventions in the trials informing the cost-effectiveness analysis were subject to heterogeneity in the methods, duration and intensity with which they were administered. Another example is that, because of our trial design, we did not control for context effects in our depression trial when comparing acupuncture with usual care, although we did control for time and attention when comparing acupuncture with counselling. In part, further trials of acupuncture would usefully expand the generalisability of data on effectiveness, for example with different populations, such as only patients with mild to moderate depression or only patients on antidepressants. Similarly, different outcomes could be addressed, for example finding out whether or not acupuncture reduces the relapse rate and sustains remission. Further research is merited in these areas related to acupuncture for chronic pain and depression.

The results of this programme of research are important on several counts. First, the data on acupuncture for chronic pain are particularly relevant as pain is the condition for which acupuncture is most commonly used. Moreover, chronic pain and depression are known to be areas in which patients and GPs consider conventional medical treatments have their limitations, whether because of perceived limitations of routinely prescribed medication or because of concerns regarding side effects and dependency. Second, we have used robust and rigorous methods throughout the studies to provide a high level of evidence. Indeed, in the systematic reviews we have purposely sought to identify and synthesise the higher-quality studies as they are known to be affected less by bias than data drawn from trials with a high risk of bias. Third, we have asked questions and used methods that have led to the synthesis of trials that between them have large numbers of participants and, when analysed, these large data sets have provided results with more precision than would be otherwise possible. Fourth, our results are relevant to important questions about resource use in a climate of limited funding within the NHS. By comparing acupuncture in an unbiased way with other 'competing' physical or psychological therapies, we are providing the very evidence on clinical effectiveness and cost-effectiveness that is of most value to policy-makers and commissioners. Finally, it is in the interests of patients to have available well-informed results based on high-quality evidence to make decisions about their health care.

# Acknowledgements

We wish to acknowledge Mike Bennett, Richard Blackwell, Peter Bower, Matthew Bowes, Sally Brabyn, Stephen Brealey, Ann Burton, Ruth Chamberlain, Ben Cross, Ben Elliot, David Geddes, Christina Giannopoulou, Simon Gilbody, Peter Hall, Pauline Holloway, Victoria Hurtado-Menses, Kamran Khan, Harriet Lansdown, Hillary Marshall, Peter Morley, Liz Newbronner, Joanne O'Conner, Karen Overend, Sara Perren, Lucy Revell, Steve Rice, Mark Roman, Micah Rose, Trevor Sheldon, Eleftherios Sideris, David Smith, Lesley Stewart, Alex Sutton, Val Wadsworth, Kerry Wheeler and Nerys Woolacott.

## Acknowledgement of related funding

## Contributions of authors

All authors made a substantial contribution to at least one area within the programme of research.

**Hugh MacPherson** (Professor of Acupuncture Research) was lead applicant and principal investigator. He designed the programme of research and was involved in all projects and their presentation in the chapters of this report.

**Andrew Vickers** (Statistician) conducted the IPD meta-analysis in *Chapter 2* and contributed to the interpretation of the results in *Chapters 4* and *5*.

**Martin Bland** (Professor of Health Statistics) was involved in the conduct of the depression trial in *Chapter 6*.

**David Torgerson** (Professor, Director of the York Trials Unit) had oversight on the conduct of the depression trial in *Chapter 6*.

**Mark Corbett** (Research Fellow, Systematic Reviewer) wrote the first draft of *Chapter 3* and contributed to writing *Chapters 1*, *5* and *7*.

**Eldon Spackman** (Research Fellow, Health Economist) took the lead in the health economic analysis of the depression trial presented in *Chapter 6*.

**Pedro Saramago** (Research Fellow, Statistician) was the lead analyst for and author of *Chapter 4*. He contributed to all aspects of development of the health economic methods, the analysis and the write-up for *Chapters 4* and *5*.

**Beth Woods** (Research Fellow, Health Economist) was the lead analyst and author of *Chapter 5*. She contributed to all aspects of development of the health economic methods, the analysis and the write-up for *Chapters 4* and *5*.

## Publications

Vickers AJ, Cronin AM, Maschino AC, Lewith G, MacPherson H, Victor N, *et al.* Individual patient data meta-analysis of acupuncture for chronic pain: protocol of the Acupuncture Triallists' Collaboration. *Trials* 2010;**11**:90.

Corbett M, Rice S, Slack R, Harden M, Madurasinghe V, Sutton A, *et al. Acupuncture and Other Physical Treatments for the Relief of Chronic Pain due to Osteoarthritis of the Knee: A Systematic Review and Network Meta-Analysis. CRD Report 40.* York: CRD, University of York; 2011.

MacPherson H, Richmond S, Bland MJ, Lansdown H, Hopton A, Kang'ombe A, *et al.* Acupuncture, Counseling, and Usual care for Depression (ACUDep): study protocol for a randomised controlled trial. *Trials* 2012;**13**:209.

Vickers AJ, Cronin AM, Maschino AC, Lewith G, MacPherson H, Foster NE, *et al.* Acupuncture for chronic pain: individual patient data meta-analysis. *Arch Intern Med* 2012;**172**:1444–53.

Woolacott NF, Corbett MS, Rice SJ. The use and reporting of WOMAC in the assessment of the benefit of physical therapies for the pain of osteoarthritis of the knee: findings from a systematic review of clinical trials. *Rheumatology* 2012;**51**:1440–6.

Corbett MS, Rice SJC, Madurasinghe V, Slack R, Fayter DA, Harden M, *et al.* Acupuncture and other physical treatments for the relief of pain due to osteoarthritis of the knee: network meta-analysis. *Osteoarthr Cartil* 2013;**2**:1290–8.

MacPherson H, Elliot B, Hopton A, Lansdown H, Richmond S. Acupuncture for depression: patterns of diagnosis and treatment within a randomised controlled trial. *Evid Based Complement Alternat Med* 2013;**2013**:286048.

MacPherson H, Maschino AC, Lewith G, Foster NE, Witt C, Vickers AJ, *et al.* Characteristics of acupuncture treatment associated with outcome: an individual patient meta-analysis of 17,922 patients with chronic pain in randomised controlled trials. *PLOS ONE* 2013;**8**:e77438.

MacPherson H, Richmond S, Bland, Brealey S, Gabe R, Hopton A, *et al.* Acupuncture and counselling for depression in primary care: a randomised controlled trial. *PLOS Med* 2013;**10**:e1001518.

Vickers AJ, Maschino AC, Lewith G, MacPherson H, Sherman KJ, Witt CM; Acupuncture Triallists' Collaboration. Responses to the Acupuncture Triallists' Collaboration individual patient data meta-analysis. *Acupunct Med* 2013;**31**:98–100.

Hopton A, Eldred J, MacPherson H. Patients' experiences of acupuncture and counselling for depression and comorbid pain: a qualitative study nested within a randomised controlled trial. *BMJ Open* 2014;**4**:e005144.

Hopton A, MacPherson H, Keding A, Morley S. Acupuncture, counselling or usual care for depression and comorbid pain: secondary analysis of a randomised controlled trial. *BMJ Open* 2014;**4**:e004964.

MacPherson H, Newbronner L, Chamberlain R, Richmond SJ, Lansdown H, Perren S, *et al.* Practitioner perspectives on strategies to promote longer-term benefits of acupuncture or counselling for depression: a qualitative study. *PLOS ONE* 2014;**9**:e104077.

MacPherson H, Vertosick E, Lewith G, Linde K, Sherman KJ, Witt CM, *et al.* Acupuncture Triallists' Collaboration. Influence of control group on effect size in trials of acupuncture for chronic pain: a secondary analysis of an individual patient data meta-analysis. *PLOS ONE* 2014;**9**:e93739.

Spackman E, Richmond S, Sculpher M, Bland M, Brealey S, Gabe R, *et al.* Cost-effectiveness analysis of acupuncture, counselling and usual care in treating patients with depression: the results of the ACUDep trial. *PLOS ONE* 2014;**9**:e113726.

Keding A, Böhnke JR, Croudace TJ, Richmond SJ, MacPherson H. Validity of single item responses to short message service texts to monitor depression: an mHealth sub-study of the UK ACUDep trial. *BMC Med Res Methodol* 2015;**15**:56.

Perren S, Richmond S, MacPherson H. The human face of an RCT: reflections on providing counselling for clients with moderate to severe depression in a randomised controlled trial. *Healthcare Couns Psychother J* January 2015, pp. 8–13.

Richmond SJ, Keding A, Hover M, Gabe R, Cross B, Torgerson D, *et al.* Feasibility, acceptability and validity of SMS text messaging for measuring change in depression during a randomised controlled trial. *BMC Psychiatry* 2015;**15**:68.

## Data sharing statement

All available data can be obtained on request from the corresponding author.

# References

1. Hopton AK, Curnoe S, Kanaan M, Macpherson H. Acupuncture in practice: mapping the providers, the patients and the settings in a national cross-sectional survey. *BMJ Open* 2012;**2**:e000456. http://dx.doi.org/10.1136/bmjopen-2011-000456

2. Thomas KJ, Nicholl JP, Coleman P. Use and expenditure on complementary medicine in England: a population based survey. *Complement Ther Med* 2001;**9**:2–11. http://dx.doi.org/10.1054/ctim.2000.0407

3. Lipman L, Dale J, MacPherson H. Attitudes of GPs towards the provision of acupuncture on the NHS. *Complement Ther Med* 2003;**11**:110–14. http://dx.doi.org/10.1016/S0965-2299(03)00042-6

4. Lim JH. Provision of medical acupuncture service in general practice under practice-based commissioning. *Acupunct Med* 2010;**28**:103–4. http://dx.doi.org/10.1136/aim.2010.002287

5. Dale J. Acupuncture practice in the UK. Part 1: report of a survey. *Complement Ther Med* 1997;**5**:215–20. http://dx.doi.org/10.1016/S0965-2299(97)80032-5

6. MacPherson H, Sinclair-Lian N, Thomas K. Patients seeking care from acupuncture practitioners in the UK: a national survey. *Complement Ther Med* 2006;**14**:20–30. http://dx.doi.org/10.1016/j.ctim.2005.07.006

7. Fisher P, van Haselen R, Hardy K, Berkovitz S, McCarney R. Effectiveness gaps: a new concept for evaluating health service and research needs applied to complementary and alternative medicine. *J Altern Complement Med* 2004;**10**:627–32. http://dx.doi.org/10.1089/acm.2004.10.627

8. Robinson N, Lorenc A, Ding W, Jia J, Bovey M, Wang XM. Exploring practice characteristics and research priorities of practitioners of traditional acupuncture in China and the EU-A survey. *J Ethnopharmacol* 2012;**140**:604–13. http://dx.doi.org/10.1016/j.jep.2012.01.052

9. Fønnebø V, Grimsgaard S, Walach H, Ritenbaugh C, Norheim AJ, MacPherson H, *et al.* Researching complementary and alternative treatments – the gatekeepers are not at home. *BMC Med Res Methodol* 2007;**7**:7. http://dx.doi.org/10.1186/1471-2288-7-7

10. Ernst E, White A. Life-threatening adverse reactions after acupuncture? A systematic review. *Pain* 1997;**71**:123–6. http://dx.doi.org/10.1016/S0304-3959(97)03368-X

11. MacPherson H, Thomas K, Walters S, Fitter M. The York acupuncture safety study: prospective survey of 34 000 treatments by traditional acupuncturists. *BMJ* 2001;**323**:486–7. http://dx.doi.org/10.1136/bmj.323.7311.486

12. White A, Hayhoe S, Hart A, Ernst E. Adverse events following acupuncture: prospective survey of 32 000 consultations with doctors and physiotherapists. *BMJ* 2001;**323**:485–6. http://dx.doi.org/10.1136/bmj.323.7311.485

13. Vincent C. The safety of acupuncture. *BMJ* 2001;**323**:467–8. http://dx.doi.org/10.1136/bmj.323.7311.467

14. Macpherson H, Scullion A, Thomas KJ, Walters S. Patient reports of adverse events associated with acupuncture treatment: a prospective national survey. *Qual Saf Health Care* 2004;**13**:349–55. http://dx.doi.org/10.1136/qshc.2003.009134

15. Hopton AK, Thomas KJ, MacPherson H. Willingness to try acupuncture again: reports from patients on their treatment reactions in a low back pain trial. *Acupunct Med* 2010;**28**:185–8. http://dx.doi.org/10.1136/aim.2010.002279

16. Melchart D, Weidenhammer W, Streng A, Reitmayr S, Hoppe A, Ernst E, *et al.* Prospective investigation of adverse effects of acupuncture in 97 733 patients. *Arch Intern Med* 2004;**164**:104–5. http://dx.doi.org/10.1001/archinte.164.1.104

17. Witt CM, Pach D, Brinkhaus B, Wruck K, Tag B, Mank S, *et al.* Safety of acupuncture: results of a prospective observational study with 229,230 patients and introduction of a medical information and consent form. *Forsch Komplementmed* 2009;**16**:91–7. http://dx.doi.org/10.1159/000209315

18. Pomeranz B. Scientific research into acupuncture for the relief of pain. *J Altern Complement Med* 1996;**2**:53–60. http://dx.doi.org/10.1089/acm.1996.2.53

19. Han JS. Acupuncture: neuropeptide release produced by electrical stimulation of different frequencies. *Trends Neurosci* 2003;**26**:17–22. http://dx.doi.org/10.1016/S0166-2236(02)00006-1

20. Kim SK, Bae H. Acupuncture and immune modulation. *Auton Neurosci* 2010;**157**:38–41. http://dx.doi.org/10.1016/j.autneu.2010.03.010

21. Cabioğlu MT, Cetin BE. Acupuncture and immunomodulation. *Am J Chin Med* 2008;**36**:25–36. http://dx.doi.org/10.1142/S0192415X08005552

22. Zhou W, Fu L-W, Tjen-A-Looi SC, Li P, Longhurst JC. Afferent mechanisms underlying stimulation modality-related modulation of acupuncture-related cardiovascular responses. *J Appl Physiol* 2005;**98**:872–80. http://dx.doi.org/10.1152/japplphysiol.01079.2004

23. Noguchi E. Acupuncture regulates gut motility and secretion via nerve reflexes. *Auton Neurosci* 2010;**156**:15–18. http://dx.doi.org/10.1016/j.autneu.2010.06.010

24. Dhond RP, Kettner N, Napadow V. Neuroimaging acupuncture effects in the human brain. *J Altern Complement Med* 2007;**13**:603–16. http://dx.doi.org/10.1089/acm.2007.7040

25. Zhao ZQ. Neural mechanism underlying acupuncture analgesia. *Prog Neurobiol* 2008;**85**:355–75. http://dx.doi.org/10.1016/j.pneurobio.2008.05.004

26. Li P, Longhurst JC. Neural mechanism of electroacupuncture's hypotensive effects. *Auton Neurosci* 2010;**157**:24–30. http://dx.doi.org/10.1016/j.autneu.2010.03.015

27. Longhurst JC. Defining meridians: a modern basis of understanding. *J Acupunct Meridian Stud* 2010;**3**:67–74. http://dx.doi.org/10.1016/S2005-2901(10)60014-3

28. Kagitani F, Uchida S, Hotta H. Afferent nerve fibers and acupuncture. *Auton Neurosci* 2010;**157**:2–8. http://dx.doi.org/10.1016/j.autneu.2010.03.004

29. Lewith GT, White PJ, Pariente J. Investigating acupuncture using brain imaging techniques: the current state of play. *Evid Based Complement Alternat Med* 2005;**2**:315–19. http://dx.doi.org/10.1093/ecam/neh110

30. Langevin HM, Yandow JA. Relationship of acupuncture points and meridians to connective tissue planes. *Anat Rec* 2002;**269**:257–65. http://dx.doi.org/10.1002/ar.10185

31. Heine H. The morphological basis of the acupuncture points. *Acupunct Sci Int J* 1990;**1**:1–6.

32. Langevin HM, Churchill DL, Wu J, Badger GJ, Yandow JA, Fox JR, *et al.* Evidence of connective tissue involvement in acupuncture. *FASEB J* 2002;**16**:872–4. http://dx.doi.org/10.1096/fj.01-0925fje

33. Becker RO, Reichmanis M, Marino AA. Electrophysiological correlates of acupuncture points and meridians. *Psychoenerg Syst* 1976;**1**:105–12.

34. Ahn AC, Martinsen OG. Electrical characterization of acupuncture points: technical issues and challenges. *J Altern Complement Med* 2007;**13**:817–24. http://dx.doi.org/10.1089/acm.2007.7193

35. Colbert AP, Spaulding KP, Ahn AC, Cutro JA. Clinical utility of electrodermal activity at acupuncture points: a narrative review. *Acupunct Med* 2011;**29**:270–5. http://dx.doi.org/10.1136/acupmed-2011-010021

36. Saku K, Mukaino Y, Ying H, Arakawa K. Characteristics of reactive electropermeable points on the auricles of coronary heart disease patients. *Clin Cardiol* 1993;**16**:415–19. http://dx.doi.org/10.1002/clc.4960160509

37. Ahn AC, Park M, Shaw JR, McManus CA, Kaptchuk TJ, Langevin HM. Electrical impedance of acupuncture meridians: the relevance of subcutaneous collagenous bands. *PLOS ONE* 2010;**5**:e11907. http://dx.doi.org/10.1371/journal.pone.0011907

38. Ahn AC, Colbert AP, Anderson BJ, Martinsen OG, Hammerschlag R, Cina S, *et al.* Electrical properties of acupuncture points and meridians: a systematic review. *Bioelectromagnetics* 2008;**29**:245–56. http://dx.doi.org/10.1002/bem.20403

39. Melzack R, Wall PD. Pain mechanisms: a new theory. *Science* 1965;**150**:971–9. http://dx.doi.org/10.1126/science.150.3699.971

40. Carlsson C. Acupuncture mechanisms for clinically relevant long-term effects – reconsideration and a hypothesis. *Acupunct Med* 2002;**20**:82–99. http://dx.doi.org/10.1136/aim.20.2-3.82

41. Clement-Jones V, McLoughlin L, Tomlin S, Besser GM, Rees LH, Wen HL. Increased beta-endorphin but not met-enkephalin levels in human cerebrospinal fluid after acupuncture for recurrent pain. *Lancet* 1980;**2**:946–9. http://dx.doi.org/10.1016/S0140-6736(80)92106-6

42. Napadow V, Liu J, Li M, Kettner N, Ryan A, Kwong KK, *et al.* Somatosensory cortical plasticity in carpal tunnel syndrome treated by acupuncture. *Hum Brain Mapp* 2007;**28**:159–71. http://dx.doi.org/10.1002/hbm.20261

43. Birch S. A review and analysis of placebo treatments, placebo effects, and placebo controls in trials of medical procedures when sham is not inert. *J Altern Complement Med* 2006;**12**:303–10. http://dx.doi.org/10.1089/acm.2006.12.303

44. Hróbjartsson A, Gøtzsche PC. Placebo interventions for all clinical conditions. *Cochrane Database Syst Rev* 2010;**1**:CD003974. http://dx.doi.org/10.1002/14651858.cd003974.pub3

45. MacPherson H, Vertosick E, Lewith G, Linde K, Sherman KJ, Witt CM, *et al.* Influence of control group on effect size in trials of acupuncture for chronic pain: a secondary analysis of an individual patient data meta-analysis. *PLOS ONE* 2014;**9**:e93739. http://dx.doi.org/10.1371/journal.pone.0093739

46. Hall H. Acupuncture's claims punctured: not proven effective for pain, not harmless. *Pain* 2011;**152**:711–12. http://dx.doi.org/10.1016/j.pain.2011.01.039

47. Ernst E. Acupuncture – a treatment to die for? *J R Soc Med* 2010;**103**:384–5. http://dx.doi.org/10.1258/jrsm.2010.100181

48. Madsen MV, Gøtzsche PC, Hróbjartsson A. Acupuncture treatment for pain: systematic review of randomised clinical trials with acupuncture, placebo acupuncture, and no acupuncture groups. *BMJ* 2009;**338**:a3115. http://dx.doi.org/10.1136/bmj.a3115

49. Ernst E, Lee MS, Choi TY. Acupuncture: does it alleviate pain and are there serious risks? A review of reviews. *Pain* 2011;**152**:755–64. http://dx.doi.org/10.1016/j.pain.2010.11.004

50. Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG, *et al.* A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol* 2009;**62**:464–75. http://dx.doi.org/10.1016/j.jclinepi.2008.12.011

51. Ezzo J, Berman B, Hadhazy VA, Jadad AR, Lao L, Singh BB. Is acupuncture effective for the treatment of chronic pain? A systematic review. *Pain* 2000;**86**:217–25. http://dx.doi.org/10.1016/S0304-3959(99)00304-8

52. Manheimer E, Linde K, Lao L, Bouter LM, Berman BM. Meta-analysis: acupuncture for osteoarthritis of the knee. *Ann Intern Med* 2007;**146**:868–77. http://dx.doi.org/10.7326/0003-4819-146-12-200706190-00008

53. Kwon YD, Pittler MH, Ernst E. Acupuncture for peripheral joint osteoarthritis: a systematic review and meta-analysis. *Rheumatology* 2006;**45**:1331–7. http://dx.doi.org/10.1093/rheumatology/kel207

54. White A, Foster NE, Cummings M, Barlas P. Acupuncture treatment for chronic knee pain: a systematic review. *Rheumatology* 2007;**46**:384–90. http://dx.doi.org/10.1093/rheumatology/kel413

55. Davis MA, Kononowech RW, Rolin SA, Spierings EL. Acupuncture for tension-type headache: a meta-analysis of randomized, controlled trials. *J Pain* 2008;**9**:667–77. http://dx.doi.org/10.1016/j.jpain.2008.03.011

56. Sun Y, Gan TJ. Acupuncture for the management of chronic headache: a systematic review. *Anesth Analg* 2008;**107**:2038–47. http://dx.doi.org/10.1213/ane.0b013e318187c76a

57. Furlan AD, van Tulder M, Cherkin D, Tsukayama H, Lao L, Koes B, *et al.* Acupuncture and dry-needling for low back pain: an updated systematic review within the framework of the Cochrane Collaboration. *Spine* 2005;**30**:944–63. http://dx.doi.org/10.1097/01.brs.0000158941.21571.01

58. Manheimer E, White A, Berman B, Forys K, Ernst E. Meta-analysis: acupuncture for low back pain. *Ann Intern Med* 2005;**142**:651–63. http://dx.doi.org/10.7326/0003-4819-142-8-200504190-00014

59. Ernst E, Lee MS. A trial design that generates only 'positive' results. *J Postgrad Med* 2008;**54**:214–16. http://dx.doi.org/10.4103/0022-3859.41806

60. Vickers AJ, Rees RW, Zollman CE, McCarney R, Smith CM, Ellis N, *et al.* Acupuncture for chronic headache in primary care: large, pragmatic, randomised trial. *BMJ* 2004;**328**:744. http://dx.doi.org/10.1136/bmj.38029.421863.EB

61. Thomas KJ, MacPherson H, Thorpe L, Brazier J, Fitter M, Campbell MJ, *et al.* Randomised controlled trial of a short course of traditional acupuncture compared with usual care for persistent non-specific low back pain. *BMJ* 2006;**333**:623–6. http://dx.doi.org/10.1136/bmj.38878.907361.7C

62. Wonderling D, Vickers AJ, Grieve R, McCarney R. Cost effectiveness analysis of a randomised trial of acupuncture for chronic headache in primary care. *BMJ* 2004;**328**:747. http://dx.doi.org/10.1136/bmj.38033.896505.EB

63. Ratcliffe J, Thomas KJ, MacPherson H, Brazier J. A randomised controlled trial of acupuncture care for persistent low back pain: cost effectiveness analysis. *BMJ* 2006;**333**:626. http://dx.doi.org/10.1136/bmj.38932.806134.7C

64. Linde K, Allais G, Brinkhaus B, Manheimer E, Vickers A, White AR. Acupuncture for tension-type headache. *Cochrane Database Syst Rev* 2009;**1**:CD007587. http://dx.doi.org/10.1002/14651858.cd007587

65. Linde K, Allais G, Brinkhaus B, Manheimer E, Vickers A, White AR. Acupuncture for migraine prophylaxis. *Cochrane Database Syst Rev* 2009;**1**:CD001218. http://dx.doi.org/10.1002/14651858.cd001218.pub2

66. Brinkhaus B, Witt CM, Jena S, Linde K, Streng A, Wagenpfeil S, *et al.* Acupuncture in patients with chronic low back pain: a randomized controlled trial. *Arch Intern Med* 2006;**166**:450–7. http://dx.doi.org/10.1001/archinte.166.4.450

67. Melchart D, Streng A, Hoppe A, Brinkhaus B, Witt C, Wagenpfeil S, *et al.* Acupuncture in patients with tension-type headache: randomised controlled trial. *BMJ* 2005;**331**:376–82. http://dx.doi.org/10.1136/bmj.38512.405440.8F

68. Linde K, Streng A, Jürgens S, Hoppe A, Brinkhaus B, Witt C, *et al.* Acupuncture for patients with migraine: a randomized controlled trial. *JAMA* 2005;**293**:2118–25. http://dx.doi.org/10.1001/jama.293.17.2118

69. Witt C, Brinkhaus B, Jena S, Linde K, Streng A, Wagenpfeil S, *et al.* Acupuncture in patients with osteoarthritis of the knee: a randomised trial. *Lancet* 2005;**366**:136–43. http://dx.doi.org/10.1016/S0140-6736(05)66871-7

70. Scharf HP, Mansmann U, Streitberger K, Witte S, Krämer J, Maier C, *et al.* Acupuncture and knee osteoarthritis: a three-armed randomized trial. *Ann Intern Med* 2006;**145**:12–20. http://dx.doi.org/10.7326/0003-4819-145-1-200607040-00005

71. Haake M, Müller HH, Schade-Brittinger C, Basler HD, Schäfer H, Maier C, *et al.* German Acupuncture Trials (GERAC) for chronic low back pain: randomized, multicenter, blinded, parallel-group trial with 3 groups. *Arch Intern Med* 2007;**167**:1892–8. http://dx.doi.org/10.1001/Archinte.167.17.1892

72. Diener HC, Kronfeld K, Boewing G, Lungenhausen M, Maier C, Molsberger A, *et al.* Efficacy of acupuncture for the prophylaxis of migraine: a multicentre randomised controlled clinical trial. *Lancet Neurol* 2006;**5**:310–16. http://dx.doi.org/10.1016/S1474-4422(06)70382-9

73. Endres HG, Böwing G, Diener HC, Lange S, Maier C, Molsberger A, *et al.* Acupuncture for tension-type headache: a multicentre, sham-controlled, patient-and observer-blinded, randomised trial. *J Headache Pain* 2007;**8**:306–14. http://dx.doi.org/10.1007/s10194-007-0416-5

74. Witt CM, Jena S, Selim D, Brinkhaus B, Reinhold T, Wruck K, *et al.* Pragmatic randomized trial evaluating the clinical and economic effectiveness of acupuncture for chronic low back pain. *Am J Epidemiol* 2006;**164**:487–96. http://dx.doi.org/10.1093/aje/kwj224

75. Witt CM, Jena S, Brinkhaus B, Liecker B, Wegscheider K, Willich SN. Acupuncture for patients with chronic neck pain. *Pain* 2006;**125**:98–106. http://dx.doi.org/10.1016/j.pain.2006.05.013

76. Jena S, Witt CM, Brinkhaus B, Wegscheider K, Willich SN. Acupuncture in patients with headache. *Cephalalgia* 2008;**28**:969–79. http://dx.doi.org/10.1111/j.1468-2982.2008.01640.x

77. Witt CM, Jena S, Brinkhaus B, Liecker B, Wegscheider K, Willich SN. Acupuncture in patients with osteoarthritis of the knee or hip: a randomized, controlled trial with an additional nonrandomized arm. *Arthritis Rheum* 2006;**54**:3485–93. http://dx.doi.org/10.1002/art.22154

78. Chalmers I. The Cochrane Collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. *Ann NY Acad Sci* 1993;**703**:156–63. http://dx.doi.org/10.1111/j.1749-6632.1993.tb26345.x

79. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Eval Health Prof* 2002;**25**:76–97. http://dx.doi.org/10.1177/0163278702025001006

80. NICE. *Headaches: Diagnosis and Management of Headaches in Young People and Adults.* NICE clinical guideline150. London: NICE; 2012. URL: http://guidance.nice.org.uk/CG150 (accessed 20 November 2014).

81. NICE. *Low Back Pain: Early Management of Persistent Non-Specific Low Back Pain*. London: NICE; 2009. pp. 4–25. URL: www.nice.org.uk/cg88 (accessed 20 November 2014).

82. NICE. *Osteoarthritis: Care and Management in Adults*. NICE clinical guideline 177. London: NICE; 2014. URL: www.nice.org.uk/nicemedia/live/14383/66527/66527.pdf (accessed 9 March 2014).

83. Latimer N. NICE guideline on osteoarthritis: is it fair to acupuncture? Yes. *Acupunct Med* 2009;**27**:72–5. http://dx.doi.org/10.1136/aim.2009.000802

84. Cummings M. Why recommend acupuncture for low back pain but not for osteoarthritis? A commentary on recent NICE guidelines. *Acupunct Med* 2009;**27**:128–9. http://dx.doi.org/10.1136/aim.2009.001214

85. Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005;**331**:897–900. http://dx.doi.org/10.1136/bmj.331.7521.897

86. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004;**23**:3105–24. http://dx.doi.org/10.1002/sim.1875

87. Han JS. Acupuncture and endorphins. *Neurosci Lett* 2004;**361**:258–61. http://dx.doi.org/10.1016/j.neulet.2003.12.019

88. Napadow V, Webb JM, Pearson N, Hammerschlag R. Neurobiological correlates of acupuncture: November 17–18, 2005. *J Altern Complement Med* 2006;**12**:931–5. http://dx.doi.org/10.1089/acm.2006.12.931

89. Napadow V, Maeda Y, Audette J, Kettner N. Neuroplasticity in Carpal Tunnel Syndrome. In Knotkova H, Cruciani RA, Merrick J. *Neural Plasticity in Chronic Pain*. New York, NY: Nova Publishers; 2011. pp. 153–84.

90. Barnes PM, Bloom B, Nahin RL. Complementary and alternative medicine use among adults and children: United States, 2007. *Natl Health Stat Report* 2008;**12**:1–23.

91. Xue CC, Zhang AL, Lin V, Myers R, Polus B, Story DF. Acupuncture, chiropractic and osteopathy use in Australia: a national population survey. *BMC Public Health* 2008;**8**:105. http://dx.doi.org/10.1186/1471-2458-8-105

92. Katona C, Peveler R, Dowrick C, Wessely S, Feinmann C, Gask L, *et al.* Pain symptoms in depression: definition and clinical significance. *Clin Med* 2005;**5**:390–5. http://dx.doi.org/10.7861/clinmedicine.5-4-390

93. Mukaino Y, Park J, White A, Ernst E. The effectiveness of acupuncture for depression – a systematic review of randomised controlled trials. *Acupunct Med* 2005;**23**:70–6. http://dx.doi.org/10.1136/aim.23.2.70

94. Smith CA, Hay PP. Acupuncture for depression. *Cochrane Database Syst Rev* 2005;**2**:CD004046.

95. Fava M. Diagnosis and definition of treatment-resistant depression. *Biol Psychiatry* 2003;**53**:649–59. http://dx.doi.org/10.1016/S0006-3223(03)00231-2

96. Gilbody S, Whitty P. *Improving the Recognition and Management of Depression in Primary Care*. York: NHS Centre for Reviews and Dissemination, University of York; 2002. URL: www.york.ac.uk/media/crd/ehc75.pdf (accessed 7 October 2016).

97. Mind. *My Choice: A Survey*. 2002. URL: www.mind.org.uk/News+policy+and+campaigns/Press+archive/Mind+launches+campaign+for+more+choice+of+mental+health+services+at+GP+level.htm (accessed 13 January 2008).

98. Schroer S, Macpherson H. Acupuncture, or non-directive counselling versus usual care for the treatment of depression: a pilot study. *Trials* 2009;**10**:3. http://dx.doi.org/10.1186/1745-6215-10-3

99. Colquhoun D, Novella SP. Acupuncture is theatrical placebo. *Anesth Analg* 2013;**116**:1360–3. http://dx.doi.org/10.1213/ANE.0b013e31828f2d5e

100. Han JS, Ho YS. Global trends and performances of acupuncture research. *Neurosci Biobehav Rev* 2011;**35**:680–7. http://dx.doi.org/10.1016/j.neubiorev.2010.08.006

101. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. *Eur J Pain* 2006;**10**:287–333. http://dx.doi.org/10.1016/j.ejpain.2005.06.009

102. Stewart LA, Parmar MK. Meta-analysis of the literature or of individual patient data: is there a difference? *Lancet* 1993;**341**:418–22. http://dx.doi.org/10.1016/0140-6736(93)93004-K

103. Vickers AJ. Statistical reanalysis of four recent randomized trials of acupuncture for pain using analysis of covariance. *Clin J Pain* 2004;**20**:319–23. http://dx.doi.org/10.1097/00002508-200409000-00006

104. Frison L, Pocock SJ. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. *Stat Med* 1992;**11**:1685–704. http://dx.doi.org/10.1002/sim.4780111304

105. Vickers AJ, Cronin AM, Maschino AC, Lewith G, Macpherson H, Victor N, *et al.* Individual patient data meta-analysis of acupuncture for chronic pain: protocol of the Acupuncture Trialists' Collaboration. *Trials* 2010;**11**:90. http://dx.doi.org/10.1186/1745-6215-11-90

106. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, *et al.* Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;**352**:609–13. http://dx.doi.org/10.1016/S0140-6736(98)01085-X

107. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;**273**:408–12. http://dx.doi.org/10.1001/jama.1995.03520290060030

108. Streitberger K, Kleinhenz J. Introducing a placebo needle into acupuncture research. *Lancet* 1998;**352**:364–5. http://dx.doi.org/10.1016/S0140-6736(97)10471-8

109. Melchart D, Linde K, Fischer P, Berman B, White A, Vickers A, *et al.* Acupuncture for idiopathic headache. *Cochrane Database Syst Rev* 2001;**1**:CD001218. http://dx.doi.org/10.1002/14651858.cd001218

110. Ezzo J, Hadhazy V, Birch S, Lao L, Kaplan G, Hochberg M, *et al.* Acupuncture for osteoarthritis of the knee: a systematic review. *Arthritis Rheum* 2001;**44**:819–25. http://dx.doi.org/10.1002/1529-0131(200104)44:4<819::AID-ANR138>3.0.CO;2-P

111. Vickers A, Goyal N, Harland R, Rees R. Do certain countries produce only positive results? A systematic review of controlled trials. *Control Clin Trials* 1998;**19**:159–66. http://dx.doi.org/10.1016/S0197-2456(97)00150-5

112. Tang JL, Zhan SY, Ernst E. Review of randomised controlled trials of traditional Chinese medicine. *BMJ* 1999;**319**:160–1. http://dx.doi.org/10.1136/bmj.319.7203.160

113. Wang G, Mao B, Xiong ZY, Fan T, Chen XD, Wang L, *et al.* The quality of reporting of randomized controlled trials of traditional Chinese medicine: a survey of 13 randomly selected journals from mainland China. *Clin Ther* 2007;**29**:1456–67. http://dx.doi.org/10.1016/j.clinthera.2007.07.023

114. Wu T, Li Y, Bian Z, Liu G, Moher D. Randomized trials published in some Chinese journals: how many are randomized? *Trials* 2009;**10**:46. http://dx.doi.org/10.1186/1745-6215-10-46

115. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988;**15**:1833–40.

116. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;**21**:1539–58. http://dx.doi.org/10.1002/sim.1186

117. Kerr DP, Walsh DM, Baxter D. Acupuncture in the management of chronic low back pain: a blinded randomized controlled trial. *Clin J Pain* 2003;**19**:364–70. http://dx.doi.org/10.1097/00002508-200311000-00004

118. Irnich D, Behrens N, Molzen H, König A, Gleditsch J, Krauss M, *et al.* Randomised trial of acupuncture compared with conventional massage and 'sham' laser acupuncture for treatment of chronic neck pain. *BMJ* 2001;**322**:1574–8. http://dx.doi.org/10.1136/bmj.322.7302.1574

119. Carlsson CP, Sjölund BH. Acupuncture for chronic low back pain: a randomized placebo-controlled study with long-term follow-up. *Clin J Pain* 2001;**17**:296–305. http://dx.doi.org/10.1097/00002508-200112000-00003

120. Vas J, Méndez C, Perea-Milla E, Vega E, Panadero MD, León JM, *et al.* Acupuncture as a complementary therapy to the pharmacological treatment of osteoarthritis of the knee: randomised controlled trial. *BMJ* 2004;**329**:1216. http://dx.doi.org/10.1136/bmj.38238.601447.3A

121. Vas J, Perea-Milla E, Méndez C, Sánchez Navarro C, León Rubio JM, Brioso M, *et al.* Efficacy and safety of acupuncture for chronic uncomplicated neck pain: a randomised controlled study. *Pain* 2006;**126**:245–55. http://dx.doi.org/10.1016/j.pain.2006.07.002

122. Vas J, Ortega C, Olmo V, Perez-Fernandez F, Hernandez L, Medina I, *et al.* Single-point acupuncture and physiotherapy for the treatment of painful shoulder: a multicentre randomized controlled trial. *Rheumatology (Oxford)* 2008;**47**:887–93. http://dx.doi.org/10.1093/rheumatology/ken040

123. MacPherson H, Maschino AC, Lewith G, Foster NE, Witt CM, Witt C, *et al.* Characteristics of acupuncture treatment associated with outcome: an individual patient meta-analysis of 17,922 patients with chronic pain in randomised controlled trials. *PLOS ONE* 2013;**8**:e77438. http://dx.doi.org/10.1371/journal.pone.0077438

124. Cherkin DC, Sherman KJ, Avins AL, Erro JH, Ichikawa L, Barlow WE, *et al.* A randomized trial comparing acupuncture, simulated acupuncture, and usual care for chronic low back pain. *Arch Intern Med* 2009;**169**:858–66. http://dx.doi.org/10.1001/archinternmed.2009.65

125. Suarez-Almazor ME, Looney C, Liu Y, Cox V, Pietz K, Marcus DM, *et al.* A randomized controlled trial of acupuncture for osteoarthritis of the knee: effects of patient-provider communication. *Arthritis Care Res* 2010;**62**:1229–36. http://dx.doi.org/10.1002/acr.20225

126. Lansdown H, Howard K, Brealey S, MacPherson H. Acupuncture for pain and osteoarthritis of the knee: a pilot study for an open parallel-arm randomised controlled trial. *BMC Musculoskelet Disord* 2009;**10**:130. http://dx.doi.org/10.1186/1471-2474-10-130

127. Molsberger AF, Schneider T, Gotthardt H, Drabik A. German Randomized Acupuncture Trial for chronic shoulder pain (GRASP) – a pragmatic, controlled, patient-blinded, multi-centre trial in an outpatient care environment. *Pain* 2010;**151**:146–54. http://dx.doi.org/10.1016/j.pain.2010.06.036

128. Coeytaux RR, Kaufman JS, Kaptchuk TJ, Chen W, Miller WC, Callahan LF, *et al.* A randomized, controlled trial of acupuncture for chronic daily headache. *Headache* 2005;**45**:1113–23. http://dx.doi.org/10.1111/j.1526-4610.2005.00235.x

129. Molsberger AF, Mau J, Pawelec DB, Winkler J. Does acupuncture improve the orthopedic management of chronic low back pain: a randomized, blinded, controlled trial with 3 months follow up. *Pain* 2002;**99**:579–87. http://dx.doi.org/10.1016/S0304-3959(02)00269-5

130. Kennedy S, Baxter GD, Kerr DP, Bradbury I, Park J, McDonough SM. Acupuncture for acute non-specific low back pain: a pilot randomised non-penetrating sham controlled trial. *Complement Ther Med* 2008;**16**:139–46. http://dx.doi.org/10.1016/j.ctim.2007.03.001

131. Cherkin DC, Eisenberg D, Sherman KJ, Barlow W, Kaptchuk TJ, Street J, *et al.* Randomized trial comparing traditional Chinese medical acupuncture, therapeutic massage, and self-care education for chronic low back pain. *Arch Intern Med* 2001;**161**:1081–8. http://dx.doi.org/10.1001/archinte.161.8.1081

132. White P, Lewith G, Prescott P, Conway J. Acupuncture versus placebo for the treatment of chronic mechanical neck pain: a randomized, controlled trial. *Ann Intern Med* 2004;**141**:911–19. http://dx.doi.org/10.7326/0003-4819-141-12-200412210-00007

133. Salter GC, Roman M, Bland MJ, MacPherson H. Acupuncture for chronic neck pain: a pilot for a randomised controlled trial. *BMC Musculoskelet Disord* 2006;**7**:99. http://dx.doi.org/10.1186/1471-2474-7-99

134. Berman BM, Lao L, Langenberg P, Lee WL, Gilpin AM, Hochberg MC. Effectiveness of acupuncture as adjunctive therapy in osteoarthritis of the knee: a randomized, controlled trial. *Ann Intern Med* 2004;**141**:901–10. http://dx.doi.org/10.7326/0003-4819-141-12-200412210-00006

135. Foster NE, Thomas E, Barlas P, Hill JC, Young J, Mason E, *et al.* Acupuncture as an adjunct to exercise based physiotherapy for osteoarthritis of the knee: randomised controlled trial. *BMJ* 2007;**335**:436. http://dx.doi.org/10.1136/bmj.39280.509803.BE

136. Williamson L, Wyatt MR, Yein K, Melton JT. Severe knee osteoarthritis: a randomized controlled trial of acupuncture, physiotherapy (supervised exercise) and standard management for patients awaiting knee replacement. *Rheumatology (Oxford)* 2007;**46**:1445–9. http://dx.doi.org/10.1093/rheumatology/kem119

137. Kleinhenz J, Streitberger K, Windeler J, Güssbacher A, Mavridis G, Martin E. Randomised clinical trial comparing the effects of acupuncture and a newly designed placebo needle in rotator cuff tendinitis. *Pain* 1999;**83**:235–41. http://dx.doi.org/10.1016/S0304-3959(99)00107-4

138. Guerra de Hoyos JA, Andrés Martín Mdel C, Bassas y Baena de Leon E, Vigára Lopez M, Molina López T, Verdugo Morilla FA, *et al.* Randomised trial of long term effect of acupuncture for shoulder pain. *Pain* 2004;**112**:289–98. http://dx.doi.org/10.1016/j.pain.2004.08.030

139. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;**315**:629–34. http://dx.doi.org/10.1136/bmj.315.7109.629

140. Green S, Buchbinder R, Hetrick S. Acupuncture for shoulder pain. *Cochrane Database Syst Rev* 2005;**2**:CD005319. http://dx.doi.org/10.1002/14651858.cd005319

141. White AR, Ernst E. A systematic review of randomized controlled trials of acupuncture for neck pain. *Rheumatology* 1999;**38**:143–7. http://dx.doi.org/10.1093/rheumatology/38.2.143

142. Manheimer E, Cheng K, Linde K, Lao L, Yoo J, Wieland S, *et al.* Acupuncture for peripheral joint osteoarthritis. *Cochrane Database Syst Rev* 2010;**1**:CD001977. http://dx.doi.org/10.1002/14651858.cd001977.pub2

143. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010;**340**:c221. http://dx.doi.org/10.1136/bmj.c221

144. Kaptchuk TJ, Stason WB, Davis RB, Legedza AR, Schnyer RN, Kerr CE, *et al.* Sham device v inert pill: randomised controlled trial of two placebo treatments. *BMJ* 2006;**332**:391–7. http://dx.doi.org/10.1136/bmj.38726.603310.55

145. Linde K, Niemann K, Meissner K. Are sham acupuncture interventions more effective than (other) placebos? A re-analysis of data from the Cochrane review on placebo effects. *Forsch Komplementrmed* 2010;**17**:259–64. http://dx.doi.org/10.1159/000320374

146. Kaptchuk TJ. The placebo effect in alternative medicine: can the performance of a healing ritual have clinical significance? *Ann Intern Med* 2002;**136**:817–25. http://dx.doi.org/10.7326/0003-4819-136-11-200206040-00011

147. Lundeberg T, Lund I, Sing A, Näslund J. Is placebo acupuncture what it is intended to be? *Evid Based Complement Alternat Med* 2011;**2011**:932407.

148. Birch S, Felt R. *Understanding Acupuncture*. Edinburgh: Churchill Livingstone; 1999.

149. Hughes JG, Goldbart J, Fairhurst E, Knowles K. Exploring acupuncturists' perceptions of treating patients with rheumatoid arthritis. *Complement Ther Med* 2007;**15**:101–8. http://dx.doi.org/10.1016/j.ctim.2006.09.008

150. Paterson C, Britten N. Acupuncture as a complex intervention: a holistic model. *J Altern Complement Med* 2004;**10**:791–801. http://dx.doi.org/10.1089/acm.2004.10.791

151. Witt CM, Lüdtke R, Wegscheider K, Willich SN. Physician characteristics and variation in treatment outcomes: are better qualified and experienced physicians more successful in treating patients with chronic pain with acupuncture? *J Pain* 2010;**11**:431–5. http://dx.doi.org/10.1016/j.jpain.2009.08.010

152. Birch S. Reflections on the German acupuncture studies. *J Chinese Med* 2007;**83**:12–17.

153. Hanfileti L. *What Licensed Acupuncturists Need to Know about the Training and Qualifications of Physicians Providing Medical Acupuncture*. Insights For Acupuncturists; 2013. URL: www.insights-for-acupuncturists.com/medical-acupuncture.html (accessed 16 September 2013).

154. Wampold BE, Brown GS. Estimating variability in outcomes attributable to therapists: a naturalistic study of outcomes in managed care. *J Consult Clin Psychol* 2005;**73**:914–23. http://dx.doi.org/10.1037/0022-006X.73.5.914

155. Lewis M, Morley S, van der Windt DA, Hay E, Jellema P, Dziedzic K, *et al.* Measuring practitioner/therapist effects in randomised trials of low back pain and neck pain interventions in primary care settings. *Eur J Pain* 2010;**14**:1033–9. http://dx.doi.org/10.1016/j.ejpain.2010.04.002

156. Price S, Mercer SW, MacPherson H. Practitioner empathy, patient enablement and health outcomes: a prospective study of acupuncture patients. *Patient Educ Couns* 2006;**63**:239–45. http://dx.doi.org/10.1016/j.pec.2005.11.006

157. Scheid V. Defining Best Practice or Cultivating the Best Practitioner. In Scheid V, MacPherson H, editors. *Integrating East Asian Medicine into Contemporary Healthcare*. Edinburgh: Churchill Livingstone; 2012. pp. 13–38.

158. Peat G, McCarney R, Croft P. Knee pain and osteoarthritis in older adults: a review of community burden and current use of primary health care. *Ann Rheum Dis* 2001;**60**:91–7. http://dx.doi.org/10.1136/ard.60.2.91

159. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. *Ann Rheum Dis* 1957;**16**:494–502. http://dx.doi.org/10.1136/ard.16.4.494

160. Lethbridge-Cejku M, Scott WW, Reichle R, Ettinger WH, Zonderman A, Costa P, *et al.* Association of radiographic features of osteoarthritis of the knee with knee pain: data from the Baltimore Longitudinal Study of Aging. *Arthritis Care Res* 1995;**8**:182–8. http://dx.doi.org/10.1002/art.1790080311

161. Jordan KM, Arden NK, Doherty M, Bannwarth B, Bijlsma JW, Dieppe P, *et al.* EULAR Recommendations 2003: an evidence based approach to the management of knee osteoarthritis: report of a Task Force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). *Ann Rheum Dis* 2003;**62**:1145–55. http://dx.doi.org/10.1136/ard.2003.011742

162. Tramèr MR, Moore RA, Reynolds DJ, McQuay HJ. Quantitative estimation of rare adverse events which follow a biological progression: a new model applied to chronic NSAID use. *Pain* 2000;**85**:169–82. http://dx.doi.org/10.1016/S0304-3959(99)00267-5

163. Pound P, Britten N, Morgan M, Yardley L, Pope C, Daker-White G, *et al.* Resisting medicines: a synthesis of qualitative studies of medicine taking. *Soc Sci Med* 2005;**61**:133–55. http://dx.doi.org/10.1016/j.socscimed.2004.11.063

164. Arthritis Care. *Osteoarthritis Nation: The Most Comprehensive UK Report of People with Osteoarthritis*. London: Arthritis Care; 2004. URL: www.arthritiscare.org.uk/Publications andResources/Forhealthprofessionals/OANation (accessed 30 November 2012).

165. Devos-Comby L, Cronan T, Roesch SC. Do exercise and self-management interventions benefit patients with osteoarthritis of the knee? A metaanalytic review. *J Rheumatol* 2006;**33**:744–56.

166. CRD. *Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care*. York: CRD, University of York; 2009.

167. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;**339**:b2535. http://dx.doi.org/10.1136/bmj.b2535

168. Corbett M, Rice S, Slack R, Harden M, Madurasinghe V, Sutton A, *et al. Acupuncture and Other Physical Treatments for the Relief of Chronic Pain due to Osteoarthritis of the Knee: A Systematic Review and Network Meta-Analysis*. CRD Report 40. York: CRD, University of York; 2011. URL: www.york.ac.uk/media/crd/crdreport40.pdf (accessed 21 December 2016).

169. Corbett MS, Rice SJ, Madurasinghe V, Slack R, Fayter DA, Harden M, *et al.* Acupuncture and other physical treatments for the relief of pain due to osteoarthritis of the knee: network meta-analysis. *Osteoarthritis Cartilage* 2013;**21**:1290–8. http://dx.doi.org/10.1016/j.joca.2013.05.007

170. Rodgers M, McKenna C, Palmer S, Chambers D, Van Hout S, Golder S, *et al.* Curative catheter ablation in atrial fibrillation and typical atrial flutter: systematic review and economic evaluation. *Health Technol Assess* 2008;**12**(34). http://dx.doi.org/10.3310/hta12340

171. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, *et al.* The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;**343**:d5928. http://dx.doi.org/10.1136/bmj.d5928

172. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med* 2010;**29**:932–44. http://dx.doi.org/10.1002/sim.3767

173. Cooper NJ, Sutton AJ, Morris D, Ades AE, Welton NJ. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Stat Med* 2009;**28**:1861–81. http://dx.doi.org/10.1002/sim.3594

174. Welton NJ, Caldwell DM, Adamopoulos E, Vedhara K. Mixed treatment comparison meta-analysis of complex interventions: psychological interventions in coronary heart disease. *Am J Epidemiol* 2009;**169**:1158–65. http://dx.doi.org/10.1093/aje/kwp014

175. Salanti G, Higgins JP, Ades AE, Ioannidis JP. Evaluation of networks of randomized trials. *Stat Methods Med Res* 2008;**17**:279–301. http://dx.doi.org/10.1177/0962280207080643

176. Ades AE, Welton N, Lu G. *Introduction to Mixed Treatment Comparisons*. Bristol: Bristol University; 2007. URL: www.bris.ac.uk/social-community-medicine/media/mpes/intro-to-mtc.pdf (accessed 19 March 2014).

177. Brooks SP, Gelman AJ. General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat* 1998;**7**:434–55.

178. Lu G, Ades AE, Sutton AJ, Cooper NJ, Briggs AH, Caldwell DM. Meta-analysis of mixed treatment comparisons at multiple follow-up times. *Stat Med* 2007;**26**:3681–99. http://dx.doi.org/10.1002/sim.2831

179. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. *J Am Stat Assoc* 2006;**101**:447–59. http://dx.doi.org/10.1198/016214505000001302

180. Spiegelhalter DJ, Best NG, Carlin BP, Van der Linden A. Bayesian measures of model complexity and fit. *J R Stat Soc Ser B* 2002;**64**:583–639. http://dx.doi.org/10.1111/1467-9868.00353

181. Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *J Clin Epidemiol* 2011;**64**:163–71. http://dx.doi.org/10.1016/j.jclinepi.2010.03.016

182. Scholten RJ, de Beurs E, Bouter LM. From effect size into number needed to treat. *Lancet* 1999;**354**:598. http://dx.doi.org/10.1016/S0140-6736(05)77952-6

183. Woolacott NF, Corbett MS, Rice SJ. The use and reporting of WOMAC in the assessment of the benefit of physical therapies for the pain of osteoarthritis of the knee: findings from a systematic review of clinical trials. *Rheumatology* 2012;**51**:1440–6. http://dx.doi.org/10.1093/rheumatology/kes043

184. Bellamy N, Carette S, Ford PM, Kean WF, le Riche NG, Lussier A, *et al.* Osteoarthritis antirheumatic drug trials. III. Setting the delta for clinical trials – results of a consensus development (Delphi) exercise. *J Rheumatol* 1992;**19**:451–7.

185. Ehrich EW, Davies GM, Watson DJ, Bolognese JA, Seidenberg BC, Bellamy N. Minimal perceptible clinical improvement with the Western Ontario and McMaster Universities osteoarthritis index questionnaire and global assessments in patients with osteoarthritis. *J Rheumatol* 2000;**27**:2635–41.

186. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, *et al.* Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. *Ann Rheum Dis* 2005;**64**:29–33. http://dx.doi.org/10.1136/ard.2004.022905

187. Dias RC, Dias JM, Ramos LR. Impact of an exercise and walking protocol on quality of life for elderly people with OA of the knee. *Physiother Res Int* 2003;**8**:121–30. http://dx.doi.org/10.1002/pri.280

188. Weiner DK, Rudy TE, Morone N, Glick R, Kwoh CK. Efficacy of periosteal stimulation therapy for the treatment of osteoarthritis-associated chronic knee pain: an initial controlled clinical trial. *J Am Geriatr Soc* 2007;**55**:1541–7. http://dx.doi.org/10.1111/j.1532-5415.2007.01314.x

189. Aglamiş B, Toraman NF, Yaman H. Change of quality of life due to exercise training in knee osteoarthritis: SF-36 and WOMAC. *J Back Musculoskelet Rehabil* 2009;**22**:43–8. http://dx.doi.org/10.3233/BMR-2009-0219

190. Börjesson M, Robertson E, Weidenhielm L, Mattsson E, Olsson E. Physiotherapy in knee osteoarthrosis: effect on pain and walking. *Physiother Res Int* 1996;**1**:89–97. http://dx.doi.org/10.1002/pri.6120010205

191. Cheing GL, Tsui AY, Lo SK, Hui-Chan CW. Optimal stimulation duration of TENS in the management of osteoarthritic knee pain. *J Rehabil Med* 2003;**35**:62–8. http://dx.doi.org/10.1080/16501970306116

192. Lu TW, Wei IP, Liu YH, Hsu WC, Wang TM, Chang CF, *et al.* Immediate effects of acupuncture on gait patterns in patients with knee osteoarthritis. *Chin Med J* 2010;**123**:165–72.

193. Yip YB, Sit JW, Fung KK, Wong DY, Chong SY, Chung LH, *et al.* Impact of an arthritis self-management programme with an added exercise component for osteoarthritic knee sufferers on improving pain, functional outcomes, and use of health care services: an experimental study. *Patient Educ Couns* 2007;**65**:113–21. http://dx.doi.org/10.1016/j.pec.2006.06.019

194. Jamtvedt G, Dahm KT, Christie A, Moe RH, Haavardsholm E, Holm I, *et al.* Physical therapy interventions for patients with osteoarthritis of the knee: an overview of systematic reviews. *Phys Ther* 2008;**88**:123–36. http://dx.doi.org/10.2522/ptj.20070043

195. Bjordal JM, Ljunggren AE, Klovning A, Slørdal L. Non-steroidal anti-inflammatory drugs, including cyclo-oxygenase-2 inhibitors, in osteoarthritic knee pain: meta-analysis of randomised placebo controlled trials. *BMJ* 2004;**329**:1317. http://dx.doi.org/10.1136/bmj.38273.626655.63

196. Vickers AJ, Cronin AM, Maschino AC, Lewith G, MacPherson H, Foster NE, *et al.* Acupuncture for chronic pain: individual patient data meta-analysis. *Arch Intern Med* 2012;**172**:1444–53. http://dx.doi.org/10.1001/archinternmed.2012.3654

197. Lund I, Näslund J, Lundeberg T. Minimal acupuncture is not a valid placebo control in randomised controlled trials of acupuncture: a physiologist's perspective. *Chin Med* 2009;**4**:1. http://dx.doi.org/10.1186/1749-8546-4-1

198. Hochberg MC, Altman RD, April KT, Benkhalti M, Guyatt G, McGowan J, *et al.* American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. *Arthritis Care Res* 2012;**64**:465–74. http://dx.doi.org/10.1002/acr.21596

199. AAOS. *Treatment of Osteoarthritis of the Knee (Non-Arthroplasty)*. Rosemont, IL: AAOS; 2008. URL: www.aaos.org/research/guidelines/OAKguideline.pdf (accessed 30 November 2012).

200. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, *et al.* OARSI recommendations for the management of hip and knee osteoarthritis, part II: OARSI evidence-based, expert consensus guidelines. *Osteoarthritis Cartilage* 2008;**16**:137–62. http://dx.doi.org/10.1016/j.joca.2007.12.013

201. Dworkin RH, Turk DC, McDermott MP, Peirce-Sandner S, Burke LB, Cowan P, *et al.* Interpreting the clinical importance of group differences in chronic pain clinical trials: IMMPACT recommendations. *Pain* 2009;**146**:238–44. http://dx.doi.org/10.1016/j.pain.2009.08.019

202. Briggs A, Claxton K, Sculpher M. *Decision Modelling for Health Economic Evaluation*. Oxford: Oxford University Press; 2006.

203. NICE. *Guide to the Methods of Technology Appraisal*. 3rd edn. London: NICE; 2012.

204. Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? *Health Econ* 2006;**15**:677–87. http://dx.doi.org/10.1002/hec.1093

205. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997;**50**:683–91. http://dx.doi.org/10.1016/S0895-4356(97)00049-8

206. Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med* 2002;**21**:2313–24. http://dx.doi.org/10.1002/sim.1201

207. Ades AE, Sculpher M, Sutton A, Abrams K, Cooper N, Welton N, *et al.* Bayesian methods for evidence synthesis in cost-effectiveness analysis. *Pharmacoeconomics* 2006;**24**:1–19. http://dx.doi.org/10.2165/00019053-200624010-00001

208. Sutton A, Ades AE, Cooper N, Abrams K. Use of indirect and mixed treatment comparisons for technology assessment. *Pharmacoeconomics* 2008;**26**:753–67. http://dx.doi.org/10.2165/00019053-200826090-00006

209. Stewart LA, Clarke MJ. Practical methodology of meta-analyses (overviews) using updated individual patient data. Cochrane Working Group. *Stat Med* 1995;**14**:2057–79. http://dx.doi.org/10.1002/sim.4780141902

210. Higgins JP, Whitehead A, Turner RM, Omar RZ, Thompson SG. Meta-analysis of continuous outcome data from individual patients. *Stat Med* 2001;**20**:2219–41. http://dx.doi.org/10.1002/sim.918

211. Riley RD, Dodd SR, Craig JV, Thompson JR, Williamson PR. Meta-analysis of diagnostic test studies using individual patient data and aggregate data. *Stat Med* 2008;**27**:6111–36. http://dx.doi.org/10.1002/sim.3441

212. Riley RD, Lambert PC, Staessen JA, Wang J, Gueyffier F, Thijs L, *et al.* Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Stat Med* 2008;**27**:1870–93. http://dx.doi.org/10.1002/sim.3165

213. Saramago P, Manca A, Sutton AJ. Deriving input parameters for cost-effectiveness modeling: taxonomy of data types and approaches to their statistical synthesis. *Value Health* 2012;**15**:639–49. http://dx.doi.org/10.1016/j.jval.2012.02.009

214. Turner RM, Omar RZ, Yang M, Goldstein H, Thompson SG. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Stat Med* 2000;**19**:3417–32. http://dx.doi.org/10.1002/1097-0258(20001230)19:24<3417::AID-SIM614>3.0.CO;2-L

215. Goldstein H, Yang M, Omar R, Turner R, Thompson S. Meta-analysis using multilevel models with an application to the study of class size effects. *J R Stat Soc Series C (Applied Stat)* 2000;**49**:399–412. http://dx.doi.org/10.1111/1467-9876.00200

216. Smith CT, Williamson PR, Marson AG. An overview of methods and empirical comparison of aggregate data and individual patient data results for investigating heterogeneity in meta-analysis of time-to-event outcomes. *J Eval Clin Pract* 2005;**11**:468–78. http://dx.doi.org/10.1111/j.1365-2753.2005.00559.x

217. Welton NJ, Willis SR, Ades AE. Synthesis of survival and disease progression outcomes for health technology assessment of cancer therapies. *Res Synth Methods* 2010;**1**:239–57. http://dx.doi.org/10.1002/jrsm.21

218. Saramago P, Sutton AJ, Cooper NJ, Manca A. Mixed treatment comparisons using aggregate- and individual-participant level data: an efficient use of evidence for cost-effectiveness modelling. *Value Health* 2011;**14**:A237–8. http://dx.doi.org/10.1016/j.jval.2011.08.024

219. Riley RD, Kauser I, Bland M, Thijs L, Staessen JA, Wang J, *et al.* Meta-analysis of randomised trials with a continuous outcome according to baseline imbalance and availability of individual participant data. *Stat Med* 2013;**32**:2747–66. http://dx.doi.org/10.1002/sim.5726

220. Vickers AJ, Altman DG. Statistics notes: analysing controlled trials with baseline and follow up measurements. *BMJ* 2001;**323**:1123–4. http://dx.doi.org/10.1136/bmj.323.7321.1123

221. Dias S, Welton N, Sutton A, Ades A. *NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials*. Sheffield: Decision Support Unit, School of Health and Related Research, University of Sheffield; April 2011. URL: www.nicedsu.org.uk/TSD2%20General%20meta%20analysis%20corrected%2015April2014.pdf (accessed 16 July 2016).

222. Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Med Decis Making* 2013;**33**:607–17. http://dx.doi.org/10.1177/0272989X12458724

223. Higgins J, Green S. *Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0*. Oxford: Wiley-Blackwell; 2011.

224. NICE. *Guide to the Methods of Technology Appraisal 2013*. London: NICE; 2013. URL: www.nice.org.uk/article/pmg9/resources/non-guidance-guide-to-the-methods-of-technology-appraisal-2013-pdf (accessed 16 July 2016).

225. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the Economic Evaluation of Health Care Programmes*. 3rd edn. Oxford: Oxford University Press; 2012.

226. Williams A. EuroQol – a new facility for the measurement of health-related quality-of-life. *Health Policy* 1990;**16**:199–208. http://dx.doi.org/10.1016/0168-8510(90)90421-9

227. Latimer NR, Bhanu AC, Whitehurst DG. Inconsistencies in NICE guidance for acupuncture: reanalysis and discussion. *Acupunct Med* 2012;**30**:182–6. http://dx.doi.org/10.1136/acupmed-2012-010152

228. NICE. *Osteoarthritis: the Care and Management of Osteoarthritis in Adults*. London: National Collaborating Centre for Chronic Conditions; 2008. URL: www.nice.org.uk/CG059 (accessed 30 November 2012).

229. Sculpher MJ, Pang FS, Manca A, Drummond MF, Golder S, Urdahl H, *et al.* Generalisability in economic evaluation studies in healthcare: a review and case studies. *Health Technol Assess* 2004;**8**(49). http://dx.doi.org/10.3310/hta8490

230. UK BEAM Trial Team. United Kingdom back pain exercise and manipulation (UK BEAM) randomised trial: effectiveness of physical treatments for back pain in primary care. *BMJ* 2004;**329**:1377. http://dx.doi.org/10.1136/bmj.38282.669225.AE

231. Underwood M, Ashby D, Cross P, Hennessy E, Letley L, Martin J, *et al.* Advice to use topical or oral ibuprofen for chronic knee pain in older people: randomised controlled trial and patient preference study. *BMJ* 2008;**336**:138–42. http://dx.doi.org/10.1136/bmj.39399.656331.25

232. Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997;**35**:1095–108. http://dx.doi.org/10.1097/00005650-199711000-00002

233. Hernández Alava M, Wailoo AJ, Ara R. Tails from the peak district: adjusted limited dependent variable mixture models of EQ-5D questionnaire health state utility values. *Value Health* 2012;**15**:550–61. http://dx.doi.org/10.1016/j.jval.2011.12.014

234. Dakin HA, Welton NJ, Ades AE, Collins S, Orme M, Kelly S. Mixed treatment comparison of repeated measurements of a continuous endpoint: an example using topical treatments for primary open-angle glaucoma and ocular hypertension. *Stat Med* 2011;**30**:2511–35. http://dx.doi.org/10.1002/sim.4284

235. Dakin H. Review of studies mapping from quality of life or clinical measures to EQ-5D: an online database. *Health Qual Life Outcomes* 2013;**11**:151. http://dx.doi.org/10.1186/1477-7525-11-151

236. Rowen D, Brazier J, Roberts J. Mapping SF-36 onto the EQ-5D index: how reliable is the relationship? *Health Qual Life Outcomes* 2009;**7**:27. http://dx.doi.org/10.1186/1477-7525-7-27

237. Gray AM, Rivero-Arias O, Clarke PM. Estimating the association between SF-12 responses and EQ-5D utility values by response mapping. *Med Decis Making* 2006;**26**:18–29. http://dx.doi.org/10.1177/0272989X05284108

238. Maund E, Craig D, Suekarran S, Neilson A, Wright K, Brealey S, *et al.* Management of frozen shoulder: a systematic review and cost-effectiveness analysis. *Health Technol Assess* 2012;**16**(11). http://dx.doi.org/10.3310/hta16110

239. Barton GR, Sach TH, Jenkinson C, Avery AJ, Doherty M, Muir KR. Do estimates of cost–utility based on the EQ-5D differ from those based on the mapping of utility scores? *Health Qual Life Outcomes* 2008;**6**:51. http://dx.doi.org/10.1186/1477-7525-6-51

240. Khan K. *Mapping Outcomes Data to Estimate Health State Utilities: An Analysis of Individual Patient Level Data from 15 Acupuncture Trials*. York: University of York; 2011.

241. Chan KKW, Willan AR, Gupta M, Pullenayegum E. Underestimation of uncertainties in health utilities derived from mapping algorithms involving health-related quality-of-life measures: statistical explanations and potential remedies. *Med Decis Making* 2014;**34**:863–72. http://dx.doi.org/10.1177/0272989x13517750

242. O'Hagan A, Luce B. *A Primer on Bayesian Statistics in Health Economics and Outcomes Research*. Sheffield: Centre for Bayesian Statistics in Health Economics; 2003.

243. Senn S. Change from baseline and analysis of covariance revisited. *Stat Med* 2006;**25**:4334–44. http://dx.doi.org/10.1002/sim.2682

244. Van Breukelen GJ. ANCOVA versus change from baseline: more power in randomized studies, more bias in nonrandomized studies. *J Clin Epidemiol* 2006;**59**:920–5. http://dx.doi.org/10.1016/j.jclinepi.2006.02.007

245. Dias S, Sutton AJ, Welton NJ, Ades AE. Evidence synthesis for decision making 3: heterogeneity – subgroups, meta-regression, bias, and bias-adjustment. *Med Decis Making* 2013;**33**:618–40. http://dx.doi.org/10.1177/0272989X13485157

246. Lunn D, Thomas A, Best N, Spiegelhalter D. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Stat Computing* 2000;**10**:325–37. http://dx.doi.org/10.1023/A:1008929526011

247. Plummer M, Best N, Cowles K, Vines K. CODA: convergence diagnosis and output analysis for MCMC. *R News* 2006;**6**:7–11.

248. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci* 1992;**7**:457–72. http://dx.doi.org/10.1214/ss/1177011136

249. Curtis L. *Unit Costs of Health and Social Care 2013*. Canterbury: Personal Social Services Research Unit, University of Kent; 2013. URL: www.pssru.ac.uk/archive/pdf/uc/uc2013/full-with-covers.pdf (accessed 16 July 2016).

250. Tudur Smith C, Dwan K, Altman DG, Clarke M, Riley R, Williamson PR. Sharing individual participant data from clinical trials: an opinion survey regarding the establishment of a central repository. *PLOS ONE* 2014;**9**:e97886. http://dx.doi.org/10.1371/journal.pone.0097886

251. Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, *et al.* Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. *Health Technol Assess* 2015;**19**(14). http://dx.doi.org/10.3310/hta19140

252. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, *et al.* Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. *J Rheumatol* 1997;**24**:799–802.

253. Drummond M. Twenty years of using economic evaluations for drug reimbursement decisions: what has been achieved? *J Health Polit Policy Law* 2013;**38**:1081–102. http://dx.doi.org/10.1215/03616878-2373148

254. Urdahl H, Manca A, Sculpher MJ. Assessing generalisability in model-based economic evaluation studies: a structured review in osteoporosis. *Pharmacoeconomics* 2006;**24**:1181–97. http://dx.doi.org/10.2165/00019053-200624120-00004

255. Caro JJ, Briggs AH, Siebert U, Kuntz KM, ISPOR-SMDM Modeling Good Research Practices Task Force. Modeling good research practices – overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. *Value Health* 2012;**15**:796–803. http://dx.doi.org/10.1016/j.jval.2012.06.012

256. National Center for Complementary and Alternative Medicine. *Complementary, Alternative, or Integrative Health: What's In a Name?* Bethesda, MD: National Center for Complementary and Alternative Medicine; 2013. URL: http://nccam.nih.gov/health/whatiscam (accessed 16 July 2016).

257. Eisenberg DM, Davis RB, Ettner SL, Appel S, Wilkey S, Van Rompay M, *et al.* Trends in alternative medicine use in the United States, 1990–1997: results of a follow-up national survey. *JAMA* 1998;**280**:1569–75. http://dx.doi.org/10.1001/jama.280.18.1569

258. Fairfield KM, Eisenberg DM, Davis RB, Libman H, Phillips RS. Patterns of use, expenditures, and perceived efficacy of complementary and alternative therapies in HIV-infected patients. *Arch Intern Med* 1998;**158**:2257–64. http://dx.doi.org/10.1001/archinte.158.20.2257

259. House of Lords. *Report of the Select Committee on Science and Technology: Complementary and Alternative Medicine*. London: The Stationery Office; 2000.

260. Lorenc A, Leach J, Robinson N. Clinical guidelines in the UK: do they mention complementary and alternative medicine (CAM) – are CAM professional bodies aware? *Eur J Integrative Med* 2014;**6**:164–75. http://dx.doi.org/10.1016/j.eujim.2013.11.003

261. CRD. *Acupuncture*. York: CRD, University of York; 2001. *Effective Health Care*, vol. 7, no. 2. URL: www.york.ac.uk/inst/crd/EHC/ehc72.pdf (accessed 19 August 2016).

262. Posadzki P, Alotaibi A, Ernst E. Prevalence of use of complementary and alternative medicine (CAM) by physicians in the UK: a systematic review of surveys. *Clin Med* 2012;**12**:505–12. http://dx.doi.org/10.7861/clinmedicine.12-6-505

263. Laufer S. Osteoarthritis therapy – are there still unmet needs? *Rheumatology* 2004;**43**(Suppl. 1):i9–15. http://dx.doi.org/10.1093/rheumatology/keh103

264. Chang J, Kauf TL, Reed SD, Friedman JY, Omar M, Kahler KH, *et al.* Productivity loss as an indicator of unmet needs for osteoarthritis patients. *Arthritis Rheum* 2003;**48**:S292–3.

265. Selfe TK, Taylor AG. Acupuncture and osteoarthritis of the knee: a review of randomized, controlled trials. *Fam Community Health* 2008;**31**:247–54. http://dx.doi.org/10.1097/01.FCH.0000324482.78577.0f

266. AAOS. *Treatment of Osteoarthritis of the Knee: Evidence-Based Guideline*. 2nd edn. Rosemont, IL: AAOS; 2013. URL: www.aaos.org/research/guidelines/TreatmentofOsteoarthritisoftheKneeGuideline.pdf (accessed 16 July 2016).

267. McAlindon TE, Bannuru RR, Sullivan MC, Arden NK, Berenbaum F, Bierma-Zeinstra SM, *et al.* OARSI guidelines for the non-surgical management of knee osteoarthritis. *Osteoarthr Cartil* 2014;**22**:363–88. http://dx.doi.org/10.1016/j.joca.2014.01.003

268. Scottish Intercollegiate Guidelines Network. *SIGN 136: Management of Chronic Pain. A National Clinical Guideline*. Edinburgh: Scottish Intercollegiate Guidelines Network; 2013. URL: www.sign.ac.uk/pdf/SIGN136.pdf (accessed 20 December 2015).

269. Abdulla A, Adams N, Bone M, Elliott AM, Gaffin J, Jones D, *et al.* Guidance on the management of pain in older people. *Age Ageing* 2013;**42**(Suppl. 1):i1–57. http://dx.doi.org/10.1093/ageing/afs199

270. Brazier JE, Harper R, Munro J, Walters SJ, Snaith ML. Generic and condition-specific outcome measures for people with osteoarthritis of the knee. *Rheumatology* 1999;**38**:870–7. http://dx.doi.org/10.1093/rheumatology/38.9.870

271. Brazier J, Deverill M, Green C, Harper R, Booth A. A review of the use of health status measures in economic evaluation. *Health Technol Assess* 1999;**3**(9).

272. Brooks R. EuroQol: the current state of play. *Health Policy* 1996;**37**:53–72. http://dx.doi.org/10.1016/0168-8510(96)00822-6

273. Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI): concepts, measurement properties and applications. *Health Qual Life Outcomes* 2003;**1**:54. http://dx.doi.org/10.1186/1477-7525-1-54

274. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002;**21**:271–92. http://dx.doi.org/10.1016/S0167-6296(01)00130-8

275. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;**30**:473–83. http://dx.doi.org/10.1097/00005650-199206000-00002

276. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996;**34**:220–33. http://dx.doi.org/10.1097/00005650-199603000-00003

277. Kirwan JR, Reeback JS. Stanford Health Assessment Questionnaire modified to assess disability in British patients with rheumatoid arthritis. *Br J Rheumatol* 1986;**25**:206–9. http://dx.doi.org/10.1093/rheumatology/25.2.206

278. Versteegh MM, Leunis A, Uyl-de Groot CA, Stolk EA. Condition-specific preference-based measures: benefit or burden? *Value Health* 2012;**15**:504–13. http://dx.doi.org/10.1016/j.jval.2011.12.003

279. Thorlund K, Walter SD, Johnston BC, Furukawa TA, Guyatt GH. Pooling health-related quality of life outcomes in meta-analysis – a tutorial and review of methods for enhancing interpretability: enhancing interpretability in continuous meta-analysis. *Res Synthesis Methods* 2011;**2**:188–203. http://dx.doi.org/10.1002/jrsm.46

280. Cummings P. Arguments for and against standardized mean differences (effect sizes). *Arch Pediatr Adolesc Med* 2011;**165**:592–6. http://dx.doi.org/10.1001/archpediatrics.2011.97

281. McKenna C, Bojke L, Manca A, Adebajo A, Dickson J, Helliwell P, *et al.* Shoulder acute pain in primary health care: is retraining GPs effective? The SAPPHIRE randomized trial: a cost-effectiveness analysis. *Rheumatology* 2009;**48**:558–63. http://dx.doi.org/10.1093/rheumatology/kep008

282. Barton GR, Sach TH, Jenkinson C, Doherty M, Avery AJ, Muir KR. Lifestyle interventions for knee pain in overweight and obese adults aged > or = 45: economic evaluation of randomised controlled trial. *BMJ* 2009;**339**:b2273. http://dx.doi.org/10.1136/bmj.b2273

283. Gu NY, Bell C, Botteman MF, Ji X, Carter JA, van Hout B. Estimating preference-based EQ-5D health state utilities or item responses from neuropathic pain scores. *Patient* 2012;**5**:185–97.

284. Fu R, Vandermeer BW, Shamliyan TA, O'Neil ME, Yazdi F, Fox SH, *et al.* Handling Continuous Outcomes in Quantitative Synthesis. In *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. Rockville, MD: Agency for Healthcare Research and Quality; 2008. URL: www.ncbi.nlm.nih.gov/books/NBK154408/ (accessed 11 August 2014).

285. Vickers AJ. Parametric versus non-parametric statistics in the analysis of randomized trials with non-normally distributed data. *BMC Med Res Methodol* 2005;**5**:35. http://dx.doi.org/10.1186/1471-2288-5-35

286. Wooldridge JM. *Introductory Econometrics: A Modern Approach*. 2nd edn. Cincinnati, OH: South-Western College Publishing; 2003. pp. 863.

287. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. *Med Decis Making* 2013;**33**:641–56. http://dx.doi.org/10.1177/0272989X12455847

288. Sturtz S, Ligges U, Gelman A. R2WinBUGS: a package for running WinBUGS from R. *J Stat Software* 2005;**12**:1–16. http://dx.doi.org/10.18637/jss.v012.i03

289. Nüesch E, Reichenbach S, Trelle S, Rutjes AW, Liewald K, Sterchi R, *et al.* The importance of allocation concealment and patient blinding in osteoarthritis trials: a meta-epidemiologic study. *Arthritis Rheum* 2009;**61**:1633–41. http://dx.doi.org/10.1002/art.24894

290. Lunn D. *The BUGS Book: A Practical Introduction to Bayesian Analysis*. Boca Raton, FL: CRC Press; 2013. pp. 381.

291. Curtis L. *Unit Costs of Health and Social Care 2012*. Canterbury: Personal Social Services Research Unit, University of Kent; 2012. URL: www.pssru.ac.uk/archive/pdf/uc/uc2012/full-with-covers.pdf (accessed 16 July 2016).

292. Castelnuovo E, Cross P, Mt-Isa S, Spencer A, Underwood M. Cost-effectiveness of advising the use of topical or oral ibuprofen for knee pain; the TOIB study [ISRCTN: 79353052]. *Rheumatology* 2008;**47**:1077–81. http://dx.doi.org/10.1093/rheumatology/ken128

293. Hawkins N, Scott DA. Cost-effectiveness analysis: discount the placebo at your peril. *Med Decis Making* 2010;**30**:536–43. http://dx.doi.org/10.1177/0272989X10362106

294. Barton GR, Briggs AH, Fenwick EAL. Optimal cost-effectiveness decisions: the role of the cost-effectiveness acceptability curve (CEAC), the cost-effectiveness acceptability frontier (CEAF), and the expected value of perfection information (EVPI). *Value Health* 2008;**11**:886–97. http://dx.doi.org/10.1111/j.1524-4733.2008.00358.x

295. Jordan K, Jinks C, Croft P. A prospective study of the consulting behaviour of older people with knee pain. *Br J Gen Pract* 2006;**56**:269–76.

296. Mitchell HL, Carr AJ, Scott DL. The management of knee pain in primary care: factors associated with consulting the GP and referrals to secondary care. *Rheumatology* 2006;**45**:771–6. http://dx.doi.org/10.1093/rheumatology/kei214

297. Office for National Statistics (ONS). *Taxonomy*. ONS; 2010. URL: www.ons.gov.uk/ons/taxonomy/index.html?nscl=Population+Estimates (accessed 16 October 2014).

298. Mitchell HL, Hurley MV. Management of chronic knee pain: a survey of patient preferences and treatment received. *BMC Musculoskelet Disord* 2008;**9**:123. http://dx.doi.org/10.1186/1471-2474-9-123

299. Claxton K, Palmer S, Longworth L, Bojke L, Griffin S, McKenna C, *et al.* Informing a decision framework for when NICE should recommend the use of health technologies only in the context of an appropriately designed programme of evidence development. *Health Technol Assess* 2012;**16**(46). http://dx.doi.org/10.3310/hta16460

300. AACP. *AACP Homepage*. URL: www.aacp.org.uk/ (accessed 20 December 2015).

301. Adedoyin RA, Olaogun MO, Fagbeja OO. Effect of interferential current stimulation in management of osteo-arthritic knee pain. *Physiotherapy* 2002;**88**:493–9. http://dx.doi.org/10.1016/S0031-9406(05)60851-6

302. Gundog M, Atamaz F, Kanyilmaz S, Kirazli Y, Celepoglu G. Interferential current therapy in patients with knee osteoarthritis: comparison of the effectiveness of different amplitude-modulated frequencies. *Am J Phys Med Rehabil* 2012;**91**:107–13. http://dx.doi.org/10.1097/PHM.0b013e3182328687

303. Burch FX, Tarro JN, Greenberg JJ, Carroll WJ. Evaluating the benefits of patterned stimulation in the treatment of osteoarthritis of the knee: a multi-center, randomized, single-blind, controlled study with an independent masked evaluator. *Osteoarthr Cartil* 2008;**16**:865–72. http://dx.doi.org/10.1016/j.joca.2007.11.013

304. Longworth L, Rowen. *NICE DSU Technical Support Document 10: The Use of Mapping Methods to Estimate Health State Utility Values*. Sheffield; 2011. URL: www.nicedsu.org.uk/TSD%2010%20mapping%20FINAL.pdf (accessed 20 December 2015).

305. Lu G, Brazier JE, Ades AE. Mapping from disease-specific to generic health-related quality-of-life scales: a common factor model. *Value Health* 2013;**16**:177–84. http://dx.doi.org/10.1016/j.jval.2012.07.003

306. Lu G, Kounali D, Ades AE. Simultaneous multioutcome synthesis and mapping of treatment effects to a common scale. *Value Health* 2014;**17**:280–7. http://dx.doi.org/10.1016/j.jval.2013.12.006

307. Shah S, Farrow A, Esnouf A. Availability and use of electrotherapy devices: a survey. *Int J Ther Rehabil* 2007;**14**:260–4. http://dx.doi.org/10.12968/ijtr.2007.14.6.23895

308. Lopez AD, Murray CC. The global burden of disease, 1990–2020. *Nat Med* 1998;**4**:1241–3. http://dx.doi.org/10.1038/3218

309. Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLOS Med* 2008;**5**:e45. http://dx.doi.org/10.1371/journal.pmed.0050045

310. Smith CA, Hay PP, MacPherson H. Acupuncture for depression. *Cochrane Database Syst Rev* 2010;**1**:CD004046. http://dx.doi.org/10.1002/14651858.cd004046.pub3

311. Mellor-Clark J, Simms-Ellis R, Burton M. National survey of counsellors working in primary care: evidence for growing professionalisation? *Occas Pap R Coll Gen Pract* 2001;**79**:vi–7.

312. Bower P, Knowles S, Coventry PA, Rowland N. Counselling for mental health and psychosocial problems in primary care. *Cochrane Database Syst Rev* 2011;**9**:CD001025. http://dx.doi.org/10.1002/14651858.cd001025.pub3

313. MacPherson H, Richmond S, Bland JM, Lansdown H, Hopton A, Kang'ombe A, *et al.* Acupuncture, Counseling, and Usual care for Depression (ACUDep): study protocol for a randomized controlled trial. *Trials* 2012;**13**:209. http://dx.doi.org/10.1186/1745-6215-13-209

314. Beck AT, Steer RA, Brown GK. *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation; 1996.

315. MacPherson H, Schroer S. Acupuncture as a complex intervention for depression: a consensus method to develop a standardised treatment protocol for a randomised controlled trial. *Complement Ther Med* 2007;**15**:92–100. http://dx.doi.org/10.1016/j.ctim.2006.09.006

316. Roth A, Hill A, Pilling S. *The Competences Required to Deliver Effective Humanistic Psychological Therapies*. London: University College London; 2009. URL: www.ucl.ac.uk/clinical-psychology/CORE/Humanistic_Competences/clinicians_guide.pdf (accessed 28 July 2014).

317. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;**16**:606–13. http://dx.doi.org/10.1046/j.1525-1497.2001.016009606.x

318. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006;**3**:77–101. http://dx.doi.org/10.1191/1478088706qp063oa

319. Charmaz K. *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. London: Sage; 2006.

320. Hopton A, Eldred J, MacPherson H. Patients' experiences of acupuncture and counselling for depression and comorbid pain: a qualitative study nested within a randomised controlled trial. *BMJ Open* 2014;**4**:e005144. http://dx.doi.org/10.1136/bmjopen-2014-005144

321. Hopton A, Macpherson H, Keding A, Morley S. Acupuncture, counselling or usual care for depression and comorbid pain: secondary analysis of a randomised controlled trial. *BMJ Open* 2014;**4**:e004964. http://dx.doi.org/10.1136/bmjopen-2014-004964

322. Vannoy SD, Arean P, Unützer J. Advantages of using estimated depression-free days for evaluating treatment efficacy. *Psychiatr Serv* 2010;**61**:160–3. http://dx.doi.org/10.1176/ps.2010.61.2.160

323. Ritchie J, Lewis J. *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. London: Sage; 2003.

324. MacPherson H, Newbronner L, Chamberlain R, Richmond SJ, Lansdown H, Perren S, *et al.* Practitioner perspectives on strategies to promote longer-term benefits of acupuncture or counselling for depression: a qualitative study. *PLOS ONE* 2014;**9**:e104077. http://dx.doi.org/10.1371/journal.pone.0104077

325. Kind P. The EuroQoL Instrument: An Index of Health-Related Quality of Life. In Spilker B, editor. *Quality of Life and Pharmacoeconomics in Clinical Trials*. Philadelphia, PA: Lippincott-Raven; 1996. pp. 191–201.

326. MacPherson H, Tilbrook H, Bland JM, Bloor K, Brabyn S, Cox H, *et al.* Acupuncture for irritable bowel syndrome: primary care based pragmatic randomised controlled trial. *BMC Gastroenterol* 2012;**12**:150. http://dx.doi.org/10.1186/1471-230X-12-150

327. NHS Choices. *Acupuncture*. URL: www.nhs.uk/Conditions/acupuncture/Pages/Introduction.aspx (accessed 16 July 2016).

328. Dolan P, Gudex C. Time preference, duration and health state valuations. *Health Econ* 1995;**4**:289–99. http://dx.doi.org/10.1002/hec.4730040405

329. Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Econ* 2005;**14**:487–96. http://dx.doi.org/10.1002/hec.944

330. Willan AR, Lin DY, Manca A. Regression methods for cost-effectiveness analysis with censored data. *Stat Med* 2005;**24**:131–45. http://dx.doi.org/10.1002/sim.1794

331. Macpherson H, Elliot B, Hopton A, Lansdown H, Richmond S. Acupuncture for depression: patterns of diagnosis and treatment within a randomised controlled trial. *Evid Based Complement Alternat Med* 2013;**2013**:286048. http://dx.doi.org/10.1155/2013/286048

332. Perren S, Richmond S, MacPherson H. The human face of an RCT: reflections on providing counselling for clients with moderate to severe depression in a randomised controlled trial. *Healthc Counsell Psychother J* January 2015, pp. 8–13.

333. Allen JJB, Schnyer R, Hitt SK. The efficacy of acupuncture in the treatment of major depression in women. *Psychol Sci* 1998;**9**:397–401. http://dx.doi.org/10.1111/1467-9280.00074

334. Allen JJ, Schnyer RN, Chambers AS, Hitt SK, Moreno FA, Manber R. Acupuncture for depression: a randomized controlled trial. *J Clin Psychiatry* 2006;**67**:1665–73. http://dx.doi.org/10.4088/JCP.v67n1101

335. Cape J, Whittington C, Buszewicz M, Wallace P, Underwood L. Brief psychological therapies for anxiety and depression in primary care: meta-analysis and meta-regression. *BMC Med* 2010;**8**:38. http://dx.doi.org/10.1186/1741-7015-8-38

336. McMillan D, Gilbody S, Richards D. Defining successful treatment outcome in depression using the PHQ-9: a comparison of methods. *J Affect Disord* 2010;**127**:122–9. http://dx.doi.org/10.1016/j.jad.2010.04.030

337. Burt DB, Zembar MJ, Niederehe G. Depression and memory impairment: a meta-analysis of the association, its pattern, and specificity. *Psychol Bull* 1995;**117**:285–305. http://dx.doi.org/10.1037/0033-2909.117.2.285

338. Hawker GA, Gignac MA, Badley E, Davis AM, French MR, Li Y, *et al.* A longitudinal study to explain the pain–depression link in older adults with osteoarthritis. *Arthritis Care Res* 2011;**63**:1382–90. http://dx.doi.org/10.1002/acr.20298

339. White DK, Wilson JC, Keysor JJ. Measures of adult general functional status: SF-36 Physical Functioning Subscale (PF-10), Health Assessment Questionnaire (HAQ), Modified Health Assessment Questionnaire (MHAQ), Katz Index of Independence in activities of daily living, Functional Independence Measure (FIM), and Osteoarthritis-Function-Computer Adaptive Test (OA-Function-CAT). *Arthritis Care Res* 2011;**63**:S297–307. http://dx.doi.org/10.1002/acr.20638

340. Eccleston C, Williams A de C, Morley S. *Psychological Therapies for the Management of Chronic Pain (Excluding Headache) in Adults*. Cochrane Database of Systematic Reviews. Chichester: John Wiley; 2009. URL: http://doi.wiley.com/10.1002/14651858.CD007407.pub2 (accessed 19 August 2016).

341. Robinson MJ, Edwards SE, Iyengar S, Bymaster F, Clark M, Katon W. Depression and pain. *Front Biosci (Landmark Ed)* 2009;**14**:5031–51. http://dx.doi.org/10.2741/3585

342. Hannes K. Chapter 4: Critical Appraisal of Qualitative Research. In Noyes J, Booth A, Hannes K, Harden A, Harris J, Lewin S, Lockwood C, editors. *Supplementary Guidance for Inclusion of Qualitative Research in Cochrane Systematic Reviews of Interventions*. Version 1 (updated August 2011). Cochrane Collaboration Qualitative Methods Group; 2011. URL: http://cqrmg.cochrane.org/supplemental-handbook-guidance (accessed 31 March 2014).

343. MacPherson H, Thorpe L, Thomas K. Beyond needling – therapeutic processes in acupuncture care: a qualitative study nested within a low-back pain trial. *J Altern Complement Med* 2006;**12**:873–80. http://dx.doi.org/10.1089/acm.2006.12.873

344. Perren S, Godfrey M, Rowland N. The long-term effects of counselling: the process and mechanisms that contribute to ongoing change from a user perspective. *Couns Psychother Res* 2009;**9**:241–9. http://dx.doi.org/10.1080/14733140903150745

345. Yeung AS, Ameral VE, Chuzi SE, Fava M, Mischoulon D. A pilot study of acupuncture augmentation therapy in antidepressant partial and non-responders with major depressive disorder. *J Affect Disord* 2011;**130**:285–9. http://dx.doi.org/10.1016/j.jad.2010.07.025

346. Mischoulon D, Brill CD, Ameral VE, Fava M, Yeung AS. A pilot study of acupuncture monotherapy in patients with major depressive disorder. *J Affect Disord* 2012;**141**:469–73. http://dx.doi.org/10.1016/j.jad.2012.03.023

347. Richmond SJ, Keding A, Hover M, Gabe R, Cross B, Torgerson D, *et al.* Feasibility, acceptability and validity of SMS text messaging for measuring change in depression during a randomised controlled trial. *BMC Psychiatry* 2015;**15**:68. http://dx.doi.org/10.1186/s12888-015-0456-3

348. Keding A, Böhnke JR, Croudace TJ, Richmond SJ, MacPherson H. Validity of single item responses to short message service texts to monitor depression: an mHealth sub-study of the UK ACUDep trial. *BMC Med Res Methodol* 2015;**15**:56. http://dx.doi.org/10.1186/s12874-015-0054-6

349. Abrahams S, Demetriou P. A comparison of the benefits of physiotherapy and anti-inflammatory drugs for osteoarthritis of the knee. *J Orthop Med* 2002;**24**:79–85. http://dx.doi.org/10.1080/1355297X.2002.11736170

350. Alcidi L, Beneforti E, Maresca M, Santosuosso U, Zoppi M. Low power radiofrequency electromagnetic radiation for the treatment of pain due to osteoarthritis of the knee. *Reumatismo* 2007;**59**:140–5.

351. An B, Dai K, Zhu Z, Wang Y, Hao Y, Tang T, *et al.* Baduanjin alleviates the symptoms of knee osteoarthritis. *J Altern Complement Med* 2008;**14**:167–74. http://dx.doi.org/10.1089/acm.2007.0600

352. Arman MI. Therapeutischer Effekt der Kryotherapie bei der aktivierten Gonarthrose. Eine untersucherblinde, kontrollierte Studie. *Z Phys Med Baln Med Klim* 1988;**17**:368–9.

353. Baker KR, Nelson ME, Felson DT, Layne JE, Sarno R, Roubenoff R. The efficacy of home based progressive strength training in older adults with knee osteoarthritis: a randomized controlled trial. *J Rheumatol* 2001;**28**:1655–65.

354. Baker K, Goggins J, Xie H, Szumowski K, LaValley M, Hunter DJ, *et al.* A randomized crossover trial of a wedged insole for treatment of knee osteoarthritis. *Arthritis Rheum* 2007;**56**:1198–203. http://dx.doi.org/10.1002/art.22516

355. Bálint GP, Buchanan WW, Adám A, Ratkó I, Poór L, Bálint PV, *et al.* The effect of the thermal mineral water of Nagybaracska on patients with knee joint osteoarthritis – a double blind study. *Clin Rheumatol* 2007;**26**:890–4. http://dx.doi.org/10.1007/s10067-006-0420-1

356. Bao F, Wu Z. Observation on therapeutic effect of knee osteoarthritis treated by electroacupuncture. *Int J Clin Acupuncture* 2007;**16**:191–5.

357. Berman BM, Singh BB, Lao L, Langenberg P, Li H, Hadhazy V, *et al.* A randomized trial of acupuncture as an adjunctive therapy in osteoarthritis of the knee. *Rheumatology* 1999;**38**:346–54. http://dx.doi.org/10.1093/rheumatology/38.4.346

358. Bezalel T, Carmeli E, Katz-Leurer M. The effect of a group education programme on pain and function through knowledge acquisition and home-based exercise among patients with knee osteoarthritis: a parallel randomised single-blind clinical trial. *Physiotherapy* 2010;**96**:137–43. http://dx.doi.org/10.1016/j.physio.2009.09.009

359. Bilgici A, Akdeniz O, Kuru O, Ulusoy H. The effect of a home-based exercise therapy versus an aerobic exercise programme on pain and functional disability in patients with knee osteoarthritis. *Ann Rheum Dis* 2004;**63**:364–5.

360. Brismée JM, Paige RL, Chyu MC, Boatright JD, Hagar JM, McCaleb JA, *et al.* Group and home-based tai chi in elderly subjects with knee osteoarthritis: a randomized controlled trial. *Clin Rehabil* 2007;**21**:99–111. http://dx.doi.org/10.1177/0269215506070505

361. Bülow PM, Jensen H, Danneskiold-Samsøe B. Low power Ga-Al-As laser treatment of painful osteoarthritis of the knee. A double-blind placebo-controlled study. *Scand J Rehabil Med* 1994;**26**:155–9.

362. Callaghan MJ, Oldham JA, Hunt J. An evaluation of exercise regimes for patients with osteoarthritis of the knee: a single-blind randomized controlled trial. *Clin Rehabil* 1995;**9**:213–18. http://dx.doi.org/10.1177/026921559500900306

363. Callaghan MJ, Whittaker PE, Grimes S, Smith L. An evaluation of pulsed shortwave on knee osteoarthritis using radioleucoscintigraphy: a randomised, double blind, controlled trial. *Joint Bone Spine* 2005;**72**:150–5. http://dx.doi.org/10.1016/j.jbspin.2004.03.010

364. Cantarini L, Leo G, Giannitti C, Cevenini G, Barberini P, Fioravanti A. Therapeutic effect of spa therapy and short wave therapy in knee osteoarthritis: a randomized, single blind, controlled trial. *Rheumatol Int* 2007;**27**:523–9. http://dx.doi.org/10.1007/s00296-006-0266-5

365. Cheing GL, Hui-Chan CW, Chan KM. Does four weeks of TENS and/or isometric exercise produce cumulative reduction of osteoarthritic knee pain? *Clin Rehabil* 2002;**16**:749–60. http://dx.doi.org/10.1191/0269215502cr549oa

366. Chen CY, Chen CL, Hsu SC, Chou SW, Wang KC. Effect of magnetic knee wrap on quadriceps strength in patients with symptomatic knee osteoarthritis. *Arch Phys Med Rehabil* 2008;**89**:2258–64. http://dx.doi.org/10.1016/j.apmr.2008.05.019

367. Christensen R, Astrup A, Bliddal H. Weight loss: the treatment of choice for knee osteoarthritis? A randomized trial. *Osteoarthr Cartil* 2005;**13**:20–7. http://dx.doi.org/10.1016/j.joca.2004.10.008

368. Clarke GR, Willis LA, Stenners L, Nichols PJ. Evaluation of physiotherapy in the treatment of osteoarthrosis of the knee. *Rheumatol Rehabil* 1974;**13**:190–7. http://dx.doi.org/10.1093/rheumatology/13.4.190

369. Defrin R, Ariel E, Peretz C. Segmental noxious versus innocuous electrical stimulation for chronic pain relief and the effect of fading sensation during treatment. *Pain* 2005;**115**:152–60. http://dx.doi.org/10.1016/j.pain.2005.02.018

370. Durmuş D, Alayli G, Cantürk F. Effects of quadriceps electrical stimulation program on clinical parameters in the patients with knee osteoarthritis. *Clin Rheumatol* 2007;**26**:674–8. http://dx.doi.org/10.1007/s10067-006-0358-3

371. Ettinger WH, Burns R, Messier SP, Applegate W, Rejeski WJ, Morgan T, *et al.* A randomized trial comparing aerobic exercise and resistance exercise with a health education program in older adults with knee osteoarthritis. The Fitness Arthritis and Seniors Trial (FAST). *JAMA* 1997;**277**:25–31. http://dx.doi.org/10.1001/jama.1997.03540250033028

372. Fioravanti A, Iacoponi F, Bellisai B, Cantarini L, Galeazzi M. Short- and long-term effects of spa therapy in knee osteoarthritis. *Am J Phys Med Rehabil* 2010;**89**:125–32. http://dx.doi.org/10.1097/PHM.0b013e3181c1eb81

373. Fischer G, Pelka RB, Barovic J. [Adjuvant treatment of knee osteoarthritis with weak pulsing magnetic fields. Results of a placebo-controlled trial prospective clinical trial]. *Z Orthop Ihre Grenzgeb* 2005;**143**:544–50. http://dx.doi.org/10.1055/s-2005-836830

374. Flusser D, Abu-Shakra M, Friger M, Codish S, Sukenik S. Therapy with mud compresses for knee osteoarthritis: comparison of natural mud preparations with mineral-depleted mud. *J Clin Rheumatol* 2002;**8**:197–203. http://dx.doi.org/10.1097/00124743-200208000-00003

375. Forestier R, Desfour H, Tessier JM, Françon A, Foote AM, Genty C, *et al.* Spa therapy in the treatment of knee osteoarthritis: a large randomised multicentre trial. *Ann Rheum Dis* 2010;**69**:660–5. http://dx.doi.org/10.1136/ard.2009.113209

376. Fukuda TY, Ovanessian V, Cunha RAD, Filho ZJ, Cazarini C, Rienzo FA, *et al.* Pulsed short wave effect in pain and function in patients with knee osteoarthritis. *J Applied Res* 2008;**8**:189–98.

377. Garland D, Holt P, Harrington JT, Caldwell J, Zizic T, Cholewczynski J. A 3-month, randomized, double-blind, placebo-controlled study to evaluate the safety and efficacy of a highly optimized, capacitively coupled, pulsed electrical stimulator in patients with osteoarthritis of the knee. *Osteoarthr Cartil* 2007;**15**:630–7. http://dx.doi.org/10.1016/j.joca.2007.01.004

378. Grimmer K. A controlled double blind study comparing the effects of strong burst mode TENS and high rate TENS on painful osteoarthritic knees. *Aust J Physiother* 1992;**38**:49–56. http://dx.doi.org/10.1016/S0004-9514(14)60551-1

379. Gür H, Cakin N, Akova B, Okay E, Küçükoğlu S. Concentric versus combined concentric-eccentric isokinetic training: effects on functional capacity and symptoms in patients with osteoarthrosis of the knee. *Arch Phys Med Rehabil* 2002;**83**:308–16. http://dx.doi.org/10.1053/apmr.2002.30620

380. Gür A, Cosut A, Sarac AJ, Cevik R, Nas K, Uyar A. Efficacy of different therapy regimes of low-power laser in painful osteoarthritis of the knee: a double-blind and randomized-controlled trial. *Lasers Surg Med* 2003;**33**:330–8. http://dx.doi.org/10.1002/lsm.10236

381. Hasegawa R, Islam MM, Nasu E, Tomiyama N, Lee SC, Koizumi D, *et al.* Effects of combined balance and resistance exercise on reducing knee pain in community-dwelling older adults. *Phys Occup Ther Geriatr* 2010;**28**:44–56. http://dx.doi.org/10.3109/02703180903381086

382. Hay EM, Foster NE, Thomas E, Peat G, Phelan M, Yates HE, *et al.* Effectiveness of community physiotherapy and enhanced pharmacy review for knee pain in people aged over 55 presenting to primary care: pragmatic randomised trial. *BMJ* 2006;**333**:995. http://dx.doi.org/10.1136/bmj.38977.590752.0B

383. Hinman MR, Ford J, Heyl H. Effects of static magnets on chronic knee pain and physical function: a double-blind study. *Altern Ther Health Med* 2002;**8**:50–5.

384. Huang MH, Lin YS, Lee CL, Yang RC. Use of ultrasound to increase effectiveness of isokinetic exercise for knee osteoarthritis. *Arch Phys Med Rehabil* 2005;**86**:1545–51. http://dx.doi.org/10.1016/j.apmr.2005.02.007

385. Hurley MV, Walsh NE, Mitchell HL, Pimm TJ, Patel A, Williamson E, *et al.* Clinical effectiveness of a rehabilitation program integrating exercise, self-management, and active coping strategies for chronic knee pain: a cluster randomized trial. *Arthritis Rheum* 2007;**57**:1211–19. http://dx.doi.org/10.1002/art.22995

386. Itoh K, Hirota S, Katsumi Y, Ochi H, Kitakoji H. A pilot study on using acupuncture and transcutaneous electrical nerve stimulation (TENS) to treat knee osteoarthritis (OA). *Chin Med* 2008;**3**:2. http://dx.doi.org/10.1186/1749-8546-3-2

387. Itoh K, Hirota S, Katsumi Y, Ochi H, Kitakoji H. Trigger point acupuncture for treatment of knee osteoarthritis – a preliminary RCT for a pragmatic trial. *Acupunct Med* 2008;**26**:17–26. http://dx.doi.org/10.1136/aim.26.1.17

388. Jacobson JI, Gorman R, Yamanashi WS, Saxena BB, Clayton L. Low-amplitude, extremely low frequency magnetic fields for the treatment of osteoarthritic knees: a double-blind clinical study. *Altern Ther Health Med* 2001;**7**:54–64, 66–9.

389. Jan MH, Lin JJ, Liau JJ, Lin YF, Lin DH. Investigation of clinical effects of high- and low-resistance training for patients with knee osteoarthritis: a randomized controlled trial. *Phys Ther* 2008;**88**:427–36. http://dx.doi.org/10.2522/ptj.20060300

390. Jenkinson CM, Doherty M, Avery AJ, Read A, Taylor MA, Sach TH, *et al.* Effects of dietary intervention and quadriceps strengthening exercises on pain and function in overweight people with knee pain: randomised controlled trial. *BMJ* 2009;**339**:b3170. http://dx.doi.org/10.1136/bmj.b3170

391. Jubb RW, Tukmachi ES, Jones PW, Dempsey E, Waterhouse L, Brailsford S. A blinded randomised trial of acupuncture (manual and electroacupuncture) compared with a non-penetrating sham for the symptoms of osteoarthritis of the knee. *Acupunct Med* 2008;**26**:69–78. http://dx.doi.org/10.1136/aim.26.2.69

392. Kang T, Kim S, Kim S. The efficacy of low power laser therapy in patients with knee osteoarthritis. *Ann Rheum Dis* 2006;**65**(Suppl. 2):233.

393. Kang RW, Lewis PB, Kramer A, Hayden JK, Cole BJ. Prospective randomized single-blinded controlled clinical trial of percutaneous neuromodulation pain therapy device versus sham for the osteoarthritic knee: a pilot study. *Orthopedics* 2007;**30**:439–45.

394. Karagülle M, Karagülle MZ, Karagülle O, Dönmez A, Turan M. A 10-day course of spa therapy is beneficial for people with severe knee osteoarthritis. A 24-week randomised, controlled pilot study. *Clin Rheumatol* 2007;**26**:2063–71. http://dx.doi.org/10.1007/s10067-007-0618-x

395. Keefe FJ, Blumenthal J, Baucom D, Affleck G, Waugh R, Caldwell DS, *et al.* Effects of spouse-assisted coping skills training and exercise training in patients with osteoarthritic knee pain: a randomized controlled study. *Pain* 2004;**110**:539–49. http://dx.doi.org/10.1016/j.pain.2004.03.022

396. Keogan F, Gilsenan C, Hussey J, O'Connell P. Open or closed chain quadriceps exercises in treatment of osteoarthritis of the knee; which is more effective? A blinded randomised controlled trial. *Physiother Ire* 2007;**28**:47–8.

397. Kovács I, Bender T. The therapeutic effects of Cserkeszölö thermal water in osteoarthritis of the knee: a double blind, controlled, follow-up study. *Rheumatol Int* 2002;**21**:218–21. http://dx.doi.org/10.1007/s00296-001-0167-6

398. Kovar PA, Allegrante JP, MacKenzie CR, Peterson MG, Gutin B, Charlson ME. Supervised fitness walking in patients with osteoarthritis of the knee. A randomized, controlled trial. *Ann Intern Med* 1992;**116**:529–34. http://dx.doi.org/10.7326/0003-4819-116-7-529

399. Kuptniratsaikul V, Tosayanonda O, Nilganuwong S, Thamalikitkul V. The efficacy of a muscle exercise program to improve functional performance of the knee in patients with osteoarthritis. *J Med Assoc Thai* 2002;**85**:33–40.

400. Law PP, Cheing GL. Optimal stimulation frequency of transcutaneous electrical nerve stimulation on people with knee osteoarthritis. *J Rehabil Med* 2004;**36**:220–5. http://dx.doi.org/10.1080/16501970410029834

401. Lee HJ, Park HJ, Chae Y, Kim SY, Kim SN, Kim ST, *et al.* Tai Chi Qigong for the quality of life of patients with knee osteoarthritis: a pilot, randomized, waiting list controlled trial. *Clin Rehabil* 2009;**23**:504–11. http://dx.doi.org/10.1177/0269215508101746

402. Lewis B, Lewis D, Cumming G. The analgesic efficacy of transcutaneous electrical nerve stimulation (TENS) compared with a non-steroidal anti-inflammatory drug (naprosyn) in painful osteoarthritis (OA) of the knee. *Aust NZ J Med* 1988;**18**(Suppl. 1):224.

403. Lewis B, Lewis D, Cumming G. The comparative analgesic efficacy of transcutaneous electrical nerve stimulation and a non-steroidal anti-inflammatory drug for painful osteoarthritis. *Br J Rheumatol* 1994;**33**:455–60. http://dx.doi.org/10.1093/rheumatology/33.5.455

404. Lewis D, Lewis B, Sturrock RD. Transcutaneous electrical nerve stimulation in osteoarthrosis: a therapeutic alternative? *Ann Rheum Dis* 1984;**43**:47–9. http://dx.doi.org/10.1136/ard.43.1.47

405. Lim BW, Hinman RS, Wrigley TV, Sharma L, Bennell KL. Does knee malalignment mediate the effects of quadriceps strengthening on knee adduction moment, pain, and function in medial knee osteoarthritis? A randomized controlled trial. *Arthritis Rheum* 2008;**59**:943–51. http://dx.doi.org/10.1002/art.23823

406. Lin D-H, Lin C-HJ, Lin Y-F, Jan M-H. Efficacy of 2 non-weight-bearing interventions, proprioception training versus strength training, for patients with knee osteoarthritis: a randomized clinical trial. *J Orthop Sports Phys Ther* 2009;**39**:450–7. http://dx.doi.org/10.2519/jospt.2009.2923

407. Lund H, Weile U, Christensen R, Rostock B, Downey A, Bartels EM, *et al.* A randomized controlled trial of aquatic and land-based exercise in patients with knee osteoarthritis. *J Rehabil Med* 2008;**40**:137–44. http://dx.doi.org/10.2340/16501977-0134

408. Mahboob N, Sousan K, Shirzad A, Amir G, Mohammad V, Reza M, *et al.* The efficacy of a topical gel prepared using Lake Urmia mud in patients with knee osteoarthritis. *J Altern Complement Med* 2009;**15**:1239–42. http://dx.doi.org/10.1089/acm.2009.0304

409. Maillefert JF, Hudry C, Baron G, Kieffert P, Bourgeois P, Lechevalier D, *et al.* Laterally elevated wedged insoles in the treatment of medial knee osteoarthritis: a prospective randomized controlled study. *Osteoarthr Cartil* 2001;**9**:738–45. http://dx.doi.org/10.1053/joca.2001.0470

410. Maurer BT, Stern AG, Kinossian B, Cook KD, Schumacher HR. Osteoarthritis of the knee: isokinetic quadriceps exercise versus an educational intervention. *Arch Phys Med Rehabil* 1999;**80**:1293–9. http://dx.doi.org/10.1016/S0003-9993(99)90032-1

411. Mazzuca SA, Page MC, Meldrum RD, Brandt KD, Petty-Saphon S. Pilot study of the effects of a heat-retaining knee sleeve on joint pain, stiffness, and function in patients with knee osteoarthritis. *Arthritis Rheum* 2004;**51**:716–21. http://dx.doi.org/10.1002/art.20683

412. McCarthy CJ, Mills PM, Pullen R, Richardson G, Hawkins N, Roberts CR, *et al.* Supplementation of a home-based exercise programme with a class-based programme for people with osteoarthritis of the knees: a randomised controlled trial and health economic analysis. *Health Technol Assess* 2004;**8**(46). http://dx.doi.org/10.3310/hta8460

413. Messier SP, Loeser RF, Miller GD, Morgan TM, Rejeski WJ, Sevick MA, *et al.* Exercise and dietary weight loss in overweight and obese older adults with knee osteoarthritis: the Arthritis, Diet, and Activity Promotion Trial. *Arthritis Rheum* 2004;**50**:1501–10. http://dx.doi.org/10.1002/art.20256

414. Miller GD, Nicklas BJ, Davis C, Loeser RF, Lenchik L, Messier SP. Intensive weight loss program improves physical function in older obese adults with knee osteoarthritis. *Obesity* 2006;**14**:1219–30. http://dx.doi.org/10.1038/oby.2006.139

415. Miranda-Filloy JA, Barbazán AC, Monteagudo Sánchez B, Graña GJ, Galdo Fernández F. [Effect of transcutaneous electrical quadriceps muscle stimulation in knee osteoarthritis symptomatology]. *Rehabilitación* 2005;**39**:167–70. http://dx.doi.org/10.1016/S0048-7120(05)74340-1

416. Molsberger A, Böwing G, Jensen KU, Lorek M. [Acupuncture treatment for the relief of gonarthrosis pain – a controlled clinical trial]. *Schmerz* 1994;**8**:37–42. http://dx.doi.org/10.1007/BF02527508

417. Moss P, Sluka K, Wright A. The initial effects of knee joint mobilization on osteoarthritic hyperalgesia. *Man Ther* 2007;**12**:109–18. http://dx.doi.org/10.1016/j.math.2006.02.009

418. Ng MML, Leung MCP, Poon DMY. The effects of electro-acupuncture and transcutaneous electrical nerve stimulation on patients with painful osteoarthritic knees: a randomized controlled trial with follow-up evaluation. *J Altern Complement Med* 2003;**9**:641–9. http://dx.doi.org/10.1089/107555303322524490

419. Nguyen M, Revel M, Dougados M. Prolonged effects of 3 week therapy in a spa resort on lumbar spine, knee and hip osteoarthritis: follow-up after 6 months. A randomized controlled trial. *Br J Rheumatol* 1997;**36**:77–81. http://dx.doi.org/10.1093/rheumatology/36.1.77

420. Ni GX, Song L, Yu B, Huang CH, Lin JH. Tai chi improves physical function in older Chinese women with knee osteoarthritis. *J Clin Rheumatol* 2010;**16**:64–7. http://dx.doi.org/10.1097/RHU.0b013e3181cf344f

421. Nigg BM, Emery C, Hiemstra LA. Unstable shoe construction and reduction of pain in osteoarthritis patients. *Med Sci Sports Exerc* 2006;**38**:1701–8. http://dx.doi.org/10.1249/01.mss.0000228364.93703.53

422. Nivbrant B, Friberg S. Therapeutic laser treatment in gonarthrosis. *Acta Orthop* 1989;**60**(Suppl. 231):33.

423. Péloquin L, Bravo G, Gauthier P, Lacombe G, Billiard JS. Effects of a cross-training exercise program in persons with osteoarthritis of the knee a randomized controlled trial. *J Clin Rheumatol* 1999;**5**:126–36. http://dx.doi.org/10.1097/00124743-199906000-00004

424. Perlman AI, Sabina A, Williams AL, Njike VY, Katz DL. Massage therapy for osteoarthritis of the knee: a randomized controlled trial. *Arch Intern Med* 2006;**166**:2533–8. http://dx.doi.org/10.1001/archinte.166.22.2533

425. Perrot S, Marty M, Kahan A, Menkes CJ. *Efficacy of Pulsed Electromagnetic Therapy in Painful Knee Osteoarthritis*. Proceedings of the 62nd Annual Meeting of the American College of Rheumatology, San Diego, CA, USA, November 1998. Abstract S357.

426. Petrou P, Winkler V, Genti G, Balint G. Double blind trial to evaluate the effect of acupuncture treatment on knee osteoarthritis. *Scand J Acupunc Electrother* 1988;**3**:112–15.

427. Pietrosimone BG, Hart JM, Saliba SA, Hertel J, Ingersoll CD. Immediate effects of transcutaneous electrical nerve stimulation and focal knee joint cooling on quadriceps activation. *Med Sci Sports Exerc* 2009;**41**:1175–81. http://dx.doi.org/10.1249/MSS.0b013e3181982557

428. Pipitone N, Scott DL. Magnetic pulse treatment for knee osteoarthritis: a randomised, double-blind, placebo-controlled study. *Curr Med Res Opin* 2001;**17**:190–6. http://dx.doi.org/10.1185/03007990152673828

429. Pollard H, Ward G, Hoskins W, Hardy K. The effect of a manual therapy knee protocol on osteoarthritic knee pain: a randomised controlled trial. *J Can Chiropr Assoc* 2008;**52**:229–42.

430. Rapp W, Boer J, Albrich C, Heitkamp HC. Efficiency of vibration or strength training for knee stability in osteoarthritis of the knee. *Aktuelle Rheumatol* 2009;**34**:240–5. http://dx.doi.org/10.1055/s-0029-1225366

431. Rattanachaiyanont M, Kuptniratsaikul V. No additional benefit of shortwave diathermy over exercise program for knee osteoarthritis in peri-/post-menopausal women: an equivalence trial. *Osteoarthr Cartil* 2008;**16**:823–8. http://dx.doi.org/10.1016/j.joca.2007.10.013

432. Røgind H, Bibow-Nielsen B, Jensen B, Møller HC, Frimodt-Møller H, Bliddal H. The effects of a physical training program on patients with osteoarthritis of the knees. *Arch Phys Med Rehabil* 1998;**79**:1421–7. http://dx.doi.org/10.1016/S0003-9993(98)90238-6

433. Rosemffet MG, Schneeberger EE, Citera G, Sgobba ME, Laiz C, Schmulevich H, *et al.* Effects of functional electrostimulation on pain, muscular strength, and functional capacity in patients with osteoarthritis of the knee. *J Clin Rheumatol* 2004;**10**:246–9. http://dx.doi.org/10.1097/01.rhu.0000141831.40350.91

434. Sangdee C, Teekachunhatean S, Sananpanich K, Sugandhavesa N, Chiewchantanakit S, Pojchamarnwiputh S, *et al.* Electroacupuncture versus diclofenac in symptomatic treatment of osteoarthritis of the knee: a randomized controlled trial. *BMC Complement Altern Med* 2002;**2**:3. http://dx.doi.org/10.1186/1472-6882-2-3

435. Schilke JM, Johnson GO, Housh TJ, O'Dell JR. Effects of muscle-strength training on the functional status of patients with osteoarthritis of the knee joint. *Nurs Res* 1996;**45**:68–72. http://dx.doi.org/10.1097/00006199-199603000-00002

436. Selfe TK, Bourguignon C, Taylor AG. Effects of noninvasive interactive neurostimulation on symptoms of osteoarthritis of the knee: a randomized, sham-controlled pilot study. *J Altern Complement Med* 2008;**14**:1075–81. http://dx.doi.org/10.1089/acm.2008.0305

437. Shen X, Zhao L, Ding G, Tan M, Gao J, Wang L, *et al.* Effect of combined laser acupuncture on knee osteoarthritis: a pilot study. *Lasers Med Sci* 2009;**24**:129–36. http://dx.doi.org/10.1007/s10103-007-0536-9

438. Sherman G, Zeller L, Avriel A, Friger M, Harari M, Sukenik S. Intermittent balneotherapy at the Dead Sea area for patients with knee osteoarthritis. *Isr Med Assoc J* 2009;**11**:88–93.

439. Smith CR, Lewith GT, Machin D. TNS and osteo-arthritic pain. Preliminary study to establish a controlled method of assessing transcutaneous nerve stimulation as a treatment for the pain caused by osteo-arthritis of the knee. *Physiotherapy* 1983;**69**:266–8.

440. Stelian J, Gil I, Habot B, Rosenthal M, Abramovici I, Kutok N, *et al.* Improvement of pain and disability in elderly patients with degenerative osteoarthritis of the knee treated with narrow-band light therapy. *J Am Geriatr Soc* 1992;**40**:23–6. http://dx.doi.org/10.1111/j.1532-5415.1992.tb01824.x

441. Sukenik S, Flusser D, Codish S, Abu-Shakra M. Balneotherapy at the Dead Sea area for knee osteoarthritis. *Isr Med Assoc J* 1999;**1**:83–5.

442. Takeda W, Wessel J. Acupuncture for the treatment of pain of osteoarthritic knees. *Arthritis Care Res* 1994;**7**:118–22. http://dx.doi.org/10.1002/art.1790070304

443. Talbot LA, Gaines JM, Ling SM, Metter EJ. A home-based protocol of electrical muscle stimulation for quadriceps muscle strength in older adults with osteoarthritis of the knee. *J Rheumatol* 2003;**30**:1571–8.

444. Tascioglu F, Armagan O, Tabak Y, Corapci I, Oner C. Low power laser treatment in patients with knee osteoarthritis. *Swiss Med Wkly* 2004;**134**:254–8.

445. Taylor P, Hallett M, Flaherty L. Treatment of osteoarthritis of the knee with transcutaneous electrical nerve stimulation. *Pain* 1981;**11**:233–40. http://dx.doi.org/10.1016/0304-3959(81)90008-7

446. Thamsborg G, Florescu A, Oturai P, Fallentin E, Tritsaris K, Dissing S. Treatment of knee osteoarthritis with pulsed electromagnetic fields: a randomized, double-blind, placebo-controlled study. *Osteoarthr Cartil* 2005;**13**:575–81. http://dx.doi.org/10.1016/j.joca.2005.02.012

447. Thorstensson CA, Roos EM, Petersson IF, Ekdahl C. Six-week high-intensity exercise program for middle-aged patients with knee osteoarthritis: a randomized controlled trial [ISRCTN20244858]. *BMC Musculoskelet Disord* 2005;**6**:27. http://dx.doi.org/10.1186/1471-2474-6-27

448. Tishler M, Rosenberg O, Levy O, Elias I, Amit-Vazina M. The effect of balneotherapy on osteoarthritis. Is an intermittent regimen effective? *Eur J Intern Med* 2004;**15**:93–6. http://dx.doi.org/10.1016/j.ejim.2004.01.002

449. Toda Y, Tsukimura N. Influence of concomitant heeled footwear when wearing a lateral wedged insole for medial compartment osteoarthritis of the knee. *Osteoarthr Cartil* 2008;**16**:244–53. http://dx.doi.org/10.1016/j.joca.2007.06.010

450. Topp R, Woolley S, Hornyak J, Khuder S, Kahaleh B. The effect of dynamic versus isometric resistance training on pain and functioning among adults with osteoarthritis of the knee. *Arch Phys Med Rehabil* 2002;**83**:1187–95. http://dx.doi.org/10.1053/apmr.2002.33988

451. Trans T, Aaboe J, Henriksen M, Christensen R, Bliddal H, Lund H. Effect of whole body vibration exercise on muscle strength and proprioception in females with knee osteoarthritis. *Knee* 2009;**16**:256–61. http://dx.doi.org/10.1016/j.knee.2008.11.014

452. Trock DH, Bollet AJ, Markoll R. The effect of pulsed electromagnetic fields in the treatment of osteoarthritis of the knee and cervical spine. Report of randomized, double blind, placebo controlled trials. *J Rheumatol* 1994;**21**:1903–11.

453. Tucker M, Brantingham JW, Myburg C. Relative effectiveness of a non-steroidal anti-inflammatory medication (Meloxicam) versus manipulation in the treatment of osteo-arthritis of the knee. *Eur J Chiropractic* 2003;**50**:163–84.

454. Tukmachi E, Jubb R, Dempsey E, Jones P. The effect of acupuncture on the symptoms of knee osteoarthritis – an open randomised controlled study. *Acupunct Med* 2004;**22**:14–22. http://dx.doi.org/10.1136/aim.22.1.14

455. Wang C, Schmid CH, Hibberd PL, Kalish R, Roubenoff R, Rones R, *et al.* Tai Chi is effective in treating knee osteoarthritis: a randomized controlled trial. *Arthritis Rheum* 2009;**61**:1545–53. http://dx.doi.org/10.1002/art.24832

456. Wigler I, Elkayam O, Paran D, Yaron M. Spa therapy for gonarthrosis: a prospective study. *Rheumatol Int* 1995;**15**:65–8. http://dx.doi.org/10.1007/BF00262710

457. Wolsko PM, Eisenberg DM, Simon LS, Davis RB, Walleczek J, Mayo-Smith M, *et al.* Double-blind placebo-controlled trial of static magnets for the treatment of osteoarthritis of the knee: results of a pilot study. *Altern Ther Health Med* 2004;**10**:36–43.

458. Wu ZH, Bao F. [Observation on therapeutic effects of electroacupuncture for the treatment of knee osteoarthritis.] *Zhongguo Gu Shang* 2008;**21**:170–2.

459. Young SL, Woodbury MG, Fryday-Field K, Donovan T, Bellamy N, Haddad R. Efficacy of interferential current stimulation alone for pain reduction in patients with osteoarthritis of the knee: a randomized placebo control clinical trial. *Phys Ther J Am Phys Ther Assoc* 1991;**71**(Suppl. 6):52.

460. Yurtkuran M, Kocagil T. TENS, electroacupuncture and ice massage: comparison of treatment for osteoarthritis of the knee. *Am J Acupunct* 1999;**27**:133–40.

461. Yurtkuran M, Yurtkuran M, Alp A, Nasircilar A, Bingöl U, Altan L, *et al.* Balneotherapy and tap water therapy in the treatment of knee osteoarthritis. *Rheumatol Int* 2006;**27**:19–27. http://dx.doi.org/10.1007/s00296-006-0158-8

462. Zizic TM, Hoffman KC, Holt PA, Hungerford DS, O'Dell JR, Jacobs MA, *et al.* The treatment of osteoarthritis of the knee with pulsed electrical stimulation. *J Rheumatol* 1995;**22**:1757–61.

463. Arazpour M, Bani MA, Maleki M, Ghomshe FT, Kashani RV, Hutchins SW. Comparison of the efficacy of laterally wedged insoles and bespoke unloader knee orthoses in treating medial compartment knee osteoarthritis. *Prosthet Orthot Int* 2013;**37**:50–7. http://dx.doi.org/10.1177/0309364612447094

464. Bennell KL, Bowles KA, Payne C, Cicuttini F, Williamson E, Forbes A, *et al.* Lateral wedge insoles for medial knee osteoarthritis: 12 month randomised controlled trial. *BMJ* 2011;**342**:d2912. http://dx.doi.org/10.1136/bmj.d2912

465. Bliddal H, Leeds AR, Stigsgaard L, Astrup A, Christensen R. Weight loss as treatment for knee osteoarthritis symptoms in obese patients: 1-year results from a randomised controlled trial. *Ann Rheum Dis* 2011;**70**:1798–803. http://dx.doi.org/10.1136/ard.2010.142018

466. Brosseau L, Wells GA, Kenny GP, Reid R, Maetzel A, Tugwell P, *et al.* The implementation of a community-based aerobic walking program for mild to moderate knee osteoarthritis: a knowledge translation randomized controlled trial: part II: clinical outcomes. *BMC Public Health* 2012;**12**:1073. http://dx.doi.org/10.1186/1471-2458-12-1073

467. Bruce-Brand RA, Walls RJ, Ong JC, Emerson BS, O'Byrne JM, Moyna NM. Effects of home-based resistance training and neuromuscular electrical stimulation in knee osteoarthritis: a randomized controlled trial. *BMC Musculoskelet Disord* 2012;**13**:118. http://dx.doi.org/10.1186/1471-2474-13-118

468. Fary RE, Carroll GJ, Briffa TG, Briffa NK. The effectiveness of pulsed electrical stimulation in the management of osteoarthritis of the knee: results of a double-blind, randomized, placebo-controlled, repeated-measures trial. *Arthritis Rheum* 2011;**63**:1333–42. http://dx.doi.org/10.1002/art.30258

469. Foroughi N, Smith RM, Lange AK, Baker MK, Fiatarone Singh MA, Vanwanseele B. Lower limb muscle strengthening does not change frontal plane moments in women with knee osteoarthritis: a randomized controlled trial. *Clin Biomech* 2011;**26**:167–74. http://dx.doi.org/10.1016/j.clinbiomech.2010.08.011

470. Fukuda TY, Alves da Cunha R, Fukuda VO, Rienzo FA, Cazarini C, Carvalho Nde A, *et al.* Pulsed shortwave treatment in women with knee osteoarthritis: a multicenter, randomized, placebo-controlled clinical trial. *Phys Ther* 2011;**91**:1009–17. http://dx.doi.org/10.2522/ptj.20100306

471. Fukuda VO, Fukuda TY, Guimarães M, Shiwa S, de Lima Bdel C, Martins RÁ, *et al.* Short-term efficacy of low-level laser therapy in patients with knee osteoarthritis: a randomized placebo-controlled, double-blind clinical trial. *Rev Bras Ortop* 2011;**46**:526–33. http://dx.doi.org/10.1590/S0102-36162011000500008

472. Giombini A, Di Cesare A, Di Cesare M, Ripani M, Maffulli N. Localized hyperthermia induced by microwave diathermy in osteoarthritis of the knee: a randomized placebo-controlled double-blind clinical trial. *Knee Surg Sports Traumatol Arthrosc* 2011;**19**:980–7. http://dx.doi.org/10.1007/s00167-010-1350-7

473. Hsieh R-L, Lo M-T, Lee W-C, Liao W-C. Therapeutic effects of short-term monochromatic infrared energy therapy on patients with knee osteoarthritis: a double-blind, randomized, placebo-controlled study. *J Orthop Sports Phys Ther* 2012;**42**:947–56. http://dx.doi.org/10.2519/jospt.2012.3881

474. Imoto AM, Peccin MS, Trevisani VF. Quadriceps strengthening exercises are effective in improving pain, function and quality of life in patients with osteoarthritis of the knee. *Acta Ortop Bras* 2012;**20**:174–9. http://dx.doi.org/10.1590/S1413-78522012000300008

475. Lim JY, Tchai E, Jang SN. Effectiveness of aquatic exercise for obese patients with knee osteoarthritis: a randomized controlled trial. *PM R* 2010;**2**:723–31. http://dx.doi.org/10.1016/j.pmrj.2010.04.004

476. Mavrommatis CI, Argyra E, Vadalouka A, Vasilakos DG. Acupuncture as an adjunctive therapy to pharmacological treatment in patients with chronic pain due to osteoarthritis of the knee: a 3-armed, randomized, placebo-controlled trial. *Pain* 2012;**153**:1720–6. http://dx.doi.org/10.1016/j.pain.2012.05.005

477. Miller E, Maimon Y, Rosenblatt Y, Mendler A, Hasner A, Barad A, *et al.* Delayed effect of acupuncture treatment in OA of the knee: a blinded, randomized, controlled trial. *Evid Based Complement Alternat Med* 2011;**2011**:792975. http://dx.doi.org/10.1093/ecam/nen080

478. Perlman AI, Ali A, Njike VY, Hom D, Davidi A, Gould-Fogerite S, *et al.* Massage therapy for osteoarthritis of the knee: a randomized dose-finding trial. *PLOS ONE* 2012;**7**:e30248. http://dx.doi.org/10.1371/journal.pone.0030248

479. Salacinski AJ, Krohn K, Lewis SF, Holland ML, Ireland K, Marchetti G. The effects of group cycling on gait and pain-related disability in individuals with mild-to-moderate knee osteoarthritis: a randomized controlled trial. *J Orthop Sports Phys Ther* 2012;**42**:985–95. http://dx.doi.org/10.2519/jospt.2012.3813

480. Salli A, Sahin N, Baskent A, Ugurlu H. The effect of two exercise programs on various functional outcome measures in patients with osteoarthritis of the knee: a randomized controlled clinical trial. *Isokinet Exerc Sci* 2010;**18**:201–9.

481. Vance CG, Rakel BA, Blodgett NP, DeSantana JM, Amendola A, Zimmerman MB, *et al.* Effects of transcutaneous electrical nerve stimulation on pain, pain sensitivity, and function in people with knee osteoarthritis: a randomized controlled trial. *Phys Ther* 2012;**92**:898–910. http://dx.doi.org/10.2522/ptj.20110183

482. Wang TJ, Lee SC, Liang SY, Tung HH, Wu SF, Lin YP. Comparing the efficacy of aquatic exercises and land-based exercises for patients with knee osteoarthritis. *J Clin Nurs* 2011;**20**:2609–22. http://dx.doi.org/10.1111/j.1365-2702.2010.03675.x

483. MacPherson H, Altman DG, Hammerschlag R, Youping L, Taixiang W, White A, *et al.* Revised STandards for Reporting Interventions in Clinical Trials of Acupuncture (STRICTA): extending the CONSORT statement. *PLOS Med* 2010;**7**:e1000261. http://dx.doi.org/10.1371/journal.pmed.1000261

484. Centre for Outcomes Research and Effectiveness (CORE). *Humanistic Psychological Therapies Competences Framework*. University College London Division of Psychology and Language Sciences; 2013. URL: www.ucl.ac.uk/pals/research/cehp/research-groups/core#,2009 (accessed 17 December 2015).

485. Joint Formulary Committee. *British National Formulary*. 64 edn. London: BMJ Group and Pharmaceutical Press; 2012.

486. Department of Health. *NHS Reference Costs 2011–2012*. London: Department of Health; 2012.

# Appendix 1 Review data related to *Chapter 1*

## MEDLINE search strategy

### *Acupuncture*
acupuncture OR electro-acupuncture OR electroacupuncture

### *Back pain*
back pain OR backache OR Intervertebral disk OR lumbar* OR sciatica

### *Neck pain*
neck OR cervic* OR spinal OR torticollis OR whiplash

### *Shoulder pain*
shoulder OR rotator cuff OR bursitis OR tendinitis OR tendonitis OR adhesive capsulitis

### *Osteoarthritis knee pain*
Knee OR Arthralgia* OR Arthriti* OR Osteoarthrit* OR Hip

### *Headache pain*
headache OR migrain* OR cephalgi* OR hemicrania

### *Randomised trials of acupuncture for pain*
Pain with "Randomized Controlled Trial" as a limit

## Outcome measure preferences

This section is to guide the selection of end points to be included in the meta-analyses. End points are classified by domain (e.g. WOMAC pain is in the 'pain intensity' domain; days of headache is in the 'pain frequency' domain). This section specifies which end point should be chosen if a trial has data on more than one end point per domain. In general, we have given preference to measures that are specific to pain types, then to the most widely used measures.

### *Pain frequency*

### Headache

- For trials with only migraine patients:

  (a) days with migraine
  (b) days with migraine symptoms
  (c) have you had a migraine during the last 4 weeks (1 = yes/0 = no)?
  (d) migraine attacks
  (e) list as below.

- For trials including non-migraine headache patients:

  (f) days with moderate/strong headache
  (g) days with headache
  (h) days with strong headache

(i)   total headache hours
(j)   days with headache from other causes
(k)   during the past 4 weeks have you experienced a TTH (1 = yes/0 = no)?
(l)   do you have other pain than a TTH (1 = yes/0 = no)?
(m)  days with impaired activity
(n)   during the past 4 weeks have you had migraine as well (as a TTH) (1 = yes/0 = no)?
(o)   do you have other pain in addition to a migraine headache (1 = yes/0 = no)?
(p)   days with at least one accompanying symptom
(q)   days with a minimum of one accompanying symptom.

## Lower back pain

(a)   Days with pain in the past month (diary 31 days maximum).

## Knee and/or hip osteoarthritis

(a)   Days with pain in the past week (diary 7 days maximum).

## Shoulder

(a)   Pain persistent at time of assessment.
(b)   New episode of pain at time of assessment.

### *Pain intensity*

1. Diary scores (if pain diaries use more than one measure of pain, then preference is as below).
2. VAS and NRS will be combined by multiplication (NRS × 10 = VAS).

    –   Average pain measured by the VAS or NRS will be first choice
    –   If a trial measures:

        –   if worst, day and night scores, we will average day and night scores
        –   if day and night scores are given, we will take the average
        –   if good and bad knee scores are given, we will average the two scores
        –   if average, highest intensity, lowest and current are given, we will use the average
        –   if pain severity in the past week, unpleasantness in the past week and severity of main problem are given, we will use severity in the past week.

3. WOMAC pain.
4. Self-Efficacy Scale pain affective/sensoric.
5. Arthritis Self-Efficacy Scale – pain subscale.
6. Shoulder Pain and Disability Index (SPADI) pain component (0–50).
7. Quality of life (SF-36): bodily pain.
8. McGill Pain Questionnaire.
9. Global headache severity.
10. CMS pain component.
11. von Korff scale Q1–3 (pain intensity).

### *Functional impairment*

1. WOMAC function.
2. CMS activity component.
3. Disability: Oswestry Disability Index.

4. Disability: Roland Morris Disability Questionnaire.
5. Back function score [FFbH-R (Hannover Functional Ability Questionnaire) 0–100%].
6. SPADI disability (0–80).
7. Days with impairment because of pain:

  – days with limited function because of pain (180 days maximum)
  – days unable to do household work
  – days with inability to work because of pain.

8. Disability: Pain Disability Index.
9. von Korff scale Q5–7 (impairment).
10. Walking ability.

### Combined pain and functional impairment

1. WOMAC index/combined.
2. Neck pain and disability score.
3. CMS total.
4. Northwick Park Neck Pain Questionnaire (measures pain and disability).
5. OARSI-OMERACT Responder Index (pain structure function scale).
6. SPADI global (0–130).
7. Oxford Knee Score.

### Mental quality of life

1. Quality of life (SF-36): MCS score.
2. Average quality of life (SF-36): social functioning, role emotional and mental health components.
3. SF-12 mental health summary score.
4. Average of Profile of Quality of Life in the Chronically Ill psychological, positive mood, negative mood, social functioning and social well-being components.
5. Average of Hospital Anxiety and Depression Scale anxiety and depression components.
6. Worry about neck pain compared with worry felt at previous questionnaire.
7. Other depression scores.
8. Other anxiety scores.

### Physical quality of life

1. Quality of life (SF-36): PCS score.
2. Average quality of life (SF-36): physical activity and role physical.
3. Profile of Quality of Life in the Chronically Ill physical component.
4. SF-12 physical health summary score.

### Overall quality of life

1. SF-36 total score (1–100).
2. Quality of life (SF-36): general health perception.
3. Quality of life: COOP/WONCA (World Organization of Colleges, Academies and Academic Associations of General Practitioners/Family Physicians) charts (30–0) (lower scores mean higher quality of life).
4. Quality of life (SF-36): vitality.
5. EQ-5D.

### Range of motion/stiffness

1. WOMAC stiffness.
2. Average of CMS range of motion, abduction, flexion, external rotation, internal rotation, strength.
3. Average of active and passive cervical mobility.
4. Range of movement.

### Health change

1. Self-evaluated improvement (NRS 0–10).
2. Assessment of treatment effect (1–5).
3. Global assessment of change (1–5).
4. Quality of life (SF-36): health change over the past year (0–10).
5. Globally improved, judged by clinician (yes/no).
6. Physician-evaluated improvement (NRS 0–10).

### Satisfaction

1. Satisfaction with treatment outcome.
2. Satisfaction with care.
3. Satisfaction with information.

### Medication

1. Days taking medication.
2. Medication quantification scale.
3. Consumption of drugs (none; less than, equal to, greater than or extra than what was prescribed).
4. Any use of pain medication.
5. Amount of trial drug taken (diclofenac) measured by weight of remaining pills.
6. Other.

## Trial-level information

### Trials evaluating acupuncture for osteoarthritis pain (n = 9)

| Trial | Patient counts for those included in primary analysis | Primary outcome | Time point | Results reported by authors | Dropout rate (in parentheses) | Assessment of blinding[a] |
|---|---|---|---|---|---|---|
| Berman 2004[134] | Total n = 391: acupuncture, n = 142; sham – both penetrating and non-penetrating needles, n = 141; no acupuncture control – non-specific advice, n = 108 | WOMAC pain subscore | 6 months | Mean (SE): acupuncture −3.79 (0.33); sham −2.92 (0.30); no acupuncture −1.69 (0.33)<br><br>Difference between groups (95% CI): acupuncture vs. sham 0.87 (0.16 to 1.58; $p = 0.003$); acupuncture vs. no acupuncture (not given) | Total 391/570 (31%): acupuncture 142/190 (25%); sham 141/191 (26%); non-specific advice 108/189 (43%) | A |
| Vas 2004[120] | Total n = 88: acupuncture, n = 47; sham – non-penetrating needles, n = 41 | WOMAC pain subscore | 3 months | Mean (SD): acupuncture 1.7 (2.6); sham 6.4 (5.8)<br><br>Difference between groups (95% CI): acupuncture vs. sham 4.7 (2.9 to 6.5; $p < 0.001$) | Total 88/97 (9%): acupuncture 47/48 (2%); sham 41/49 (16%) | A |
| Witt 2005[69] | Total n = 285: acupuncture, n = 145; sham – penetrating needles, n = 73; no acupuncture control – usual care, n = 67 | WOMAC index | 2 months | Mean (SE): acupuncture 26.9 (1.4); sham 35.8 (1.9); no acupuncture 49.6 (2.0)<br><br>Difference between groups (95% CI): acupuncture vs. sham 8.8 (4.2 to 13.5; $p < 0.001$); acupuncture vs. no acupuncture 22.7 (17.9 to 27.5; $p < 0.001$) | Total 285/300 (5%): acupuncture 145/150 (3%); sham 73/76 (4%); usual care 67/74 (9%) | A |
| Scharf 2006[70] | Total n = 985:[b] acupuncture, n = 318; sham – penetrating needles, n = 360; no acupuncture control – ancillary care, n = 307 | WOMAC pain subscore | 6 months | Mean (95% CI): acupuncture 2.9 (2.65 to 3.17); sham 3.2 (2.93 to 3.43); no acupuncture 4.0 (3.69 to 4.22)<br><br>Difference between groups (95% CI): acupuncture vs. sham 0.3 (−0.05 to 0.59; no $p$-value given); acupuncture vs. no acupuncture 1.0 (0.71 to 1.38; no $p$-value given) | Total 985/1039 (5%): acupuncture 318/330 (4%); sham 360/367 (2%); usual care 307/342 (10%) | A |
| Witt 2006[77] | Total n = 579:[b] acupuncture, n = 300; no acupuncture control – usual care, n = 279 | WOMAC index | 3 months | Mean (SEM): acupuncture 30.5 (1.0); no acupuncture 47.3 (1.0)<br><br>Difference between groups (SEM): acupuncture vs. no acupuncture 16.7 (1.4; $p < 0.001$) | Total 579/712 (19%): acupuncture 300/357 (16%); usual care 279/355 (21%) | NA |

| Trial | Patient counts for those included in primary analysis | Primary outcome | Time point | Results reported by authors | Dropout rate (in parentheses) | Assessment of blinding[a] |
|---|---|---|---|---|---|---|
| Foster 2007[135] | Total $n = 325$: acupuncture, $n = 108$; sham – non-penetrating needles, $n = 112$; no acupuncture control – ancillary care, $n = 105$ | WOMAC pain subscore | 6 months | Mean (SD): acupuncture 7.07 (4.4); sham 6.50 (4.8); no acupuncture 6.78 (4.5)<br><br>Difference between groups (95% CI): acupuncture vs. sham: not given; acupuncture vs. no acupuncture 0.08 (−1.0 to 0.9; $p = 0.9$) | Total 325/352 (8%): acupuncture 108/117 (8%); sham 112/119 (6%); usual care 105/116 (9%) | A |
| Williamson 2007[136] | Total $n = 181$:[c] acupuncture, $n = 60$; no acupuncture control – non-specific advice, $n = 61$; physiotherapy, $n = 60$ (not analysed) | Oxford Knee Score | 2 months | Mean (SD): acupuncture 36.8 (7.20); no acupuncture 40.3 (8.48)<br><br>Difference between groups (95% CI): acupuncture vs. no acupuncture 3.5 (0.66 to 6.33; Bonferroni $p = 0.016$) | Total 161/181 (11%): acupuncture 59/60 (2%); usual care 49/61 (20%); physiotherapy 53/60 (12%) | NA |
| Lansdown 2009[126] | Total $n = 30$: acupuncture, $n = 15$; no acupuncture control – usual care, $n = 15$ | WOMAC pain subscore | 3 months | Mean (SD): acupuncture 3.6 (2.92); no acupuncture 6.57 (4.54)<br><br>Difference between groups (95% CI): acupuncture vs. no acupuncture −2.62 (−0.77 to −4.47; $p = 0.007$) | Total 28/30 (7%): acupuncture 14/15 (7%); usual care 14/15 (7%) | (Not evaluated) |
| Suarez-Almazor 2010[125] | Total $n = 494$: acupuncture, $n = 139$; sham – penetrating needles, $n = 283$; no acupuncture control – ancillary care, $n = 72$ | WOMAC pain subscore | 3 months | Mean (SD): acupuncture 30.8 (17.9); sham 31.0 (19.1); no acupuncture 42.4 (16.8)<br><br>Difference between groups: acupuncture vs. no acupuncture $p = 0.0002$ (no estimate given); acupuncture vs. sham $p > 0.20$ (no estimate given) | Acupuncture 139/153 (91%); sham 283/302 (94%); usual care 72/72 (100%) | (Not evaluated) |

NA, not available; SEM, standard error of the mean.
a A, low likelihood of bias; NA, open trials with no patient blinding.
b Patient counts lower in the forest plots because of missing baseline scores for some patients.
c Patient counts lower in the forest plots as number reported in paper includes imputed data.

**Note**

Ancillary care – programme of care received by both acupuncture and non-acupuncture groups (e.g. trial comparing physiotherapy plus acupuncture with physiotherapy alone); usual care – protocol did not specify treatments received in control group (e.g. trials with 'waiting list controls'); non-specific advice – patients in control group received general advice and support ('attention control').

## Trials evaluating acupuncture for chronic headache pain (n = 7)

| Pain type | Trial | Patient counts for those included in primary analysis | Primary outcome | Time point | Results reported by author | Dropout rate | Assessment of blinding[a] |
|---|---|---|---|---|---|---|---|
| Migraine n = 2 | Linde 2005[68] | Total n = 272: acupuncture, n = 132; sham – penetrating needles, n = 76; no acupuncture control – usual care, n = 64 | Moderate to severe pain days | 3 months | Mean (SD): acupuncture 2.8 (2.3); sham 2.6 (2.4); no acupuncture 4.3 (2.2)<br><br>Difference between groups (95% CI): acupuncture vs. sham 0.0 (−0.7 to 0.7; p > 0.9); acupuncture vs. no acupuncture 1.4 (0.8 to 2.1; p < 0.001) | Total 272/302 (10%): acupuncture 132/145 (9%); sham 76/81 (6%); usual care 64/76 (16%) | A |
|  | Diener 2006[72] | Total n = 794:[b] acupuncture, n = 290; sham – penetrating needles, n = 317; no acupuncture control – guidelined care, n = 187 | Migraine days | 6 months | Change from baseline, mean (SD): acupuncture −2.8 (3.8); sham −2.0 (3.9); no acupuncture −2.7 (4.2)<br><br>Difference between groups (95% CI): acupuncture vs. sham 0.57 (0.09 to 1.05; p = 0.021); acupuncture vs. no acupuncture 0.50 (−0.06 to 1.05; p = 0.4) | Total 794/960 (7%): acupuncture 290/313 (7%); sham 317/339 (6%); guidelined care 187/308 (39%) | B |
| TTH (n = 3) | Coeytaux 2005[128] | Total n = 71: acupuncture, n = 34; no acupuncture control – ancillary care, n = 37 | Headache Impact Test | 1 month | Change from baseline, mean (95% CI): acupuncture −3.9 (−6.5 to −1.2); no acupuncture −0.4 (−1.8 to 1.0)<br><br>Difference between groups (95% CI): acupuncture vs. no acupuncture 3.0 (1.0 to 4.9; no p-value given) | Total 71/74 (4%): acupuncture 34/35 (3%); usual care 37/39 (5%) | NA |
|  | Melchart 2005[67] | Total n = 238: acupuncture, n = 118; sham – penetrating needles, n = 57; no acupuncture control – usual, care n = 63 | Migraine days | 3 months | Mean (SD): acupuncture 9.9 (8.7); sham 10.8 (8.3); no acupuncture 16.3 (7.4)<br><br>Difference between groups (95% CI): acupuncture vs. sham 0.06 (−1.2 to 2.4; p = 0.5); acupuncture vs. no acupuncture 5.8 (4.0 to 7.6; p < 0.001) | Total 238/270 (12%): acupuncture 118/132 (11%); sham 57/63 (10%); usual care 63/75 (16%) | A |
|  | Endres 2007[73] | Total n = 398: acupuncture, n = 204; sham – penetrating needles, n = 194 | Headache days | 6 months | Mean (SD): acupuncture 6.0 (6.2); sham 8.4 (7.9)<br><br>Difference between groups (95% CI): acupuncture vs. sham 1.94 (0.69 to 3.18; p = 0.002) | Total 398/413 (3%): acupuncture 204/209 (2%); sham 194/200 (3%) | A |

| Pain type | Trial | Patient counts for those included in primary analysis | Primary outcome | Time point | Results reported by author | Dropout rate | Assessment of blinding[a] |
|---|---|---|---|---|---|---|---|
| Both (n = 2) | Vickers 2004[60] | Total n = 301: acupuncture, n = 161; no acupuncture control – usual care, n = 140 | Severity score | 12 months | Mean (SD): acupuncture 16.2 (13.7); no acupuncture 22.3 (17.0)<br><br>Difference between groups (95% CI): acupuncture vs. no acupuncture 4.6 (2.2 to 7.0; p = 0.0002) | Total 301/401 (25%): acupuncture 161/205 (21%); usual care 140/196 (29%) | NA |
| | Jena 2008[76] | Total n = 2871:[c] acupuncture, n = 1447; no acupuncture control – usual care, n = 1424 | Headache days | 3 months | Percent reduction, mean (95% CI): acupuncture 43.0 (41.0 to 45.1); no acupuncture 15.2 (13.3 to 17.0)<br><br>Difference between groups (95% CI): acupuncture vs. no acupuncture 27.9 (25.1 to 30.6; p < 0.001) | Total 2871/3404 (16%): acupuncture 1447/1711 (15%); usual care 1424/1693 (16%) | NA |

NA, not available.
a  A, low likelihood of bias; B, intermediate likelihood of bias; NA, open trials with no patient blinding.
b  These differ from the patient counts in the forest plot. Authors confirmed that this was an error on their part and have published an erratum.
c  Patient counts lower in the forest plots because of missing baseline scores for some patients.
**Note**
Ancillary care – programme of care received by both acupuncture and non-acupuncture groups (e.g. trial comparing physiotherapy plus acupuncture with physiotherapy alone); usual care – usual care did not specify treatments received in control group (e.g. trials with 'waiting list controls'); guidelined care – patients in control group received care according to national guidelines.

**Trials evaluating acupuncture for non-specific musculoskeletal pain (n = 15)**

| Pain type | Trial | Patient counts for those included in primary analysis | Primary outcome | Time point | Results reported by author | Dropout rate | Assessment of blinding[a] |
|---|---|---|---|---|---|---|---|
| Back (n = 10) | Carlsson 2001[119] | Total n = 27: acupuncture, n = 21; sham – non-needle, n = 6 | Pain VAS | 6 months | Mean weekly VAS in percent of baseline, mean (SD):[b] acupuncture – morning VAS 75 (33), night VAS 68 (31); sham – morning VAS 132 (76), night VAS 101 (48) <br><br> Difference between groups: acupuncture vs. sham – morning VAS p = 0.13 (no estimate given), night VAS p = 0.056 (no estimate given) | Total 27/50 (46%): acupuncture 21/34 (38%); sham 6/16 (63%) | B |
| | Cherkin 2001[131] | Total n = 249:[c] acupuncture, n = 89; no acupuncture control – non-specific advice, n = 83; massage, n = 77 (not analysed) | Roland Morris Disability Questionnaire | 2 months | Mean (95% CI): acupuncture 7.9 (6.5 to 9.3); no acupuncture 8.8 (7.4 to 10.2) <br><br> Difference between groups: acupuncture vs. no acupuncture – adjusted p = 0.75 (no estimate given) | Total 249/262 (7%): acupuncture 89/94 (5%); usual care 83/90 (8%); massage 77/78 (1%) | NA |
| | Molsberger 2002[129] | Total n = 124: acupuncture, n = 47; sham – penetrating needles, n = 41; no acupuncture control – ancillary care, n = 36 | Pain VAS | 3 months | Mean (SD): acupuncture 23 (20); sham 43 (23); no acupuncture control 52 (19) <br><br> Difference between groups:[d] acupuncture vs. sham p < 0.001 (no estimate given); acupuncture vs. no acupuncture control p < 0.001 (no estimate given) | Total 124/186 (33%): acupuncture 47/65 (28%); sham 41/61 (33%); usual care 36/60 (40%) | (Not evaluated) |
| | Kerr 2003[117] | Total n = 46: acupuncture, n = 26; sham – non-needle, n = 20 | Pain VAS | 1 month | Mean (SD): acupuncture 51.3 (22.4); sham 61.7 (30.6) <br><br> Difference between groups: acupuncture vs. sham p = 0.2 (no estimate given) | Total 46/60 (23%): acupuncture, 26/30 (13%); sham 20/30 (33%) | B |

| Pain type | Trial | Patient counts for those included in primary analysis | Primary outcome | Time point | Results reported by author | Dropout rate | Assessment of blinding[a] |
|---|---|---|---|---|---|---|---|
| | Brinkhaus 2006[66] | Total n = 284: acupuncture, n = 140; sham – penetrating needles, n = 70; no acupuncture control – usual care, n = 74 | Pain VAS | 2 months | Mean change from baseline (SD): acupuncture 28.7 (30.3); sham 23.6 (31.0); no acupuncture 6.9 (22.0)<br><br>Difference between groups (95% CI): acupuncture vs. sham 5.1 (–3.7 to 13.9; p = 0.3); acupuncture vs. no acupuncture 21.7 (13.9 to 30.0; p < 0.001) | Total 284/301 (6%): acupuncture 140/147 (5%); sham 70/75 (7%); usual care 74/79 (6%) | A |
| | Thomas 2006[61] | Total n = 182: acupuncture, n = 123; no acupuncture control – usual care, n = 59 | SF-36 bodily pain | 24 months | Mean (SD): acupuncture 67.8 (24.1); no acupuncture 59.5 (23.4)<br><br>Difference between groups (95% CI): acupuncture vs. no acupuncture 8.0 (2.8 to 13.2; p = 0.003) | Total 182/241 (24%): acupuncture 123/160 (23%); usual care 59/81 (27%) | NA |
| | Witt 2006[74] | Total n = 2594:[e] acupuncture, n = 1350; no acupuncture control – usual care, n = 1244 | Back function measured with the Hannover Functional Ability Questionnaire | 3 months | Mean (SE): acupuncture 12.1 (0.4); no acupuncture 2.7 (0.4)<br><br>Difference between groups (95% CI): acupuncture vs. no acupuncture 9.4 (8.3 to 10.5; p < 0.001) | Total 2594/3093 (16%): acupuncture 1350/1549 (13%); usual care 1244/1544 (19%) | A |
| | Haake 2007[71] | Total n = 1117:[f] acupuncture, n = 377; sham – penetrating needles, n = 376; no acupuncture control – guidelined care, n = 364 | von Korff chronic pain scale | 6 months | Mean (SD) of von Korff chronic pain scale: acupuncture 40.2 (22.5); sham 43.3 (23.0); no acupuncture 52.3 (21.2)<br><br>Difference between groups in treatment success (95% CI) (%):[9] acupuncture vs. no acupuncture 20.2 (13.4 to 26.7; p < 0.001); acupuncture vs. sham 3.4 (–3.7 to 10.3; p = 0.4) | Total 1117/1162 (4%): acupuncture 377/387 (3%); sham 376/387 (3%); usual care 364/388 (6%) | A |
| | Kennedy 2008[130] | Total n = 40: acupuncture, n = 22; sham – needle, non-penetrating, n = 18 | Roland Morris Disability Questionnaire | 3 months | Mean (SEM): acupuncture 5.0 (1.0); sham 7.7 (1.5)<br><br>Difference between groups (95% CI): acupuncture vs. sham 2.6 (–0.7 to 5.9; p = 0.12) | Total 40/48 (17%): acupuncture 22/24 (8%); sham 18/24 (25%) | A |

| Pain type | Trial | Patient counts for those included in primary analysis | Primary outcome | Time point | Results reported by author | Dropout rate | Assessment of blinding[a] |
|---|---|---|---|---|---|---|---|
| | Cherkin 2009[124] | Total $n = 606$: acupuncture, $n = 299$[h] (individualised $n = 147$, standardised $n = 152$); sham – non-needle, $n = 159$; no acupuncture control – usual care, $n = 148$ | Roland Morris Disability Questionnaire | 2 months | Mean (SD): individualised acupuncture 6.4 (5.3), standardised acupuncture 6.3 (5.7); sham 5.4 (4.9); no acupuncture 8.9 (6.0)<br><br>Difference between groups (95% CI): individualised vs. standardised 0.16 (–0.90 to 1.22; $p \geq 0.05$); individualised vs. sham 0.45 (–0.61 to 1.50; $p \geq 0.05$); individualised vs. no acupuncture –2.47 (–3.53 to –1.40; $p < 0.05$) | Total 606/638 (5%): individualised acupuncture 147/157 (6%); standardised acupuncture 152/158 (4%); sham 159/162 (2%); usual care 148/161 (8%) | (Not evaluated) |
| Neck ($n = 5$) | Irnich 2001[118] | Total $n = 108$: acupuncture, $n = 51$; sham – non-needle, $n = 57$; massage, $n = 57$ (not analysed) | Pain VAS | 1 month | Change from baseline, mean (95% CI): acupuncture 24.22 (16.5 to 31.9); sham 17.28 (10.0 to 24.6)<br><br>Difference between groups (95% CI): acupuncture vs. sham 6.9 (–5.0 to 18.9; $p = 0.3$) | Total 165/177 (8%): acupuncture 51/56 (9%); sham 57/61 (7%); massage 57/60 (5%) | B |
| | White 2004[132] | Total $n = 124$;[i] acupuncture, $n = 63$; sham – non-needle, $n = 61$ | Pain VAS | 1 month | Mean (SD): acupuncture 20.39 (20.26); sham 30.69 (22.00)<br><br>Difference between groups (95% CI): acupuncture vs. sham 6.3 (1.4 to 11.3; $p = 0.012$) | Total 124/135 (8%): acupuncture 63/70 (10%); sham 61/65 (6%) | A |
| | Salter 2006[133] | Total $n = 21$: acupuncture, $n = 9$; no acupuncture control – usual care, $n = 12$ | Northwick Park Neck Pain Questionnaire | 3 months | Mean (SD): acupuncture 22.73 (18.64); no acupuncture 25.72 (16.29)<br><br>Difference between groups: acupuncture vs. no acupuncture 1.75 (no CI given; $p = 0.8$) | Total 21/24 (13%): acupuncture 9/10 (10%); usual care 12/14 (14%) | NA |
| | Vas 2006[121] | Total $n = 123$;[j] acupuncture, $n = 61$; sham – non-needle, $n = 62$ | Pain VAS | 1 month | Change from baseline, mean (SD): acupuncture 42.1 (21.1); sham 14.0 (15.7)<br><br>Difference between groups (95% CI): acupuncture vs. sham 28.1 (21.4 to 34.7; $p < 0.001$) | Total 115/123 (7%): acupuncture 58/61 (5%); sham 57/62 (8%) | A |

| Pain type | Trial | Patient counts for those included in primary analysis | Primary outcome | Time point | Results reported by author | Dropout rate | Assessment of blinding[a] |
|---|---|---|---|---|---|---|---|
| | Witt 2006[75] | Total $n = 3162$:[e] acupuncture, $n = 1618$; no acupuncture control – usual care, $n = 1544$ | Neck Pain and Disability Scale | 3 months | Change from baseline, mean (SE): acupuncture 16.2 (0.4); no acupuncture 3.9 (0.4)<br><br>Difference between groups (95% CI): acupuncture vs. no acupuncture 12.3 (11.3 to 13.3; $p < 0.001$) | Total 3162/3766 (16%): acupuncture 1618/1880 (14%); usual care 1544/1886 (18%) | NA |

NA, not available.

a A, low likelihood of bias; B, intermediate likelihood of bias; NA, open trials with no patient blinding.
b These numbers were taken from data provided; can be estimated only from what is given in the paper.
c One person in the no acupuncture control group was missing Roland Morris Disability Questionnaire data but this was not reported in the paper.
d Pain relief ≥ 50%.
e Patient counts lower in the forest plots because of missing baseline scores for some patients.
f Lower patient counts in our analyses are due to missing randomisation stratification variables: baseline von Korff scores, chronification, fear avoidance belief, levels of activity, patient expectations or trial centre.
g Values are given as percentage of patients (95% CI). Success was defined as ≥ 33% improvement on three pain-related items on the Chronic Pain Grade Scale.
h We combined the individualised and standardised acupuncture estimates in our analyses.
i We averaged weeks 4, 5 and 6 to obtain a 1-month score.
j Patient counts lower in the forest plots as number reported in paper includes imputed data.

**Note**

Ancillary care – programme of care received by both acupuncture and non-acupuncture groups (e.g. trial comparing physiotherapy plus acupuncture with physiotherapy alone); usual care – protocol did not specify treatments received in control group (e.g. trials with 'waiting list controls'); non-specific advice – patients in control group received general advice and support ('attention control'); guidelined care – patients in control group received care according to national guidelines.

## Trials evaluating acupuncture for specific shoulder pain (n = 4)

| Trial | Patient counts for those included in primary analysis | Primary outcome | Time point | Results reported by author | Dropout rate | Assessment of blinding[a] |
|---|---|---|---|---|---|---|
| Kleinhenz 1999[137] | Total $n$ = 45: acupuncture, $n$ = 22; sham – non-penetrating needles, $n$ = 23 | CMS | 1 month | Mean change from baseline (SD): acupuncture 19.2 (16.1), sham 8.4 (14.6)<br><br>Difference between groups (95% CI): acupuncture vs. sham (no estimate given) (2.3 to 19.4; $p$ = 0.001) | Total 45/52 (13%): acupuncture 22/25 (12%); sham 23/27 (15%) | A |
| Guerra de Hoyos 2004[138] | Total $n$ = 110: acupuncture, $n$ = 55; sham – non-penetrating needles, $n$ = 55 | Pain VAS | 6 months | Mean (SD): acupuncture 3.5 (3.0); sham 1.2 (1.9)<br><br>Difference between groups (95% CI): acupuncture vs. sham 2.0 (1.2 to 2.9; $p$ < 0.0005) | Total 110/130 (15%): acupuncture 55/65 (15%); sham 55/65 (15%) | A |
| Vas 2008[122] | Total $n$ = 425:[b] acupuncture, $n$ = 205; sham – non-needle, $n$ = 220 | CMS | 1 month | Mean change from baseline (SD): acupuncture 16.6 (15.6); sham 10.6 (13.5)<br><br>Difference between groups (95% CI): acupuncture vs. sham 6.0 (3.2 to 8.8; $p$ < 0.001) | Total 409/425 (4%): acupuncture 202/205 (1%); sham 207/220 (6%) | A |
| Molsberger 2010[127] | Total $n$ = 308: acupuncture, $n$ = 128; sham – non-penetrating needles, $n$ = 74; no acupuncture control – usual care, $n$ = 106 | VAS | 6 months | Mean (SD): acupuncture 19 (23.3); sham 33 (29.6); no acupuncture 33 (26.6)<br><br>Difference between groups (95% CI): acupuncture vs. sham 14 (7.87 to 20.13; $p$ < 0.001); acupuncture vs. no acupuncture 14 (8.22 to 19.78; $p$ < 0.001) | Total 308/424 (27%): acupuncture 128/154 (17%); sham 74/135 (45%); usual care 106/135 (21%) | (Not evaluated) |

a  A, low likelihood of bias.
b  Patient counts lower in the forest plots as number reported in paper includes imputed data.
**Note**
Usual care – protocol did not specify treatments received in control group (e.g. trials with 'waiting list controls').

# Appendix 2 Acupuncture characteristics related to trials in *Chapter 2*

## Trial- and patient-level information when analysed

| Pain type | Trial | Patient counts for those included in primary analysis | Results and data reported by author |
|---|---|---|---|
| *Chronic headache/migraine* (n = 6) | | | |
| Migraine (*n* = 2) | Linde 2005[68] | Total *n* = 272: acupuncture, *n* = 132; sham – penetrating needles, *n* = 76; no acupuncture control – usual care, *n* = 64 | Difference between groups (95% CI): acupuncture vs. sham 0.0 (–0.7 to 0.7; *p* > 0.9); acupuncture vs. no acupuncture control 1.4 (0.8 to 2.1; *p* < 0.001) |
| | | | Style of acupuncture: combination of traditional Chinese and Western |
| | | | Point prescription: flexible formula |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: no |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |
| | | | Deqi elicited: yes |
| | | | Acupuncture-specific patient–practitioner interactions: no |
| | | | Minimum years of experience required: 3 |
| | | | Maximum number of sessions: 12 |
| | | | Frequency of sessions (per week): 1.5 |
| | | | Duration of sessions (minutes): 29 |
| | | | Number of needles used: 17 |
| | Diener 2006[72] | Total *n* = 794:[a] acupuncture, *n* = 290; sham – penetrating needles, *n* = 317; no acupuncture control – guidelined care, *n* = 187 | Difference between groups (95% CI): acupuncture vs. sham 0.57 (0.09 to 1.05; *p* = 0.021); acupuncture vs. no acupuncture control 0.50 (–0.06 to 1.05; *p* = 0.4) |
| | | | Style of acupuncture: traditional Chinese |
| | | | Point prescription: flexible formula |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: no |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |

| Pain type | Trial | Patient counts for those included in primary analysis | Results and data reported by author |
|---|---|---|---|
| | | | Deqi elicited: yes |
| | | | Acupuncture-specific patient–practitioner interactions: no |
| | | | Minimum years of experience required: 2 |
| | | | Maximum number of sessions: 10 |
| | | | Frequency of sessions (per week): 2 |
| | | | Duration of sessions (minutes): 32 (patient-level data analysed) |
| | | | Number of needles used: 16 (patient-level data analysed) |
| TTH (*n* = 2) | Melchart 2005[67] | Total *n* = 238: acupuncture, *n* = 118; sham – penetrating needles, *n* = 57; no acupuncture control – usual care, *n* = 63 | Difference between groups (95% CI): acupuncture vs. sham 0.06 (–1.2 to 2.4; *p* = 0.5); acupuncture vs. no acupuncture control 5.8 (4.0 to 7.6; *p* < 0.001) |
| | | | Style of acupuncture: combination of traditional Chinese and Western |
| | | | Point prescription: flexible formula |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: no |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |
| | | | Deqi elicited: yes |
| | | | Acupuncture-specific patient–practitioner interactions: no |
| | | | Minimum years of experience required: 3 |
| | | | Maximum number of sessions: 12 |
| | | | Frequency of sessions (per week): 1.5 |
| | | | Duration of sessions (minutes): 30 |
| | | | Number of needles used: 15 |
| | Endres 2007[73] | Total *n* = 398: acupuncture, *n* = 204; sham – penetrating needles, *n* = 194 | Difference between groups (95% CI): acupuncture vs. sham 1.94 (0.69 to 3.18; *p* = 0.002) |
| | | | Style of acupuncture: traditional Chinese |
| | | | Point prescription: flexible formula |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: no |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |

| Pain type | Trial | Patient counts for those included in primary analysis | Results and data reported by author |
|---|---|---|---|
| | | | Deqi elicited: yes |
| | | | Acupuncture-specific patient–practitioner interactions: no |
| | | | Minimum years of experience required: 0 |
| | | | Maximum number of sessions: 10 |
| | | | Frequency of sessions (per week): 2 |
| | | | Duration of sessions (minutes): 32 |
| | | | Number of needles used: 16 |
| Both migraine and headache ($n = 2$) | Vickers 2004[60] | Total $n = 301$: acupuncture, $n = 161$; no acupuncture control – usual care, $n = 140$ | Difference between groups (95% CI): acupuncture vs. no acupuncture control 4.6 (2.2 to 7.0; $p = 0.0002$) |
| | | | Style of acupuncture: traditional Chinese |
| | | | Point prescription: fully individualised |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: no |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: yes |
| | | | Deqi elicited: yes |
| | | | Acupuncture-specific patient–practitioner interactions: yes |
| | | | Minimum years of experience required: 0 |
| | | | Maximum number of sessions: 24 (patient-level data analysed) |
| | | | Frequency of sessions (per week): 1 |
| | | | Duration of sessions (minutes): 30 |
| | | | Number of needles used: not provided |
| | Jena 2008[76] | Total $n = 2871$:[b] acupuncture, $n = 1447$; no acupuncture control – usual care, $n = 1424$ | Difference between groups (95% CI): acupuncture vs. no acupuncture control 27.9 (25.1 to 30.6; $p < 0.001$) |
| | | | Style of acupuncture: combination of traditional Chinese and Western |
| | | | Point prescription: fully individualised |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: no |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |

| Pain type | Trial | Patient counts for those included in primary analysis | Results and data reported by author |
|---|---|---|---|
| | | | Deqi elicited: not provided |
| | | | Acupuncture-specific patient–practitioner interactions: yes |
| | | | Minimum years of experience required: 0 |
| | | | Age of acupuncturists: yes (patient-level data analysed) |
| | | | Sex of acupuncturists: yes (patient-level data analysed) |
| | | | Maximum number of sessions: 30 (patient-level data analysed) |
| | | | Frequency of sessions (per week): 1 |
| | | | Duration of sessions (minutes): not provided |
| | | | Number of needles used: not provided |
| *Non-specific musculoskeletal* (n = *13*) | | | |
| Back (*n* = 8) | Carlsson 2001[119] | Total *n* = 27: acupuncture, *n* = 21; sham – non-needle, *n* = 6 | Difference between groups: acupuncture vs. sham – morning VAS *p* = 0.13 (no estimate given), night VAS *p* = 0.056 (no estimate given) |
| | | | Style of acupuncture: combination of traditional Chinese and Western |
| | | | Point prescription: flexible formula |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: yes |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |
| | | | Deqi elicited: yes |
| | | | Acupuncture-specific patient–practitioner interactions: no |
| | | | Minimum years of experience required: 5 |
| | | | Maximum number of sessions: 10 |
| | | | Frequency of sessions (per week): 1 |
| | | | Duration of sessions (minutes): 20 |
| | | | Number of needles used: 16 |
| | Cherkin 2001[131] | Total *n* = 172:[c] acupuncture, *n* = 89; no acupuncture control, non-specific advice, *n* = 83 | Difference between groups: acupuncture vs. no acupuncture control – adjusted *p* = 0.75 (no estimate given) |
| | | | Style of acupuncture: traditional Chinese |
| | | | Point prescription: fully individualised |

| Pain type | Trial | Patient counts for those included in primary analysis | Results and data reported by author |
|---|---|---|---|
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: yes |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: yes |
| | | | Deqi elicited: yes |
| | | | Acupuncture-specific patient–practitioner interactions: yes |
| | | | Minimum years of experience required: 3 |
| | | | Maximum number of sessions: 10 |
| | | | Frequency of sessions (per week): 1.5 |
| | | | Duration of sessions (minutes): 28 |
| | | | Number of needles used: 12 |
| | Kerr 2003[117] | Total $n = 46$: acupuncture, $n = 26$; sham – non-needle, $n = 20$ | Difference between groups: acupuncture vs. sham $p = 0.2$ (no estimate given) |
| | | | Style of acupuncture: Western |
| | | | Point prescription: fixed formula |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: no |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |
| | | | Deqi elicited: yes |
| | | | Acupuncture-specific patient–practitioner interactions: yes |
| | | | Minimum years of experience required: 0 |
| | | | Maximum number of sessions: 6 |
| | | | Frequency of sessions (per week): 1 |
| | | | Duration of sessions (minutes): 30 |
| | | | Number of needles used: 11 |
| | Brinkhaus 2006[66] | Total $n = 284$: acupuncture, $n = 140$; sham – penetrating needles, $n = 70$; no acupuncture control – usual care, $n = 74$ | Difference between groups (95% CI): acupuncture vs. sham 5.1 (–3.7 to 13.9; $p = 0.3$); acupuncture vs. no acupuncture control 21.7 (13.9 to 30.0; $p < 0.001$) |
| | | | Style of acupuncture: traditional Chinese |
| | | | Point prescription: flexible formula |

| Pain type | Trial | Patient counts for those included in primary analysis | Results and data reported by author |
|---|---|---|---|
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: no |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |
| | | | Deqi elicited: yes |
| | | | Acupuncture-specific patient–practitioner interactions: yes |
| | | | Minimum years of experience required: 3 |
| | | | Age of acupuncturists: yes (patient-level data analysed) |
| | | | Sex of acupuncturists: yes (patient-level data analysed) |
| | | | Maximum number of sessions: 12 (patient-level data analysed) |
| | | | Frequency of sessions (per week): 1.5 |
| | | | Duration of sessions (minutes): 28 (patient-level data analysed) |
| | | | Number of needles used: 17 (patient-level data analysed) |
| | Thomas 2006[61] | Total $n = 182$: acupuncture, $n = 123$; no acupuncture control – usual care, $n = 59$ | Difference between groups (95% CI): acupuncture vs. no acupuncture control 8.0 (2.8 to 13.2; $p = 0.003$) |
| | | | Style of acupuncture: traditional Chinese |
| | | | Point prescription: fully individualised |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: yes |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: yes |
| | | | Deqi elicited: yes |
| | | | Acupuncture-specific patient–practitioner interactions: yes |
| | | | Minimum years of experience required: 3 |
| | | | Maximum number of sessions: 10 |
| | | | Frequency of sessions (per week): 1 |
| | | | Duration of sessions (minutes): 25 |
| | | | Number of needles used: 10 |

| Pain type | Trial | Patient counts for those included in primary analysis | Results and data reported by author |
|---|---|---|---|
| | Witt 2006[74] | Total $n = 2594$:[b] acupuncture, $n = 1350$; no acupuncture control – usual care, $n = 1244$ | Difference between groups (95% CI): acupuncture vs. no acupuncture control 9.4 (8.3 to 10.5; $p < 0.001$) |
| | | | Style of acupuncture: combination of traditional Chinese and Western |
| | | | Point prescription: fully individualised |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: no |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |
| | | | Deqi elicited: not provided |
| | | | Acupuncture-specific patient–practitioner interactions: yes |
| | | | Minimum years of experience required: 0 |
| | | | Age of acupuncturists: yes (patient-level data analysed) |
| | | | Sex of acupuncturists: yes (patient-level data analysed) |
| | | | Maximum number of sessions: 25 (patient-level data analysed) |
| | | | Frequency of sessions (per week): 1 |
| | | | Duration of sessions (minutes): not provided |
| | | | Number of needles used: not provided |
| | Haake 2007[71] | Total $n = 1117$:[d] acupuncture, $n = 377$; sham – penetrating needles, $n = 376$; no acupuncture control – guidelined care, $n = 364$ | Difference between groups in treatment success (95% CI):[e] acupuncture vs. no acupuncture control 20.2% (13.4% to 26.7%; $p < 0.001$); acupuncture vs. sham 3.4% (–3.7% to 10.3%; $p = 0.4$) |
| | | | Style of acupuncture: traditional Chinese |
| | | | Point prescription: flexible formula |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: no |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |
| | | | Deqi elicited: yes |
| | | | Acupuncture-specific patient–practitioner interactions: no |
| | | | Minimum years of experience required: 2 |

| Pain type | Trial | Patient counts for those included in primary analysis | Results and data reported by author |
|---|---|---|---|
| | | | Maximum number of sessions: 10 |
| | | | Frequency of sessions (per week): 2 |
| | | | Duration of sessions (minutes): 32 (patient-level data analysed) |
| | | | Number of needles used: 16 (patient-level data analysed) |
| | Kennedy 2008[130] | Total $n = 40$: acupuncture, $n = 22$; sham – non-penetrating needles, $n = 18$ | Difference between groups (95% CI): acupuncture vs. sham 2.6 (–0.7 to 5.9; $p = 0.12$) |
| | | | Style of acupuncture: Western |
| | | | Point prescription: flexible formula |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: no |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |
| | | | Deqi elicited: yes |
| | | | Acupuncture-specific patient–practitioner interactions: no |
| | | | Minimum years of experience required: 10 |
| | | | Maximum number of sessions: 12 |
| | | | Frequency of sessions (per week): 1.5 |
| | | | Duration of sessions (minutes): 30 |
| | | | Number of needles used: 11 |
| Neck ($n = 5$) | Irnich 2001[118] | Total $n = 108$: acupuncture, $n = 51$; sham – non-needle, $n = 57$ | Difference between groups (95% CI): acupuncture vs. sham 6.9 (–5.0 to 18.9; $p = 0.3$) |
| | | | Style of acupuncture: combination of traditional Chinese and Western |
| | | | Point prescription: fully individualised |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: no |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |
| | | | Deqi elicited: yes |
| | | | Acupuncture-specific patient–practitioner interactions: no |
| | | | Minimum years of experience required: 0 |

| Pain type | Trial | Patient counts for those included in primary analysis | Results and data reported by author |
|---|---|---|---|
| | | | Maximum number of sessions: 5 |
| | | | Frequency of sessions (per week): 2 |
| | | | Duration of sessions (minutes): 30 |
| | | | Number of needles used: 9 |
| | White 2004[132] | Total $n = 124$:[f] acuncture, $n = 63$; sham – non-needle, $n = 61$ | Difference between groups (95% CI): acupuncture vs. sham 6.3 (1.4 to 11.3; $p = 0.012$) |
| | | | Style of acupuncture: Western |
| | | | Point prescription: flexible formula |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: no |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |
| | | | Deqi elicited: yes |
| | | | Acupuncture-specific patient–practitioner interactions: no |
| | | | Minimum years of experience required: 5 |
| | | | Maximum number of sessions: 8 |
| | | | Frequency of sessions (per week): 2 |
| | | | Duration of sessions (minutes): 30 |
| | | | Number of needles used: 6 |
| | Salter 2006[133] | Total $n = 21$: acupuncture, $n = 9$; no acupuncture control – usual care, $n = 12$ | Difference between groups: acupuncture vs. no acupuncture control 1.75 (no CI given; $p = 0.8$) |
| | | | Style of acupuncture: traditional Chinese |
| | | | Point prescription: fully individualised |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: yes |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: yes |
| | | | Deqi elicited: yes |
| | | | Acupuncture-specific patient–practitioner interactions: yes |
| | | | Minimum years of experience required: 3 |
| | | | Maximum number of sessions: 10 |

| Pain type | Trial | Patient counts for those included in primary analysis | Results and data reported by author |
|---|---|---|---|
| | | | Frequency of sessions (per week): 1 |
| | | | Duration of sessions (minutes): 25 |
| | | | Number of needles used: 13 |
| | Vas 2006[121] | Total $n = 123$:[9] acupuncture, $n = 61$; sham – non-needle, $n = 62$ | Difference between groups (95% CI): acupuncture vs. sham 28.1 (21.4 to 34.7; $p < 0.001$) |
| | | | Style of acupuncture: traditional Chinese |
| | | | Point prescription: flexible formula |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: no |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |
| | | | Deqi elicited: yes |
| | | | Acupuncture-specific patient–practitioner interactions: no |
| | | | Minimum years of experience required: 3 |
| | | | Maximum number of sessions: 5 |
| | | | Frequency of sessions (per week): 1.5 |
| | | | Duration of sessions (minutes): 30 |
| | | | Number of needles used: 10 |
| | Witt 2006[75] | Total $n = 3162$:[b] acupuncture, $n = 1618$; no acupuncture control – usual care, $n = 1544$ | Difference between groups (95% CI): acupuncture vs. no acupuncture control 12.3 (11.3 to 13.3; $p < 0.001$) |
| | | | Style of acupuncture: combination of traditional Chinese and Western |
| | | | Point prescription: fully individualised |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: no |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |
| | | | Deqi elicited: not provided |
| | | | Acupuncture-specific patient–practitioner interactions: yes |
| | | | Minimum years of experience required: 0 |
| | | | Age of acupuncturists: yes (patient-level data analysed) |

| Pain type | Trial | Patient counts for those included in primary analysis | Results and data reported by author |
|---|---|---|---|
| | | | Sex of acupuncturists: yes (patient-level data analysed) |
| | | | Maximum number of sessions: 30 (patient-level data analysed) |
| | | | Frequency of sessions (per week): 1 |
| | | | Duration of sessions (minutes): not provided |
| | | | Number of needles used: not provided |
| *Osteoarthritis (n = 7)* | | | |
| | Berman 2004[134] | Total $n = 391$: acupuncture, $n = 142$; sham – both penetrating and non-penetrating needles, $n = 141$; no acupuncture control – non-specific advice, $n = 108$ | Difference between groups (95% CI): acupuncture vs. sham 0.87 (0.16 to 1.58; $p = 0.003$); acupuncture vs. no acupuncture control not given |
| | | | Style of acupuncture: traditional Chinese |
| | | | Point prescription: fixed formula |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: yes |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |
| | | | Deqi elicited: yes |
| | | | Acupuncture-specific patient–practitioner interactions: no |
| | | | Minimum years of experience required: 2 |
| | | | Maximum number of sessions: 23 |
| | | | Frequency of sessions (per week): 0.88 |
| | | | Duration of sessions (minutes): 20 |
| | | | Number of needles used: 11 |
| | Vas 2004[120] | Total $n = 88$: acupuncture, $n = 47$; sham – non-penetrating needles, $n = 41$ | Difference between groups (95% CI): acupuncture vs. sham 4.7 (2.9 to 6.5; $p < 0.001$) |
| | | | Style of acupuncture: traditional Chinese |
| | | | Point prescription: flexible formula |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: yes |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |
| | | | Deqi elicited: yes |

| Pain type | Trial | Patient counts for those included in primary analysis | Results and data reported by author |
|---|---|---|---|
| | | | Acupuncture-specific patient–practitioner interactions: no |
| | | | Minimum years of experience required: 3 |
| | | | Maximum number of sessions: 13 |
| | | | Frequency of sessions (per week): 1 |
| | | | Duration of sessions (minutes): 30 |
| | | | Number of needles used: 8 |
| | Witt 2005[69] | Total $n = 285$: acupuncture, $n = 145$; sham – penetrating needles, $n = 73$, no acupuncture control – usual care, $n = 67$ | Difference between groups (95% CI): acupuncture vs. sham 8.8 (4.2 to 13.5; $p < 0.001$); acupuncture vs. no acupuncture control 22.7 (17.9 to 27.5; $p < 0.001$) |
| | | | Style of acupuncture: combination of traditional Chinese and Western |
| | | | Point prescription: flexible formula |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: no |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |
| | | | Deqi elicited: yes |
| | | | Acupuncture-specific patient–practitioner interactions: yes |
| | | | Minimum years of experience required: 3 |
| | | | Age of acupuncturists: yes (patient-level data analysed) |
| | | | Sex of acupuncturists: yes (patient-level data analysed) |
| | | | Maximum number of sessions: 12 (patient-level data analysed) |
| | | | Frequency of sessions (per week): 1.5 |
| | | | Duration of sessions (minutes): 29 (patient-level data analysed) |
| | | | Number of needles used: 17 (patient-level data analysed) |
| | Scharf 2006[70] | Total $n = 985$:[b] acupuncture, $n = 318$; sham – penetrating needles, $n = 360$; no acupuncture control – ancillary care, $n = 307$ | Difference between groups (95% CI): acupuncture vs. sham: 0.3 (–0.05 to 0.59; no $p$-value given); acupuncture vs. no acupuncture control 1.0 (0.71 to 1.38; no $p$-value given) |
| | | | Style of acupuncture: traditional Chinese |
| | | | Point prescription: flexible formula |

| Pain type | Trial | Patient counts for those included in primary analysis | Results and data reported by author |
|---|---|---|---|
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: no |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |
| | | | Deqi elicited: yes |
| | | | Acupuncture-specific patient–practitioner interactions: no |
| | | | Minimum years of experience required: 2 |
| | | | Maximum number of sessions: 10 |
| | | | Frequency of sessions (per week): 1.67 |
| | | | Duration of sessions (minutes): 30 (patient-level data analysed) |
| | | | Number of needles used: 13 (patient-level data analysed) |
| | Witt 2006[77] | Total $n = 579$:[b] acupuncture, $n = 300$; no acupuncture control – usual care, $n = 279$ | Difference between groups (SEM): acupuncture vs. no acupuncture control 16.7 (1.4) ($p < 0.001$) |
| | | | Style of acupuncture: combination of traditional Chinese and Western |
| | | | Point prescription: fully individualised |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: no |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |
| | | | Deqi elicited: not provided |
| | | | Acupuncture-specific patient–practitioner interactions: yes |
| | | | Minimum years of experience required: 0 |
| | | | Age of acupuncturists: yes (patient-level data analysed) |
| | | | Sex of acupuncturists: yes (patient-level data analysed) |
| | | | Maximum number of sessions: 15 (patient-level data analysed) |
| | | | Frequency of sessions (per week): 1 |
| | | | Duration of sessions (minutes): not provided |
| | | | Number of needles used: not provided |

| Pain type | Trial | Patient counts for those included in primary analysis | Results and data reported by author |
|---|---|---|---|
| | Foster 2007[135] | Total $n = 325$: acupuncture, $n = 108$; sham – non-penetrating needles, $n = 112$; no acupuncture control – ancillary care, $n = 105$ | Difference between groups (95% CI): acupuncture vs. sham not given; acupuncture vs. no acupuncture control 0.08 (–1.0 to 0.9; $p = 0.9$) |
| | | | Style of acupuncture: traditional Chinese |
| | | | Point prescription: flexible formula |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: no |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |
| | | | Deqi elicited: yes |
| | | | Acupuncture-specific patient–practitioner interactions: yes |
| | | | Minimum years of experience required: 0 |
| | | | Maximum number of sessions: 9 (patient-level data analysed) |
| | | | Frequency of sessions (per week): 1 |
| | | | Duration of sessions (minutes): 30 |
| | | | Number of needles used: 8 |
| | Williamson 2007[136] | Total $n = 121$:[9] acupuncture, $n = 60$; no acupuncture control – non-specific advice, $n = 61$ | Difference between groups (95% CI): acupuncture vs. no acupuncture control 3.5 (0.66 to 6.33; Bonferroni $p = 0.016$) |
| | | | Style of acupuncture: Western |
| | | | Point prescription: flexible formula |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: no |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |
| | | | Deqi elicited: yes |
| | | | Acupuncture-specific patient–practitioner interactions: no |
| | | | Minimum years of experience required: 0 |
| | | | Maximum number of sessions: 6 |
| | | | Frequency of sessions (per week): 1 |
| | | | Duration of sessions (minutes): 20 |
| | | | Number of needles used: 9 |

| Pain type | Trial | Patient counts for those included in primary analysis | Results and data reported by author |
|---|---|---|---|
| **Specific shoulder (n = 3)** | | | |
| | Kleinhenz 1999[137] | Total $n = 45$: acupuncture, $n = 22$; sham – non-penetrating needles, $n = 23$ | Difference between groups (95% CI): acupuncture vs. sham no estimate given (2.3 to 19.4; $p = 0.001$) |
| | | | Style of acupuncture: traditional Chinese |
| | | | Point prescription: flexible formula |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: no |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |
| | | | Deqi elicited: yes |
| | | | Acupuncture-specific patient–practitioner interactions: no |
| | | | Minimum years of experience required: 0 |
| | | | Maximum number of sessions: 8 |
| | | | Frequency of sessions (per week): 2 |
| | | | Duration of sessions (minutes): 20 |
| | | | Number of needles used: 10 |
| | Guerra de Hoyos 2004[138] | Total $n = 110$: acupuncture, $n = 55$; sham – non-penetrating needles, $n = 55$ | Difference between groups (95% CI): acupuncture vs. sham 2.0 (1.2 to 2.9; $p < 0.0005$) |
| | | | Style of acupuncture: traditional Chinese |
| | | | Point prescription: fixed formula |
| | | | Location of needles: both local and distal points |
| | | | Electrical stimulation allowed: yes |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |
| | | | Deqi elicited: yes |
| | | | Acupuncture-specific patient–practitioner interactions: no |
| | | | Minimum years of experience required: 0 |
| | | | Maximum number of sessions: 7 |
| | | | Frequency of sessions (per week): 1 |
| | | | Duration of sessions (minutes): 15 |
| | | | Number of needles used: 7 |

| Pain type | Trial | Patient counts for those included in primary analysis | Results and data reported by author |
|---|---|---|---|
| | Vas 2008[122] | Total $n = 425$:[9] acupuncture, $n = 205$; sham – non-needle, $n = 220$ | Difference between groups (95% CI): acupuncture vs. sham 6.0 (3.2 to 8.8; $p < 0.001$) |
| | | | Style of acupuncture: traditional Chinese |
| | | | Point prescription: fixed formula |
| | | | Location of needles: distal points only |
| | | | Electrical stimulation allowed: no |
| | | | Manual stimulation allowed: yes |
| | | | Moxibustion allowed: no |
| | | | Deqi elicited: yes |
| | | | Acupuncture-specific patient–practitioner interactions: no |
| | | | Minimum years of experience required: 0.5 |
| | | | Maximum number of sessions: 3 |
| | | | Frequency of sessions (per week): 1 |
| | | | Duration of sessions (minutes): 30 |
| | | | Number of needles used: 1 |

a  This differs from the patient counts in the forest plots of the primary research.[196] The authors confirmed that this was an error on their part and have published an erratum.
b  Patient counts lower in the forest plots of the primary research[196] because of missing baseline scores for some patients.
c  One person in the no-acupuncture control group was missing Roland Morris Disability Questionnaire data but this was not reported in the paper.
d  Lower patient counts in our analyses of the primary research[196] are due to missing randomisation stratification variables: baseline von Korff scores, chronification, fear avoidance belief, levels of activity, patient expectations or trial centre.
e  Values are given as percentage of patients (95% CI). Success was defined as a ≥ 33% improvement on three pain-related items on the Chronic Pain Grade Scale.
f  We averaged weeks 4–6 to obtain a 1-month score.
g  Patient counts lower in the forest plots of the primary research[196] as number reported in paper includes imputed data.

**Note**
Ancillary care – programme of care received by both acupuncture and non-acupuncture groups (e.g. trial comparing physiotherapy plus acupuncture with physiotherapy alone); usual care – protocol did not specify treatments received in control group (e.g. trials with 'waiting list controls'); non-specific advice – patients in control group received general advice and support ('attention control'); guidelined care – patients in control group received care according to national guidelines.

# Appendix 3 Review data related to *Chapter 3*

## Databases searched

- MEDLINE.
- MEDLINE In-Process & Other Non-Indexed Citations.
- EMBASE.
- Allied and Complementary Medicine Database.
- Cochrane Database of Systematic Reviews.
- Database of Abstracts of Reviews of Effects.
- Health Technology Assessment database.
- Cochrane Central Register of Controlled Trials.
- Cumulative Index to Nursing and Allied Health Literature.
- MANTIS.
- PASCAL.
- Inside Conferences.
- Conference Proceedings Citation Index – Science.
- PEDro.
- CAMbase.
- LILACS.
- ClinicalTrials.gov.

## MEDLINE search strategy

The following search strategy was developed in Ovid MEDLINE and adapted for use in the other databases searched. The strategy incorporates the sensitivity-maximising version of the Cochrane Highly Sensitive Search Strategy for randomised trials in MEDLINE.

1. Osteoarthritis, Knee/ (8978)
2. (gonarthrosis or gonarthritis).ti,ab. (912)
3. 1 or 2 (9707)
4. Osteoarthritis/ (27,166)
5. (Osteoarthriti$ or OA or osteo arthriti$ or osteoarthros$ or osteo arthros$ or arthropath$ or arthrosis or arthroses).ti,ab. (50,583)
6. degenerative arthriti$.ti,ab. (969)
7. degenerative joint disease.ti,ab. (1515)
8. 4 or 5 or 6 or 7 (61,760)
9. Knee/ (10,013)
10. Knee Joint/ (36,003)
11. (knee$ or patella$ or knee cap$ or kneecap$ or femorotibial or femoro tibial or tibiofemoral or tibio femoral or patellofemoral or patello femoral).ti,ab. (87,341)
12. 9 or 10 or 11 (98,496)
13. 8 and 12 (16,792)
14. Arthralgia/ (4180)
15. (arthralgi$ or (joint$ adj3 pain$)).ti,ab. (11,723)
16. Chronic Pain/ (1255)
17. chronic pain$.ti,ab. (17,043)
18. 14 or 15 or 16 or 17 (31,979)

19. 12 and 18 (2629)
20. 3 or 13 or 19 (21,357)
21. Acupuncture/ (1115)
22. exp Acupuncture Therapy/ (15,218)
23. acupuncture$.ti,ab. (12,290)
24. (electroacupuncture$ or electro acupuncture$).ti,ab. (2514)
25. (osteopuncture$ or osteo puncture$).ti,ab. (2)
26. (perioste$ adj3 (stimulati$ or therap$ or needling)).ti,ab. (74)
27. exp Physical Therapy Modalities/ (115,890)
28. (physiotherap$ or physio therap$ or physical therap$ or manual therap$).ti,ab. (24,353)
29. (massage$ or acupressure or shiatsu or shiatzu or zhi ya or chih ya).ti,ab. (6745)
30. Chiropractic/ (2970)
31. Traction/ (5704)
32. (chiropractic or manipulat$ or traction or kinesiolog$ or mobilis$ or mobiliz$).ti,ab. (165,625)
33. Osteopathic medicine/ (2472)
34. osteopath$.ti,ab. (3568)
35. (hydrotherap$ or hydro therap$ or water therap$ or pool therap$).ti,ab. (801)
36. exp Exercise/ (99,442)
37. exp Sports/ (101,842)
38. Physical Fitness/ (20,358)
39. (exercise$ or workout$ or work out$ or train$ or physical$ activ$ or kinesiotherap$ or keep$ fit or aerobics).ti,ab. (450,237)
40. (muscle$ adj3 (stretch$ or strengthen$)).ti,ab. (3418)
41. (walk$ adj3 (fitness or aerobic or program$ or intervention$ or session$ or regime$)).ti,ab. (1398)
42. pedometer$.ti,ab. (1102)
43. (bicycl$ or cycle$ or cycling).ti,ab. (373,541)
44. (run$ or jog$ or treadmill$).ti,ab. (122,114)
45. (swim$ or water sport$ or aquatic$ or water aerobic$ or aqua aerobic$ or water gymnastics).ti,ab. (41,769)
46. (tai ji or taiji or taijiquan or tai chi or t ai chi or taichi or shadow boxing).ti,ab. (666)
47. (yoga or yogic or pilates or danc$).ti,ab. (4740)
48. (qigong or qi gong or chi kung or chikung or ch i kung).ti,ab. (347)
49. (CPM or (passive adj (motion or movement))).ti,ab. (5468)
50. vibration/ (18,338)
51. (vibrati$ or mechanical stimul$).ti,ab. (32,959)
52. (balneology or balneotherap$ or balneo therap$ or bath$ or crenobalneotherap$ or thalassotherap$ or spa or spas).ti,ab. (42,994)
53. (thermotherap$ or thermo therap$ or hypertherm$ or hyper therm$ or diatherm$ or short wave or shortwave or ultrasonic or cryotherap$ or cryo therap$).ti,ab. (64,704)
54. (heat or hot or ice or cold).ti,ab. (227,320)
55. exp Cryotherapy/ (19,170)
56. exp Orthotic Devices/ (9055)
57. (brace$ or bracing or orthotic$ or orthoses).ti,ab. (8251)
58. (insert$ or insole$).ti,ab. (180,523)
59. (TENS or ALTENS).ti,ab. (6246)
60. (transcutaneous adj2 nerve stimulation).ti,ab. (1425)
61. (electroanalgesia or electro analgesia).ti,ab. (215)
62. (electric$ nerve stimulation or electrostimulation or electro stimulation).ti,ab. (4261)
63. EMS.ti,ab. (7219)
64. ((muscle or electric$) adj3 stimulat$).ti,ab. (55,060)
65. (neuromodulation or neuro modulation or neurostimulation or neuro stimulation).ti,ab. (3051)

66. interferential.ti,ab. (256)
67. Electromagnetic Fields/ (13,738)
68. Magnetic Field Therapy/ (516)
69. PEMF.ti,ab. (308)
70. ((electromagnetic$ or magnetic$) adj3 field$).ti,ab. (20,877)
71. (biomagnetic$ or bio magnetic$ or pulsed signal).ti,ab. (391)
72. Laser Therapy, Low-Level/ (2660)
73. laser.ti,ab. (124,281)
74. phototherapy/ (5215)
75. (light or phototherap$ or photo therap$).ti,ab. (347,805)
76. or/21-75 (2,221,346)
77. 20 and 76 (4608)
78. exp Obesity/ (125,953)
79. Overweight/ (9213)
80. body mass index/ (69,416)
81. (obese or obesity).ti,ab. (141,323)
82. (overweight or over weight).ti,ab. (29,555)
83. Weight Loss/ (21,542)
84. Weight Gain/ (20,613)
85. (weight adj3 los$).ti,ab. (52,781)
86. (weight adj3 reduc$).ti,ab. (20,417)
87. (weight adj3 decreas$).ti,ab. (13,520)
88. (weight adj3 gain$).ti,ab. (41,860)
89. (weight adj3 increas$).ti,ab. (27,215)
90. (weight adj3 chang$).ti,ab. (14,373)
91. (BMI or body mass index).ti,ab. (102,658)
92. or/78-91 (367,063)
93. Bariatrics/ (209)
94. exp Diet/ (178,778)
95. exp Diet Therapy/ (38,718)
96. (diet$ or slim or slimming).ti,ab. (336,098)
97. (weight adj3 control$).ti,ab. (10,820)
98. (weight adj3 manage$).ti,ab. (2468)
99. low calorie$.ti,ab. (1958)
100. calorie control$.ti,ab. (64)
101. (calorie adj3 count$).ti,ab. (79)
102. (caloric adj3 restrict$).ti,ab. (2805)
103. (calorie$ adj3 restrict$).ti,ab. (1998)
104. (energy adj3 restrict$).ti,ab. (1894)
105. (protein adj3 restrict$).ti,ab. (3973)
106. (weight watchers or weightwatchers or slimfast or nutrition class$ or meal replacement$).ti,ab. (470)
107. or/93-106 (436,733)
108. exp Anti-Obesity Agents/ (10,515)
109. ((anti obes$ or antiobes$) adj3 (agent$ or drug$ or therap$ or medicine$)).ti,ab. (877)
110. appetite suppressant$.ti,ab. (492)
111. appetite depressant$.ti,ab. (108)
112. (orlistat or xenical).ti,ab,rn. (1224)
113. phentermine.ti,ab,rn. (794)
114. Phentermine/ (659)
115. (sibutramine or reductil).ti,ab,rn. (1067)
116. (rimonabant or acomplia).ti,ab,rn. (2109)

117. or/108-116 (13,527)
118. 107 or 117 (447,143)
119. 20 and 92 and 118 (119)
120. 77 or 119 (4647)
121. randomized controlled trial.pt. (342,334)
122. controlled clinical trial.pt. (85,694)
123. randomized.ab. (244,919)
124. placebo.ab. (136,550)
125. drug therapy.fs. (1,588,363)
126. randomly.ab. (175,193)
127. trial.ab. (253,825)
128. groups.ab. (1,145,730)
129. 121 or 122 or 123 or 124 or 125 or 126 or 127 or 128 (2,960,405)
130. exp animals/ not humans/ (3,812,817)
131. 129 not 130 (2,515,366)
132. 120 and 131 (1614)

## Methods for rating study quality and risk of bias

### Method for rating overall study quality
See column headings in following table for description of questions.

- Excellent – answers 'yes' for questions 1–3, 5, 6, 8–14 and 'yes' or 'not applicable' for question 4.
- Good – answers 'yes' for questions 1, 5, 8, 9, 11–14 and 'yes' or 'unclear/not stated' for questions 2–4 (or 'not applicable' for question 4).
- Satisfactory – answers 'yes' for questions 1, 5, 11–13 and not 'no' for question 4.
- Poor – the answer is not 'yes' for one of the criteria required for 'satisfactory'.

### Comparability of quality assessment and risk of bias
In applying the risk-of-bias tool we included the level of patient blinding when assessing the blinding of the outcome assessment. This was necessary given that patients recorded their pain level and response to treatment. Therefore, even in studies in which the outcome assessor was reported as blinded, if patient were not blinded the risk of bias was recorded as high. Note that both patients and outcome assessor had to be blinded for a trial to be rated as being of 'good' quality.

All of the studies rated as being of 'poor' quality were found to have a high or unclear risk of bias. Of the 27 trials rated 'satisfactory', 21 were found to have a high risk of bias, three an unclear risk of bias and three a low risk of bias. Of the 21 rated as being at high risk of bias, the explanation in most cases was that the study was not blinded: the above method for rating study quality made allowances for the difficulties of blinding treatments in trials of physical therapies, so a trial could be 'satisfactory' (but not 'good') if it was not blinded. In addition, this method accounted for poor reporting of randomisation and allocation concealment methods, meaning that a study could be rated as 'satisfactory' (but not 'good') if these were unclear, providing the study groups appeared to have similar baseline characteristics (i.e. the aim of randomisation/allocation concealment appeared to have been achieved). Of the four 'good' studies, two had an unclear risk of bias and two a low risk of bias.

## Study quality assessment/risk-of-bias details for original review in 2011

For algorithm to produce overall study quality ratings see previous section.

| Study | 1. Number of patients randomised to treatment stated? | 2. Method of randomisation appropriate? | 3. Appropriate allocation concealment? | 4. Appropriate type of placebo used? | 5. Group baseline characteristics comparable? | 6. Described as double blind? | 7. Treatment giver blinded? | 8. Patient blinded? | 9. Outcome assessor blinded? | 10. Use of a power calculation reported? | 11. Eligibility criteria adequately reported? | 12. Clear reporting of losses to follow-up? | 13. Intention-to-treat data reported/analysed? | 14. At least 90% full follow-up achieved? | Overall study quality/risk of bias |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Abrahams 2002[349] | Yes | UC/NS | UC/NS | N/A | No | Yes | No | No | UC/NS | No | Yes | No | UC | UC | Poor |
|  |  | UC | UC |  | High |  |  |  | UC |  |  |  | UC |  | High |
| Adedoyin 2002[301] | Yes | No | UC/NS | Yes | Yes | No | No | Yes | Yes | No | Yes | No | UC | UC | Poor |
|  |  | High | UC |  | Low |  |  |  | Low |  |  |  | UC |  | High |
| Aglamiş 2009[189] | Yes | Yes | UC/NS | N/A | No | No | UC/NS | No | No | No | Yes | Yes | No | No | Poor |
|  |  | Low | UC |  | High |  |  |  | High |  |  |  | High |  | High |
| Alcidi 2007[350] | Yes | UC/NS | UC/NS | N/A | UC/NS | No |  | No | No | No | No | No | UC | UC | Poor |
|  |  | UC | UC |  | UC |  |  |  | High |  |  |  | High |  | High |
| An 2008[351] | Yes | UC/NS | UC/NS | N/A | No | No | No | No | No | No | Yes | Yes | No | No | Poor |
|  |  | UC | UC |  | High |  |  |  | High |  |  |  | High |  | High |
| Aman 1988[352] | Yes | UC/NS | UC/NS | N/A | UC/NS | No | No | No | No | No | No | No | UC | UC | Poor |
|  |  | UC | UC |  | UC |  |  |  | High |  |  |  | UC |  | High |
| Baker 2001[353] | Yes | Yes | Yes | No | Yes | No | No | No | No | No | Yes | Yes | No | No | Poor |
|  |  | Low | Low |  | low |  |  |  | High |  |  |  | High |  | High |
| Baker 2007[354] | Yes | Yes | UC/NS | Yes | Yes | Yes | UC/NS | UC/NS | UC/NS | Yes | Yes | Yes | No | Yes | Poor |
|  |  | Low | UC |  | Low |  |  |  | UC |  |  |  | UC |  | UC |
| Bálint 2007[355] | Yes | No | Yes | Yes | Yes | Yes | No | Yes | Yes | No | Yes | Yes | No | No | Poor |
|  |  | High | Low |  | Low |  |  |  | Low |  |  |  | High |  | High |
| Bao 2007[356] | Yes | UC/NS | UC/NS | N/A | Yes | No | No | No | UC/NS | No | Yes | No | Yes | UC | Poor |
|  |  | UC | UC |  | Low |  |  |  | UC |  |  |  | UC |  | UC |
| Berman 1999[357] | Yes | Yes | Yes | N/A | No | No | No | No | No | Yes | Yes | Yes | Yes | No | Poor |
|  |  | Low | Low |  | High |  |  |  | High |  |  |  | Low |  | High |
| Berman 2004[134] | Yes | Yes | Yes | Yes | Yes | No | No | Yes | Yes | Yes | Yes | Yes | Yes | No | Satisfactory |
|  |  | Low | Low |  | Low |  |  |  | Low |  |  |  | Low |  | Low |

| Study | 1. Number of patients randomised to treatment stated? | 2. Method of randomisation appropriate? | 3. Appropriate allocation concealment? | 4. Appropriate type of placebo used? | 5. Group baseline characteristics comparable? | 6. Described as double blind? | 7. Treatment giver blinded? | 8. Patient blinded? | 9. Outcome assessor blinded? | 10. Use of a power calculation reported? | 11. Eligibility criteria adequately reported? | 12. Clear reporting of losses to follow-up? | 13. Intention-to-treat data reported/analysed? | 14. At least 90% full follow-up achieved? | Overall study quality/risk of bias |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bezalel 2010[358] | Yes | Yes / Low | UC/NS / UC | N/A | No / High | No | No | No | Yes / High | Yes | Yes | Yes | Yes / Low | No | Poor / High |
| Bilgici 2004[359] | No | UC/NS / UC | UC/NS / UC | N/A | UC/NS / UC | No | UC/NS | UC/NS | UC/NS / UC | No | No | No | UC / UC | UC | Poor / UC |
| Börjesson 1996[190] | Yes | Yes / Low | UC/NS / UC | N/A | Yes / Low | No | No | No | Yes / High | No | Yes | Yes | Yes / Low | Yes | Satisfactory / High |
| Brismée 2007[360] | Yes | Yes / Low | UC/NS / UC | N/A | Yes / Low | No | No | No | Yes / High | Yes | Yes | Yes | Yes / Low | No | Satisfactory / High |
| Bülow 1994[361] | Yes | UC/NS / UC | Yes / Low | Yes | UC/NS / UC | Yes | Yes | Yes | Yes / Low | Yes | Yes | Yes | Yes / Low | Yes | Poor / UC |
| Burch 2008[303] | Yes | Yes / Low | Yes / Low | N/A | No / High | No | Yes | Yes | Yes / Low | Yes | Yes | Yes | No / High | No | Poor / High |
| Callaghan 1995[362] | Yes | UC/NS / UC | UC/NS / UC | N/A | No / High | No | No | No | No / High | No | No | No | UC / UC | UC | Poor / High |
| Callaghan 2005[363] | Yes | UC/NS / UC | UC/NS / UC | Yes | Yes / Low | Yes | No | No | Yes / High | Yes | Yes | Yes | No / High | No | Poor / High |
| Cantarini 2007[364] | Yes | UC/NS / UC | UC/NS / UC | N/A | UC/NS / UC | No | No | No | Yes / High | No | Yes | Yes | Yes / Low | Yes | Poor / High |
| Cheing 2002[365] | Yes | UC/NS / UC | UC/NS / UC | N/A | UC/NS / UC | No | No | No | No / High | No | Yes | Yes | No / High | Yes | Poor / High |
| Cheing 2003[191] | Yes | UC/NS / UC | UC/NS / UC | No | No / High | UC/NS | No | No | UC/NS / High | No | Yes | Yes | No / High | Yes | Poor / High |
| Chen 2008[366] | Yes | Yes / Low | Yes / Low | Yes | Yes / Low | Yes | Yes | Yes | Yes / Low | Yes | Yes | Yes | No / High | No | Poor / High |
| Christensen 2005[367] | Yes | UC/NS / UC | UC/NS / UC | N/A | Yes / Low | No | No | No | No / High | Yes | Yes | Yes | No / High | No | Poor / High |

| Study | 1. Number of patients randomised to treatment stated? | 2. Method of randomisation appropriate? | 3. Appropriate allocation concealment? | 4. Appropriate type of placebo used? | 5. Group baseline characteristics comparable? | 6. Described as double blind? | 7. Treatment giver blinded? | 8. Patient blinded? | 9. Outcome assessor blinded? | 10. Use of a power calculation reported? | 11. Eligibility criteria adequately reported? | 12. Clear reporting of losses to follow-up? | 13. Intention-to-treat data reported/analysed? | 14. At least 90% full follow-up achieved? | Overall study quality/risk of bias |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Clarke 1974[368] | Yes | UC/NS (UC) | UC/NS (UC) | Yes | Yes (Low) | No | UC/NS | UC/NS | UC/NS (UC) | No | No | Yes | Yes (Low) | Yes | Poor |
| Defrin 2005[369] | Yes | UC/NS (UC) | UC/NS (UC) | Yes | Yes (Low) | No | No | Yes | UC/NS (UC) | No | Yes | No | UC (UC) | UC | Poor |
| Dias 2003[187] | Yes | UC/NS (UC) | UC/NS (UC) | N/A | UC/NS (UC) | No | No | No | No (High) | No | No | Yes | No (High) | Yes | Poor |
| Durmuş 2007[370] | Yes | UC/NS (UC) | UC/NS (UC) | N/A | Yes (Low) | No | No | No | No (High) | No | Yes | Yes | Yes (Low) | Yes | Satisfactory |
| Ettinger 1997[371] | Yes | Yes (Low) | Yes (Low) | N/A | UC/NS (UC) | No | No | No | Yes (High) | Yes | Yes | Yes | Yes (Low) | No | Poor |
| Fioravanti 2010[372] | Yes | Yes (Low) | UC/NS (UC) | N/A | Yes (Low) | No | No | No | UC/NS (High) | No | Yes | Yes | Yes (Low) | Yes | Satisfactory |
| Fischer 2005[373] | Yes | No (High) | No (High) | Yes | Yes (Low) | Yes | Yes | Yes | Yes (Low) | No | Yes | Yes | No (High) | Yes | Poor |
| Flusser 2002[374] | Yes | UC/NS (UC) | UC/NS (UC) | Yes | Yes (Low) | Yes | Yes | Yes | Yes (Low) | No | Yes | No | UC (UC) | UC | Poor |
| Forestier 2010[375] | Yes | UC/NS (UC) | UC/NS (UC) | N/A | Yes (Low) | No | No | No | UC/NS (High) | Yes | Yes | Yes | No (High) | No | Poor |
| Fukuda 2008[376] | Yes | UC/NS (UC) | Yes (Low) | UC | Yes (Low) | No | No | UC/NS | Yes (Low) | No | Yes | Yes | No (High) | No | Poor |
| Garland 2007[377] | Yes | Yes (Low) | Yes (Low) | Yes | No (High) | Yes | Yes | Yes | Yes (Low) | No | Yes | Yes | No (High) | Yes | Poor |
| Grimmer 1992[378] | Yes | Yes (Low) | Yes (Low) | Yes | UC/NS (UC) | Yes | No | Yes | Yes (Low) | No | Yes | Yes | UC (UC) | Yes | Poor |
| Gür 2002[379] | Yes | UC/NS (UC) | UC/NS (UC) | N/A | Yes (Low) | No | UC/NS | UC/NS | UC/NS (UC) | No | No | No | UC (UC) | Yes | Poor |
| Gür 2003[380] | Yes | UC/NS (UC) | UC/NS (UC) | Yes | Yes (Low) | Yes | Yes | Yes | No (High) | No | Yes | Yes | Yes (Low) | Yes | Satisfactory |

| Study | 1. Number of patients randomised to treatment stated? | 2. Method of randomisation appropriate? | 3. Appropriate allocation concealment? | 4. Appropriate type of placebo used? | 5. Group baseline characteristics comparable? | 6. Described as double blind? | 7. Treatment giver blinded? | 8. Patient blinded? | 9. Outcome assessor blinded? | 10. Use of a power calculation reported? | 11. Eligibility criteria adequately reported? | 12. Clear reporting of losses to follow-up? | 13. Intention-to-treat data reported/analysed? | 14. At least 90% full follow-up achieved? | Overall study quality/risk of bias |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Hasegawa 2010[381] | Yes | UC/NS | UC/NS | N/A | Yes | No | No | No | No | No | No | Yes | Yes | Yes | Poor |
|  |  | UC | UC |  | Low |  |  |  | High |  |  |  | Low |  | High |
| Hay 2006[382] | Yes | Yes | Yes | N/A | Yes | No | No | No | Yes | Yes | Yes | Yes | No | No | Poor |
|  |  | Low | Low |  | Low |  |  |  | High |  |  |  | High |  | High |
| Hinman 2002[383] | No | UC/NS | Yes | Yes | UC/NS | Yes | Yes | UC/NS | Yes | No | Yes | Yes | No | Yes | Poor |
|  |  | UC | Low |  | UC |  |  |  | UC |  |  |  | High |  | High |
| Huang 2005[384] | Yes | No | Yes | N/A | Yes | No | No | UC/NS | Yes | No | No | Yes | UC | Yes | Poor |
|  |  | High | Low |  | Low |  |  |  | UC |  |  |  |  |  | High |
| Hurley 2007[385] | Yes | UC/NS | UC/NS | N/A | Yes | No | No | No | Yes | Yes | Yes | Yes | Yes | No | Satisfactory |
|  |  | UC | UC |  | Low |  |  |  | High |  |  |  | Low |  | High |
| Itoh 2008[386] | Yes | Yes | UC/NS | N/A | UC/NS | No | No | No | No | No | Yes | Yes | No | No | Poor |
|  |  | Low | UC |  | UC |  |  |  | High |  |  |  | High |  | High |
| Itoh 2008[387] | Yes | UC/NS | No | No | Yes | No | No | No | Yes | No | Yes | Yes | No | No | Poor |
|  |  | UC | High |  | Low |  |  |  | High |  |  |  | High |  | High |
| Jacobson 2001[388] | No | UC/NS | Yes | Yes | UC/NS | Yes | Yes | Yes | Yes | No | No | No | No | Yes | Poor |
|  |  | UC | Low |  | UC |  |  |  | Low |  |  |  | High |  | High |
| Jan 2008[389] | Yes | Yes | UC/NS | N/A | Yes | No | No | No | No | Yes | Yes | Yes | No | Yes | Satisfactory |
|  |  | Low | UC |  | Low |  |  |  | High |  |  |  | High |  | High |
| Jenkinson 2009[390] | Yes | Yes | Yes | N/A | Yes | No | No | No | No | Yes | Yes | Yes | Yes | No | Satisfactory |
|  |  | Low | Low |  | Low |  |  |  | High |  |  |  | Low |  | High |
| Jubb 2008[391] | Yes | Yes | UC/NS | Yes | No | No | No | Yes | Yes | Yes | Yes | Yes | No | Yes | Poor |
|  |  | Low | UC |  | High |  |  |  | Low |  |  |  | High |  | High |
| Kang 2006[392] | Yes | UC/NS | UC/NS | N/A | UC/NS | No | UC/NS | UC/NS | UC/NS | No | No | No | UC | | Poor |
|  |  | UC | UC |  | UC |  |  |  | UC |  |  |  | UC |  |  |
| Kang 2007[393] | Yes | UC/NS | UC/NS | Yes | Yes | No | No | Yes | Yes | Yes | Yes | Yes | No | Yes | Poor |
|  |  | UC | UC |  | Low |  |  |  | Low |  |  |  | High |  | High |

| Study | 1. Number of patients randomised to treatment stated? | 2. Method of randomisation appropriate? | 3. Appropriate allocation concealment? | 4. Appropriate type of placebo used? | 5. Group baseline characteristics comparable? | 6. Described as double blind? | 7. Treatment giver blinded? | 8. Patient blinded? | 9. Outcome assessor blinded? | 10. Use of a power calculation reported? | 11. Eligibility criteria adequately reported? | 12. Clear reporting of losses to follow-up? | 13. Intention-to-treat data reported/analysed? | 14. At least 90% full follow-up achieved? | Overall study quality/risk of bias |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Karagülle 2007[394] | Yes | Yes | Yes | N/A | Yes | No | No | No | Yes | No | Yes | Yes | Yes | No | Satisfactory |
|  |  | Low | Low |  | Low |  |  |  | High |  |  |  | Low |  | High |
| Keefe 2004[395] | Yes | UC/NS | UC/NS | N/A | Yes | No | No | No | No | No | Yes | Yes | UC | Yes | Poor |
|  |  | UC | UC |  | Low |  |  |  | High |  |  |  | UC |  | High |
| Keogan 2007[396] | Yes | UC/NS | UC/NS | N/A | Yes | No | UC/NS | UC/NS | UC/NS | No | No | No | UC | UC | Poor |
|  |  | UC | UC |  | Low |  |  |  | UC |  |  |  | UC |  | UC |
| Kovács 2002[397] | No | Yes | UC/NS | Yes | UC/NS | Yes | UC/NS | Yes | Yes | No | No | Yes | UC | No | Poor |
|  |  | Low | UC |  | UC |  |  |  | Low |  |  |  | UC |  | UC |
| Kovar 1992[398] | Yes | Yes | UC/NS | N/A | No | No | No | No | No | No | Yes | Yes | No | Yes | Poor |
|  |  | Low | UC |  | High |  |  |  | High |  |  |  | High |  | High |
| Kupthiratsaikul 2002[399] | Yes | UC/NS | UC/NS | N/A | Yes | No | No | No | UC/NS | No | No | Yes | No | Yes | Poor |
|  |  | UC | UC |  | Low |  |  |  | High |  |  |  | High |  | High |
| Lansdown 2009[126] | Yes | Yes | Yes | Yes | UC/NS | No | No | No | No | No | Yes | Yes | No | Yes | Poor |
|  |  | Low | Low |  | UC |  |  |  | High |  |  |  | High |  | High |
| Law 2004[400] | No | Yes | UC/NS | N/A | Yes | Yes | No | Yes | Yes | No | Yes | Yes | Yes | UC | Poor |
|  |  | Low | UC |  | Low |  |  |  | Low |  |  |  | Low |  | UC |
| Lee 2009[401] | Yes | Yes | UC/NS | N/A | Yes | No | No | No | Yes | No | Yes | Yes | Yes | Yes | Satisfactory |
|  |  | Low | UC |  | Low |  |  |  | Low |  |  |  | Low |  | High |
| Lewis 1988[402] | Yes | UC/NS | UC/NS | UC/NS | UC/NS | Yes | UC/NS | UC/NS | UC/NS | No | No | No | UC | No | Poor |
|  |  | UC | UC |  | UC |  |  |  | UC |  |  |  | UC |  | UC |
| Lewis 1994[403] | Yes | UC/NS | UC/NS | Yes | UC/NS | No | Yes | Yes | Yes | Yes | Yes | Yes | No | No | Poor |
|  |  | UC | UC |  | UC |  |  |  | Low |  |  |  | High |  | High |
| Lewis 1984[404] | Yes | UC/NS | UC/NS | Yes | UC/NS | Yes | Yes | Yes | UC/NS | No | Yes | Yes | No | Yes | Poor |
|  |  | UC | UC |  | UC |  |  |  | UC |  |  |  | High |  | High |
| Lim 2008[405] | Yes | Yes | Yes | N/A | Yes | No | No | No | No | Yes | Yes | Yes | Yes | Yes | Satisfactory |
|  |  | Low | Low |  | Low |  |  |  | High |  |  |  | Low |  | High |
| Lin 2009[406] | Yes | Yes | Yes | N/A | Yes | No | No | No | Yes | Yes | Yes | Yes | Yes | Yes | Satisfactory |
|  |  | Low | Low |  | Low |  |  |  | High |  |  |  | Low |  | High |

| Study | 1. Number of patients randomised to treatment stated? | 2. Method of randomisation appropriate? | 3. Appropriate allocation concealment? | 4. Appropriate type of placebo used? | 5. Group baseline characteristics comparable? | 6. Described as double blind? | 7. Treatment giver blinded? | 8. Patient blinded? | 9. Outcome assessor blinded? | 10. Use of a power calculation reported? | 11. Eligibility criteria adequately reported? | 12. Clear reporting of losses to follow-up? | 13. Intention-to-treat data reported/analysed? | 14. At least 90% full follow-up achieved? | Overall study quality/risk of bias |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Lu 2010[192] | Yes | UC/NS | UC/NS | Yes | Yes | No | No | Yes | No | No | Yes | Yes | Yes | Yes | Satisfactory |
| | | UC | UC | | Low | | | | High | | | | Low | | High |
| Lund 2008[407] | Yes | UC/NS | Yes | N/A | No | No | No | No | No | Yes | Yes | Yes | Yes | Yes | Poor |
| | | UC | Low | | High | | | | High | | | | Low | | High |
| Mahboob 2009[408] | Yes | UC/NS | Yes | Yes | UC/NS | Yes | Yes | Yes | Yes | No | No | No | Yes | UC | Poor |
| | | UC | Low | | UC | | | | Low | | | | UC | | UC |
| Maillefert 2001[409] | Yes | UC/NS | UC/NS | Yes | Yes | No | No | UC/NS | UC/NS | Yes | Yes | Yes | Yes | Yes | Satisfactory |
| | | UC | UC | | Low | | | | UC | | | | Low | | UC |
| Maurer 1999[410] | Yes | Yes | UC/NS | N/A | Yes | No | No | No | Yes | Yes | Yes | No | UC | No | Poor |
| | | Low | UC | | Low | | | | High | | | | UC | | High |
| Mazzuca 2004[411] | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | No | Yes | Poor |
| | | Low | Low | | Low | | | | Low | | | | High | | High |
| McCarthy 2004[412] | Yes | Yes | Yes | N/A | Yes | No | No | No | Yes | Yes | Yes | Yes | UC | No | Satisfactory |
| | | Low | Low | | Low | | | | High | | | | UC | | High |
| Messier 2004[413] | Yes | Yes | Yes | N/A | Yes | No | No | No | Yes | Yes | Yes | Yes | Yes | No | Satisfactory |
| | | Low | Low | | Low | | | | High | | | | Low | | High |
| Miller 2006[414] | Yes | UC/NS | UC/NS | N/A | Yes | No | No | No | No | No | Yes | Yes | No | No | Poor |
| | | UC | UC | | Low | | | | High | | | | High | | High |
| Miranda-Filloy 2005[415] | Yes | UC/NS | UC/NS | Yes | UC/NS | No | UC/NS | Yes | UC/NS | No | No | No | UC | UC | Poor |
| | | UC | UC | | UC | | | | UC | | | | UC | | UC |
| Molsberger 1994[416] | Yes | UC/NS | UC/NS | Yes | Yes | Yes | No | Yes | Yes | No | Yes | No | UC | Yes | Poor |
| | | UC | UC | | Low | | | | Low | | | | UC | | UC |
| Moss 2007[417] | Yes | UC/NS | UC/NS | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Good |
| | | UC | UC | | Low | | | | Low | | | | Low | | UC |
| Ng 2003[418] | Yes | Yes | No | N/A | Yes | No | No | No | Yes | No | Yes | No | UC | UC | Poor |
| | | Low | High | | Low | | | | High | | | | UC | | High |

| Study | 1. Number of patients randomised to treatment stated? | 2. Method of randomisation appropriate? | 3. Appropriate allocation concealment? | 4. Appropriate type of placebo used? | 5. Group baseline characteristics comparable? | 6. Described as double blind? | 7. Treatment giver blinded? | 8. Patient blinded? | 9. Outcome assessor blinded? | 10. Use of a power calculation reported? | 11. Eligibility criteria adequately reported? | 12. Clear reporting of losses to follow-up? | 13. Intention-to-treat data reported/analysed? | 14. At least 90% full follow-up achieved? | Overall study quality/risk of bias |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nguyen 1997[419] | Yes | Yes / Low | UC/NS / UC | N/A | UC/NS / UC | No | No | No | No / High | No | Yes | Yes | Yes / Low | Yes | Poor / High |
| Ni 2010[420] | Yes | Yes / Low | Yes / Low | N/A | Yes / Low | No | No | No | Yes / High | No | Yes | Yes | No / High | No | Poor / High |
| Nigg 2006[421] | Yes | Yes / Low | UC/NS / UC | N/A | Yes / Low | No | No | No | No / High | No | Yes | Yes | Yes / Low | Yes | Satisfactory / High |
| Nivbrant 1989[422] | Yes | UC/NS / UC | UC/NS / UC | Yes | UC/NS / UC | No | UC/NS | UC/NS | UC/NS | No | No | No | UC | Yes | Poor / UC |
| Péloquin 1999[423] | Yes | Yes / Low | UC/NS / UC | N/A | Yes / Low | No | No | No | Yes / High | Yes | Yes | Yes | No / High | Yes | Poor / High |
| Perlman 2006[424] | Yes | Yes / Low | UC/NS / UC | N/A | No / High | No | No | No | No / High | Yes | Yes | Yes | Yes / Low | No | Poor / High |
| Perrot 1998[425] | Yes | UC/NS / UC | UC/NS / UC | UC/NS | UC/NS / UC | Yes | UC/NS | UC/NS | UC/NS / UC | No | No | Yes | UC | No | Poor / UC |
| Petrou 1988[426] | Yes | UC/NS / UC | UC/NS / UC | UC/NS | Yes / Low | No | No | Yes | Yes / Low | No | No | Yes | UC | Yes | Poor / UC |
| Pietrosimone 2009[427] | Yes | UC/NS / UC | Yes / Low | N/A | Yes / Low | No | No | No | No / High | No | No | Yes | Yes / Low | Yes | Poor / High |
| Pipitone 2001[428] | Yes | Yes / Low | Yes / Low | Yes | No / High | Yes | Yes | Yes | Yes / Low | Yes | Yes | Yes | Yes / Low | No | Poor / High |
| Pollard 2008[429] | Yes | Yes / Low | Yes / Low | Yes | UC/NS / UC | No | No | Yes | Yes / Low | No | Yes | Yes | Yes / Low | Yes | Poor / UC |

| Study | 1. Number of patients randomised to treatment stated? | 2. Method of randomisation appropriate? | 3. Appropriate allocation concealment? | 4. Appropriate type of placebo used? | 5. Group baseline characteristics comparable? | 6. Described as double blind? | 7. Treatment giver blinded? | 8. Patient blinded? | 9. Outcome assessor blinded? | 10. Use of a power calculation reported? | 11. Eligibility criteria adequately reported? | 12. Clear reporting of losses to follow-up? | 13. Intention-to-treat data reported/analysed? | 14. At least 90% full follow-up achieved? | Overall study quality/risk of bias |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Rapp 2009[430] | Yes | UC/NS | UC/NS | N/A | UC/NS | No | No | No | No | No | Yes | No | UC | UC | Poor |
| | | UC | UC | | UC | | | | High | | | | UC | | High |
| Rattanachaiyanont 2008[431] | Yes | Yes | Yes | Yes | No | Yes | No | Yes | Yes | Yes | Yes | Yes | No | Yes | Poor |
| | | Low | Low | | High | | | | Low | | | | High | | High |
| Røgind 1998[432] | Yes | Yes | UC/NS | N/A | UC/NS | No | No | No | Yes | No | Yes | Yes | No | Yes | Poor |
| | | Low | UC | | UC | | | | High | | | | High | | High |
| Rosemffet 2004[433] | Yes | Yes | UC/NS | N/A | No | No | No | No | No | No | Yes | No | No | No | Poor |
| | | Low | UC | | High | | | | High | | | | High | | High |
| Sangdee 2002[434] | Yes | UC/NS | UC/NS | No | Yes | No | No | No | Yes | Yes | Yes | Yes | No | Yes | Poor |
| | | UC | UC | | Low | | | | High | | | | High | | High |
| Schilke 1996[435] | Yes | Yes | UC/NS | N/A | No | No | No | No | No | No | No | No | Yes | No | Poor |
| | | Low | UC | | High | | | | High | | | | UC | | High |
| Selfe 2008[436] | Yes | Yes | Yes | Yes | Yes | No | No | Yes | Yes | Yes | Yes | Yes | No | Yes | Poor |
| | | Low | Low | | Low | | | | Low | | | | High | | High |
| Shen 2009[437] | Yes | Yes | Yes | Yes | Yes | No | No | Yes | Yes | No | Yes | Yes | Yes | No | Satisfactory |
| | | Low | Low | | Low | | | | Low | | | | Low | | Low |
| Sherman 2009[438] | No | Yes | UC/NS | No | No | No | UC/NS | No | Yes | No | Yes | Yes | No | Yes | Poor |
| | | Low | UC | | High | | | | High | | | | High | | High |
| Smith 1983[439] | Yes | Yes | Yes | Yes | Yes | No | No | Yes | Yes | No | Yes | Yes | No | Yes | Poor |
| | | Low | Low | | Low | | | | Low | | | | High | | High |
| Stelian 1992[440] | Yes | UC/NS | UC/NS | Yes | UC/NS | Yes | Yes | Yes | UC/NS | No | No | Yes | UC | Yes | Poor |
| | | UC | UC | | UC | | | | UC | | | | UC | | UC |
| Sukenik 1999[441] | Yes | UC/NS | UC/NS | Yes | Yes | No | No | No | Yes | No | No | Yes | No | Yes | Poor |
| | | UC | UC | | Low | | | | High | | | | High | | High |

| Study | 1. Number of patients randomised to treatment stated? | 2. Method of randomisation appropriate? | 3. Appropriate allocation concealment? | 4. Appropriate type of placebo used? | 5. Group baseline characteristics comparable? | 6. Described as double blind? | 7. Treatment giver blinded? | 8. Patient blinded? | 9. Outcome assessor blinded? | 10. Use of a power calculation reported? | 11. Eligibility criteria adequately reported? | 12. Clear reporting of losses to follow-up? | 13. Intention-to-treat data reported/analysed? | 14. At least 90% full follow-up achieved? | Overall study quality/risk of bias |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Takeda 1994[442] | Yes | UC/NS | UC/NS | Yes | Yes | No | No | Yes | Yes | Yes | No | No | Yes | Yes | Poor |
|  |  | UC | UC |  | Low |  |  |  | Low |  |  |  | UC |  | UC |
| Talbot 2003[443] | Yes | UC/NS | UC/NS | N/A | No | No | No | No | No | No | Yes | Yes | No | No | Poor |
|  |  | UC | UC |  | High |  |  |  | High |  |  |  | High |  | High |
| Tascioglu 2004[444] | Yes | UC/NS | UC/NS | Yes | Yes | No | Yes | No | No | No | Yes | Yes | Yes | Yes | Satisfactory |
|  |  | UC | UC |  | Low |  |  |  | High |  |  |  | Low |  | High |
| Taylor 1981[445] | Yes | UC/NS | UC/NS | Yes | Yes | Yes | Yes | Yes | Yes | No | No | Yes | No | No | Poor |
|  |  | UC | UC |  | Low |  |  |  | Low |  |  |  | High |  | High |
| Thamsborg 2005[446] | Yes | UC/NS | UC/NS | UC/NS | Yes | Yes | UC/NS | Yes | Yes | Yes | Yes | Yes | No | Yes | Poor |
|  |  | UC | UC |  | Low |  |  |  | Low |  |  |  | High |  | High |
| Thorstensson 2005[447] | Yes | No | UC/NS | N/A | Yes | No | No | No | No | Yes | Yes | Yes | No | Yes | Poor |
|  |  | High |  |  | Low |  |  |  | High |  |  |  | High |  | High |
| Tishler 2004[448] | Yes | UC/NS | No | N/A | Yes | No | UC/NS | No | No | No | Yes | Yes | No | Yes | Poor |
|  |  | UC | High |  | Low |  |  |  | High |  |  |  | High |  | High |
| Toda 2008[449] | Yes | Yes | Yes | Yes | Yes | No | No | No | Yes | No | Yes | Yes | Yes | Yes | Satisfactory |
|  |  | Low | Low |  | Low |  |  |  | High |  |  |  | Low |  | High |
| Topp 2002[450] | Yes | UC/NS | UC/NS | N/A | Yes | No | No | No | UC/NS | No | Yes | No | Yes | Yes | Poor |
|  |  | UC | UC |  | Low |  |  |  | High |  |  |  | UC |  | High |
| Trans 2009[451] | Yes | UC/NS | Yes | N/A | Yes | No | No | UC/NS | UC/NS | Yes | Yes | Yes | Yes | No | Satisfactory |
|  |  | UC | Low |  | Low |  |  |  | UC |  |  |  | Low |  | UC |
| Trock 1994[452] | No | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | No | Yes | Yes | No | UC | Poor |
|  |  | Low | Low |  | Low |  |  |  | Low |  |  |  | High |  | High |

| Study | 1. Number of patients randomised to treatment stated? | 2. Method of randomisation appropriate? | 3. Appropriate allocation concealment? | 4. Appropriate type of placebo used? | 5. Group baseline characteristics comparable? | 6. Described as double blind? | 7. Treatment giver blinded? | 8. Patient blinded? | 9. Outcome assessor blinded? | 10. Use of a power calculation reported? | 11. Eligibility criteria adequately reported? | 12. Clear reporting of losses to follow-up? | 13. Intention-to-treat data reported/analysed? | 14. At least 90% full follow-up achieved? | Overall study quality/risk of bias |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tucker 2003[453] | Yes | UC/NS | UC/NS | N/A | Yes | No | No | No | No | No | Yes | Yes | No | Yes | Poor |
| | | UC | UC | | Low | | | | High | | | | High | | High |
| Tukmachi 2004[454] | Yes | Yes | Yes | N/A | Yes | No | No | No | Yes | No | Yes | Yes | Yes | Yes | Satisfactory |
| | | Low | Low | | Low | | | | High | | | | Low | | High |
| Vas 2004[120] | Yes | Yes | Yes | Yes | Yes | No | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Good |
| | | Low | Low | | Low | | | | Low | | | | Low | | Low |
| Wang 2009[455] | Yes | Yes | Yes | N/A | No | No | No | No | UC/NS | No | Yes | Yes | Yes | Yes | Poor |
| | | Low | Low | | High | | | | UC | | | | Low | | High |
| Weiner 2007[188] | Yes | UC/NS | Yes | Yes | Yes | No | No | Yes | Yes | No | Yes | Yes | Yes | Yes | Good |
| | | UC | Low | | Low | | | | Low | | | | Low | | Low |
| Wigler 1995[456] | Yes | UC/NS | UC/NS | Yes | UC/NS | Yes | Yes | Yes | UC/NS | No | Yes | No | UC | Yes | Poor |
| | | UC | UC | | UC | | | | UC | | | | UC | | UC |
| Williamson 2007[136] | Yes | Yes | Yes | N/A | Yes | No | No | No | Yes | Yes | Yes | Yes | Yes | No | Satisfactory |
| | | Low | Low | | Low | | | | High | | | | Low | | High |
| Witt 2005[69] | Yes | Yes | Yes | Yes | Yes | No | No | Yes | UC/NS | Yes | Yes | Yes | Yes | Yes | Satisfactory |
| | | Low | Low | | Low | | | | UC | | | | Low | | UC |
| Witt 2006[77] | Yes | Yes | Yes | N/A | Yes | No | No | No | No | Yes | Yes | Yes | Yes | Yes | Satisfactory |
| | | Low | Low | | Low | | | | High | | | | Low | | High |
| Wolsko 2004[457] | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes | Poor |
| | | Low | Low | | High | | | | Low | | | | Low | | High |

| Study | 1. Number of patients randomised to treatment stated? | 2. Method of randomisation appropriate? | 3. Appropriate allocation concealment? | 4. Appropriate type of placebo used? | 5. Group baseline characteristics comparable? | 6. Described as double blind? | 7. Treatment giver blinded? | 8. Patient blinded? | 9. Outcome assessor blinded? | 10. Use of a power calculation reported? | 11. Eligibility criteria adequately reported? | 12. Clear reporting of losses to follow-up? | 13. Intention-to-treat data reported/analysed? | 14. At least 90% full follow-up achieved? | Overall study quality/ risk of bias |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Wu 2008[458] | Yes | No | UC/NS | N/A | Yes | No | No | No | No | No | Yes | No | UC | UC | Poor |
| | | High | | | Low | | | | High | | | | UC | | High |
| Yip 2007[193] | Yes | Yes | UC/NS | N/A | UC/NS | No | No | No | No | Yes | Yes | Yes | Yes | No | Poor |
| | | Low | UC | | UC | | | | High | | | | Low | | High |
| Young 1991[459] | No | UC/NS | UC/NS | Yes | UC/NS | UC/NS | UC/NS | UC/NS | UC/NS | No | Yes | No | UC | UC | Poor |
| | | UC | UC | | UC | | | | UC | | | | UC | | UC |
| Yurtkuran 1999[460] | Yes | UC/NS | UC/NS | Yes | UC/NS | No | No | Yes | Yes | No | Yes | Yes | Yes | Yes | Poor |
| | | UC | UC | | UC | | | | Low | | | | Low | | UC |
| Yurtkuran 2006[461] | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes | No | Yes | Poor |
| | | Low | Low | | Low | | | | Low | | | | High | | High |
| Zizic 1995[462] | Yes | UC/NS | UC/NS | UC/NS | Yes | Yes | UC/NS | UC/NS | UC/NS | No | Yes | Yes | No | Yes | Poor |
| | | UC | UC | | Low | | | | UC | | | | High | | High |

N/A, not applicable; NS, not specified; UC, unclear.

## Study quality assessment/risk-of-bias details for the 22 studies identified for the 2013 update

For algorithm to produce overall study quality rating see earlier.

| Study | 1. Number of patients randomised to treatment stated? | 2. Method of randomisation appropriate? | 3. Appropriate allocation concealment? | 4. Appropriate type of placebo used? | 5. Group baseline characteristics comparable? | 6. Described as double blind? | 7. Treatment giver blinded? | 8. Patient blinded? | 9. Outcome assessor blinded? | 10. Use of a power calculation reported? | 11. Eligibility criteria adequately reported? | 12. Clear reporting of losses to follow-up? | 13. Intention-to-treat data reported/analysed? | 14. At least 90% full follow-up achieved? | Overall study quality/risk of bias |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Arazpour 2013[463] | Yes | UC/NS | UC/NS | N/A | Yes | No | No | No | No | No | Yes | No | UC | UC | Poor |
|  |  | UC | UC |  | Low |  |  |  | High |  |  |  |  |  | High |
| Bennell 2011[464] | Yes | Yes | Yes | Yes | Yes | Yes | UC/NS | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Good |
|  |  | Low | Low |  | Low |  |  |  | Low |  |  |  | Low |  | Low |
| Bliddal 2011[465] | Yes | Yes | Yes | N/A | Yes | No | No | No | Yes | Yes | Yes | Yes | Yes | No | Satisfactory |
|  |  | Low | Low |  | Low |  |  |  | Low |  |  |  | Low |  | Low |
| Brosseau 2012[466] | Yes | Yes | UC/NS | N/A | Yes | No | No | No | Yes | No | Yes | Yes | No | No | Poor |
|  |  | Low | UC |  | Low |  |  |  | Low |  |  |  | High |  | High |
| Bruce-Brand 2012[467] | Yes | Yes | UC/NS | N/A | No | No | No | No | Yes | No | Yes | Yes | No | No | Poor |
|  |  | Low | UC |  | High |  |  |  | Low |  |  |  | High |  | High |
| Fary 2011[468] | Yes | Yes | Yes | Yes | No[a] | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Satisfactory[a] |
|  |  | Low | Low |  | UC |  |  |  | Low |  |  |  | Low |  | Low/UC[a] |
| Foroughi 2011[469] | Yes | UC/NS | UC/NS | UC/NS | Yes | No | No | Yes | UC/NS | Yes | Yes | Yes | No | No | Poor |
|  |  | UC | UC |  | Low |  |  |  | UC |  |  |  | High |  | High |
| Fukuda 2011[470] | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes | No | Yes | Poor |
|  |  | Low | Low |  | Low |  |  |  | Low |  |  |  | High |  | High |
| Fukuda 2011[471] | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes | Good |
|  |  | Low | Low |  | Low |  |  |  | Low |  |  |  | Low |  | Low |
| Giombini 2011[472] | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes | Yes | No | Satisfactory |
|  |  | Low | Low |  | Low |  |  |  | Low |  |  |  | Low |  | Low |
| Gundog 2012[302] | Yes | Yes | UC/NS | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Good |
|  |  | Low | UC |  | Low |  |  |  | Low |  |  |  | Low |  | UC |

| Study | 1. Number of patients randomised to treatment stated? | 2. Method of randomisation appropriate? | 3. Appropriate allocation concealment? | 4. Appropriate type of placebo used? | 5. Group baseline characteristics comparable? | 6. Described as double blind? | 7. Treatment giver blinded? | 8. Patient blinded? | 9. Outcome assessor blinded? | 10. Use of a power calculation reported? | 11. Eligibility criteria adequately reported? | 12. Clear reporting of losses to follow-up? | 13. Intention-to-treat data reported/analysed? | 14. At least 90% full follow-up achieved? | Overall study quality/risk of bias |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Hsieh 2012[473] | Yes | UC/NS | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Good |
|  |  | UC | Low |  | Low |  |  |  | Low |  |  |  | Low |  | UC |
| Imoto 2012[474] | Yes | Yes | Yes | N/A | Yes | No | No | No | Yes | Yes | Yes | Yes | Yes | No | Satisfactory |
|  |  | Low | Low |  | Low |  |  |  | Low |  |  |  | Low |  | Low |
| Lim 2010[475] | Yes | Yes | Yes | N/A | Yes | No | No | No | Yes | Yes | Yes | Yes | No | Yes | Poor |
|  |  | Low | Low |  | Low |  |  |  | Low |  |  |  | High |  | High |
| Mavrommatis 2012[476] | Yes | Yes | UC/NS | Yes | Yes | No | No | Yes | Yes | Yes | Yes | Yes | No | Yes | Poor |
|  |  | Low | UC |  | Low |  |  |  | Low |  |  |  | High |  | High |
| Miller 2011[477] | Yes | Yes | Yes | Yes | Yes | UC/NS | No | Yes | Yes | No | Yes | Yes | Yes | No | Satisfactory |
|  |  | Low | Low |  | Low |  |  |  | Low |  |  |  | Low |  | Low |
| Perlman 2012[478] | Yes | Yes | Yes | N/A | Yes | No | No | No | Yes | Yes | Yes | Yes | Yes | Yes | Satisfactory |
|  |  | Low | Low |  | Low |  |  |  | Low |  |  |  | Low |  | Low |
| Salacinski 2012[479] | Yes | Yes | Yes | N/A | No | No | No | No | No | Yes | Yes | Yes | No | No | Poor |
|  |  | Low | Low |  | High |  |  |  | High |  |  |  | High |  | High |
| Salli 2010[480] | Yes | UC/NS | UC/NS | N/A | Yes | No | No | No | Yes | No | Yes | Yes | Yes | Yes | Satisfactory |
|  |  | UC | UC |  | Low |  |  |  | Low |  |  |  | Low |  | UC |
| Suarez-Almazor 2010[125] | Yes | Yes | Yes | Yes | Yes | No | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Good |
|  |  | Low | Low |  | Low |  |  |  | Low |  |  |  | Low |  | Low |
| Vance 2012[481] | Yes | UC/NS | Yes | Yes | No | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes | Yes | Poor[b] |
|  |  | UC | Low |  | High |  |  |  | Low |  |  |  | Low |  | High |
| Wang 2011[482] | Yes | Yes | UC/NS | N/A | Yes | No | No | No | No | Yes | Yes | Yes | No | Yes | Poor |
|  |  | Low | UC |  | Low |  |  |  | High |  |  |  | High |  | High |

N/A, not applicable; NS, not specified; UC, unclear.
a Lack of baseline comparability on BMI could reduce the overall score from excellent to poor. However, as the BMI actual mean difference was not large (27 kg/m² vs. 29 kg/m²) and groups were comparable on all other important characteristics, including pain, WOMAC score and SF-36 score, and because the authors stated that 'adjustment for covariates did not alter any of the findings', an overall rating of satisfactory was considered appropriate for this study.
b Baseline characteristics not comparable and also duration of treatment too short.

## List of individual studies included in the systematic review by intervention

### *Acupuncture studies*

| Study | Treatments (if suitable for NMA, adjunct code in parentheses[a]) | Number analysed at end of treatment, data suitable for pain analyses? | Country | % Female | Age (years), mean unless stated | Weight (kg) | BMI (kg/m$^2$) | Method of diagnosis | Kellgren and Lawrence score |
|---|---|---|---|---|---|---|---|---|---|
| Bao 2007[356] | Acupuncture (2), standard care (4) | 40, yes | China | 63 | 62 | NR | NR | Clinical | NR |
| Berman 1999[357] | Acupuncture (2), standard care (2) | 73, yes | USA | 60 | 65 | NR | Mean 32 | Clinical and radiological | ≥ 2 |
| Berman 2004[134] | Acupuncture, sham acupuncture | 330, yes | USA | 63 | 66 | NR | NR | Clinical and radiological | ≥ 2 |
| Itoh 2008[386] | Acupuncture (1), TENS (1), standard care (2) | 18, yes | Japan | 66 | Range 62–83 | NR | NR | Clinical and radiological | ≥ 2 |
| Itoh 2008[387] | Acupuncture (2), sham acupuncture (2) | 24, yes | Japan | 77 | 73 | NR | NR | Clinical and radiological | ≥ 2 |
| Jubb 2008[391] | Acupuncture, sham acupuncture | 62, yes | UK | 81 | 65 | NR | Mean 32 | Clinical and radiological | 2 or 3 |
| Lansdown 2009[126] | Acupuncture (2), standard care (2) | 30, yes | UK | 60 | 64 | NR | NR | Clinical | NR/unclear |
| Lu 2010[192] | Acupuncture (2), sham acupuncture (2) | 20, yes | China | NR | 64 | Mean 66 | NR | Clinical and radiological | 2 or 3 |
| *Mavrommatis 2012[476]* | *Acupuncture, sham acupuncture, standard care* | *117, yes* | *Greece* | *76* | *62* | *NR* | *Mean 31* | *Clinical and radiological* | *≥ 2* |
| *Miller 2011[477]* | *Acupuncture, sham acupuncture* | *55, yes* | *Israel* | *69* | *71* | *NR* | *NR* | *Unclear/not stated* | *NR/unclear* |
| Molsberger 1994[416] | Acupuncture, sham acupuncture | 97, no[b] | Germany | 63 | 60 | NR | NR | Clinical and radiological | NR (Wirth classification) |
| Ng 2003[418] | Acupuncture (3), TENS (3), standard care (3) | 14 (imputed), yes (only vs. TENS) | China | 96 | 85 | NR | NR | Clinical | NR/unclear |
| Petrou 1988[426] | Acupuncture (1), sham acupuncture (1) | 31, yes | Hungary | 74 | 62 | Mean 80 | NR | Clinical | NR/unclear |

| Overall study quality | Duration of individual sessions (minutes) | Total number of sessions | Duration of treatment period (weeks) | Number of points needled | Depth of insertion | Electrical stimulation used? | Pain outcomes assessed | Overall WOMAC score reported? |
|---|---|---|---|---|---|---|---|---|
| Poor | 20 | 12 | 4 | 7 | Unclear | Yes | Lysholm scores | No |
| Poor | 20 | 16 | 8 | 5 | 0.4–0.6 inches | Yes | WOMAC pain 5-point Likert scale | Yes |
| Satisfactory | 20 | 23 | 26 | 9 (11 for sham: 9 knee, 2 abdomen) | 0.3–1.0 inches; unclear for sham | Yes; no for sham | WOMAC pain 5-point Likert scale, range 0–20 | Other overall score: patient global assessment |
| Poor | 15–25; extra 10 minutes when deqi achieved | 5 | 5 | 6 | 10 mm | No | Other pain VAS, 10-cm VAS scale (0–100) | Yes |
| Poor | 30 | 5 | 5 | 6 for standard acupuncture; mean of 3.3 for trigger acupuncture; 3.1 for sham | 10 mm for standard acupuncture; 10–30 mm for trigger acupuncture; 0 mm for sham | No | Other pain VAS, 100 mm | Yes probably using Likert scale |
| Poor | Manual acupuncture: 10; electrical acupuncture: anterior part 10, then posterior part 10 | 10 | 5 | 9 | Varied between 1 and 1.5 cm; needle not inserted for sham | Yes; no for sham | WOMAC pain subscale VAS 0–100; other pain VAS, overall knee pain (0–100) | Individual WOMAC subscores reported |
| Poor | 10–30 | Up to 10 (as necessary) | Around 10 | Varied from 4 to 24 | Varied from 3 to 30 mm | Unclear/not stated | WOMAC pain 5-point Likert scale | Yes |
| Satisfactory | 30 | 1 | Unclear/not stated | 5 | 1–1.5 cm for both acupuncture and sham | Yes; no for sham | Other pain VAS 1–10 (note: SD for post-treatment pain read off graph) | No |
| Poor | Unclear/not stated | 16 | 8 | 10 | Unclear/not stated; sham: retractable needles (Streitberger) | Yes | WOMAC pain subscale VAS 0–100; other pain VAS 0–100 | Yes |
| Satisfactory | 20; manually manipulated every 5 minutes | 16 | 8 | 7 | Unclear/not stated; sham: not inserted into skin | No | Other pain VAS: Knee Society knee pain score | No |
| Poor | 20 | 10 | 5 | 9 | 0.5–1.5 cm; 0.5 cm for sham | Unclear/not stated | Other pain VAS 0–10 | No |
| Poor | 20 | 8 | 2 | 2 | 10–15 mm | Yes | NRS of pain | No |
| Poor | 20 | 8 | 2 | 12 | Unclear/not stated; 'superficially' for sham | No | Graded standard pain scale (4 points = severe, 3 points = moderate, 2 points = mild, 1 point = none) | No |

| Study | Treatments (if suitable for NMA, adjunct code in parentheses[a]) | Number analysed at end of treatment, data suitable for pain analyses? | Country | % Female | Age (years), mean unless stated | Weight (kg) | BMI (kg/m²) | Method of diagnosis | Kellgren and Lawrence score |
|---|---|---|---|---|---|---|---|---|---|
| Sangdee 2002[434] | Acupuncture, sham acupuncture, standard care | 186, yes | Thailand | 78 | 63 | Mean 60 | NR | Clinical and radiological | ≥ 1 |
| *Suarez-Almazor 2010[125]* | *Acupuncture, sham acupuncture, standard care* | *455, yes* | *USA* | *64* | *64* | *NR* | *NR* | *Clinical and radiological* | *NR/unclear* |
| Takeda 1994[442] | Acupuncture (2), sham acupuncture (2) | 40, yes | Canada | 50 | 62 | Mean 90 | Mean 33 | Clinical and radiological | NR/unclear |
| Tukmachi 2004[454] | Acupuncture (1 and 2), standard care (2) | 29, yes | UK | 83 | 61 | NR | NR | Clinical and radiological | 2 or 3 |
| Vas 2004[120] | Acupuncture (4), sham acupuncture (4) | 88, yes | Spain | 84 | 67 | NR | Mean 33 | Clinical and radiological | NR - used Ahlbäck grade 1 or higher |
| Weiner 2007[188] | Acupuncture (2), sham acupuncture (2) | 88, yes | USA | 55 | 71 | NR | Mean 32 | Clinical and radiological | ≥ 2 |
| Williamson 2007[136] | Acupuncture (2), MSE (2), standard care (3) | 181, yes | UK | 54 | 71 | NR | Mean 32 | Clinical and radiological (patients awaiting arthroplasty) | NR/unclear |
| Witt 2005[69] | Acupuncture (2), sham acupuncture (2), standard care (2) | 285, yes | Germany | 66 | 64 | NR | Mean 29 | Clinical and radiological | ≥ 2 |
| Witt 2006[77] | Acupuncture (2), standard care (2) | 342, yes | Germany | 60 | 61 | NR | NR | Clinical and radiological | NR/unclear |
| Wu 2008[458] | Acupuncture (2), standard care (4) | 34 (imputed), yes | China | 63 | 62 | NR | NR | Clinical and radiological | NR/unclear |
| Yurtkuran 1999[460] | Acupuncture (2), ice/cooling treatment (2), TENS (2), placebo TENS (2) | 100, yes | Turkey | 91 | 58 | NR | NR | Clinical and radiological | NR/unclear |

MSE, muscle-strengthening exercise; NMA, network meta-analysis; NR, not reported.
a 1, no medication; 2, treatment as usual; 3, treatment as usual + exercise/education; 4, treatment as usual + analgesia; 5, no medication + exercise/education.
b No means.

**Note**
Studies identified from the 2013 update searches are italicised.

| Overall study quality | Duration of individual sessions (minutes) | Total number of sessions | Duration of treatment period (weeks) | Number of points needled | Depth of insertion | Electrical stimulation used? | Pain outcomes assessed | Overall WOMAC score reported? |
|---|---|---|---|---|---|---|---|---|
| Poor | 20 | 12 | 4 | 4 | 0.5 inches; patched electrodes used for sham | Yes; sound generated for sham | WOMAC pain 5-point Likert scale; other pain VAS; patient's global pain as 100-mm VAS | Yes |
| *Good* | *20* | *12* | *6* | *7* | *0.7–4.0 cm (0.2–1.2 cun); sham: 0.7–1.7 cm (0.2–0.5 cun)* | *Yes* | *WOMAC pain subscale VAS 0–100* | *No* |
| Poor | 30 | 9 | 3 | 5 | 30 mm; sham: 'just enough to puncture skin' | No | WOMAC pain subscale VAS 0–10; Pain Rating Index of McGill Pain Questionnaire | Individual WOMAC subscores reported: pain, stiffness and function indices |
| Satisfactory | 20–30 | 10 | 5 | 9 | 1–1.5 cm | Yes | WOMAC pain, 5-point Likert scale; other pain VAS 0–10 cm | Other overall score: global assessment VAS 0–10 |
| Good | 20 | 12 | 12 | 8 | Unclear/not stated | Yes for acupuncture and sham | WOMAC pain, 5-point Likert scale; other pain VAS 0–100 | Yes |
| Good | 30 | 6 | 6 | 6 | Until just touched the bone | Yes, four needles for 30 minutes, two for 2 minutes; sham: two needles for 1 minute | WOMAC pain 5-point Likert scale | Individual WOMAC subscores reported: WOMAC function |
| Satisfactory | 20 | 6 | 6 | 7–10 | Unclear/not stated | No | Other pain VAS; VAS pain 0–10 cm | Yes |
| Good for acupuncture vs. sham; satisfactory for acupuncture vs. standard care | 30 | 12 | 8 | At least 8 for unilateral pain and at least 16 for bilateral pain | Unclear/not stated; 'superficial' for sham | No | WOMAC pain subscale VAS 0–100 | Yes |
| Satisfactory | Individually prescribed | On average 11 | 12 | Individually prescribed | Unclear/not stated | No | WOMAC pain subscale VAS 0–10 | Yes |
| Poor | 20 | 12 | 4 | At least 8 | Unclear/not stated | Yes | Lysholm score | No |
| Poor | Unclear | 10 | 2 | 4 | 0.5–1.0 inches | Yes | 5-point Likert scale | No |

## *Muscle-strengthening exercise studies*

| Study | Treatments (if suitable for NMA, adjunct code in parentheses[a]) | Number analysed at end of treatment, data suitable for pain analyses? | Country | % Female | Age (years) | Weight (kg) | BMI (kg/m$^2$), mean | Method of diagnosis |
|---|---|---|---|---|---|---|---|---|
| Abrahams 2002[349] | MSE, standard care | 56, no[b] | UK | NR | Range 22–83 | NR | NR | Clinical and radiological |
| An 2008[351] | MSE (2), standard care (2) | 21, yes | China | 100 | Mean 65 | NR | 26 | Clinical |
| Baker 2001[353] | MSE (2), placebo (2) | 38, yes | USA | 78 | Mean 69 | NR | 32 | Clinical and radiological |
| Bezalel 2010[358] | MSE (2), heat treatment (2) | 50, yes | Israel | 74 | Mean 74 | NR | NR | Unclear/NR |
| Börjesson 1996[190] | MSE (2), standard care (2) | 68, yes | Sweden | 50 | Mean 64 | Mean 83 | NR | Clinical and radiological |
| *Bruce-Brand 2012[467]* | *MSE, NMES, standard care* | *26, yes* | *Republic of Ireland* | *42* | *Mean 64* | *NR* | *33* | *Clinical and radiological* |
| Callaghan 1995[362] | MSE, standard care, placebo | 27, no[b] | UK | 31 | Median 53 | NR | NR | Radiological |
| Cheing 2002[365] | MSE (2), TENS (2), placebo TENS (2) | 47, yes | China | 89 | Mean 63 | Mean 67 | 28 | Clinical and radiological |
| Durmuş 2007[370] | MSE (2), PES (2) | 50, yes | Turkey | 100 | Mean 55 | NR | 33 | Clinical and radiological |
| Ettinger 1997[371] | MSE (2), aerobic exercise (2), standard care (3) | 364, yes | USA | 70 | Mean 69 | NR | 53% > 30 kg/m$^2$ | Clinical and radiological |
| *Foroughi 2011[469]* | *MSE, placebo* | *45, yes* | *Australia* | *100* | *Mean 66* | *Mean 83* | *32* | *Clinical* |
| Gür 2002[379] | MSE (1), no treatment (1) | 23, yes | Turkey | NR | Mean 56 | Mean 79 | NR | Radiological |

| Kellgren and Lawrence score | Overall study quality | Duration of individual sessions (minutes) | Total number of sessions | Duration of treatment period (weeks) | Intervention also incorporated home exercise? | Pain outcomes used | Overall WOMAC score reported? |
|---|---|---|---|---|---|---|---|
| NR/unclear | Poor | Unclear/NR | Unclear/NR | 12 | Unclear/NR | Other pain VAS, VAS slide indicator version (0—100) but no data reported | No |
| Other method used to classify osteoarthritis severity: ACR criteria | Poor | 30 | 40 | 8 | No | WOMAC pain subscale VAS 0–100 | No |
| NR/unclear, median 3 | Poor | Unclear/NR | 48 (patients exercised three times per week for 16 weeks at home, with 12 supervised visits) | 16 | Yes | WOMAC pain subscale VAS 0–100 | Pain and physical function subscores reported |
| NR/unclear | Poor | 45 | 4 | 4 | Yes | WOMAC pain five-point Likert scale; data for post treatment and follow-up read off graph – not very precise | Yes |
| Other method used to classify osteoarthritis severity: Ahlbäck grades I–III | Satisfactory | 40 | 15 | 5 | Yes | Borg scale (11-grade category scale, 0 = no pain to 10 = worse pain) | No |
| ≥ 3 | Poor | 30 | 18 | 6 | Yes | WOMAC pain five-point Likert scale | No |
| NR/unclear | Poor | 20 | 8 | 4 | No | Other pain VAS 0–10 scale | No |
| ≥ 2 | Poor | 20 | 20 | 4 | No | Other pain VAS 0–100, with baseline score standardised to 100 | No |
| ≤ 3 | Satisfactory | 20 | 20 | 4 | No | WOMAC pain subscale VAS 0–10; other pain VAS 0–10 | All three individual WOMAC subscores reported |
| NR/unclear | Poor | 60 | 216 | 72 | Yes | Disability score five-point Likert scale; pain score: 1 (no pain) to 6 (excruciating pain) | No |
| Other method used: modified Outerbridge classification | Poor | Not stated | 78 | 26 | No | WOMAC pain five-point Likert scale | Yes |
| 2 or 3 | Poor | Unclear/NR | 24 | 8 | No | Other pain VAS, NRS 0–10 (10 unbearable pain); score is the sum of scores for pain at night, after inactivity, on sitting, on rising from a chair, climbing stairs, descending stairs | No |

| Study | Treatments (if suitable for NMA, adjunct code in parentheses[a]) | Number analysed at end of treatment, data suitable for pain analyses? | Country | % Female | Age (years) | Weight (kg) | BMI (kg/m²), mean | Method of diagnosis |
|---|---|---|---|---|---|---|---|---|
| Hasegawa 2010[381] | MSE (2), standard care (2) | 28, yes | Japan | 64 | Mean 77 | Mean 55 | 24 | Clinical diagnosis was of knee joint pain |
| Hay 2006[382] | MSE, standard care | 217, yes | UK | 65 | Mean 68 | NR | NR | Clinical patients with knee pain – osteoarthritis of the knee not specified |
| Huang 2005[384] | MSE (2), standard care (2) | 98, yes | Taiwan | 81 | Mean 62 | NR | NR | Clinical and radiological |
| Hurley 2007[385] | MSE, standard care | 53, no[b] | UK | 70 | Mean 67 | Mean 81 | 30 | Clinical |
| *Imoto 2012[474]* | *MSE, standard care* | *100, yes* | *Brazil* | *92* | *Mean 60* | *NR* | *30* | *Clinical and radiological* |
| Jan 2008[389] | MSE (1), no treatment (1) | 98, yes | Taiwan | 81 | Mean 63 | Mean 63 | NR | Clinical and radiological |
| Keogan 2007[396] | MSE, aerobic exercise, standard care | 80, no[c] | Republic of Ireland | 64 | Mean 66 | NR | 30.6 | Clinical and radiological |
| Kuptniratsaikul 2002[399] | MSE (2), standard care (2) | 366, yes | Thailand | 78 | Mean 68 | NR | 11% were obese | Radiological |
| Lim 2008[405] | MSE (2), standard care (2) | 55, yes | Australia | 62 | Mean 62 | Mean 78 | 29 | Clinical and radiological |
| Lin 2009[406] | MSE (2), standard care (2) | 72, yes | Taiwan | 69 | Mean 63 | Mean 62 | NR | Clinical and radiological |
| Lund 2008[407] | MSE (2), standard care (2) | 79, yes | Denmark | 78 | Mean 68 | Mean 75 | NR | Clinical and radiological |
| Maurer 1999[410] | MSE (2), standard care (3) | 98, yes | USA | 42 | Mean 65 | Mean 85 | NR | Clinical and radiological |
| McCarthy 2004[412] | MSE (3), standard care (3) | 172, yes | UK | 58 | Mean 65 | NR | 30 | Clinical and radiological |

| Kellgren and Lawrence score | Overall study quality | Duration of individual sessions (minutes) | Total number of sessions | Duration of treatment period (weeks) | Intervention also incorporated home exercise? | Pain outcomes used | Overall WOMAC score reported? |
|---|---|---|---|---|---|---|---|
| NR/unclear | Poor | 80 (60 exercise plus 20 warm-up/-down) | 12 (also a minimum of 24 home exercise sessions) | 12 | Yes | Other pain VAS; 0–10 NRS (pain on movement) | No |
| NR/unclear | Poor | 20 | 3–6 | 10 | Yes | WOMAC pain Likert 5; other pain VAS; pain severity over previous 7 days on scale 0–10 | No |
| NR; patients with Altman grade II were included | Poor | Unclear/NR | 24 | 8 | Yes, but only after 8 weeks of treatment (i.e. during post-treatment follow-up) | Other pain VAS 1–10 | No |
| NR/unclear | Satisfactory | 35–45 | 12 | 6 | No, but did include self-management and coping skills | WOMAC pain five-point Likert scale | Yes |
| ≥ 2 | *Satisfactory* | *30–40* | *16* | *8* | *No* | *WOMAC pain five-point Likert scale; NRS 0–10* | *No* |
| ≤ 3 | Satisfactory | 30 high-resistance exercise, 50 low-resistance exercise | 24 | 8 | No | WOMAC pain five-point Likert scale | Only physical subscale reported |
| NR/unclear | Poor | Unclear/NR | Unclear/NR | 6 | Unclear/NR | Other pain VAS, NR – likely 0–10 | No |
| 2 or 3 | Poor | 60 | 16 | 8 | Unclear/NR | Other AIMS pain subscale (0–10) | No |
| ≥ 2 | Satisfactory | Unclear/NR | 60 (7 with physiotherapist at weeks 1–5, 7 and 10) | 12 | Yes, exercise was home based but participants also visited the physiotherapist seven times during the 12-week period | WOMAC pain five-point Likert scale (transformed to a 0–100 scale) | No |
| ≤ 3 | Satisfactory | Unclear/NR | 24 | 8 | No, told to cease other exercises | WOMAC pain five-point Likert scale | Pain and function subscores reported |
| Other method used to classify osteoarthritis severity: Lequesne (1 –26) score. Mean score of participants was 11 | Poor | 50 | 16 | 8 | Unclear/NR | Other pain VAS 0–100 | No |
| ≤ 3 | Poor | Unclear/NR | 24 | 8 | Unclear/NR | WOMAC pain subscale VAS 0–100 | Pain and function subscores reported |
| | Satisfactory | 45 | 16 | 8 | Yes | WOMAC pain five-point Likert scale; other pain VAS, VAS pain score 0–100 | No (only at baseline) |

| Study | Treatments (if suitable for NMA, adjunct code in parentheses[a]) | Number analysed at end of treatment, data suitable for pain analyses? | Country | % Female | Age (years) | Weight (kg) | BMI (kg/m$^2$), mean | Method of diagnosis |
|---|---|---|---|---|---|---|---|---|
| Péloquin 1999[423] | MSE (3), standard care (3) | 124, yes | Canada | 70 | Mean 66 | NR | 30 | Clinical and radiological |
| Rapp 2009[430] | MSE (2), manual therapy (2), standard care (2) | 39, yes | Germany | 64 | Mean 60 | Mean 83 | NR | Clinical and radiological |
| Røgind 1998[432] | MSE, standard care | 23, no[b] | Denmark | 91 | Mean 71 | Mean 71 | 27 | Clinical and radiological |
| Rosemffet 2004[433] | MSE, NMES | 18, no[b] | Argentina | 77 | Median 60 | NR | 30.9 | Clinical and radiological |
| *Salli 2010[480]* | *MSE, standard care* | *71, yes* | *Turkey* | *82* | *Mean 57* | *NR* | *32* | *Clinical and radiological* |
| Schilke 1996[435] | MSE (2), standard care (2) | 20, yes | USA | 85 | Mean 66 | NR | NR | Clinical |
| Topp 2002[450] | MSE (2), standard care (2) | 102, yes | USA | 73 | Mean 63 | Mean 89 | NR | Clinical and radiological |
| Trans 2009[451] | MSE (2), standard care (2) | 52, yes | Denmark | 100 | Mean 60 | Mean 81 | 30 | Clinical and radiological |
| Williamson 2007[136] | MSE (2), acupuncture (2), standard care (3) | 181, yes | UK | 54 | Mean 71 | NR | 32 | Clinical and radiological, patients on NHS arthroplasty waiting list |

MSE, muscle-strengthening exercise; NMA, network meta-analysis; NR, not reported.
a  1, no medication; 2, treatment as usual; 3, treatment as usual + exercise/education; 4, treatment as usual + analgesia; 5, no medication + exercise/education.
b  No means.
c  Only change from baseline scores reported.
**Note**
Studies identified from the 2013 update searches are italicised.

| Kellgren and Lawrence score | Overall study quality | Duration of individual sessions (minutes) | Total number of sessions | Duration of treatment period (weeks) | Intervention also incorporated home exercise? | Pain outcomes used | Overall WOMAC score reported? |
|---|---|---|---|---|---|---|---|
| ≤ 3 | Poor | 60 | 36 | 12 | No | Other AIMS2 subscale for pain | No |
| ≥ 2 | Poor | 45 | 16 | 8 | Unclear/NR | Other pain VAS 0–10 | No |
| ≥ 3 | Poor | Unclear/NR | 24 or 26 (two per week for 3 months) | 12 | Yes | Other pain VAS, pain on an 11-point scale. Separate scores for pain at night, at rest and on weight bearing | No |
| ≥ 2 | Poor | 75 | 16 | 8 | Unclear/NR | Other pain VAS 20–80 mm | Yes |
| ≤ 2 | *Satisfactory* | *Not stated* | *24* | *8* | *No* | *Other VAS 0–10* | *No* |
| NR/unclear | Poor | Unclear/NR | 24 | 8 | No | Other pain VAS: Osteoarthritis Screening Index 10-cm VAS | No |
| NR/unclear | Poor | 50 | 48 | 16 | Yes | WOMAC pain five-point Likert scale | No |
| NR/unclear | Satisfactory | Up to 10.5 | 16 | 8 | No, both interventions | WOMAC pain subscale VAS 0–100 | All three individual WOMAC subscores reported |
| NR/unclear | Satisfactory | 60 | 6 | 6 | No | Other pain VAS; VAS pain 0–10 cm | Yes |

## Aerobic exercise studies

| Study | Treatments (if suitable for NMA, adjunct code in parentheses[a]) | Number analysed at end of treatment, data suitable for pain analyses? | Country | % Female | Age (years) | Weight (kg) | BMI (kg/m²), mean | Method of diagnosis |
|---|---|---|---|---|---|---|---|---|
| Aglamiş 2009[189] | Aerobic exercise (2), standard care (2) | 25, yes | Turkey | 100 | Mean 56 | NR | 33 | Clinical and radiological |
| Bilgici 2004[359] | Aerobic exercise, standard care | NR, no[b] | Turkey | NR | Mean 54 | NR | NR | Clinical |
| Brosseau 2012[466] | Aerobic exercise, standard care | 84, yes | Canada | 69 | Mean 63 | Mean 82 | 30 | Clinical and radiological |
| Dias 2003[187] | Aerobic exercise, standard care | NR, no[b] | Brazil | 88 | Median 75 | NR | NR | Clinical and radiological |
| Ettinger 1997[371] | Aerobic exercise (2), MSE (2), standard care (3) | 364, yes | USA | 70 | Mean 69 | NR | 53% > 30 | Clinical and radiological |
| Keefe 2004[395] | Aerobic exercise (2), standard care (2) | 30 (imputed), yes | USA | 50 | Mean 59 | NR | NR | Clinical |
| Keogan 2007[396] | Aerobic exercise, MSE, standard care | 80, no[c] | Republic of Ireland | 64 | Mean 66 | NR | 31 | Clinical and radiological |
| Kovar 1992[398] | Aerobic exercise (2), standard care (3) | 92, yes | USA | 83 | Mean 69 | Mean 77 | NR | Clinical and radiological |
| Lim 2010[475] | Aerobic exercise, standard care | 66, yes | Korea | 87 | Mean 66 | NR | 28 | Clinical and radiological |
| Messier 2004[413] | Aerobic exercise (2), weight loss (2), standard care (3) | 240, yes | USA | 72 | Mean 69 | Mean 94 | 34 | Clinical and radiological |
| Salacinski 2012[479] | Aerobic exercise, standard care | 28, yes | USA | 73 | Mean 58 | NR | 24 | Clinical and radiological |
| Thorstensson 2005[447] | Aerobic exercise, standard care | 56, no[c] | Sweden | 51 | Mean 56 | NR | 30 | Clinical and radiological |
| Wang 2011[482] | Aerobic exercise, standard care | 78, yes | Taiwan | 86 | Mean 68 | NR | 26 | Clinical and radiological |

KOOS, Knee injury and Osteoarthritis Outcome Score; MSE, muscle-strengthening exercise; NMA, network meta-analysis; NR, not reported.
a 1, no medication; 2, treatment as usual; 3, treatment as usual + exercise/education; 4, treatment as usual + analgesia; 5, no medication + exercise/education.
b No means.
c Only change from baseline scores reported.
**Note**
Studies identified from the 2013 update searches are italicised.

| Kellgren and Lawrence score | Overall study quality | Duration of individual sessions (minutes) | Total number of sessions | Duration of treatment period (weeks) | Individual or group sessions? | Also incorporated home exercise? | Pain outcomes used | Overall WOMAC score reported? |
|---|---|---|---|---|---|---|---|---|
| ≥ 2 | Poor | 20 | 36 | 12 | Unclear/NR | No | WOMAC pain subscale VAS 0–10 | Yes |
| NR/unclear | Poor | 60 | 16 | 8 | Unclear/NR | Unclear/NR | WOMAC pain subscale VAS 0–10; unclear what form of WOMAC was used: no data reported | No |
| NR/unclear | Poor | 65 | Approximately 18 | 52 | Group | No | WOMAC pain five-point Likert scale | Yes |
| NR/unclear | Poor | 40 | 12 | 6 (followed by 6 weeks of home exercise) | Sessions given to groups | Yes for 6 weeks, following the 6-week intervention | SF-36 bodily pain; insufficient pain outcome data reported for data extraction | No |
| NR/unclear | Poor | 60 | 216 | 72 | Sessions given to groups | Yes | Other pain 5-point Likert scale; Likert scale 1 (no pain) to 6 (excruciating pain) | No |
| NR/unclear | Poor | 60 | 36 | 12 | Sessions given to groups | No | Other; AIMS four-item pain scale | No, but Arthritis Self-efficacy Scale used |
| NR/unclear | Poor | Unclear/NR | Unclear/NR | 6 | Unclear/NR | Unclear/NR | Other pain VAS; VAS 0–10 most likely to have been used | No |
| NR/unclear | Poor | 90 | 24 | 8 | Sessions given to groups | No | Other: AIMS VAS 10 | No |
| ≥ 2 | Poor | Aquatic group 40; land group 40 | Aquatic 24; land NR | Both 8 | Groups | Unclear/NR | Brief pain inventory (0–10) | Yes |
| ≤ 3 | Satisfactory | 60 | 234 | 78 | Sessions given to individuals | Yes | WOMAC pain five-point Likert scale | Pain and physical function subscores reported |
| ≤ 3 | Poor | 40–60 | 24 | 12 | Groups | No | WOMAC pain five-point Likert scale | Yes |
| ≥ 3 | Poor | 60 | 12 | 6 | Sessions given to groups | Yes | Other pain VAS: KOOS | No |
| NR/unclear | Poor | 60 | 36 | 12 | Groups | Unclear | KOOS pain (0–100) | No |

## T'ai chi studies

| Study | Treatments (if suitable for NMA, adjunct code in parentheses[a]) | Number analysed at end of treatment, data suitable for pain analyses? | Country | % Female | Age (years), mean | Weight (kg), mean | BMI (kg/m²), mean | Method of diagnosis |
|---|---|---|---|---|---|---|---|---|
| Brismée 2007[360] | T'ai chi (2); standard care (3) | 41, yes | USA | 83 | 70 | 73 | 28 | Clinical |
| Lee 2009[401] | T'ai chi (2); standard care (2) | 44, yes | Korea | 93 | 69 | 61 | 26 | Clinical and radiological |
| *Ni 2010[420] (retracted study)* | *T'ai chi (2); standard care (3)* | *29, yes (but excluded – retracted study)* | *China* | *100* | *63* | *66* | *27* | *Clinical* |
| Wang 2009[455] | T'ai chi; standard care | 40, yes | USA | 75 | 65 | NR | 30 | Clinical and radiological |
| Yip 2007[193] | T'ai chi (1); standard care (2) | 182, yes | Hong Kong | 84 | 65 | NR | NR | Clinical |

NMA, network meta-analysis; NR, not reported.
a   1, no medication; 2, treatment as usual; 3, treatment as usual + exercise/education; 4, treatment as usual + analgesia; 5, no medication + exercise/education.
**Note**
Studies identified from the 2013 update searches are italicised.

| Kellgren and Lawrence score | Overall study quality | Duration of individual sessions (minutes) | Total number of sessions | Duration of treatment period (weeks) | Pain outcomes used | Overall WOMAC score reported? |
|---|---|---|---|---|---|---|
| NR | Satisfactory | 40 | 36 (18 classes and 18 home sessions) | 12 | WOMAC pain subscale VAS 0–100 (reported as 7–35); knee pain VAS 0–10 | Yes (scale 26–130) |
| ≥ 2 | Satisfactory | 60 | 16 | 8 | WOMAC used but reported as 26–130 | Yes (scale 26–130); all three individual WOMAC subscores reported |
| *NR* | *Poor* | *40* | *72* | *24* | *WOMAC pain five-point Likert scale* | *Yes* |
| ≥ 2 | Poor | 60 | 24 | 12 | WOMAC pain subscale VAS 0–100; global knee pain VAS 0–10 | All three individual WOMAC subscores reported |
| NR | Poor | 15 (within a 2-hour self-management session) | 6 | 6 | VAS 0–100 | No |

## Weight-loss studies

| Study | Treatments (if suitable for NMA, adjunct code in parentheses[a]) | Number analysed at end of treatment, data suitable for pain analyses? | Country | % Female | Age (years), mean | Weight (kg) | BMI (kg/m²) | Method of diagnosis |
|---|---|---|---|---|---|---|---|---|
| *Bliddal 2011[465]* | *Weight loss, standard care* | *89, yes* | *Denmark* | *89* | *63* | *Mean 96* | *Mean 36* | *Clinical and radiological* |
| Christensen 2005[367] | Weight loss, standard care | 78, yes | Denmark | 89 | 63 | Mean 97 | Mean 36 | Clinical and radiological |
| Jenkinson 2009[390] | Weight loss (2 and 3), standard care (3) | 389, yes | UK | 66 | 61 | Median 93 | Median 33 | Clinical |
| Messier 2004[413] | Weight loss (2), aerobic exercise (2), standard care (3) | 240, yes | USA | 72 | 69 | Mean 94 | Mean 34 | Clinical and radiological |
| Miller 2006[414] | Weight loss (3), standard care (3) | 74, yes | USA | 26 | 70 (all ≥ 60) | Mean 98 | Mean 35 | Clinical |

NMA, network meta-analysis; NR, not reported.

a  1, no medication; 2, treatment as usual; 3, treatment as usual + exercise/education; 4, treatment as usual + analgesia; 5, no medication + exercise/education.

**Note**

Studies identified from the 2013 update searches are italicised.

## Balneotherapy studies

| Study | Treatments (if suitable for NMA, adjunct code in parentheses[a]) | Number analysed at end of treatment, data suitable for pain analyses? | Country | % Female | Age (years) | Weight (kg), mean | BMI (kg/m²), mean | Method of diagnosis |
|---|---|---|---|---|---|---|---|---|
| Bálint 2007[355] | Balneotherapy (3), placebo (3) | 52, yes | Hungary | 63 | Range 50–75 | NR | NR | Clinical |
| Cantarini 2007[364] | Balneotherapy, heat treatment, standard care | 74, no[b] | Italy | 63 | Mean 64 | 71 | NR | Clinical and radiological |
| Fioravanti 2010[372] | Balneotherapy (2), standard care (2) | 80, yes | Italy | 75 | Mean 70 | NR | 26 | Clinical and radiological |
| Flusser 2002[374] | Balneotherapy (2), placebo (2) | 58, yes | Israel | 85 | Mean 65 | 76 | NR | Clinical and radiological |
| Forestier 2010[375] | Balneotherapy, standard care | 309, no[b] | France | 47 | Mean 64 | NR | 30 | Clinical and radiological |

| Kellgren and Lawrence score | Overall study quality | Duration of dieting period (weeks) | Difference in weight loss between treatment groups (kg) | Weight-loss cut-off used to analyse effect on pain or WOMAC (e.g. 10% of body weight) | Pain outcomes used | Overall WOMAC score reported? |
|---|---|---|---|---|---|---|
| *2 or 3* | *Satisfactory* | *52* | *7.3* | 8% | *WOMAC pain subscale VAS 0–100* | *Yes* |
| 2 or 3 | Poor | 8 | 6.6 (95% CI 5.3 to 7.9; $p < 0.0001$); percentage of patients achieving 10% body weight loss was 50% in weight-loss group and 0% in standard care group | 10% | WOMAC pain subscale VAS 0–100 | Yes |
| ≤ 4; 41% had a score of 0 | Satisfactory | 104 | NR/unclear | Cut-off used; dietary interventions aimed for a weight loss of 0.5–1.0 kg per week; weight loss NR | WOMAC pain five-point Likert scale | No |
| ≤ 3 | Satisfactory | 78 | Mean weight loss by group: weight-loss group 4.61 kg, exercise group 3.46 kg, standard care group 1.1 kg | Cut-off used; average weight-loss goal was 5% | WOMAC pain five-point Likert scale | Individual WOMAC subscores reported: pain, physical function |
| NR/ unclear | Poor | 26 | 8 kg | Cut off used; 10% weight loss was goal for intervention group | WOMAC pain five-point Likert scale | Yes |

| Kellgren and Lawrence score | Overall study quality | Duration of individual sessions (minutes) | Number of sessions | Duration of treatment period (weeks) | Temperature (°C) | Type of treatment | Pain outcomes used | Overall WOMAC score reported? |
|---|---|---|---|---|---|---|---|---|
| NR | Poor | 30 | 20 | 4 | 36 | Bathing mineral water | WOMAC pain subscale | Yes |
| ≤ 3 | Poor | 35 | 15 | 3 | Water 38, mud packs 45 | Bathing mineral water and mud packs | Pain VAS 0–100 | No |
| ≤ 3 | Satisfactory | 35 | 12 | 2 | Mud packs 45, mineral bath 38 | Bathing mineral water and mud packs | WOMAC pain subscale VAS 0–100 | No |
| 2 or 3 | Poor | 20 | 15 | 3 | 30–35 | Mud packs | Pain VAS 0–10 | Lequesne index |
| ≥ 1 | Poor | 65 | 18 | 3 | Mineral hydrojet 37, massages by physiotherapist 38, mineral matured mud 45, collective mineral water pool 32 | Bathing mineral water and mud packs | WOMAC pain subscale VAS 0–100; other pain VAS 0–100 | WOMAC function |

| Study | Treatments (if suitable for NMA, adjunct code in parentheses[a]) | Number analysed at end of treatment, data suitable for pain analyses? | Country | % Female | Age (years) | Weight (kg), mean | BMI (kg/m$^2$), mean | Method of diagnosis |
|---|---|---|---|---|---|---|---|---|
| Karagülle 2007[394] | Balneotherapy, standard care | 20, no[b] | Turkey | 85 | Mean 60 | NR | NR | Clinical and radiological |
| Kovács 2002[397] | Balneotherapy, placebo | 68, no[b] | Hungary | Range 58–78 | NR | NR | NR | Clinical and radiological |
| Mahboob 2009[408] | Balneotherapy (2), placebo (2) | 50, yes | Iran | 100 | Range 44–79 | NR | NR | Clinical |
| Nguyen 1997[419] | Balneotherapy, standard care | 64, no[b] | France | 81 | NR | NR | 61% > 24 | Clinical |
| Sherman 2009[438] | Balneotherapy (2), placebo (2) | 44, yes | Israel | 80 | Mean 67 | NR | NR | Clinical and radiological |
| Sukenik 1999[441] | Balneotherapy (2), placebo (2) | 36, yes | Israel | 89 | Mean 63 | NR | NR | Clinical and radiological |
| Tishler 2004[448] | Balneotherapy, standard care | 68, yes | Israel | 78 | Mean 64 | NR | NR | Clinical |
| Wigler 1995[456] | Balneotherapy (2), placebo (2) | 33, yes | Israel | 88 | Mean 65 | NR | NR | Clinical and radiological |
| Yurtkuran 2006[461] | Balneotherapy (5), placebo (5) | 52, yes | Turkey | 97 | Mean 54 | 77 | 32 | Clinical and radiological |

NMA, network meta-analysis; NR, not reported.
a  1, no medication; 2, treatment as usual; 3, treatment as usual + exercise/education; 4, treatment as usual + analgesia; 5, no medication + exercise/education.
b  No means.

| Kellgren and Lawrence score | Overall study quality | Duration of individual sessions (minutes) | Number of sessions | Duration of treatment period (weeks) | Temperature (°C) | Type of treatment | Pain outcomes used | Overall WOMAC score reported? |
|---|---|---|---|---|---|---|---|---|
| ≥ 2 | Satisfactory | 30 | 20 | 10 days | 38 ± 1 | Bathing mineral water | Pain VAS 0–10 | Lequesne Algofunctional Index |
| NR/unclear | Poor | 30 | 15 | 15 days | 36 | Bathing mineral water | Pain VAS | No |
| NR | Poor | 20 | 30 | 30 days | NR | 50 g of mud gel | WOMAC pain subscale VAS | Individual WOMAC subscores for pain, function and stiffness |
| NR | Poor | Unclear | Unclear | 3 | NR | Bathing mineral water | Pain VAS 0–100 | No |
| ≥ 1 | Poor | 20 | 12 | 6 | 35–36 | Bathing mineral water | WOMAC pain subscale VAS 0–10; pain VAS 0–100 | Individual WOMAC subscores for pain, function and stiffness |
| Lequesne index of severity | Poor | 20 | 14 | 2 | Dead sea sulphur pools 37, sweet water control 24–25 | Bathing mineral water | VAS 0–10 | No |
| NR | Poor | 30 | 6 | 6 | 37 | Bathing mineral water | WOMAC pain five-point Likert scale; other pain VAS 0–100 | WOMAC subscores reported for pain, function and stiffness |
| NR | Poor | 40 | 7 | 2 | Water 38, mud pack 45 | Bathing mineral water and mud packs | Pain VAS 0–10 | No |
| 2 or 3 | Poor | 20 | 10 | 2 | 37 | Bathing mineral water | WOMAC pain five-point Likert scale; pain VAS 0–100 | Yes |

## Insoles studies

| Study | Treatments (if suitable for NMA, adjunct code in parentheses)[a] | Number analysed at end of treatment, data suitable for pain analyses? | Country | % Female | Age (years), mean | Weight (kg), mean | BMI (kg/m²), mean | Method of diagnosis |
|---|---|---|---|---|---|---|---|---|
| *Arazpour 2013[463]* | *Insoles, braces* | *24, yes* | *Iran* | *63* | *59.5* | *NR* | *27* | *Clinical and radiological* |
| Baker 2007[354] | Insoles, placebo | 172, no[b] | USA | 59 | 68 | NR | 33 | Clinical or radiological |
| *Bennell 2011[464]* | *Insoles, placebo* | *200, yes* | *Australia* | *59* | *64* | *NR* | *29* | *Clinical and radiological* |
| Maillefert 2001[409] | Insoles (4), placebo (4) | 156, yes | France | 74 | 65 | NR | 29 | Clinical and radiological |
| Nigg 2006[421] | Insoles, placebo | 123, no[c] | Canada | 54 | 58 | 85 | 30 | Clinical and radiological |
| Toda 2008[449] | Insoles, placebo | 122, no[c] | Japan | 88 | 64 | NR | 25 | Radiological |

NMA, network meta-analysis; NR, not reported.
a   1, no medication; 2, treatment as usual; 3, treatment as usual + exercise/education; 4, treatment as usual + analgesia; 5, no medication + exercise/education.
b   No means.
c   Only change from baseline scores reported.
**Note**
Studies identified from the 2013 update searches are italicised.

## Static magnets studies

| Study | Treatments | Number analysed at end of treatment, data suitable for pain analyses? | Country | % Female | Age (years) | Weight (kg), mean | BMI (kg/m²), mean | Method of diagnosis |
|---|---|---|---|---|---|---|---|---|
| Chen 2008[366] | Static magnets, placebo | 42, no[a] | Taiwan | 79 | Mean 65 | 64 | 27 | Clinical and radiological |
| Hinman 2002[383] | Static magnets, placebo | 43, yes | USA | 60 | Mean 63 | NR | NR | Clinical |
| Wolsko 2004[457] | Static magnets, placebo | 46, yes | USA | 69 | Median 63 | NR | 30 | Clinical and radiological |

NMA, network meta-analysis; NR, not reported.
a   No means.

| Kellgren and Lawrence score | Overall study quality | Time spent wearing (per day) (minutes) | Number of days worn per week | Duration of treatment period (weeks) | Ankle support also worn? | Pain outcome | Overall WOMAC score reported? |
|---|---|---|---|---|---|---|---|
| ≤ 2 | Poor | 7.6 hours | Unclear | 6 weeks | No | Other pain VAS 0–10 | No |
| ≥ 1 | Poor | Approximately 420 (7 hours) | Unclear | 6; phase 1: 6, phase 2: 6, washout period: 4 | No | WOMAC pain subscale VAS 0–100 | No |
| 2 or 3 | Good | Unclear | 7 | 12 months | Unclear | WOMAC pain five-point Likert scale | No |
| ≥ 2 | Satisfactory | Wear permanently | Wear permanently | Unclear: up to 24 months | No | WOMAC pain subscale VAS 0–100 | No |
| ≥ 2 | Satisfactory | As much as possible | 7 | 12 | No | WOMAC pain subscale VAS 0–100 | Yes |
| ≤ 4 | Satisfactory | 300–600 | 7 | 12 | Sock-type ankle support | Other pain VAS 0–100% | No |

| Kellgren and Lawrence score | Overall study quality | Strength of magnetic field (gauss or tesla) | When to wear magnets | Duration of treatment period (weeks) | Pain outcomes used | Overall WOMAC score reported? |
|---|---|---|---|---|---|---|
| Ahlbäck classification (grade I) | Poor | 35 mT, as measured with a Lakeshore 430 gaussmeter | Only during waking hours | 12 | HAQ pain scale 0–100 | No |
| NR | Poor | 1.08 T | Magnet group: 1.0–23.5 hours per day, mean number of hours magnet worn 116.33 (range 25–235); placebo group: 1–24 hours per day, the mean number of hours magnet worn 85.12 (range 6.5–213) | 2 | WOMAC pain subscale VAS 0–10 | No |
| ≥ 1 | Poor | 40–850 G | At least 6 hours a day | 6 | WOMAC pain subscale VAS 0–100 and VAS 0–100 with five scales (maximum 500) | No |

## Braces studies

| Study | Treatments | Number analysed at end of treatment, data suitable for pain analyses? | Country | % Female | Age (years), mean | Weight (kg), mean | BMI (kg/m²), mean | Method of diagnosis |
|---|---|---|---|---|---|---|---|---|
| *Arazpour 2013*[463] | *Braces, insoles* | *24, yes* | *Iran* | *63* | *59.5* | *NR* | *27* | *Clinical and radiological* |

**Note**
Studies identified from the 2013 update searches are italicised.

## Transcutaneous electrical nerve stimulation studies

| Study | Treatments (if suitable for NMA, adjunct code in parentheses)[a] | Number analysed at end of treatment, data suitable for pain analyses? | Country | % Female | Age (years) | Weight (kg), mean | BMI (kg/m²), mean | Method of diagnosis | Kellgren and Lawrence score |
|---|---|---|---|---|---|---|---|---|---|
| Alcidi 2007[350] | TENS (2), heat treatment (2) | 40, yes | Italy | 85 | Mean 66 | NR | NR | Clinical and radiological | NR |
| Burch 2008[303] | TENS, interferential therapy | 116, no[b] | USA | 72 | Mean 62 | 86 | 31 | Clinical and radiological | ≥ 1 |
| Cheing 2002[365] | TENS (2), MSE (2), placebo (2) | 47, yes | China | 89 | Mean 63 | 67 | 28 | Clinical and radiological | ≥ 2 |
| Cheing 2003[191] | TENS (2), placebo (2) | 38, yes | China | 89 | Mean 66 | 66 | NR | Clinical and radiological | ≥ 2 |
| Grimmer 1992[378] | TENS (1), placebo (1) | 60, yes | Australia | 62 | Mean 67 | NR | NR | Clinical and radiological | NR |
| Itoh 2008[386] | TENS (1), acupuncture (1), standard care (2) | 12, yes (only one study) | Japan | 66 | Range 62–83 | NR | NR | Clinical and radiological | ≥ 2 |
| Kang 2007[393] | TENS (2), placebo (2) | 63, yes | USA | 71 | Mean 57 | NR | NR | Clinical and radiological | NR |
| Law 2004[400] | TENS, placebo | 34, no[c] | Hong Kong | 97 | Mean 83 | 57 | 26 | Clinical and radiological | ≥ 2 |
| Lewis 1988[402] | TENS, placebo | Unknown, no[c] | Australia | NR | NR | NR | NR | Unclear/not stated | NR |
| Lewis 1994[403] | TENS (1), standard care (4) | 56, yes | Australia | 58 | Mean 66 | NR | NR | Clinical and radiological | NR |
| Lewis 1984[404] | TENS, placebo | 30, no[c] | UK | 73 | Median 61 | NR | NR | Unclear/not stated | NR |
| Ng 2003[418] | TENS, electroacupuncture, standard care | 14, yes | China | 96 | Mean 85 | NR | NR | Clinical | NR |
| Pietrosimone 2009[427] | TENS (1), ice/cooling treatment (1), no treatment (1) | 33, yes | USA | 48 | Mean 56 | 88 | 30 | Radiological | NR |
| Selfe 2008[436] | TENS (2), placebo (2) | 37, yes | USA | 68 | Mean 67 | NR | 31 | Clinical and radiological | NR |

| Kellgren and Lawrence score | Overall study quality | Time spent wearing (per day) (hours) | Number of days worn per week | Duration of treatment period (weeks) | Pain outcomes used | Overall WOMAC score reported? |
|---|---|---|---|---|---|---|
| ≤ 2 | Poor | 7.3 | Unclear | 6 | Other pain VAS 0–10 | No |

| Overall study quality | Duration of individual sessions (minutes) | Number of sessions | Duration of treatment period (weeks) | Frequency of current (Hz) | Intensity of current (mA) | Pulse width/duration (microseconds) | Electrodes placed at acupuncture points? | Pain outcomes used | Overall WOMAC score reported? |
|---|---|---|---|---|---|---|---|---|---|
| Poor | 20 | 5 | 1 | 50 | 'Well tolerated tingling' | 0.5 | No | Other pain VAS 1–100 | No |
| Poor | 35 | 56 | 8 | 0.2 | 60 | 300 | Unclear/not stated | WOMAC pain five-point Likert scale | All three individual WOMAC subscores reported |
| Poor | 60 | 20 | 4 | 80 | To achieve a tingling sensation | 140 | Yes | Other pain VAS 0–100 | No |
| Poor | 20–60 | 10 | 2 | 100 | NR | 200 | Yes | Other pain VAS 0–10 | No |
| Poor | 30 | 1 | 1 session | 80 | NR | NR | Yes | Other pain VAS | No |
| Poor | 15 | 5 | 5 | 4 | NR | NR | No | Other pain VAS 0–100 | Yes |
| Poor | 30 | 1 | 1 session | NR | 16% at start, increased to 23% at 15 minutes | NR | No | Other pain VAS 0–10 | All three individual WOMAC subscores reported |
| Poor | 40 | 10 | 2 | 2, 100 or alternating 2/100 | 25–35 | 2 Hz: 576, 100 Hz: 200 | Yes | Other pain VAS 0–100 | No |
| Poor | NR | NR | 9 | Unclear/not stated | Unclear/not stated | Unclear/not stated | Unclear/not stated | McGill Pain Questionnaire, Pain Index | No |
| Poor | 30–60 | 63 | 3 | 70 | Adjusted until comfortable | 100 | Spleen 9 and 10, stomach 34 and 35 | VAS 0–100 | No |
| Poor | 30–60 | 63 | 7 | 70 | NR | NR | Yes | VAS | No |
| Poor | 20 | 8 | 2 | 2 | NR | 200 | Yes | NRS | No |
| Poor | 45 | 1 | 1 day | 150 | NR | 150 | No | VAS | No |
| Poor | 20–30 | 17 | 8 | NR | NR | NR | Sometimes | WOMAC pain VAS 0–10; NRS | Yes |

| Study | Treatments (if suitable for NMA, adjunct code in parentheses)[a] | Number analysed at end of treatment, data suitable for pain analyses? | Country | % Female | Age (years) | Weight (kg), mean | BMI (kg/m$^2$), mean | Method of diagnosis | Kellgren and Lawrence score |
|---|---|---|---|---|---|---|---|---|---|
| Smith 1983[439] | TENS, placebo | 30, no[c] | UK | 67 | Mean 68 | NR | NR | Clinical | NR |
| Taylor 1981[445] | TENS, placebo | 20, no[b] | USA | 90 | 72 | NR | NR | Clinical and radiological | NR |
| *Vance 2012[481]* | *TENS, placebo* | *75, yes* | *USA* | *61* | *56* | *NR* | *36* | *Clinical and radiological* | *NR/ unclear* |
| Yurtkuran 1999[460] | TENS (2), acupuncture (2), ice/cooling (2), placebo (2) | 100, yes | Turkey | 91 | 58 | NR | NR | Clinical and radiological | NR |

MSE, muscle-strengthening exercise; NMA, network meta-analysis; NR, not reported.
a 1, no medication; 2, treatment as usual; 3, treatment as usual + exercise/education; 4, treatment as usual + analgesia; 5, no medication + exercise/education.
b No means.
c Only change from baseline scores reported.
d Defined as any of the following two criteria being fulfilled (and no worsening in the remaining criterion): (1) a 10-point or 50% decrease in weekly pain score, (2) a 50% decrease in analgesic intake compared with baseline and (3) a 5-point improvement in the weekly sleep disturbance score.
**Note**
Studies identified from the 2013 update searches are italicised.

| Overall study quality | Duration of individual sessions (minutes) | Number of sessions | Duration of treatment period (weeks) | Frequency of current (Hz) | Intensity of current (mA) | Pulse width/duration (microseconds) | Electrodes placed at acupuncture points? | Pain outcomes used | Overall WOMAC score reported? |
|---|---|---|---|---|---|---|---|---|---|
| Poor | 20 | 8 | 4 | 32–50 | Adjusted until comfortable | 80 | Yes | Other significant pain relief[d] | No |
| Poor | 30 | NR | 2 | NR | NR | NR | No | Five-point Likert scale | No |
| *Poor* | *40–50* | *1* | *3 hours* | *100* | *High frequency 27, low frequency 24* | *100* | *No* | *Other pain VAS 0–100* | *No* |
| Poor | 20 | 10 | 2 | 4 | 0.4–2.5 V | 1000 ms | Yes | Five-point Likert scale | No |

## *Pulsed electrical stimulation studies*

| Study | Treatments (if suitable for NMA, adjunct code in parentheses)[a] | Number analysed at end of treatment, data suitable for pain analyses? | Country | % Female | Age (years), mean | Weight (kg), mean | BMI (kg/m²), mean | Method of diagnosis |
|---|---|---|---|---|---|---|---|---|
| Callaghan 2005[363] | PES (2), placebo (2) | 21, yes | UK | 48 | 60 | NR | 28 | Radiological |
| Durmuş 2007[370] | PES (2), MSE (2) | 50, yes[b] | Turkey | 100 | 55 | NR | 33 | Clinical and radiological |
| *Fary 2011[468]* | *PES, placebo* | *70, yes* | *Australia* | *47* | *70* | *NR* | *28.1* | *Clinical and radiological* |
| *Fukuda 2008[376]* | *PES, standard care, placebo* | *NR, no[d]* | *Brazil* | *100* | *61* | *67* | *NR* | *Clinical and radiological* |
| Fukuda 2011[470] | PES, placebo, no treatment | 112, yes | Brazil | 100 | 60 | 71 | 28 | Radiological |
| Garland 2007[377] | PES (2), placebo (2) | 58, yes | USA | 66 | 66 | NR | 31 | Clinical and radiological |
| Miranda-Filloy 2005[415] | PES (2), placebo (2) | 18 (imputed), yes | Spain | 80 | > 40 | NR | NR | Clinical and radiological |
| Zizic 1995[462] | PES, placebo | 71, no[d] | USA | 46 | NR | NR | NR | Clinical and radiological |

MSE, muscle-strengthening exercise; NMA, network meta-analysis; NR, not reported.
a  1, no medication; 2, treatment as usual; 3, treatment as usual + exercise/education; 4, treatment as usual + analgesia; 5, no medication + exercise/education.
b  Only one study of this comparison.
c  Lack of baseline comparability on BMI reduced overall score from excellent to poor. However, as the BMI actual mean difference was not too large (27 kg/m² vs. 29 kg/m²) and groups were comparable on all other important characteristics, including pain, WOMAC score and SF-36, and as the authors also stated that 'adjustment for covariates did not alter any of the findings', an overall rating of satisfactory has been considered appropriate for this study.
d  Only change from baseline scores reported.
**Note**
Studies identified from the 2013 update searches are italicised.

| Kellgren and Lawrence score | Overall study quality | Duration of individual sessions (minutes) | Total number of sessions | Duration of treatment period (weeks) | Frequency of current (Hz) | Pain outcomes used | Overall WOMAC score reported? |
|---|---|---|---|---|---|---|---|
| ≥ 3 | Poor | 20 | 6 | 2 | High-frequency group 27 MHz; low-frequency group 27 MHz | VAS 0–10 and pain subscale of AIMS | No |
| ≤ 3 | Satisfactory | 20 | 20 | 4 | 50 | WOMAC pain subscale VAS 0–10; pain VAS 0–10 | All three individual WOMAC subscores reported |
| ≤ 4 | Satisfactory[c] | 7 hours | 182 | 26 | 100 | Other pain VAS | Yes |
| NR; grade II or III using Gupta criteria | Poor | 38 for 33-KJ dose group; 19 for 17-KJ dose group and placebo group | 39 (3 per week for 3 months) (13 weeks) | 13 | 33 kJ vs. 17 kJ (27.12 Hz with a pulse frequency of 145 Hz) | Other pain VAS; VAS (unclear what scoring) | Other overall score: Lequesne |
| NR/unclear | Poor | Low dose 19; high dose 38 | 9 | 3 | 145 | NRS 0–10 | No |
| ≥ 3 | Poor | ≥ 6 hours | Unclear/NR; each day | 13 | 100 | WOMAC pain subscale VAS 0–100 | Yes |
| ≥ 3 | Poor | 20 | 24 | 8 | 35 | WOMAC pain subscale VAS 0–100; other pain VAS; VAS 0-100 | No |
| NR/unclear | Poor | 6–10 hours a day | Unclear/NR | 4 | 100 | Other pain VAS 0–10 | No |

## *Pulsed electromagnetic fields studies*

| Study | Treatments (if suitable for NMA, adjunct code in parentheses)[a] | Number analysed at end of treatment, data suitable for pain analyses? | Country | % Female | Age (years) | Weight (kg), mean | BMI (kg/m²), mean | Method of diagnosis |
|---|---|---|---|---|---|---|---|---|
| Fischer 2005[373] | PEMF (2), placebo (2) | 69, yes | Slovenia | 72 | Mean 60 | NR | 29.3 | Radiological |
| Jacobson 2001[388] | PEMF (1), placebo (1) | 176, yes | USA | NR | NR | NR | NR | Unclear/not stated |
| Perrot 1998[425] | PEMF (2), placebo (2) | 40, yes | France | 80 | Mean 69 | NR | NR | Unclear/not stated |
| Pipitone 2001[428] | PEMF, placebo | 69, no[b] | UK | 28 | Median around 63 (range 40–84) | NR | NR | Clinical and radiological |
| Thamsborg 2005[446] | PEMF (2), placebo (2) | 83, yes | Denmark | 54 | Mean 60 | NR | 27 | Clinical and radiological |
| Trock 1994[452] | PEMF, placebo | 84, yes | USA | 70 | Mean 67 | 80 | NR | Clinical and radiological |

NMA, network meta-analysis; NR, not reported.
a   1, no medication; 2, treatment as usual; 3, treatment as usual + exercise/education; 4, treatment as usual + analgesia;
    5, no medication + exercise/education.
b   No means.

## *Neuromuscular electrical stimulation studies*

| Study | Treatments (if suitable for NMA, adjunct code in parentheses)[a] | Number analysed at end of treatment, data suitable for pain analyses? | Country | % Female | Age (years) | Weight (kg), mean | BMI (kg/m²), mean | Method of diagnosis | Kellgren and Lawrence score |
|---|---|---|---|---|---|---|---|---|---|
| *Bruce-Brand 2012[467]* | *MSE, NMES, standard care* | *26, yes* | *Republic of Ireland* | *42* | *Mean 64* | *NR* | *33* | *Clinical and radiological* | *≥ 3* |
| Rosemffet 2004[433] | NMES, MSE | 18, no[b] | Argentina | 77 | Median 60 | NR | 30.9 | Clinical and radiological | ≥ 2 |
| Talbot 2003[443] | NMES (3), standard care (3) | 34, yes | USA | 79 | Mean 71 | NR | 30 | Clinical and radiological | ≥ 1 |

MSE, muscle-strengthening exercise; NMA, network meta-analysis; NR, not reported.
a   1, no medication; 2, treatment as usual; 3, treatment as usual + exercise/education; 4, treatment as usual + analgesia;
    5, no medication + exercise/education.
b   No means.
**Note**
Studies identified from the 2013 update searches are italicised.

| Kellgren and Lawrence score | Overall study quality | Duration of individual sessions (minutes) | Total number of sessions | Duration of treatment period (weeks) | Frequency of pulse (Hz) | Strength of magnetic field (gauss or tesla) | Pain outcomes used | Overall WOMAC score reported? |
|---|---|---|---|---|---|---|---|---|
| NR/unclear | Poor | 16 | 30 | 6 | Unclear/not stated; low frequency | Varied between 3.4 and 13.6 μT | VAS 0–100 and 5-point Likert scale | No (Knee Society score) |
| NR/unclear | Poor | 6 | 8 | 2 | 0.976–7.7 | $2.74 \times 10^{-7}$–$3.4 \times 10^{-8}$ G | Scale from 1 to 10 (no further details) | No |
| NR | Poor | 60 | 9 | 1.3 | Unclear/not stated | Unclear/not stated | Other Pain VAS 0–100 | No |
| NR/unclear | Poor | Unclear/not stated | 147 | 6 | 7.8 in morning and afternoon, 3 in the evening | < 0.5 G | WOMAC pain five-point Likert scale | Yes |
| ≥ 1 | Poor | 120 | 30 | 6 | 50 | Unclear/not stated | WOMAC pain five-point Likert scale | All three individual WOMAC subscores reported |
| NR/unclear | Poor | 30 | 18 | Around 4–6 | 5 for 10 minutes, 10 for 10 minutes and 12 for 10 minutes | 10–15 G for 10 minutes and 15–25 G for 20 minutes | Other pain VAS 0–10 | No |

| Overall study quality | Duration of individual sessions (minutes) | Total number of sessions | Duration of treatment period (weeks) | Frequency of current (Hz) | Intensity of current | Pulse width/duration (microseconds) | Electrodes placed at acupuncture points? | Pain outcomes used | Overall WOMAC score reported? |
|---|---|---|---|---|---|---|---|---|---|
| *Poor* | *20* | *30* | *6* | *50* | *18 mA* | *100–400* | *No* | *WOMAC pain five-point Likert scale* | *No* |
| Poor | 30 | 24 | 8 | 25 | According to patient tolerance | Unclear/NR | Unclear/NR | 20–80 VAS scale | Yes |
| Poor | 15 | 36 | 12 | Unclear/NR | Unclear/NR | 300 | Unclear/NR | McGill Pain Rating Index | No |

## Interferential therapy studies

| Study | Treatments (if suitable for NMA, adjunct code in parentheses)[a] | Number analysed at end of treatment, data suitable for pain analyses? | Country | % Female | Age (years) | Weight (kg), mean | BMI (kg/m²), mean | Method of diagnosis |
|---|---|---|---|---|---|---|---|---|
| Adedoyin 2002[301] | Interferential therapy (5), placebo (5) | 26 (imputed), yes | Nigeria | 67 | Mean 59 | 78 | 28 | Clinical and radiological |
| Burch 2008[303] | Interferential TENS | 106, no[b] | USA | 72 | Mean 62 | 86 | 31 | Clinical and radiological |
| Defrin 2005[369] | Interferential therapy (2), placebo (2), no treatment (2) | 55 (imputed), yes | Israel | NR | Mean 67 | NR | NR | Clinical |
| *Gundog 2012[302]* | *Interferential therapy, placebo* | *60, yes* | *Turkey* | *80* | *Mean 60* | *NR* | *29* | *Clinical and radiological* |
| Young 1991[459] | Interferential therapy, placebo | NR, no[c] | Canada | NR | Range 18–75 | NR | NR | Clinical and radiological |

NMA, network meta-analysis.
a  1, no medication; 2, treatment as usual; 3, treatment as usual + exercise/education; 4, treatment as usual + analgesia; 5, no medication + exercise/education.
b  Only change from baseline scores reported.
c  No means.
**Note**
Studies identified from the 2013 update searches are italicised.

| Kellgren and Lawrence score | Overall study quality | Duration of individual sessions (minutes) | Total number of sessions | Duration of treatment period (weeks) | Frequency (Hz) | Pulse width/duration | Pain outcomes used | Overall WOMAC score reported? |
|---|---|---|---|---|---|---|---|---|
| NR/unclear | Poor | 20 | 8 | 4 | 100 for 15 minutes, reduced to 80 for a further 5 minutes | 0.033 seconds | VAS 0–10 | No |
| ≥ 1 | Poor | 15 of interferential followed by 20 of patterned stimulation | 56 | 8 | Base of 5000, and a pre-modulated beat frequency sweeping between 1 and 150 | Between 3 and 102 microseconds | WOMAC pain five-point Likert scale | All three individual WOMAC subscores reported |
| NR/unclear | Poor | 20 | 12 | 4 | Constantly ranged between 30 and 60 | NR | VAS 0–10 | No |
| *2 or 3* | *Good* | *20* | *15* | *3* | *Group 1 40, group 2 100, group 3 180* | *NR* | *WOMAC Likert scale* | *No* |
| NR/unclear | Poor | 25 | 10 | 1.7 | NR/unclear | NR/unclear | Reported WOMAC but unclear how assessed | No |

## *Heat treatment studies*

| Study | Treatments (if suitable for NMA, adjunct code in parentheses)[a] | Number analysed at end of treatment, data suitable for pain analyses? | Country | % Female | Age (years), mean | Weight (kg), mean | BMI (kg/m²), mean | Method of diagnosis |
|---|---|---|---|---|---|---|---|---|
| Alcidi 2007[350] | Heat (2), TENS (2) | 40, yes | Italy | 85 | 66 | NR | NR | Clinical and radiological |
| Bezalel 2010[358] | Heat (2), MSE (2) | 50, yes | Israel | 74 | 74 | NR | NR | NR |
| Cantarini 2007[364] | Heat, balneotherapy, standard care | 74, no[b] | Italy | 63 | 64 | 71 | NR | Clinical and radiological |
| Clarke 1974[368] | Heat (2), ice/cooling treatment (2), placebo (2) | 30, yes | UK | 69 | 61 | NR | NR | Radiological |
| *Giombini 2011[472]* | *Heat treatment, placebo* | *63, yes* | *Italy* | *67* | *67* | *NR* | *27* | *Clinical and radiological* |
| Mazzuca 2004[411] | Heat (2), placebo (2) | 51, yes | USA | 77 | 63 | NR | NR | Clinical and radiological |
| Rattanachaiyanont 2008[431] | Heat, placebo | 104, no[c] | Thailand | 100 | 63 | NR | 26 | Clinical |

MSE, muscle-strengthening exercise; N/A, not applicable; NMA, network meta-analysis; NR, not reported.
a 1, no medication; 2, treatment as usual; 3, treatment as usual + exercise/education; 4, treatment as usual + analgesia; 5, no medication + exercise/education.
b No means.
c Only change from baseline scores reported.
**Note**
Studies identified from the 2013 update searches are italicised.

## *Ice/cooling studies*

| Study | Treatments (if suitable for NMA, adjunct code in parentheses)[a] | Number analysed at end of treatment, data suitable for pain analyses? | Country | % Female | Age (years) | Weight (kg), mean | BMI (kg/m²), mean |
|---|---|---|---|---|---|---|---|
| Arman 1988[352] | Ice/cooling, standard care | 33, no[b] | Turkey | NR | NR | NR | NR |
| Clarke 1974[368] | Ice/cooling (2), heat treatment (2), placebo (2) | 45, yes | UK | 69 | Mean 61 | NR | NR |
| Pietrosimone 2009[427] | Ice/cooling (1), TENS (1), no treatment (1) | 33, yes | USA | 48 | Mean 56 | 88 | 30 |
| Yurtkuran 1999[460] | Acupuncture (2), ice/cooling (2), TENS (2), placebo TENS (2) | 100, yes | Turkey | 91 | Mean 58 | NR | NR |

NMA, network meta-analysis; NR, not reported.
a 1, no medication; 2, treatment as usual; 3, treatment as usual + exercise/education; 4, treatment as usual + analgesia; 5, no medication + exercise/education.
b No means.

| Kellgren and Lawrence score | Overall study quality | Duration of sessions (minutes) | Number of sessions | Duration of treatment period (weeks) | Type of treatment | Pain outcomes used | Overall WOMAC score reported? |
|---|---|---|---|---|---|---|---|
| NR | Poor | 20 | 5 | 0.7 | Radiofrequency electromagnetic radiation | VAS 1–100 | No |
| NR | Poor | 20 | 6 | 4 | Short-wave diathermy | WOMAC pain five-point Likert scale | Yes |
| ≤ 3 | Poor | 15 | 10 | 3 | Short-wave diathermy | VAS 0–100 | No |
| NR | Poor | NR | 9 | 3 | Short-wave diathermy | Likert scale 0–3; maximum score 17 | No |
| *2 or 3* | *Satisfactory* | *30* | *12* | *4* | *Short-wave diathermy* | *WOMAC pain* | *Yes* |
| ≥ 2 | Poor | N/A | N/A | 4 | Heat-retaining knee sleeve | WOMAC pain five-point Likert scale | Individual WOMAC subscores |
| NR | Poor | 20 | 9 | 3 | Short-wave diathermy | WOMAC pain VAS 0–10 | Yes |

| Method of diagnosis | Kellgren and Lawrence score | Overall study quality | Duration of sessions (minutes) | Number of sessions | Duration of treatment period (weeks) | Pain outcomes used | Overall WOMAC score reported? |
|---|---|---|---|---|---|---|---|
| Clinical and radiological | NR | Poor | 10 | 40 | 3 | NR | No |
| Radiological | NR | Poor | NR | 9 | 3 | Likert scale 0–3 (maximum score 17) | No |
| Radiological | NR | Poor | 20 | 1 | 1 day | VAS | No |
| Clinical and radiological | NR | Poor | 20 | 10 | 2 | Likert scale 1–5 | No |

## Laser/light therapy studies

| Study | Treatments (if suitable for NMA, adjunct code in parentheses)[a] | Number analysed at end of treatment, data suitable for pain analyses? | Country | % Female | Age (years), mean | Weight (kg), mean | BMI (kg/m²), mean | Method of diagnosis |
|---|---|---|---|---|---|---|---|---|
| Bülow 1994[361] | Laser, placebo | 29, no[b] | Denmark | 83 | ≥ 74 | NR | NR | Clinical and radiological |
| *Fukuda 2011[471]* | *Laser/light therapy, placebo* | *47, yes* | *Brazil* | *72* | *63* | *75* | *29* | *Clinical and radiological* |
| Gür 2003[380] | Laser, placebo | 90, no[b] | Turkey | 80 | 60 | NR | 30 | Clinical and radiological |
| *Hsieh 2012[473]* | *Laser/light therapy, placebo* | *72, yes* | *Taiwan* | *86* | *60* | *NR* | *26* | *Clinical and radiological* |
| Kang 2006[392] | Laser, standard care | Unclear, no[c] | Korea | NR | NR | NR | NR | Clinical and radiological |
| Nivbrant 1989[422] | Laser, placebo | Unclear, no[c] | Sweden | NR | NR | NR | NR | Clinical and radiological |
| Shen 2009[437] | Laser (2), placebo (2) | 40, yes | China | 90 | 58 | NR | NR | Clinical and radiological |
| Stelian 1992[440] | Laser (2), placebo (2) | 50, yes | Israel | 68 | 68 | NR | NR | NR |
| Tascioglu 2004[444] | Laser (4), placebo (4) | 60, yes | Turkey | 70 | 62 | NR | 29 | Clinical and radiological |

NMA, network meta-analysis; NR, not reported.
a 1, no medication; 2, treatment as usual; 3, treatment as usual + exercise/education; 4, treatment as usual + analgesia; 5, no medication + exercise/education.
b No means.
c Only change from baseline scores reported.
**Note**
Studies identified from the 2013 update searches are italicised.

| Kellgren and Lawrence score | Overall study quality | Duration of sessions (minutes) | Number of sessions | Duration of treatment period (weeks) | Light dose | Pain outcomes used | Overall WOMAC score reported? |
|---|---|---|---|---|---|---|---|
| NR | Poor | 15 | 9 | 3 | Accumulated dose 202.5 J | Likert scale 0–3 (maximum score 126) | No |
| ≥ 2 | *Good* | *NR* | *9* | *3* | *27 J per session* | *Visual numerical pain scale 0–10* | *No* |
| ≥ 2 | Satisfactory | 5 | 10 | 2 | Group 1, 30 J accumulated dose; group 2, 20 J accumulated dose | VAS 0–10 | Yes |
| ≥ 2 | *Good* | *40* | *6* | *2* | *2.08 J/cm$^2$/minute per session* | *Knee injury and Osteoarthritis Outcome Score VAS* | *No* |
| ≥ 2 | Poor | 30 | 56 | 8 | 402 laser diodes, 650 nm, 50 mW | WOMAC pain subscale VAS 0–10 | Individual WOMAC subscores reported |
| Osteoarthritis severity, radiographic grades 2–4 | Poor | 20 | 6 | 6 | NR | VAS | No |
| ≥ 2 | Satisfactory | 20 | 12 | 4 | 283 J | WOMAC pain five-point Likert scale | All three individual WOMAC subscale scores |
| NR | Poor | 30 | 20 | 10 days | Red light 10.3 J; infrared light 11.1 J | Other pain VAS 0–10; pain Likert scale 0–5 | No |
| 2 or 3 | Satisfactory | 10 minutes for first group, 5 minutes for second group | 10 | 2 | Group 1, 3 J per tender joint (five joints in all), dose per treatment 15 J; group 2, 1.5 J per tender joint (five joints in all), dose per treatment 7.5 J | WOMAC pain five-point Likert scale; other pain VAS 0–100 | All three individual WOMAC subscale scores |

## Manual therapy studies

| Study | Treatments (if suitable for NMA, adjunct code in parentheses)[a] | Number analysed at end of treatment, data suitable for pain analyses? | Country | % Female | Age (years), mean | Weight (kg), mean | BMI (kg/m²), mean |
|---|---|---|---|---|---|---|---|
| Moss 2007[417] | Manual therapy, placebo | 114, no[b] | Australia | 66 | 65 | NR | NR |
| Perlman 2006[424] | Manual therapy, standard care | 68, no[b] | USA | 78 | 68 | NR | 29 |
| *Perlman 2012[478]* | *Manual therapy, standard care* | *119, yes* | *USA* | *70* | *64* | *NR* | *32* |
| Pollard 2008[429] | Manual therapy (2), placebo (2) | 43, yes (only study) | Australia | NR | 56 | NR | NR |
| Rapp 2009[430] | Manual therapy (2), MSE (2), standard care (2) | 39, yes | Germany | 64 | 60 | 83 | NR |
| Tucker 2003[453] | Manual therapy (2), standard care (4) | 103, yes | South Africa | 63 | 59 | 82 | NR |

MSE, muscle-strengthening exercise; NMA, network meta-analysis; NR, not reported.
a 1, no medication; 2, treatment as usual; 3, treatment as usual + exercise/education; 4, treatment as usual + analgesia; 5, no medication + exercise/education.
b Only change from baseline scores reported.

| Method of diagnosis | Kellgren and Lawrence score | Overall study quality | Duration of sessions (minutes) | Number of sessions | Duration of treatment period (weeks) | Pain outcomes used | Overall WOMAC score reported? |
|---|---|---|---|---|---|---|---|
| Clinical and radiological | NR | Good | 10 | 1 | 10 minutes | WOMAC five-point Likert scale; VAS 0–10 | No |
| Clinical and radiological | NR | Poor | 60 | 12 | 8 | WOMAC VAS 0–100; pain VAS 0–100 | Yes |
| *Clinical and radiological* | *NR/unclear* | *Satisfactory* | *Varied across four groups (30 or 60)* | *8 or 12 depending on group* | *8* | *WOMAC VAS 0–100* | *Yes* |
| Clinical and radiological | NR | Poor | NR | 6 | 2 | VAS 0–10 | No |
| Clinical and radiological | ≥ 2 | Poor | 20 | 16 | 8 | VAS 0–10 | No |
| Clinical and radiological | NR | Poor | NR | 8 | 3 | VAS 0–100; NRS | No |

# Appendix 4 Review data related to *Chapter 4*

## Mapping and standardised pain descriptive results

| | | | | | Score, mean (SD) | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | Predicted EQ-5D | | | Standardised pain | |
| ID | Study | Pain group (type) | Treatment | HRQoL mapped outcome | $t_0$ – baseline | $t_1$ – 3 months | Pain standard outcome | $t_0$ – baseline | $t_1$ – 3 months |
| 1 | Diener 2006[72] | H | Usual care | SF-12 | 0.68 (0.268) | 0.759 (0.213) | Migraine days | 1.795 (1.142) | 1.222 (1.114) |
| | | | Sham acupuncture | | 0.696 (0.248) | 0.76 (0.218) | | 1.629 (0.948) | 1.08 (1.036) |
| | | | Acupuncture | | 0.683 (0.251) | 0.783 (0.203) | | 1.705 (0.95) | 0.996 (0.865) |
| 2 | Endres 2007[73] | H | Sham acupuncture | | 0.666 (0.263) | 0.765 (0.237) | TTH days | 2.863 (1.065) | 1.258 (1.103) |
| | | | Acupuncture | | 0.633 (0.281) | 0.757 (0.219) | | 2.728 (0.931) | 0.943 (0.865) |
| 10 | Haake 2007[71] | MSK | Usual care | | 0.633 (0.216) | 0.648 (0.231) | von Korff pain intensity score | 4.877 (1.051) | 2.809 (0.944) |
| | | | Sham acupuncture | | 0.656 (0.208) | 0.666 (0.212) | | 4.86 (0.946) | 2.485 (1.001) |
| | | | Acupuncture | | 0.631 (0.236) | 0.668 (0.199) | | 4.87 (1.002) | 2.325 (0.995) |
| 24 | Scharf 2006[70] | OAK | Usual care | | 0.533 (0.309) | 0.536 (0.325) | WOMAC total score | 3.609 (0.992) | 2.148 (0.999) |
| | | | Sham acupuncture | | 0.494 (0.34) | 0.593 (0.318) | | 3.646 (1.021) | 1.697 (0.978) |
| | | | Acupuncture | | 0.514 (0.323) | 0.596 (0.32) | | 3.591 (0.986) | 1.566 (0.933) |
| 3 | Jena 2008[76] | H | Usual care | SF-36 | 0.725 (0.198) | 0.754 (0.191) | Headache days | 1.171 (0.97) | 1.225 (1.033) |
| | | | Acupuncture | | 0.726 (0.202) | 0.826 (0.159) | | 1.2 (1.029) | 0.765 (0.911) |
| 4 | Linde 2005[68] | H | Usual care | | 0.741 (0.175) | 0.744 (0.13) | Days of moderate to severe pain | 2.082 (1.132) | 1.797 (0.9) |
| | | | Sham acupuncture | | 0.767 (0.128) | 0.815 (0.133) | | 1.933 (0.927) | 1.099 (1.018) |
| | | | Acupuncture | | 0.737 (0.142) | 0.809 (0.138) | | 2.001 (0.97) | 1.178 (0.963) |
| 5 | Melchart 2005[67] | H | Usual care | | 0.737 (0.164) | 0.721 (0.159) | Headache days | 2.534 (1.017) | 1.883 (0.855) |
| | | | Sham acupuncture | | 0.75 (0.167) | 0.846 (0.126) | | 2.59 (0.982) | 1.248 (0.954) |
| | | | Acupuncture | | 0.766 (0.135) | 0.834 (0.144) | | 2.56 (1.006) | 1.145 (1.001) |
| 6 | Vickers 2004[60] | H | Usual care | | 0.834 (0.148) | 0.806 (0.182) | Severity score | 1.732 (1.088) | 1.48 (1.05) |
| | | | Acupuncture | | 0.789 (0.171) | 0.809 (0.189) | | 1.593 (0.915) | 1.125 (0.927) |

| ID | Study | Pain group (type) | Treatment | HRQoL mapped outcome | Score, mean (SD) | | Pain standard outcome | | |
|----|-------|-------------------|-----------|---------------------|--------------------|--------------------|----------------------|--------------------|--------------------|
| | | | | | **Predicted EQ-5D** | | | **Standardised pain** | |
| | | | | | $t_0 -$ baseline | $t_1 - 3$ months | | $t_0 -$ baseline | $t_1 - 3$ months |
| 7 | Brinkhaus 2006[66] | MSK | Usual care | | 0.612 (0.185) | 0.635 (0.24) | VAS pain score | 4.714 (0.97) | 1.979 (0.847) |
| | | | Sham acupuncture | | 0.598 (0.168) | 0.728 (0.187) | | 4.749 (1.121) | 1.474 (1.005) |
| | | | Acupuncture | | 0.616 (0.192) | 0.763 (0.196) | | 4.508 (0.944) | 1.165 (0.963) |
| 17 | Vas 2006[121] | MSK | Sham acupuncture | | 0.556 (0.226) | 0.611 (0.203) | VAS pain score | 4.842 (1.033) | 2.326 (0.724) |
| | | | Acupuncture | | 0.605 (0.271) | 0.752 (0.215) | | 4.603 (0.959) | 0.918 (0.697) |
| 20 | Witt 2006[75] | MSK | Usual care | | 0.663 (0.217) | 0.684 (0.22) | Neck pain and disability score | 3.394 (1.004) | 2.698 (0.966) |
| | | | Acupuncture | | 0.65 (0.209) | 0.769 (0.188) | | 3.46 (0.995) | 2.079 (0.937) |
| 21 | Witt 2006[74] | MSK | Usual care | | 0.596 (0.232) | 0.638 (0.241) | Hanover functional ability score | −3.027 (0.994) | −3.051 (1.012) |
| | | | Acupuncture | | 0.585 (0.235) | 0.732 (0.225) | | −2.957 (1.005) | −3.453 (0.95) |
| 27 | Witt 2005[69] | OAK | Usual care | | 0.564 (0.185) | 0.59 (0.2) | WOMAC total score | 2.675 (1.042) | 2.084 (0.813) |
| | | | Sham acupuncture | | 0.561 (0.2) | 0.626 (0.218) | | 2.751 (0.991) | 1.57 (1.001) |
| | | | Acupuncture | | 0.564 (0.218) | 0.71 (0.19) | | 2.664 (1.032) | 1.119 (0.921) |
| 28 | Witt 2006[77] | OAK | Usual care | | 0.56 (0.213) | 0.57 (0.225) | WOMAC total score | 2.057 (1.002) | 1.876 (0.921) |
| | | | Acupuncture | | 0.576 (0.234) | 0.682 (0.234) | | 2.062 (0.999) | 1.185 (0.956) |
| 8 | Carlsson 2001[119] | MSK | Sham acupuncture | VAS pain | 0.604 (0.222) | 0.434 (0.314) | VAS pain score | 2.386 (1.022) | 2.251 (1.154) |
| | | | Acupuncture | | 0.481 (0.197) | 0.685 (0.246) | | 2.79 (0.978) | 1.684 (0.911) |
| 9 | Guerra de Hoyos 2004[138] | MSK | Sham acupuncture | | 0.526 (0.229) | 0.632 (0.256) | VAS pain score | 2.857 (0.871) | 1.134 (1.07) |
| | | | Acupuncture | | 0.525 (0.202) | 0.666 (0.229) | | 2.732 (1.118) | 0.488 (0.814) |
| 11 | Irnich 2001[118] | MSK | Sham acupuncture | | 0.551 (0.236) | 0.642 (0.259) | VAS pain score | 2.381 (1.113) | 1.527 (1.055) |
| | | | Acupuncture | | 0.562 (0.225) | 0.636 (0.233) | | 2.256 (0.913) | 1.277 (0.918) |
| 12 | Kennedy 2008[130] | MSK | Sham acupuncture | | 0.61 (0.25) | 0.653 (0.242) | Roland Morris disability score | 2.382 (1.033) | 1.265 (1.155) |
| | | | Acupuncture | | 0.568 (0.22) | 0.562 (0.247) | | 2.359 (0.988) | 0.813 (0.825) |
| 13 | Kerr 2003[117] | MSK | Sham acupuncture | | 0.523 (0.178) | 0.548 (0.204) | VAS pain score | 3.981 (0.924) | 2.324 (1.154) |
| | | | Acupuncture | | 0.5 (0.132) | 0.514 (0.266) | | 4.176 (1.065) | 1.943 (0.849) |

| ID | Study | Pain group (type) | Treatment | HRQoL mapped outcome | Score, mean (SD) Predicted EQ-5D $t_0$ – baseline | $t_1$ – 3 months | Pain standard outcome | Standardised pain $t_0$ – baseline | $t_1$ – 3 months |
|---|---|---|---|---|---|---|---|---|---|
| 18 | Vas 2008[122] | MSK | Sham acupuncture | | 0.556 (0.171) | 0.585 (0.254) | CMS | −3.123 (0.971) | −3.346 (0.975) |
| | | | Acupuncture | | 0.563 (0.155) | 0.623 (0.249) | | −3.281 (1.026) | −4.102 (0.874) |
| 19 | White 2004[132] | MSK | Sham acupuncture | | 0.548 (0.129) | 0.679 (0.202) | VAS pain score | 3.971 (1.073) | 1.18 (1.039) |
| | | | Acupuncture | | 0.572 (0.133) | 0.653 (0.228) | | 3.64 (0.906) | 0.871 (0.943) |
| 22 | Foster 2007[135] | OAK | Usual care | | 0.543 (0.214) | 0.566 (0.279) | WOMAC pain score | 2.463 (1.004) | 1.626 (0.999) |
| | | | Sham acupuncture | | 0.531 (0.205) | 0.621 (0.234) | | 2.428 (0.903) | 1.418 (1.029) |
| | | | Acupuncture | | 0.576 (0.213) | 0.559 (0.266) | | 2.52 (1.092) | 1.513 (0.969) |
| 25 | Vas 2004[120] | OAK | Sham acupuncture | WOMAC total | 0.37 (0.381) | 0.565 (0.352) | WOMAC total score | 3.308 (1.068) | 1.589 (1.037) |
| | | | Acupuncture | | 0.345 (0.366) | 0.76 (0.233) | | 3.273 (0.937) | 0.333 (0.452) |
| 26 | Williamson 2007[136] | OAK | Usual care | | 0.424 (0.304) | 0.487 (0.314) | Oxford Knee Score | 4.886 (1.039) | 5.104 (1.02) |
| | | | Acupuncture | | 0.405 (0.278) | 0.508 (0.262) | | 4.841 (0.922) | 4.774 (0.861) |
| 14 | Kleinhenz 1999[137] | MSK | Sham acupuncture | CMS and VAS pain | 0.558 (0.127) | 0.577 (0.071) | CMS | −3.986 (1.037) | −3.462 (0.965) |
| | | | Acupuncture | | 0.566 (0.101) | 0.578 (0.109) | | −4.483 (0.909) | −4.417 (0.795) |
| 15 | Salter 2006[133] | MSK | Usual care | EQ-5D available – no mapping necessary | 0.546 (0.322) | 0.67 (0.254) | Northwick Park pain score | 2.395 (1.156) | 1.518 (0.961) |
| | | | Acupuncture | | 0.614 (0.268) | 0.734 (0.269) | | 2.138 (0.73) | 1.342 (1.1) |
| 16 | Thomas 2006[61] | MSK | Usual care | | 0.532 (0.284) | 0.655 (0.274) | SF-36 bodily pain score | −1.815 (1.075) | −2.316 (1.064) |
| | | | Acupuncture | | 0.534 (0.293) | 0.753 (0.19) | | −1.839 (0.964) | −2.545 (0.962) |
| 23 | Berman 2004[134] | OAK | Usual care | | 0.65 (0.196) | 0.628 (0.24) | WOMAC pain score | 2.577 (1.059) | 1.851 (1.098) |
| | | | Sham acupuncture | | 0.615 (0.246) | 0.693 (0.169) | | 2.547 (0.969) | 1.587 (0.981) |
| | | | Acupuncture | | 0.608 (0.232) | 0.682 (0.184) | | 2.552 (0.979) | 1.345 (0.892) |

H, headache; MSK, musculoskeletal; OAK, osteoarthritis of the knee.

## WinBUGS code (models 1 and 2)

The WinBUGS modelling code is provided in the following section followed by a summary table of all variables included in the data sets and the R-code describing the specification of initial values for two chains.

## WinBUGS model code

```
model {

sw[1]<- 0

for(i in 1:n.subjects) {
#likelihood
        y[i] ~ dnorm(theta[i], prec[study[i]])
        theta[i] <- mu[study[i]] + delta[index[i]] * (1 - equals(treat[i],baseline[i])) +
        b_basey * bl_y[i] +
        b_treat_pain1[index[i]] * (1 - equals(treat[i],baseline[i])) * pain1[i] +
        b_treat_pain2[index[i]] * (1 - equals(treat[i],baseline[i])) * pain2[i]
#residual deviance
        dev[i]<- (y[i] - theta[i]) * (y[i] - theta[i])
        }

for(l in 1:n.arms) {
#consistency model for treatment effects and pain type interactions
        delta[l]~dnorm(md[l], precd[l])
        md[l]<- d[treat1[l]] - d[baseline1[l]] + equals(m[l],3) * sw[l]
        precd[l]<- pre * (1 + equals(m[l],3) / 3)
        b_treat_pain1[l] <- btp1[treat1[l]] - btp1[baseline1[l]]
        b_treat_pain2[l] <- btp2[treat1[l]] - btp2[baseline1[l]]
}

#correction for multi-arm trials
for(k in 2:n.arms) {
        sw[k]<- (delta[k-1] - d[treat1[k-1]] + d[baseline1[k-1]]) / 2
}

for(j in 1:n.trials) {
#priors on study-specific baseline outcome and precision parameter
        mu[j]~dnorm(0,1.0E-6)
        prec[j] ~ dgamma(0.001, 0.001)
# summed residual deviance contribution for each trial
        resdev[j]<- sum(dev[start[j]:end[j]]) * prec[j]
}

#total Residual Deviance
```

```
totresdev <- sum(resdev[])


for (k in 2:NT) {
#random effect on pain interaction effects
        btp1[k] ~ dnorm(m.btp1,prec.btp1)
        btp2[k] ~ dnorm(m.btp2,prec.btp2)
#prior on treatment effects
        d[k] ~ dnorm(0,1.0E-6)
        }


#prior on random treatment effect variance
tau ~ dunif(0,10)
tau.sq <- tau * tau
pre<- 1 / (tau.sq)
#priors on mean and random effect variance for pain type interaction
m.btp1 ~ dnorm(0,1.0E-6)
tau.btp1 ~ dunif(0,2)
tau.sq.btp1 <- tau.btp1 * tau.btp1
prec.btp1 <- 1 / (tau.sq.btp1)


m.btp2 ~ dnorm(0,1.0E-6)
tau.btp2 ~ dunif(0,2)
tau.sq.btp2 <- tau.btp2 * tau.btp2
prec.btp2 <- 1 / (tau.sq.btp2)


#prior on impact of baseline outcome on final outcome
b_basey ~ dnorm(0,1.0E-6)


d[1]<-0
btp1[1]<-0
btp2[1]<-0


}
```

## Description of data sets and variables

| Data set | Variable | Description |
|---|---|---|
| Data set descriptors, constants | n.subjects | Number of subjects in data set |
| | n.arms | Number of arms in studies in data set |
| | n.trials | Number of trials in data set |
| | NT | Number of treatments in data set |
| Data at individual patient level | study | Study code |
| | treat | Treatment code |
| | baseline | Baseline (lowest index) treatment in study |
| | y | EQ-5D or standardised pain at follow-up |
| | bl_y | EQ-5D or standardised pain at baseline |
| | index | Sequential numbering of study arms in data set |
| Data at study arm level | pain1 | Dummy variable indicating whether or not study was conducted in patients with headache pain |
| | pain2 | Dummy variable indicating whether or not study was conducted in patients with musculoskeletal pain |
| | treat1 | Treatment code |
| | baseline1 | Baseline (lowest index) treatment in study |
| | m | Code for each study arm indicating whether it is the first, second, third, fourth or fifth arm in the study |
| Data at study level | start | Number of the first line of IPD data set including study data |
| | end | Number of the last line of IPD data set including study data |

## R-code used to generate initial values (two sets)

```
list(

list(d = c(NA,0,0), mu = c(0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0), delta = c(0,0,0,0,0,
0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0,
0,0,0,0,0, 0,0,0,0,0), b_basey=c(0), tau = c(1), btp1 = c(NA,0,0), btp2 = c(NA,0,0), prec=c(1,1,1,1,1,
1,1,1,1,1, 1,1,1,1,1, 1,1,1,1,1, 1,1,1,1,1, 1,1,1), tau.btp1 = c(1), tau.btp2 = c(1)),

list(d = c(NA,-1,1), mu = c(-3,-3,-3,3,3, -3,-3,-3,3,3, -3,-3,-3,3,3, -3,-3,-3,3,3, -3,-3,-3,3,3, -3,-3,-3), delta
= c(0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0,
0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0), b_basey=c(1), tau = c(1), btp1 = c(NA,0,0), btp2 = c(NA,0,0),
prec=c(1,1,1,1,1, 1,1,1,1,1, 1,1,1,1,1, 1,1,1,1,1, 1,1,1,1,1, 1,1,1), tau.btp1 = c(1), tau.btp2 = c(1)))
```

# Appendix 5 Review data related to *Chapter 5*

## Data requirements

Studies were included in the data set if they provided post-baseline mean outcomes for all dimensions of any of the following HRQoL instruments: EQ-5D, SF-36 (individual dimensions or PCS and MCS scores), SF-12 (individual dimensions or PCS or MCS scores), WOMAC, pain VAS and pain NRS. Although a wide range of additional HRQoL data was reported, mapping algorithms were not available to translate these data to EQ-5D preference scores. [Note: for the Nottingham Health Profile (NHP) an algorithm was available but this was estimated from a small sample of patients in a very different disease area (hepatitis C). Data from the NHP were not therefore included in the analysis. One study reported Assessment of Quality of Life preference scores but as these were reported to only one decimal place these data were not included.]

Absolute values and change from baseline values were included. When reported, baseline data were also extracted. When change from baseline data were reported studies were included only if baseline data were also reported, to enable calculation of absolute scores at follow-up time points. SEs were extracted directly or derived from reported SDs or CIs. Studies for which no measure of variance was reported were retained within the analysis and the missing SDs were imputed (see *Chapter 5, Health-related quality of life for cost-effectiveness analysis*). Methods for deriving final value variances from baseline value and change score variances are not currently available. In cases in which change from baseline data were reported the variance at follow-up was therefore assumed to be equal to the variance at baseline.

The WOMAC instrument can be administered in a variety of ways resulting in a range of possible scoring ranges. Mapping from the WOMAC score to the EQ-5D, however, requires the WOMAC score range to be 0–96. Reporting of the WOMAC score range in the data was often ambiguous;[183] only data in which the scoring range was sufficiently clear to allow the data to be transformed onto the 0–96 range were therefore included in the data set. Where all subcomponents of the WOMAC (pain, stiffness and physical function) were reported, they were aggregated to produce a total WOMAC score. Calculation of the variance for the resulting total WOMAC score required the covariance of the subcomponents to be accounted for. The covariance for each study was therefore imputed using correlations between subcomponents derived from the IPD studies.

## Study characteristics for the analysis data set

| Study | Intervention | *n* analysed | Quality-of-life instrument | Analysis time point (weeks) | IPD available | Adequate allocation concealment | Adequate allocation concealment 3–13 weeks |
|---|---|---|---|---|---|---|---|
| Adedoyin 2002[301] | Interferential therapy | 15 | Pain VAS | 4 | No | No | No |
| Adedoyin 2002[301] | Placebo | 15 | Pain VAS | 4 | No | No | No |
| Aglamiş 2009[189] | Exercise – aerobic | 16 | SF-36 dimensions | 6 | No | No | No |
| Aglamiş 2009[189] | Usual care | 9 | SF-36 dimensions | 6 | No | No | No |
| Alcidi 2007[350] | Heat treatment | 20 | Pain VAS | 0.71 | No | No | No |
| Alcidi 2007[350] | TENS | 20 | Pain VAS | 0.71 | No | No | No |

| Study | Intervention | *n* analysed | Quality-of-life instrument | Analysis time point (weeks) | IPD available | Adequate allocation concealment | Adequate allocation concealment 3–13 weeks |
|---|---|---|---|---|---|---|---|
| An 2008[351] | Exercise – muscle strengthening | 11 | WOMAC | 8 | No | No | No |
| An 2008[351] | Usual care | 10 | WOMAC | 8 | No | No | No |
| Arazpour 2013[463] | Braces | 12 | Pain VAS | 6 | No | No | No |
| Arazpour 2013[463] | Insoles | 12 | Pain VAS | 6 | No | No | No |
| Baker 2001[353] | Exercise – muscle strengthening | 19 | SF-36 dimensions | 16 | No | Yes | No |
| Baker 2001[353] | Placebo | 19 | SF-36 dimensions | 16 | No | Yes | No |
| Bálint 2007[355] | Balneotherapy | 27 | WOMAC | 4 | No | Yes | Yes |
| Bálint 2007[355] | Placebo | 25 | WOMAC | 4 | No | Yes | Yes |
| Bennell 2011[464] | Insoles | 90 | WOMAC | 52 | No | Yes | No |
| Bennell 2011[464] | Usual care | 89 | WOMAC | 52 | No | Yes | No |
| Berman 1999[357] | Acupuncture | 36 | WOMAC | 8 | No | Yes | Yes |
| Berman 1999[357] | Usual care | 37 | WOMAC | 8 | No | Yes | Yes |
| Berman 2004[134] | Acupuncture | 154 | EQ-5D | 8 | Yes | Yes | Yes |
| Berman 2004[134] | Sham acupuncture | 145 | EQ-5D | 8 | Yes | Yes | Yes |
| Brismée 2007[360] | Usual care | 19 | WOMAC | 9 | No | No | No |
| Brismée 2007[360] | T'ai chi | 22 | WOMAC | 9 | No | No | No |
| Brosseau 2012[466] | Exercise – aerobic | 44 | SF-36 dimensions | 52.18 | No | No | No |
| Brosseau 2012[466] | Exercise – aerobic | 44 | SF-36 dimensions | 52.18 | No | No | No |
| Brosseau 2012[466] | Usual care | 41 | SF-36 dimensions | 52.18 | No | No | No |
| Bruce-Brand 2012[467] | Exercise – muscle strengthening | 10 | SF-36 MCS and PCS | 8 | No | No | No |
| Bruce-Brand 2012[467] | NMES | 10 | SF-36 MCS and PCS | 8 | No | No | No |
| Bruce-Brand 2012[467] | Usual care | 6 | SF-36 MCS and PCS | 8 | No | No | No |
| Burch 2008[303] | Interferential therapy | 52 | WOMAC | 8 | No | Yes | Yes |
| Burch 2008[303] | TENS | 53 | WOMAC | 8 | No | Yes | Yes |
| Callaghan 2005[363] | Placebo | 9 | Pain VAS | 2 | No | No | No |
| Callaghan 2005[363] | PES | 9 | Pain VAS | 2 | No | No | No |
| Callaghan 2005[363] | PES | 9 | Pain VAS | 2 | No | No | No |
| Cheing 2003[191] | Placebo | 8 | Pain VAS | 1.43 | No | No | No |
| Cheing 2003[191] | TENS | 10 | Pain VAS | 1.43 | No | No | No |
| Cheing 2003[191] | TENS | 10 | Pain VAS | 1.43 | No | No | No |
| Cheing 2003[191] | TENS | 10 | Pain VAS | 1.43 | No | No | No |

| Study | Intervention | n analysed | Quality-of-life instrument | Analysis time point (weeks) | IPD available | Adequate allocation concealment | Adequate allocation concealment 3–13 weeks |
|---|---|---|---|---|---|---|---|
| Durmuş 2007[370] | Exercise – muscle strengthening | 25 | Pain VAS | 4 | No | No | No |
| Durmuş 2007[370] | PES | 25 | Pain VAS | 4 | No | No | No |
| Fary 2011[468] | Placebo | 36 | SF-36 MCS and PCS | 4 | No | Yes | Yes |
| Fary 2011[468] | PES | 34 | SF-36 MCS and PCS | 4 | No | Yes | Yes |
| Fischer 2005[373] | Placebo | 35 | Pain VAS | 6 | No | No | No |
| Fischer 2005[373] | PEMFs | 34 | Pain VAS | 6 | No | No | No |
| Flusser 2002[374] | Balneotherapy | 40 | Pain VAS | 3 | No | No | No |
| Flusser 2002[374] | Placebo | 18 | Pain VAS | 3 | No | No | No |
| Foroughi 2011[469] | Exercise – muscle strengthening | 20 | WOMAC | 26.09 | No | No | No |
| Foroughi 2011[469] | Placebo | 25 | WOMAC | 26.09 | No | No | No |
| Fukuda 2011[470] | Placebo | 21 | Pain NRS | 3 | No | Yes | Yes |
| Fukuda 2011[470] | PES | 30 | Pain NRS | 3 | No | Yes | Yes |
| Fukuda 2011[470] | PES | 29 | Pain NRS | 3 | No | Yes | Yes |
| Fukuda 2011[470] | Usual care | 32 | Pain NRS | 3 | No | Yes | Yes |
| Fukuda 2011[471] | Laser/light therapy | 25 | Pain VAS | 6 | No | Yes | Yes |
| Fukuda 2011[471] | Placebo | 22 | Pain VAS | 6 | No | Yes | Yes |
| Garland 2007[377] | Placebo | 19 | WOMAC | 13.04 | No | Yes | Yes |
| Garland 2007[377] | PES | 39 | WOMAC | 13.04 | No | Yes | Yes |
| Grimmer 1992[378] | Placebo | 20 | Pain VAS | 0.14 | No | Yes | No |
| Grimmer 1992[378] | TENS | 20 | Pain VAS | 0.14 | No | Yes | No |
| Grimmer 1992[378] | TENS | 20 | Pain VAS | 0.14 | No | Yes | No |
| Gundog 2012[302] | Interferential therapy | 15 | WOMAC | 4.35 | No | No | No |
| Gundog 2012[302] | Interferential therapy | 15 | WOMAC | 4.35 | No | No | No |
| Gundog 2012[302] | Interferential therapy | 15 | WOMAC | 4.35 | No | No | No |
| Gundog 2012[302] | Placebo | 15 | WOMAC | 4.35 | No | No | No |
| Hasegawa 2010[381] | Exercise – muscle strengthening | 14 | Pain NRS | 12 | No | No | No |
| Hasegawa 2010[381] | Usual care | 14 | Pain NRS | 12 | No | No | No |
| Huang 2005[384] | Exercise – muscle strengthening | 25 | Pain VAS | 8 | No | Yes | Yes |
| Huang 2005[384] | Usual care | 28 | Pain VAS | 8 | No | Yes | Yes |
| Imoto 2012[474] | Exercise – muscle strengthening | 38 | SF-36 dimensions | 8 | No | Yes | Yes |

| Study | Intervention | *n* analysed | Quality-of-life instrument | Analysis time point (weeks) | IPD available | Adequate allocation concealment | Adequate allocation concealment 3–13 weeks |
|---|---|---|---|---|---|---|---|
| Imoto 2012[474] | Usual care | 43 | SF-36 dimensions | 8 | No | Yes | Yes |
| Itoh 2008[386] | Acupuncture | 6 | WOMAC | 5 | No | No | No |
| Itoh 2008[386] | Usual care | 6 | WOMAC | 5 | No | No | No |
| Itoh 2008[386] | TENS | 6 | WOMAC | 5 | No | No | No |
| Itoh 2008[387] | Acupuncture | 9 | WOMAC | 5 | No | No | No |
| Itoh 2008[387] | Acupuncture | 8 | WOMAC | 5 | No | No | No |
| Itoh 2008[387] | Sham acupuncture | 7 | WOMAC | 5 | No | No | No |
| Jubb 2008[391] | Acupuncture | 31 | WOMAC | 5 | No | No | No |
| Jubb 2008[391] | Sham acupuncture | 31 | WOMAC | 5 | No | No | No |
| Kang 2007[393] | Placebo | 28 | Pain VAS | 0.29 | No | No | No |
| Kang 2007[393] | TENS | 35 | Pain VAS | 0.29 | No | No | No |
| Kovar 1992[398] | Exercise – aerobic | 47 | Pain VAS | 8 | No | No | No |
| Kovar 1992[398] | Usual care | 45 | Pain VAS | 8 | No | No | No |
| Kuptniratsaikul 2002[399] | Exercise – muscle strengthening | 173 | Pain VAS | 8 | No | No | No |
| Kuptniratsaikul 2002[399] | Usual care | 193 | Pain VAS | 8 | No | No | No |
| Lansdown 2009[126] | Acupuncture | 15 | EQ-5D | 13.04 | No | Yes | Yes |
| Lansdown 2009[126] | Usual care | 15 | EQ-5D | 13.04 | No | Yes | Yes |
| Lee 2009[401] | Usual care | 15 | SF-36 MCS and PCS | 8 | No | No | No |
| Lee 2009[401] | T'ai chi | 29 | SF-36 MCS and PCS | 8 | No | No | No |
| Lewis 1994[403] | Placebo | 28 | Pain VAS | 3 | No | No | No |
| Lewis 1994[403] | TENS | 28 | Pain VAS | 3 | No | No | No |
| Lim 2010[475] | Exercise – muscle strengthening | 22 | SF-36 MCS and PCS | 8 | No | Yes | Yes |
| Lim 2010[475] | Exercise – muscle strengthening | 24 | SF-36 MCS and PCS | 8 | No | Yes | Yes |
| Lim 2010[475] | Usual care | 20 | SF-36 MCS and PCS | 8 | No | Yes | Yes |
| Lu 2010[192] | Acupuncture | 10 | Pain VAS | 0.14 | No | No | No |
| Lu 2010[192] | Sham acupuncture | 10 | Pain VAS | 0.14 | No | No | No |
| Lund 2008[407] | Exercise – aerobic | 25 | Pain VAS | 8 | No | Yes | Yes |
| Lund 2008[407] | Exercise – muscle strengthening | 27 | Pain VAS | 8 | No | Yes | Yes |
| Lund 2008[407] | Usual care | 27 | Pain VAS | 8 | No | Yes | Yes |

| Study | Intervention | *n* analysed | Quality-of-life instrument | Analysis time point (weeks) | IPD available | Adequate allocation concealment | Adequate allocation concealment 3–13 weeks |
|---|---|---|---|---|---|---|---|
| Maillefert 2001[409] | Insoles | 82 | WOMAC | 4.35 | No | No | No |
| Maillefert 2001[409] | Placebo | 74 | WOMAC | 4.35 | No | No | No |
| Maurer 1999[410] | Exercise – muscle strengthening | 49 | WOMAC | 8 | No | No | No |
| Maurer 1999[410] | Usual care | 49 | WOMAC | 8 | No | No | No |
| Mavrommatis 2012[476] | Acupuncture | 39 | SF-36 MCS and PCS | 8 | No | No | No |
| Mavrommatis 2012[476] | Sham acupuncture | 40 | SF-36 MCS and PCS | 8 | No | No | No |
| Mavrommatis 2012[476] | Usual care | 38 | SF-36 MCS and PCS | 8 | No | No | No |
| Mazzuca 2004[411] | Heat treatment | 26 | WOMAC | 4 | No | Yes | Yes |
| Mazzuca 2004[411] | Placebo | 25 | WOMAC | 4 | No | Yes | Yes |
| McCarthy 2004[412] | Exercise – muscle strengthening | 80 | EQ-5D | 4 | No | Yes | Yes |
| McCarthy 2004[412] | Usual care | 71 | EQ-5D | 4 | No | Yes | Yes |
| Messier 2004[413] | Exercise – aerobic | 69 | SF-36 MCS and PCS | 39.13 | No | Yes | No |
| Messier 2004[413] | Usual care | 68 | SF-36 MCS and PCS | 39.13 | No | Yes | No |
| Miranda-Filloy 2005[415] | Placebo | 10 | WOMAC | 8 | No | No | No |
| Miranda-Filloy 2005[415] | PES | 10 | WOMAC | 8 | No | No | No |
| Ng 2003[418] | Acupuncture | 8 | Pain NRS | 4 | No | No | No |
| Ng 2003[418] | TENS | 8 | Pain NRS | 4 | No | No | No |
| Nguyen 1997[419] | Balneotherapy | 91 | Pain VAS | 4 | No | No | No |
| Nguyen 1997[419] | Usual care | 97 | Pain VAS | 4 | No | No | No |
| Nigg 2006[421] | Insoles | 57 | WOMAC | 12 | No | No | No |
| Nigg 2006[421] | Usual care | 66 | WOMAC | 12 | No | No | No |
| Perlman 2006[424] | Manual therapy | 34 | WOMAC | 8 | No | No | No |
| Perlman 2006[424] | Usual care | 34 | WOMAC | 8 | No | No | No |
| Perlman 2012[478] | Manual therapy | 22 | WOMAC | 8 | No | Yes | Yes |
| Perlman 2012[478] | Manual therapy | 24 | WOMAC | 8 | No | Yes | Yes |
| Perlman 2012[478] | Manual therapy | 24 | WOMAC | 8 | No | Yes | Yes |
| Perlman 2012[478] | Manual therapy | 25 | WOMAC | 8 | No | Yes | Yes |
| Perlman 2012[478] | Usual care | 24 | WOMAC | 8 | No | Yes | Yes |
| Perrot 1998[425] | Placebo | 19 | Pain VAS | 1.29 | No | No | No |

| Study | Intervention | n analysed | Quality-of-life instrument | Analysis time point (weeks) | IPD available | Adequate allocation concealment | Adequate allocation concealment 3–13 weeks |
|---|---|---|---|---|---|---|---|
| Perrot 1998[425] | PEMFs | 21 | Pain VAS | 1.29 | No | No | No |
| Pietrosimone 2009[427] | Ice/cooling treatment | 11 | Pain VAS | 0.14 | No | Yes | No |
| Pietrosimone 2009[427] | Usual care | 12 | Pain VAS | 0.14 | No | Yes | No |
| Pietrosimone 2009[427] | TENS | 10 | Pain VAS | 0.14 | No | Yes | No |
| Pipitone 2001[428] | Placebo | 35 | WOMAC | 6 | No | Yes | Yes |
| Pipitone 2001[428] | PEMFs | 34 | WOMAC | 6 | No | Yes | Yes |
| Pollard 2008[429] | Manual therapy | 26 | Pain VAS | 2 | No | Yes | No |
| Pollard 2008[429] | Placebo | 17 | Pain VAS | 2 | No | Yes | No |
| Rapp 2009[430] | Exercise – muscle strengthening | 9 | Pain VAS | 8 | No | No | No |
| Rapp 2009[430] | Manual therapy | 15 | Pain VAS | 8 | No | No | No |
| Rapp 2009[430] | Usual care | 15 | Pain VAS | 8 | No | No | No |
| Rattanachaiyanont 2008[431] | Heat treatment | 50 | WOMAC | 3 | No | Yes | Yes |
| Rattanachaiyanont 2008[431] | Placebo | 54 | WOMAC | 3 | No | Yes | Yes |
| Salacinski 2012[479] | Exercise – aerobic | 13 | WOMAC | 12 | No | Yes | Yes |
| Salacinski 2012[479] | Usual care | 15 | WOMAC | 12 | No | Yes | Yes |
| Salli 2010[480] | Exercise – muscle strengthening | 24 | SF-36 MCS and PCS | 8 | No | No | No |
| Salli 2010[480] | Exercise – muscle strengthening | 23 | SF-36 MCS and PCS | 8 | No | No | No |
| Salli 2010[480] | Usual care | 24 | SF-36 MCS and PCS | 8 | No | No | No |
| Sangdee 2002[434] | Acupuncture | 46 | WOMAC | 4 | No | No | No |
| Sangdee 2002[434] | Acupuncture | 46 | WOMAC | 4 | No | No | No |
| Sangdee 2002[434] | Placebo | 45 | WOMAC | 4 | No | No | No |
| Sangdee 2002[434] | Usual care | 49 | WOMAC | 4 | No | No | No |
| Selfe 2008[436] | Placebo | 19 | WOMAC | 8 | No | Yes | Yes |
| Selfe 2008[436] | TENS | 18 | WOMAC | 8 | No | Yes | Yes |
| Shen 2009[437] | Laser/light therapy | 18 | WOMAC | 4 | No | Yes | Yes |
| Shen 2009[437] | Placebo | 9 | WOMAC | 4 | No | Yes | Yes |
| Sherman 2009[438] | Balneotherapy | 24 | WOMAC | 7 | No | No | No |
| Sherman 2009[438] | Placebo | 20 | WOMAC | 7 | No | No | No |
| Stelian 1992[440] | Laser/light therapy | 15 | Pain VAS | 1.43 | No | No | No |

| Study | Intervention | n analysed | Quality-of-life instrument | Analysis time point (weeks) | IPD available | Adequate allocation concealment | Adequate allocation concealment 3–13 weeks |
|---|---|---|---|---|---|---|---|
| Stelian 1992[440] | Laser/light therapy | 18 | Pain VAS | 1.43 | No | No | No |
| Stelian 1992[440] | Placebo | 17 | Pain VAS | 1.43 | No | No | No |
| Suarez-Almazor 2010[125] | Acupuncture | 153 | SF-12 MCS and PCS | 6 | No | Yes | Yes |
| Suarez-Almazor 2010[125] | Sham acupuncture | 302 | SF-12 MCS and PCS | 6 | No | Yes | Yes |
| Sukenik 1999[441] | Balneotherapy | 10 | Pain VAS | 2 | No | No | No |
| Sukenik 1999[441] | Balneotherapy | 10 | Pain VAS | 2 | No | No | No |
| Sukenik 1999[441] | Balneotherapy | 7 | Pain VAS | 2 | No | No | No |
| Sukenik 1999[441] | Placebo | 9 | Pain VAS | 2 | No | No | No |
| Tascioglu 2004[444] | Laser/light therapy | 20 | WOMAC | 3 | No | No | No |
| Tascioglu 2004[444] | Laser/light therapy | 20 | WOMAC | 3 | No | No | No |
| Tascioglu 2004[444] | Placebo | 20 | WOMAC | 3 | No | No | No |
| Thamsborg 2005[446] | Placebo | 41 | WOMAC | 6 | No | No | No |
| Thamsborg 2005[446] | PEMFs | 42 | WOMAC | 6 | No | No | No |
| Thorstensson 2005[447] | Exercise – aerobic | 28 | SF-36 MCS and PCS | 6 | No | No | No |
| Thorstensson 2005[447] | Usual care | 28 | SF-36 MCS and PCS | 6 | No | No | No |
| Tishler 2004[448] | Balneotherapy | 44 | Pain VAS | 6 | No | No | No |
| Tishler 2004[448] | Usual care | 24 | Pain VAS | 6 | No | No | No |
| Toda 2008[449] | Insoles | 43 | Pain VAS | 12 | No | Yes | Yes |
| Toda 2008[449] | Insoles | 41 | Pain VAS | 12 | No | Yes | Yes |
| Toda 2008[449] | Placebo | 38 | Pain VAS | 12 | No | Yes | Yes |
| Topp 2002[450] | Exercise – muscle strengthening | 35 | WOMAC | 16 | No | No | No |
| Topp 2002[450] | Exercise – muscle strengthening | 32 | WOMAC | 16 | No | No | No |
| Topp 2002[450] | Usual care | 35 | WOMAC | 16 | No | No | No |
| Trock 1994[452] | Placebo | 44 | Pain VAS | 5 | No | Yes | Yes |
| Trock 1994[452] | PEMFs | 40 | Pain VAS | 5 | No | Yes | Yes |
| Tucker 2003[453] | Manual therapy | 30 | Pain VAS | 3 | No | No | No |
| Tucker 2003[453] | Usual care | 30 | Pain VAS | 3 | No | No | No |
| Tukmachi 2004[454] | Acupuncture | 9 | Pain VAS | 5 | No | Yes | Yes |
| Tukmachi 2004[454] | Acupuncture | 10 | Pain VAS | 5 | No | Yes | Yes |
| Tukmachi 2004[454] | Usual care | 10 | Pain VAS | 5 | No | Yes | Yes |
| Vance 2012[481] | Placebo | 25 | Pain VAS | 0.14 | No | Yes | No |

| Study | Intervention | *n* analysed | Quality-of-life instrument | Analysis time point (weeks) | IPD available | Adequate allocation concealment | Adequate allocation concealment 3–13 weeks |
|---|---|---|---|---|---|---|---|
| Vance 2012[481] | TENS | 25 | Pain VAS | 0.14 | No | Yes | No |
| Vance 2012[481] | TENS | 25 | Pain VAS | 0.14 | No | Yes | No |
| Vas 2004[120] | Acupuncture | 47 | WOMAC | 12 | Yes | Yes | Yes |
| Vas 2004[120] | Sham acupuncture | 43 | WOMAC | 12 | Yes | Yes | Yes |
| Wang 2009[455] | Usual care | 20 | SF-36 MCS and PCS | 12 | No | Yes | Yes |
| Wang 2009[455] | T'ai chi | 20 | SF-36 MCS and PCS | 12 | No | Yes | Yes |
| Wigler 1995[456] | Balneotherapy | 11 | Pain VAS | 4 | No | No | No |
| Wigler 1995[456] | Balneotherapy | 10 | Pain VAS | 4 | No | No | No |
| Wigler 1995[456] | Placebo | 12 | Pain VAS | 4 | No | No | No |
| Williamson 2007[136] | Acupuncture | 59 | WOMAC | 8 | Yes | Yes | Yes |
| Williamson 2007[136] | Exercise – muscle strengthening | 53 | WOMAC | 8 | Yes | Yes | Yes |
| Williamson 2007[136] | Usual care | 49 | WOMAC | 8 | Yes | Yes | Yes |
| Witt 2005[69] | Acupuncture | 145 | SF-36 dimensions | 8 | Yes | Yes | Yes |
| Witt 2005[69] | Sham acupuncture | 73 | SF-36 dimensions | 8 | Yes | Yes | Yes |
| Witt 2005[69] | Usual care | 67 | SF-36 dimensions | 8 | Yes | Yes | Yes |
| Witt 2006[77] | Acupuncture | 250 | SF-36 dimensions | 12 | Yes | Yes | Yes |
| Witt 2006[77] | Usual care | 244 | SF-36 dimensions | 12 | Yes | Yes | Yes |
| Wolsko 2004[457] | Placebo | 13 | WOMAC | 6 | No | Yes | Yes |
| Wolsko 2004[457] | Static magnets | 13 | WOMAC | 6 | No | Yes | Yes |
| Yip 2007[193] | Usual care | 70 | Pain VAS | 1 | No | No | No |
| Yip 2007[193] | T'ai chi | 79 | Pain VAS | 1 | No | No | No |
| Yurtkuran 2006[461] | Balneotherapy | 27 | WOMAC | 2 | No | Yes | No |
| Yurtkuran 2006[461] | Placebo | 28 | WOMAC | 2 | No | Yes | No |
| Zizic 1995[462] | Placebo | 33 | Pain VAS | 4 | No | No | No |
| Zizic 1995[462] | PES | 38 | Pain VAS | 4 | No | No | No |

## WinBUGS code

The WinBUGS modelling code is provided in the following section followed by a summary table of all variables included in the data sets and the R-code describing the specification of initial values for two chains.

### WinBUGS model code

```
model {
### Model for ipd ###
for(i in 1:n.subjects.ipd) {
        #likelihood
        y.ipd[i] ~ dnorm(theta[i], prec.ipd[study.ipd[i]])
        #model
        theta[i] <- mu.ipd[study.ipd[i]] +    b_basey[study.ipd[i]] *
        basey.ipd[i] +
        b_age.ipd[study.ipd[i]]*age.ipd[i]*mod1+b_bmi.ipd[study.ipd[i]]*bmi.
        ipd[i]*mod2 +
        (delta[index.ipd[i]] +
(b_tx_age*age.ipd[i]*mod1+b_tx_bmi*bmi.ipd[i]*mod2)*
        equals(baseline.ipd[i],1))*(1 -equals(treat.ipd[i],baseline.ipd[i]))

        #residual deviance
        dev.ipd[i]<- (y.ipd[i] - theta[i]) * (y.ipd[i] - theta[i])

        #impute missing data
        age.ipd[i]~dnorm(mu.age.ipd,p.age.ipd)
        bmi.ipd[i]~dnorm(mu.bmi.ipd,p.bmi.ipd)
}

for(l in 1:(n.arms.ipd)) {
        #consistency model
        delta[l]~dnorm(md[l], prec.re.ipd[l])
        md[l]<- d[treat.ipd.sum[l]] - d[baseline.ipd.sum[l]] + sw.ipd[l]
        prec.re.ipd[l]<- prec.re * 2 * (m.ipd.sum[l] - 1)/m.ipd.sum[l]

        #residual deviance
        resdev.ipd[l]<- sum(dev.ipd[start.ipd.sum[l]:end.ipd.sum[l]]) *
prec.ipd[study.ipd.sum[l]]
}

        #multi-arm adjustment for up to 4 arm studies
        sw.ipd[1]<-0
        sw.ipd[2]<-0
for (l in 3:n.arms.ipd) {
        sw.ipd[l]<- ((delta[l - 1] - d[treat.ipd.sum[l - 1]] +
        d[baseline.ipd.sum[l - 1]]) *
        (equals(m.ipd.sum[l],3)+equals(m.ipd.sum[l],4)) +
        (delta[l - 2] - d[treat.ipd.sum[l - 2]] + d[baseline.ipd.sum[l -
        2]]) * equals(m.ipd.sum[l],4))/ (m.ipd.sum[l]-1)
}

        #priors for missing data imputation parameters
        mu.age.ipd~dunif(-50,50)
        sd.age.ipd~dunif(0,30)
        p.age.ipd<-1/pow(sd.age.ipd,2)
        mu.bmi.ipd~dunif(-40,40)
```

```
        sd.bmi.ipd~dunif(0,20)
        p.bmi.ipd<-1/pow(sd.bmi.ipd,2)

for(j in 1:n.trials.ipd) {
        #random effect on baseline outcome coefficient
        b_basey[j]~dnorm(b_basey_mean,prec_basey)
        #prior on covariate main effect coefficients, study-specific
baseline term, outcome se
        b_age.ipd[j]~dnorm(0,1.0E-4)
        b_bmi.ipd[j]~dnorm(0,1.0E-4)
        mu.ipd[j]~dnorm(0,1.0E-4)
        se.ipd[j]~dunif(0,2)
        var.ipd[j]<-se.ipd[j]*se.ipd[j]
        prec.ipd[j]<-1/var.ipd[j]
}


### Model for aggregate data ###
        #impute standard deviation and remove feedback
        sdev.imp~dgamma(alpha,beta)
        sdev.imp.cut<-cut(sdev.imp)
        #parameter to adjust for baseline imbalance, remove feedback
        b_basey_new~dnorm(b_basey_mean,prec_basey)
        ancova.cut<-cut(b_basey_new)

for (j in 1:n.arms.ad) {
        #standard deviation imputed for study arm
        sdev[j]<- sd.y.ad[j]*(1-equals(sd.y.ad[j],0)) +
sdev.imp.cut*equals(sd.y.ad[j],0)
        se[j]<-sdev[j]/sqrt(n.ad[j])
        prec.ad[j]<-1/(se[j]*se[j])

        #likelihood
        y.ad[j] ~ dnorm(theta.ad[j], prec.ad[j])

        #model
        theta.ad[j] <- mu.ad[study.ad[j]] +
                (delta[j + n.arms.ipd]+ancova.cut*imb.ad[j] +
                        (b_tx_age*age.ad[study.ad[j]]*mod1 +
        b_tx_bmi*bmi.ad[study.ad[j]]*mod2)*
                (equals(baseline.ad[j],1))) *(1-
        equals(treat.ad[j],baseline.ad[j]))

        #consistency model
        delta[j + n.arms.ipd] ~ dnorm(md.ad[j], prec.re.ad[j])
        md.ad[j]<- d[treat.ad[j]] - d[baseline.ad[j]] + sw.ad[j]
        prec.re.ad[j]<- prec.re * 2 * (m.ad[j] - 1)/m.ad[j]

        #residual deviance
```

```
        resdev.ad[j] <- ((y.ad[j] - theta.ad[j]) * (y.ad[j] - theta.ad[j]) *
prec.ad[j])
}

        #multi-arm adjustment for up to 5 arm studies
        sw.ad[1]<-0
        sw.ad[2]<-0
for (i in 3:3) {
        sw.ad[i]<- ((delta[(i+n.arms.ipd) -1] - d[treat.ad[i-1]] +
        d[baseline.ad[i - 1]]) * (equals(m.ad[i],3)+equals(m.ad[i],4)) +
        (delta[(i + n.arms.ipd) - 2] - d[treat.ad[i - 2]] + d[baseline.ad[i
        - 2]]) * (equals(m.ad[i],4))) / (m.ad[i]-1)
}

for (i in 4:n.arms.ad) {
        sw.ad[i]<- ((delta[(i + n.arms.ipd) - 1] - d[treat.ad[i - 1]] +
        d[baseline.ad[i - 1]]) *
        (equals(m.ad[i],3)+equals(m.ad[i],4)+equals(m.ad[i],5)) +
         (delta[(i + n.arms.ipd) - 2] - d[treat.ad[i - 2]] + d[baseline.ad[i
        - 2]]) * (equals(m.ad[i],4)+equals(m.ad[i],5))+(delta[(i +
        n.arms.ipd) - 3] - d[treat.ad[i - 3]] + d[baseline.ad[i - 3]]) *
        (equals(m.ad[i],5))) / (m.ad[i]-1)
}

        #priors for missing data imputation parameters
        mu.age.ad~dunif(-50,50)
        sd.age.ad~dunif(0,30)
        p.age.ad<-1/pow(sd.age.ad,2)
        mu.bmi.ad~dunif(-40,40)
        sd.bmi.ad~dunif(0,20)
        p.bmi.ad<-1/pow(sd.bmi.ad,2)

for(j in 1:n.trials.ad) {
        #prior on study-specific baseline term
        mu.ad[j]~dnorm(0,1.0E-4)
        #impute missing data
        age.ad[j]~dnorm(mu.age.ad,p.age.ad)
        bmi.ad[j]~dnorm(mu.bmi.ad,p.bmi.ad)
}


        #sum of residual deviance for aggregate and individual patient data
        totresdev.ad<- sum(resdev.ad[])
        totresdev.ipd<- sum(resdev.ipd[])


###Priors for shared parameters###
        #priors on treatment effects
        d[1]<-0
for (k in 2:n.treatments) {
```

```
        d[k] ~ dnorm(0,1.0E-3)
}

        #prior on random treatment effect
        tau~ dunif(0,2)
        tau.sq <- tau *
        prec.re<- 1 / tau.sq

        #prior on mean and random effect variance for coefficient on
baseline outcome
        b_basey_mean~dnorm(0,1.0E-6)
        tau_basey~dunif(0,2)
        tau_basey.sq<-tau_basey*tau_basey
        prec_basey<-1 / tau_basey.sq

        #prior on covariate-treatment interaction coefficients
        b_tx_age~dnorm(0,1.0E-4)
        b_tx_bmi~dnorm(0,1.0E-4)

###Additional statistical summaries
# pairwise comparisons
for (c in 1: n.treatments) {
        for (k in 1: n.treatments) {
        ef[c,k] <- d[c]     - d[k]
        }
}

# ranking and prob{treatment k is the best}
for (k in 1: n.treatments) {
        rk[k]<- n.treatments + 1 - rank(d[],k)
        best[k]<- equals(rk[k],1)
}

}
```

## Description of data sets and variables

| Data set | Variable | Description |
|---|---|---|
| Data set descriptors, constants | n.subjects.ipd | Number of subjects in IPD |
| | n.arms.ipd | Number of arms in IPD trials |
| | n.trials.ipd | Number of IPD trials |
| | n.trials.ad | Number of AD trials |
| | n.arms.ad | Number of arms in AD trials |
| | n.treatments | Total number of treatments |
| | alpha | Alpha and beta parameters for gamma distribution, used to impute missing SDs |
| | beta | |
| | mod1 | Dummy variable indicating whether or not to include age covariate |
| | mod2 | Dummy variable indicating whether or not to include BMI covariate |

| Data set | Variable | Description |
|---|---|---|
| Data at individual patient level (IPD studies) | study.ipd | Study code |
| | treat.ipd | Treatment code |
| | baseline.ipd | Baseline (lowest index) treatment in study |
| | y.ipd | EQ-5D at follow-up |
| | basey.ipd | Baseline EQ-5D |
| | age.ipd | Centred baseline age |
| | bmi.ipd | Centred baseline BMI |
| | index.ipd | Sequential numbering of study arms in IPD set |
| Data at study arm level (IPD studies) | study.ipd.sum | Study code |
| | treat.ipd.sum | Treatment code |
| | baseline.ipd.sum | Baseline (lowest index) treatment in study |
| | start.ipd.sum | Number of the first line of data including study arm |
| | end.ipd.sum | Number of the last line of data including study arm |
| | m.ipd.sum | Code for each study arm indicating whether it is the first, second, third, fourth or fifth arm in the study. Labelled as m = 2, 2, 3, 4, 5, respectively |
| Data at study arm level (AD studies) | study.ad | Study code |
| | treat.ad | Treatment code |
| | baseline.ad | Baseline (lowest index) treatment in study |
| | y.ad | Mean EQ-5D at follow-up |
| | sd.y.ad | SD for EQ-5D at follow-up (set to zero if missing) |
| | n.ad | Number of patients |
| | imb.ad | Difference in baseline EQ-5D from baseline treatment (imbalance) |
| | m.ad | Code for each study arm indicating whether it is the first, second, third, fourth or fifth arm in the study. Labelled as m = 2, 2, 3, 4, 5, respectively |
| Data at study level (AD studies) | age.ad | Centred baseline age |
| | bmi.ad | Centred baseline BMI |

AD, aggregate data.

# R-code used to generate initial values

```
#Initial values set 1
inits1<-list(d = c(NA,rep(0,n.treatments-1)), mu.ipd =
rep(0.7,n.trials.ipd), mu.ad = rep(0.4,n.trials.ad), delta =
rep(0.2,n.arms.ad+n.arms.ipd), b_basey_mean=0,tau_basey=1,
mu.age.ipd=0,sd.age.ipd=10,mu.bmi.ipd=0,sd.bmi.ipd=10,
mu.age.ad=0,sd.age.ad=10,mu.bmi.ad=0,sd.bmi.ad=10,tau = c(1), se.ipd =
rep(2,n.trials.ipd), b_age.ipd=0.1,b_bmi.ipd=0.1,b_tx_age=0.1,b_tx_bmi=0.1)
#Initial values set 2
inits2<-list(d = c(NA,rep(0.2,n.treatments-1)), mu.ipd =
rep(0.4,n.trials.ipd), mu.ad = rep(0.7,n.trials.ad), delta =
rep(0,n.arms.ad+n.arms.ipd), b_basey_mean=0.5,tau_basey=0.2,
mu.age.ipd=5,sd.age.ipd=5,mu.bmi.ipd=5,sd.bmi.ipd=5,
mu.age.ad=0,sd.age.ad=5,mu.bmi.ad=0,sd.bmi.ad=5,tau = c(0.5), se.ipd =
rep(0.05,n.trials.ipd),  b_age.ipd=0.5,b_bmi.ipd=0.5,b_tx_age=0,b_tx_bmi=0)
```

# Mapped EQ-5D values

| Study | Intervention | Quality-of-life instrument | EQ-5D score, mean (SE) | | |
| --- | --- | --- | --- | --- | --- |
| | | | Baseline | Follow-up | Change in |
| Adedoyin 2002[301] | Interferential therapy | Pain VAS | 0.48 (N/A) | 0.69 (N/A) | 0.22 (N/A) |
| Adedoyin 2002[301] | Placebo | Pain VAS | 0.49 (N/A) | 0.63 (N/A) | 0.13 (N/A) |
| Aglamiş 2009[189] | Usual care | SF-36 dimensions | 0.41 (0.08) | 0.50 (0.08) | 0.08 (0.07) |
| Aglamiş 2009[189] | Exercise – aerobic | SF-36 dimensions | 0.68 (0.05) | 0.87 (0.03) | 0.19 (0.04) |
| Alcidi 2007[350] | Heat treatment | Pain VAS | 0.53 (0.04) | 0.59 (0.06) | 0.06 (0.05) |
| Alcidi 2007[350] | TENS | Pain VAS | 0.56 (0.05) | 0.60 (0.05) | 0.03 (0.05) |
| An 2008[351] | Usual care | WOMAC | 0.70 (0.03) | 0.62 (0.08) | −0.08 (0.07) |
| An 2008[351] | Exercise – muscle strengthening | WOMAC | 0.66 (0.06) | 0.68 (0.06) | 0.02 (0.06) |
| Arazpour 2013[463] | Insoles | Pain VAS | 0.52 (0.02) | 0.61 (0.02) | 0.09 (0.02) |
| Arazpour 2013[463] | Braces | Pain VAS | 0.54 (0.02) | 0.61 (0.02) | 0.07 (0.02) |
| Baker 2001[353] | Placebo | SF-36 dimensions | 0.76 (0.04) | 0.75 (0.05) | −0.02 (0.03) |
| Baker 2001[353] | Exercise – muscle strengthening | SF-36 dimensions | 0.69 (0.05) | 0.78 (0.04) | 0.09 (0.04) |
| Bálint 2007[355] | Balneotherapy | WOMAC | 0.56 (0.01) | 0.60 (0.01) | 0.04 (0.01) |
| Bálint 2007[355] | Placebo | WOMAC | 0.56 (0.01) | 0.65 (0.01) | 0.10 (0.01) |
| Bennell 2011[464] | Usual care | WOMAC | 0.61 (0.02) | 0.64 (0.02) | 0.03 (0.02) |
| Bennell 2011[464] | Insoles | WOMAC | 0.60 (0.02) | 0.63 (0.02) | 0.03 (0.02) |
| Berman 1999[357] | Acupuncture | WOMAC | 0.47 (0.05) | 0.64 (0.03) | 0.17 (0.05) |
| Berman 1999[357] | Usual care | WOMAC | 0.46 (0.04) | 0.46 (0.05) | 0.00 (0.04) |
| Berman 2004[134] | Acupuncture | EQ-5D | 0.61 (0.02) | 0.68 (0.01) | 0.07 (0.02) |

| Study | Intervention | Quality-of-life instrument | EQ-5D score, mean (SE) | | |
|-------|-------------|---------------------------|------------------------|-----------|-----------|
| | | | Baseline | Follow-up | Change in |
| Berman 2004[134] | Sham acupuncture | EQ-5D | 0.62 (0.02) | 0.69 (0.01) | 0.07 (0.02) |
| Brismée 2007[360] | T'ai chi | WOMAC | 0.34 (0.08) | 0.48 (0.08) | 0.14 (0.08) |
| Brismée 2007[360] | Usual care | WOMAC | 0.41 (0.07) | 0.38 (0.07) | −0.02 (0.07) |
| Brosseau 2012[466] | Usual care | SF-36 dimensions | 0.81 (0.02) | 0.82 (0.02) | 0.02 (0.02) |
| Brosseau 2012[466] | Exercise – aerobic | SF-36 dimensions | 0.82 (0.02) | 0.81 (0.02) | −0.01 (0.02) |
| Brosseau 2012[466] | Exercise – aerobic | SF-36 dimensions | 0.79 (0.02) | 0.81 (0.02) | 0.03 (0.02) |
| Bruce-Brand 2012[467] | Usual care | SF-36 MCS and PCS | N/A (N/A) | 0.42 (0.26) | 0.42 (0.26) |
| Bruce-Brand 2012[467] | Exercise – muscle strengthening | SF-36 MCS and PCS | N/A (N/A) | 0.46 (0.17) | 0.46 (0.17) |
| Bruce-Brand 2012[467] | NMES | SF-36 MCS and PCS | N/A (N/A) | 0.44 (0.15) | 0.44 (0.15) |
| Burch 2008[303] | TENS | WOMAC | 0.48 (0.04) | 0.57 (0.04) | 0.09 (0.04) |
| Burch 2008[303] | Interferential therapy | WOMAC | 0.49 (0.04) | 0.64 (0.02) | 0.15 (0.03) |
| Callaghan 2005[363] | PES | Pain VAS | 0.54 (0.08) | 0.56 (0.09) | 0.03 (0.08) |
| Callaghan 2005[363] | PES | Pain VAS | 0.57 (0.11) | 0.58 (0.12) | 0.01 (0.11) |
| Callaghan 2005[363] | Placebo | Pain VAS | 0.55 (0.02) | 0.54 (0.07) | −0.01 (0.06) |
| Cheing 2003[191] | TENS | Pain VAS | 0.59 (0.04) | 0.67 (0.04) | 0.08 (0.04) |
| Cheing 2003[191] | TENS | Pain VAS | 0.57 (0.03) | 0.71 (0.02) | 0.14 (0.03) |
| Cheing 2003[191] | TENS | Pain VAS | 0.56 (0.04) | 0.68 (0.03) | 0.12 (0.03) |
| Cheing 2003[191] | Placebo | Pain VAS | 0.58 (0.03) | 0.59 (0.03) | 0.01 (0.03) |
| Durmuş 2007[370] | PES | Pain VAS | 0.57 (N/A) | 0.68 (N/A) | 0.10 (N/A) |
| Durmuş 2007[370] | Exercise – muscle strengthening | Pain VAS | 0.55 (N/A) | 0.67 (N/A) | 0.12 (N/A) |
| Fary 2011[468] | Placebo | SF-36 MCS and PCS | 0.58 (0.04) | 0.64 (N/A) | 0.06 (0.04) |
| Fary 2011[468] | PES | SF-36 MCS and PCS | 0.59 (0.04) | 0.66 (N/A) | 0.07 (0.04) |
| Fischer 2005[373] | PEMFs | Pain VAS | 0.56 (0.02) | 0.58 (0.03) | 0.02 (0.03) |
| Fischer 2005[373] | Placebo | Pain VAS | 0.55 (0.02) | 0.62 (0.03) | 0.07 (0.03) |
| Flusser 2002[374] | Balneotherapy | Pain VAS | 0.53 (0.03) | 0.55 (0.03) | 0.02 (0.03) |
| Flusser 2002[374] | Placebo | Pain VAS | 0.54 (0.05) | 0.53 (0.05) | −0.02 (0.04) |
| Foroughi 2011[469] | Exercise – muscle strengthening | WOMAC | 0.65 (0.04) | 0.70 (0.02) | 0.05 (0.03) |
| Foroughi 2011[469] | Placebo | WOMAC | 0.62 (0.04) | 0.66 (0.03) | 0.04 (0.04) |
| Fukuda 2011[470] | Usual care | Pain NRS | 0.58 (0.04) | 0.60 (0.04) | 0.02 (0.04) |
| Fukuda 2011[470] | Placebo | Pain NRS | 0.49 (0.04) | 0.54 (0.05) | 0.05 (0.04) |
| Fukuda 2011[470] | PES | Pain NRS | 0.53 (0.05) | 0.68 (0.03) | 0.15 (0.04) |
| Fukuda 2011[470] | PES | Pain NRS | 0.55 (0.05) | 0.64 (0.04) | 0.10 (0.04) |
| Fukuda 2011[471] | Placebo | Pain VAS | 0.54 (0.02) | 0.57 (0.04) | 0.03 (0.04) |
| Fukuda 2011[471] | Laser/light therapy | Pain VAS | 0.54 (0.04) | 0.59 (0.02) | 0.05 (0.03) |
| Garland 2007[377] | PES | WOMAC | 0.46 (0.05) | 0.54 (0.06) | 0.08 (0.05) |
| Garland 2007[377] | Placebo | WOMAC | 0.53 (0.05) | 0.52 (0.07) | −0.01 (0.06) |
| Grimmer 1992[378] | TENS | Pain VAS | 0.52 (0.06) | 0.65 (0.05) | 0.13 (0.05) |
| Grimmer 1992[378] | TENS | Pain VAS | 0.55 (0.05) | 0.68 (0.03) | 0.13 (0.04) |
| Grimmer 1992[378] | Placebo | Pain VAS | 0.54 (0.05) | 0.62 (0.06) | 0.08 (0.05) |
| Gundog 2012[302] | Interferential therapy | WOMAC | 0.31 (0.06) | 0.73 (0.01) | 0.42 (0.05) |

| Study | Intervention | Quality-of-life instrument | EQ-5D score, mean (SE) | | |
|---|---|---|---|---|---|
| | | | Baseline | Follow-up | Change in |
| Gundog 2012[302] | Interferential therapy | WOMAC | 0.24 (0.08) | 0.74 (0.01) | 0.50 (0.07) |
| Gundog 2012[302] | Interferential therapy | WOMAC | 0.22 (0.06) | 0.71 (0.03) | 0.48 (0.05) |
| Gundog 2012[302] | Placebo | WOMAC | 0.27 (0.06) | 0.41 (0.05) | 0.14 (0.05) |
| Hasegawa 2010[381] | Exercise – muscle strengthening | Pain NRS | 0.64 (0.05) | 0.72 (0.05) | 0.08 (0.05) |
| Hasegawa 2010[381] | Usual care | Pain NRS | 0.68 (0.05) | 0.67 (0.06) | −0.01 (0.05) |
| Huang 2005[384] | Exercise – muscle strengthening | Pain VAS | 0.58 (0.03) | 0.61 (0.01) | 0.04 (0.02) |
| Huang 2005[384] | Usual care | Pain VAS | 0.58 (0.03) | 0.60 (0.03) | 0.02 (0.03) |
| Imoto 2012[474] | Usual care | SF-36 dimensions | 0.52 (0.04) | 0.60 (0.04) | 0.09 (0.03) |
| Imoto 2012[474] | Exercise – muscle strengthening | SF-36 dimensions | 0.52 (0.03) | 0.66 (0.04) | 0.14 (0.03) |
| Itoh 2008[386] | Acupuncture | WOMAC | 0.45 (0.08) | 0.59 (0.05) | 0.14 (0.06) |
| Itoh 2008[386] | TENS | WOMAC | 0.44 (0.13) | 0.60 (0.04) | 0.16 (0.11) |
| Itoh 2008[386] | Usual care | WOMAC | 0.49 (0.06) | 0.52 (0.06) | 0.03 (0.06) |
| Itoh 2008[387] | Acupuncture | WOMAC | 0.39 (0.08) | 0.54 (0.03) | 0.16 (0.07) |
| Itoh 2008[387] | Acupuncture | WOMAC | 0.46 (0.08) | 0.63 (0.05) | 0.17 (0.06) |
| Itoh 2008[387] | Sham acupuncture | WOMAC | 0.45 (0.06) | 0.48 (0.05) | 0.03 (0.05) |
| Jubb 2008[391] | Sham acupuncture | WOMAC | 0.41 (0.07) | 0.48 (0.06) | 0.07 (0.06) |
| Jubb 2008[391] | Acupuncture | WOMAC | 0.37 (0.06) | 0.54 (0.05) | 0.17 (0.05) |
| Kang 2007[393] | TENS | Pain VAS | 0.56 (N/A) | 0.59 (N/A) | 0.02 (N/A) |
| Kang 2007[393] | Placebo | Pain VAS | 0.55 (N/A) | 0.57 (N/A) | 0.01 (N/A) |
| Kovar 1992[398] | Exercise – aerobic | Pain VAS | 0.57 (0.03) | 0.61 (0.02) | 0.04 (0.02) |
| Kovar 1992[398] | Usual care | Pain VAS | 0.58 (0.03) | 0.58 (0.03) | 0.00 (0.03) |
| Kuptniratsaikul 2002[399] | Exercise – muscle strengthening | Pain VAS | 0.56 (0.01) | 0.60 (0.02) | 0.04 (0.01) |
| Kuptniratsaikul 2002[399] | Usual care | Pain VAS | 0.55 (0.01) | 0.57 (0.02) | 0.02 (0.01) |
| Lansdown 2009[126] | Usual care | EQ-5D | 0.67 (0.04) | 0.66 (0.06) | −0.01 (0.06) |
| Lansdown 2009[126] | Acupuncture | EQ-5D | 0.61 (0.06) | 0.71 (0.07) | 0.10 (0.08) |
| Lee 2009[401] | Usual care | SF-36 MCS and PCS | 0.58 (0.10) | 0.56 (0.12) | −0.02 (0.06) |
| Lee 2009[401] | T'ai chi | SF-36 MCS and PCS | 0.51 (0.09) | 0.49 (0.09) | −0.01 (0.10) |
| Lewis 1994[403] | TENS | Pain VAS | N/A (N/A) | 0.58 (N/A) | 0.58 (N/A) |
| Lewis 1994[403] | Placebo | Pain VAS | N/A (N/A) | 0.56 (N/A) | 0.56 (N/A) |
| Lim 2010[475] | Usual care | SF-36 MCS and PCS | 0.51 (0.07) | 0.54 (0.08) | 0.04 (0.02) |
| Lim 2010[475] | Exercise – muscle strengthening | SF-36 MCS and PCS | 0.57 (0.04) | 0.65 (0.03) | 0.08 (0.02) |
| Lim 2010[475] | Exercise – muscle strengthening | SF-36 MCS and PCS | 0.53 (0.06) | 0.63 (0.03) | 0.10 (0.03) |
| Lu 2010[192] | Acupuncture | Pain VAS | 0.56 (0.03) | 0.60 (0.03) | 0.03 (0.03) |
| Lu 2010[192] | Sham acupuncture | Pain VAS | 0.57 (0.04) | 0.58 (0.04) | 0.02 (0.04) |
| Lund 2008[407] | Usual care | Pain VAS | 0.58 (0.05) | 0.55 (0.04) | −0.03 (0.04) |
| Lund 2008[407] | Exercise – muscle strengthening | Pain VAS | 0.54 (0.03) | 0.56 (0.04) | 0.01 (0.03) |
| Lund 2008[407] | Exercise – aerobic | Pain VAS | 0.57 (0.06) | 0.57 (0.04) | 0.00 (0.05) |

| Study | Intervention | Quality-of-life instrument | EQ-5D score, mean (SE) | | |
| | | | Baseline | Follow-up | Change in |
|---|---|---|---|---|---|
| Maillefert 2001[409] | Insoles | WOMAC | 0.48 (0.04) | 0.45 (0.04) | –0.02 (0.04) |
| Maillefert 2001[409] | Placebo | WOMAC | 0.47 (0.04) | 0.49 (0.04) | 0.01 (0.04) |
| Maurer 1999[410] | Usual care | WOMAC | 0.60 (N/A) | 0.64 (N/A) | 0.04 (N/A) |
| Maurer 1999[410] | Exercise – muscle strengthening | WOMAC | 0.62 (N/A) | 0.67 (N/A) | 0.05 (N/A) |
| Mavrommatis 2012[476] | Usual care | SF-36 MCS and PCS | 0.46 (0.04) | 0.60 (0.02) | 0.14 (0.02) |
| Mavrommatis 2012[476] | Sham acupuncture | SF-36 MCS and PCS | 0.43 (0.05) | 0.61 (0.02) | 0.18 (0.03) |
| Mavrommatis 2012[476] | Acupuncture | SF-36 MCS and PCS | 0.48 (0.03) | 0.72 (0.02) | 0.23 (0.02) |
| Mazzuca 2004[411] | Heat treatment | WOMAC | 0.46 (0.07) | 0.52 (0.06) | 0.06 (0.06) |
| Mazzuca 2004[411] | Placebo | WOMAC | 0.47 (0.07) | 0.53 (0.05) | 0.07 (0.06) |
| McCarthy 2004[412] | Usual care | EQ-5D | 0.50 (N/A) | 0.52 (N/A) | 0.02 (N/A) |
| McCarthy 2004[412] | Exercise – muscle strengthening | EQ-5D | 0.54 (N/A) | 0.60 (N/A) | 0.06 (N/A) |
| Messier 2004[413] | Usual care | SF-36 MCS and PCS | 0.55 (0.03) | 0.61 (0.02) | 0.06 (0.01) |
| Messier 2004[413] | Exercise – aerobic | SF-36 MCS and PCS | 0.58 (0.02) | 0.63 (0.02) | 0.06 (0.01) |
| Miranda-Filloy 2005[415] | PES | WOMAC | 0.51 (N/A) | 0.69 (N/A) | 0.18 (N/A) |
| Miranda-Filloy 2005[415] | Placebo | WOMAC | 0.46 (N/A) | 0.55 (N/A) | 0.09 (N/A) |
| Ng 2003[418] | Acupuncture | Pain NRS | 0.65 (0.05) | 0.70 (0.04) | 0.05 (0.05) |
| Ng 2003[418] | TENS | Pain NRS | 0.67 (0.04) | 0.69 (0.04) | 0.02 (0.04) |
| Nguyen 1997[419] | Usual care | Pain VAS | 0.58 (0.02) | 0.58 (0.02) | 0.00 (0.02) |
| Nguyen 1997[419] | Balneotherapy | Pain VAS | 0.57 (0.02) | 0.62 (0.02) | 0.05 (0.02) |
| Nigg 2006[421] | Insoles | WOMAC | 0.61 (0.03) | 0.66 (0.02) | 0.05 (0.03) |
| Nigg 2006[421] | Usual care | WOMAC | 0.60 (0.03) | 0.65 (0.03) | 0.06 (0.03) |
| Perlman 2006[424] | Manual therapy | WOMAC | 0.49 (0.06) | 0.65 (0.04) | 0.16 (0.05) |
| Perlman 2006[424] | Usual care | WOMAC | 0.42 (0.06) | 0.47 (0.06) | 0.05 (0.06) |
| Perlman 2012[478] | Manual therapy | WOMAC | 0.44 (0.08) | 0.60 (0.05) | 0.16 (0.07) |
| Perlman 2012[478] | Manual therapy | WOMAC | 0.47 (0.07) | 0.62 (0.05) | 0.15 (0.06) |
| Perlman 2012[478] | Manual therapy | WOMAC | 0.44 (0.07) | 0.64 (0.04) | 0.20 (0.06) |
| Perlman 2012[478] | Manual therapy | WOMAC | 0.49 (0.07) | 0.66 (0.04) | 0.17 (0.06) |
| Perlman 2012[478] | Usual care | WOMAC | 0.45 (0.06) | 0.52 (0.05) | 0.06 (0.05) |
| Perrot 1998[425] | PEMFs | Pain VAS | 0.52 (N/A) | 0.60 (N/A) | 0.08 (N/A) |
| Perrot 1998[425] | Placebo | Pain VAS | 0.50 (N/A) | 0.54 (N/A) | 0.05 (N/A) |
| Pietrosimone 2009[427] | TENS | Pain VAS | 0.68 (0.04) | 0.70 (0.02) | 0.03 (0.04) |
| Pietrosimone 2009[427] | Ice/cooling treatment | Pain VAS | 0.67 (0.04) | 0.69 (0.04) | 0.02 (0.04) |
| Pietrosimone 2009[427] | Usual care | Pain VAS | 0.67 (0.04) | 0.67 (0.05) | 0.00 (0.04) |
| Pipitone 2001[428] | Placebo | WOMAC | 0.44 (0.05) | 0.46 (0.05) | 0.01 (0.05) |
| Pipitone 2001[428] | PEMFs | WOMAC | 0.44 (0.07) | 0.49 (0.06) | 0.05 (0.06) |
| Pollard 2008[429] | Manual therapy | Pain VAS | 0.63 (0.04) | 0.67 (0.03) | 0.04 (0.03) |
| Pollard 2008[429] | Placebo | Pain VAS | 0.62 (0.06) | 0.63 (0.05) | 0.01 (0.05) |
| Rapp 2009[430] | Exercise – muscle strengthening | Pain VAS | N/A (N/A) | 0.67 (0.05) | 0.67 (0.05) |
| Rapp 2009[430] | Manual therapy | Pain VAS | N/A (N/A) | 0.66 (0.04) | 0.66 (0.04) |

| Study | Intervention | Quality-of-life instrument | EQ-5D score, mean (SE) | | |
| | | | Baseline | Follow-up | Change in |
|---|---|---|---|---|---|
| Rapp 2009[430] | Usual care | Pain VAS | N/A (N/A) | 0.64 (0.05) | 0.64 (0.05) |
| Rattanachaiyanont 2008[431] | Heat treatment | WOMAC | 0.60 (0.03) | 0.66 (0.02) | 0.07 (0.02) |
| Rattanachaiyanont 2008[431] | Placebo | WOMAC | 0.58 (0.03) | 0.66 (0.02) | 0.07 (0.03) |
| Salacinski 2012[479] | Usual care | WOMAC | 0.66 (0.04) | 0.64 (0.05) | −0.02 (0.04) |
| Salacinski 2012[479] | Exercise – aerobic | WOMAC | 0.65 (0.04) | 0.71 (0.02) | 0.06 (0.04) |
| Salli 2010[480] | Usual care | SF-36 MCS and PCS | 0.54 (0.07) | 0.59 (0.05) | 0.04 (0.03) |
| Salli 2010[480] | Exercise – muscle strengthening | SF-36 MCS and PCS | 0.55 (0.08) | 0.62 (0.05) | 0.07 (0.08) |
| Salli 2010[480] | Exercise – muscle strengthening | SF-36 MCS and PCS | 0.53 (0.09) | 0.60 (0.08) | 0.07 (0.11) |
| Sangdee 2002[434] | Acupuncture | WOMAC | 0.42 (0.06) | 0.65 (0.03) | 0.23 (0.05) |
| Sangdee 2002[434] | Acupuncture | WOMAC | 0.42 (0.06) | 0.65 (0.03) | 0.23 (0.05) |
| Sangdee 2002[434] | Usual care | WOMAC | 0.44 (0.05) | 0.63 (0.03) | 0.19 (0.04) |
| Sangdee 2002[434] | Placebo | WOMAC | 0.44 (0.05) | 0.60 (0.03) | 0.16 (0.04) |
| Selfe 2008[436] | Placebo | WOMAC | 0.45 (0.08) | 0.63 (0.05) | 0.18 (0.07) |
| Selfe 2008[436] | TENS | WOMAC | 0.44 (0.10) | 0.64 (0.05) | 0.21 (0.08) |
| Shen 2009[437] | Laser/light therapy | WOMAC | 0.58 (0.06) | 0.69 (0.03) | 0.11 (0.06) |
| Shen 2009[437] | Placebo | WOMAC | 0.66 (0.03) | 0.73 (0.01) | 0.07 (0.04) |
| Sherman 2009[438] | Placebo | WOMAC | 0.40 (0.09) | 0.50 (0.08) | 0.10 (0.08) |
| Sherman 2009[438] | Balneotherapy | WOMAC | 0.36 (0.08) | 0.49 (0.08) | 0.13 (0.08) |
| Stelian 1992[440] | Laser/light therapy | Pain VAS | 0.53 (0.06) | 0.62 (0.06) | 0.09 (0.05) |
| Stelian 1992[440] | Laser/light therapy | Pain VAS | 0.51 (0.05) | 0.63 (0.05) | 0.11 (0.05) |
| Stelian 1992[440] | Placebo | Pain VAS | 0.54 (0.05) | 0.54 (0.05) | 0.00 (0.05) |
| Suarez-Almazor 2010[125] | Sham acupuncture | SF-12 MCS and PCS | 0.61 (0.02) | 0.69 (0.02) | 0.08 (0.02) |
| Suarez-Almazor 2010[125] | Acupuncture | SF-12 MCS and PCS | 0.61 (0.02) | 0.71 (0.02) | 0.10 (0.02) |
| Sukenik 1999[441] | Balneotherapy | Pain VAS | 0.54 (0.05) | 0.55 (0.04) | 0.01 (0.04) |
| Sukenik 1999[441] | Balneotherapy | Pain VAS | 0.53 (0.03) | 0.56 (0.05) | 0.03 (0.04) |
| Sukenik 1999[441] | Balneotherapy | Pain VAS | 0.52 (0.03) | 0.61 (0.06) | 0.09 (0.05) |
| Sukenik 1999[441] | Placebo | Pain VAS | 0.54 (0.05) | 0.54 (0.04) | 0.00 (0.04) |
| Tascioglu 2004[444] | Laser/light therapy | WOMAC | 0.46 (0.05) | 0.48 (0.05) | 0.02 (0.04) |
| Tascioglu 2004[444] | Laser/light therapy | WOMAC | 0.42 (0.05) | 0.45 (0.05) | 0.03 (0.05) |
| Tascioglu 2004[444] | Placebo | WOMAC | 0.43 (0.06) | 0.46 (0.05) | 0.03 (0.05) |
| Thamsborg 2005[446] | PEMFs | WOMAC | 0.46 (0.04) | 0.54 (0.03) | 0.08 (0.04) |
| Thamsborg 2005[446] | Placebo | WOMAC | 0.42 (0.05) | 0.48 (0.05) | 0.05 (0.05) |
| Thorstensson 2005[447] | Usual care | SF-36 MCS and PCS | 0.72 (N/A) | 0.73 (N/A) | 0.01 (N/A) |
| Thorstensson 2005[447] | Exercise – aerobic | SF-36 MCS and PCS | 0.71 (N/A) | 0.74 (N/A) | 0.02 (N/A) |
| Tishler 2004[448] | Balneotherapy | Pain VAS | 0.53 (0.03) | 0.66 (0.03) | 0.13 (0.03) |
| Tishler 2004[448] | Usual care | Pain VAS | 0.52 (0.04) | 0.53 (0.04) | 0.01 (0.04) |
| Toda 2008[449] | Placebo | Pain VAS | 0.59 (0.03) | 0.58 (0.02) | −0.01 (0.03) |
| Toda 2008[449] | Insoles | Pain VAS | 0.61 (0.03) | 0.60 (0.03) | −0.01 (0.03) |

| Study | Intervention | Quality-of-life instrument | EQ-5D score, mean (SE) | | |
|---|---|---|---|---|---|
| | | | Baseline | Follow-up | Change in |
| Toda 2008[449] | Insoles | Pain VAS | 0.59 (0.03) | 0.60 (0.04) | 0.01 (0.03) |
| Topp 2002[450] | Exercise – muscle strengthening | WOMAC | 0.35 (0.06) | 0.40 (0.09) | 0.05 (0.08) |
| Topp 2002[450] | Exercise – muscle strengthening | WOMAC | 0.40 (0.06) | 0.45 (0.05) | 0.04 (0.05) |
| Topp 2002[450] | Usual care | WOMAC | 0.41 (0.06) | 0.39 (0.06) | −0.01 (0.05) |
| Trock 1994[452] | Placebo | Pain VAS | 0.53 (0.03) | 0.57 (0.03) | 0.04 (0.03) |
| Trock 1994[452] | PEMFs | Pain VAS | 0.51 (0.03) | 0.59 (0.03) | 0.08 (0.03) |
| Tucker 2003[453] | Manual therapy | Pain VAS | 0.62 (0.03) | 0.69 (0.02) | 0.07 (0.02) |
| Tucker 2003[453] | Usual care | Pain VAS | 0.63 (0.04) | 0.69 (0.02) | 0.06 (0.03) |
| Tukmachi 2004[454] | Acupuncture | Pain VAS | 0.54 (0.06) | 0.63 (0.08) | 0.08 (0.07) |
| Tukmachi 2004[454] | Acupuncture | Pain VAS | 0.54 (0.05) | 0.67 (0.06) | 0.12 (0.05) |
| Tukmachi 2004[454] | Usual care | Pain VAS | 0.51 (0.04) | 0.52 (0.06) | 0.01 (0.05) |
| Vance 2012[481] | TENS | Pain VAS | 0.65 (0.04) | 0.68 (0.04) | 0.02 (0.04) |
| Vance 2012[481] | TENS | Pain VAS | 0.64 (0.04) | 0.66 (0.04) | 0.02 (0.04) |
| Vance 2012[481] | Placebo | Pain VAS | 0.64 (0.04) | 0.68 (0.03) | 0.04 (0.04) |
| Vas 2004[120] | Acupuncture | WOMAC | 0.40 (0.06) | 0.75 (0.04) | 0.35 (0.08) |
| Vas 2004[120] | Sham acupuncture | WOMAC | 0.35 (0.06) | 0.56 (0.06) | 0.21 (0.07) |
| Wang 2009[455] | Usual care | SF-36 MCS and PCS | 0.50 (0.06) | 0.56 (0.05) | 0.06 (0.02) |
| Wang 2009[455] | T'ai chi | SF-36 MCS and PCS | 0.58 (0.05) | 0.71 (0.04) | 0.13 (0.03) |
| Wigler 1995[456] | Balneotherapy | Pain VAS | 0.51 (N/A) | 0.55 (N/A) | 0.04 (N/A) |
| Wigler 1995[456] | Balneotherapy | Pain VAS | 0.51 (N/A) | 0.56 (N/A) | 0.05 (N/A) |
| Wigler 1995[456] | Placebo | Pain VAS | 0.51 (N/A) | 0.54 (N/A) | 0.03 (N/A) |
| Williamson 2007[136] | Acupuncture | WOMAC | 0.44 (0.04) | 0.43 (0.04) | −0.01 (0.05) |
| Williamson 2007[136] | Exercise – muscle strengthening | WOMAC | 0.43 (0.05) | 0.46 (0.05) | 0.03 (0.06) |
| Williamson 2007[136] | Usual care | WOMAC | 0.43 (0.04) | 0.36 (0.04) | −0.06 (0.05) |
| Witt 2005[69] | Acupuncture | SF-36 dimensions | 0.57 (0.02) | 0.70 (0.02) | 0.14 (0.02) |
| Witt 2005[69] | Sham acupuncture | SF-36 dimensions | 0.55 (0.02) | 0.62 (0.03) | 0.07 (0.02) |
| Witt 2005[69] | Usual care | SF-36 dimensions | 0.56 (0.03) | 0.57 (0.02) | 0.01 (0.02) |
| Witt 2006[77] | Acupuncture | SF-36 dimensions | 0.57 (0.01) | 0.67 (0.02) | 0.10 (0.02) |
| Witt 2006[77] | Usual care | SF-36 dimensions | 0.57 (0.01) | 0.56 (0.02) | −0.01 (0.01) |
| Wolsko 2004[457] | Static magnets | WOMAC | 0.47 (0.12) | 0.64 (0.07) | 0.17 (0.10) |
| Wolsko 2004[457] | Placebo | WOMAC | 0.48 (N/A) | 0.66 (N/A) | 0.18 (N/A) |
| Yip 2007[193] | Usual care | Pain VAS | 0.59 (0.02) | 0.59 (0.03) | 0.00 (0.03) |
| Yip 2007[193] | T'ai chi | Pain VAS | 0.57 (0.02) | 0.61 (0.02) | 0.04 (0.02) |
| Yurtkuran 2006[461] | Placebo | WOMAC | 0.44 (0.04) | 0.58 (0.04) | 0.14 (0.04) |
| Yurtkuran 2006[461] | Balneotherapy | WOMAC | 0.47 (0.06) | 0.64 (0.03) | 0.17 (0.05) |
| Zizic 1995[462] | PES | Pain VAS | 0.53 (N/A) | 0.59 (N/A) | 0.06 (N/A) |
| Zizic 1995[462] | Placebo | Pain VAS | 0.54 (N/A) | 0.58 (N/A) | 0.04 (N/A) |

N/A, not applicable.

## Results of sensitivity analysis using upper and lower values for intervention costs and private costs for acupuncture sessions

(See *Chapter 5, Cost-effectiveness analysis* for values used.)

### *All trials*

| Treatment | Cost-effective intervention at threshold of £20,000 per QALY (ICER, £ per QALY) | |
| --- | --- | --- |
| | Use of minimum cost | Use of maximum cost |
| Standard care | TENS (2691) | TENS (2695) |
| Acupuncture | Acupuncture (19,033) | TENS (2694) |
| Braces | TENS (2693) | TENS (2691) |
| Heat treatment | TENS (2694) | TENS (2691) |
| Insoles | TENS (2690) | TENS (2694) |
| Interferential therapy | Interferential therapy (7626) | TENS (2687) |
| Laser light therapy | TENS (2690) | TENS (2694) |
| Manual therapy | TENS (2695) | TENS (2692) |
| NMES | TENS (2689) | TENS (2691) |
| PES | TENS (2690) | TENS (2693) |
| PEMFs | TENS (2688) | TENS (2695) |
| Static magnets | TENS (2696) | TENS (2689) |
| TENS | TENS (1426) | TENS (3959) |

### *Trials with a low risk of bias for allocation concealment*

| Treatment | Cost-effective intervention at threshold of £20,000 per QALY (ICER, £ per QALY) | |
| --- | --- | --- |
| | Use of minimum cost | Use of maximum cost |
| Usual care | Acupuncture (13,501) | Acupuncture (13,500) |
| Acupuncture | Acupuncture (5842) | Acupuncture (17,690) |
| Heat treatment | Acupuncture (13,499) | Acupuncture (13,500) |
| Insoles | Acupuncture (13,501) | Acupuncture (13,501) |
| Interferential therapy | Acupuncture (13,502) | Acupuncture (13,500) |
| Laser light therapy | Acupuncture (13,503) | Acupuncture (13,499) |
| Manual therapy | Acupuncture (13,500) | Acupuncture (13,503) |
| PES | Acupuncture (13,503) | Acupuncture (13,502) |
| PEMFs | Acupuncture (13,498) | Acupuncture (13,500) |
| Static magnets | Acupuncture (13,499) | Acupuncture (13,501) |
| TENS | Acupuncture (14,701) | Acupuncture (12,297) |

### *Trials with a low risk of bias for allocation concealment reporting at 3–13 weeks*

| Treatment | Cost-effective intervention at threshold of £20,000 per QALY (ICER, £ per QALY) | |
|---|---|---|
| | Use of minimum cost | Use of maximum cost |
| Usual care | Acupuncture (14,271) | Acupuncture (14,272) |
| Acupuncture | Acupuncture (6518) | Acupuncture (18,697) |
| Heat treatment | Acupuncture (14,278) | Acupuncture (14,275) |
| Insoles | Acupuncture (14,271) | Acupuncture (14,272) |
| Interferential therapy | Acupuncture (14,274) | Acupuncture (14,275) |
| Laser light therapy | Acupuncture (14,273) | Acupuncture (14,280) |
| Manual therapy | Manual therapy (9500) | Acupuncture (14,276) |
| PES | Acupuncture (14,272) | Acupuncture (14,276) |
| PEMFs | Acupuncture (14,276) | Acupuncture (14,276) |
| Static magnets | Acupuncture (14,273) | Acupuncture (14,269) |
| TENS | Acupuncture (15,544) | Acupuncture (13,704) |

When the private cost was used for acupuncture sessions the conclusions of the analysis of all trials were unaffected. The results of the analysis restricted to trials with adequate allocation concealment were affected and TENS became the cost-effective comparator (ICER £6142 per QALY vs. usual care). The results of the analysis restricted to trials with adequate allocation concealment reporting between 3 and 13 weeks were also affected and manual therapy became the cost-effective comparator (ICER £19,942 per QALY vs. TENS).

## Results of sensitivity analysis using shorter treatment durations and a range of assumptions around efficacy

| Relationship between EQ-5D treatment effect and session duration | Cost-effective intervention at threshold of £20,000 per QALY (ICER, £ per QALY) | | |
|---|---|---|---|
| | All trials | Trials with adequate allocation concealment | Trials with adequate allocation concealment, reporting at 3–13 weeks |
| Scenario 1: linear relationship | TENS (2695) | Acupuncture (17,669) | Acupuncture (19,253) |
| Scenario 2: linear up to 60 minutes followed by plateau | TENS (2693) | Acupuncture (17,664) | Acupuncture (19,240) |
| Scenario 3: 75% benefit in 0–30 minutes, remaining benefit 30–60 minutes | Interferential therapy (18,159) | Acupuncture (13,516) | Acupuncture (14,423) |
| Scenario 4: all benefit achieved in first 20/30 minutes | Interferential therapy (4324) | Interferential therapy (8150) | Interferential therapy (8356) |

## Results of sensitivity analysis using upper and lower credible intervals from network meta-analysis

Bold text indicates scenarios in which the optimal decision remains as per the base case (assuming a cost-effectiveness threshold of £20,000 per QALY). Black text indicates scenarios in which the optimal decision switches. The last column indicates the effect size required to alter the decision and the univariate posterior probability of observing a value of the treatment effect this extreme. This was calculated only when use of the upper or lower 95% CrI altered the decision.

### All trials

| Intervention | Cost-effective treatment at threshold of £20,000 per QALY (ICER, £ per QALY) | | Parameter value required to alter decision (probability) |
| | Lower 95% CrI | Upper 95% CrI | |
| --- | --- | --- | --- |
| Acupuncture | **TENS (2691)** | Acupuncture (£19,154) | Effect > 0.12 (0.05) |
| Braces | **TENS (2692)** | Braces (2509) | Effect > 0.08 (0.08) |
| Heat treatment | **TENS (2689)** | **TENS (2692)** | – |
| Insoles | **TENS (2693)** | **TENS (5515)** | – |
| Interferential therapy | **TENS (2689)** | **TENS (2687)** | – |
| Laser/light therapy | **TENS (2690)** | **TENS (2689)** | – |
| Manual therapy | **TENS (2689)** | **TENS (2692)** | – |
| NMES | **TENS (2693)** | NMES (8043) | Effect > 0.22 (0.18) |
| PES | **TENS (2686)** | **TENS (2694)** | – |
| PEMFs | **TENS (2692)** | **TENS (2697)** | – |
| Static magnets | **TENS (2694)** | Static magnets (325) | Effect > 0.07 (0.25) |
| TENS | Acupuncture (15,070) | **TENS (1740)** | Effect < 0.05 (0.15) |

### Trials with adequate allocation concealment

| Intervention | Cost-effective treatment at threshold of £20,000 per QALY (ICER, £ per QALY) | | Parameter value required to alter decision (probability) |
| | Lower 95% CrI | Upper 95% CrI | |
| --- | --- | --- | --- |
| Acupuncture | TENS (6143) | **Acupuncture (9273)** | Effect < 0.08 (0.05) |
| Heat treatment | **Acupuncture (13,507)** | **Acupuncture (13,504)** | – |
| Insoles | **Acupuncture (13,500)** | Insoles (1389) | Effect > 0.05 (0.09) |
| Interferential therapy | **Acupuncture (13,500)** | **Acupuncture (13,500)** | – |
| Laser/light therapy | **Acupuncture (13,499)** | **Acupuncture (13,501)** | – |
| Manual therapy | **Acupuncture (13,501)** | Manual therapy (11,583) | Effect > 0.14 (0.10) |
| PES | **Acupuncture (13,504)** | **Acupuncture (13,500)** | – |
| PEMFs | **Acupuncture (13,502)** | **Acupuncture (13,501)** | – |
| Static magnets | **Acupuncture (13,500)** | Static magnets (335) | Effect > 0.05 (0.29) |
| TENS | **Acupuncture (11,687)** | TENS (2002) | Effect > 0.06 (0.21) |

### *Trials with adequate allocation concealment reporting at 3–13 weeks*

| Intervention | Cost-effective treatment at threshold of £20,000 per QALY (ICER, £ per QALY) | | Parameter value required to alter decision (probability) |
| --- | --- | --- | --- |
| | Lower 95% CrI | Upper 95% CrI | |
| Acupuncture | TENS (9784) | **Acupuncture (9568)** | Effect < 0.09 (0.14) |
| Heat treatment | **Acupuncture (14,271)** | Heat treatment (11,425) | Effect > 0.12 (0.05) |
| Insoles | **Acupuncture (14,274)** | Insoles (676) | Effect > 0.05 (0.35) |
| Interferential therapy | **Acupuncture (14,272)** | **Acupuncture (14,274)** | – |
| Laser/light therapy | **Acupuncture (14,274)** | **Acupuncture (14,272)** | – |
| Manual therapy | **Acupuncture (14,271)** | Manual therapy (8484) | Effect > 0.14 (0.35) |
| PES | **Acupuncture (14,278)** | **Acupuncture (14,277)** | – |
| PEMFs | **Acupuncture (14,272)** | **Acupuncture (14,273)** | – |
| Static magnets | **Acupuncture (14,271)** | Static magnets (315) | Effect > 0.05 (0.30) |
| TENS | **Acupuncture (13,709)** | TENS (1034) | Effect > 0.06 (0.40) |

## Results of sensitivity analysis of extending the duration of treatment benefit by 50%

| Intervention | Cost-effective treatment at threshold of £20,000 per QALY (ICER, £ per QALY) | | |
| --- | --- | --- | --- |
| | All trials | Trials with adequate allocation concealment | Trials with adequate allocation concealment, reporting at 3–13 weeks |
| Standard care | TENS (2693) | Acupuncture (13,500) | Acupuncture (14,276) |
| Acupuncture | Acupuncture (14,732) | Acupuncture (8000) | Acupuncture (8373) |
| Braces | TENS (2691) | NA | NA |
| Heat treatment | TENS (2689) | Acupuncture (13,499) | Acupuncture (14,270) |
| Insoles | TENS (2691) | Acupuncture (13,501) | Acupuncture (16,061) |
| Interferential therapy | Interferential therapy (19,286) | Acupuncture (13,503) | Acupuncture (14,270) |
| Laser/light therapy | TENS (2689) | Acupuncture (13,502) | Acupuncture (14,277) |
| Manual therapy | TENS (2690) | Acupuncture (13,507) | Manual therapy (11,678) |
| NMES | TENS (2689) | NA | NA |
| PES | TENS (2694) | Acupuncture (13,499) | Acupuncture (14,270) |
| PEMFs | TENS (2692) | Acupuncture (13,494) | Acupuncture (14,280) |
| Static magnets | TENS (2697) | Acupuncture (13,505) | Acupuncture (14,269) |
| TENS | TENS (1849) | Acupuncture (16,857) | Acupuncture (18,794) |

NA, not available because of intervention not providing information to network.

# Appendix 6 Trial data related to *Chapter 6*

## Details of acupuncture treatment based on the Revised STandards for Reporting Interventions in Clinical Trials of Acupuncture (STRICTA) reporting guidelines,[483] an official extension to the Consolidated Standards of Reporting Trials (CONSORT)

| STRICTA item | Details of acupuncture treatment within the trial ($n = 266$) |
|---|---|
| 1. (a) Style | Traditional Chinese medicine |
| 1. (b) Reasoning for treatment provided | Systematic Cochrane review[312] and consensus methods[315] were used to develop a treatment protocol in a pragmatic trial designed to evaluate acupuncture as provided in routine care |
| 1. (c) Variation | Individualised treatments using common theoretical frameworks of TCM as reported separately |
| 2. (a) Number of needles per treatment | On average, 13 needles were inserted per session (range 3–26) |
| 2. (b) Names | 246 different points were used. Common points were SP-6, LIV-3, ST-36 and LI-4, which were used within a course of treatment on 91%, 89%, 83% and 74% of patients, respectively |
| 2. (c) Depth of insertion | Range of minimum mean depths was 0.49 cm, maximum 1.41 cm. Total range 0.05–2.5 cm |
| 2. (d) Response sought | The response sought varied, most commonly deqi by 96% of acupuncturists |
| 2. (e) Needle stimulation | Manual methods: tonifying (68%), reducing (43%) and even method (55%) |
| 2. (f) Retention | Average 22.5 minutes (range 15–45 minutes) |
| 2. (g) Needle type | Needle length ranged from 5 mm to 40 mm and needle diameter ranged from 0.13 mm to 0.35 mm |
| 3. (a) Number of sessions | Patients were offered 12 sessions and completed an average of 10.3 sessions (median 12, range 1–12) |
| 3. (b) Frequency and duration | Average number of sessions per week: 0.74. Average number of weeks between sessions: 1.36. Average duration of session: 53 minutes (range 28–95 minutes) |
| 4. (a) Other components of treatment | Acupuncturists were allowed to use moxa, electroacupuncture, ear seeds, cupping, acupressure and heat lamps. Most commonly used were acupressure (used with 13% of patients), moxa (12%) and electroacupuncture (7%). Acupuncturists were allowed to provide acupuncture theory-based lifestyle advice. In total, 66% of patients received lifestyle advice, most commonly diet (42%), exercise (30%) and relaxation (25%) advice. Advice unrelated to acupuncture theory as well as herbs and magnets were proscribed |
| 4. (b) Setting and context | Provision or treatments in independent clinics. Acupuncturists encouraged to practice as closely as possible to their normal practice and 96% reported applying acupuncture exactly the same way or similarly as they did routinely |
| 5. Participating acupuncturists | BAcC members with > 3 years of post-qualification experience |
| 6. Control or comparator interventions | Patients continued to receive usual care as an adjunct from their GP, as well as over-the-counter treatments according to need, based on a need to evaluate the impact of acupuncture plus usual care compared with the impact of usual care alone |

# Details of the competences that underpinned the delivery of humanistic counselling

The counselling was designed to be humanistic in approach and based on relevant competences. These competences were developed independently by Anthony D Roth, Andrew Hill and Stephen Pilling,[316] as commissioned by Skills for Health, supported by an Expert Reference Group and now documented by the Centre for Outcomes Research and Effectiveness at University College London.[484] A manualised protocol was developed using an evidence-based methodology. These competences for the delivery of high-quality humanistic psychological therapies covered the following three domains:

1. *Basic competences in humanistic psychological therapies*. Humanistic counselling interventions are underpinned by a range of more specific humanistic techniques. These basic competences set out a range of activities that humanistically oriented therapists should be able to acknowledge as fundamental to their practice and that should be delivered to patients within the trial.
2. *Metacompetences*. These are the competences needed to implement higher-order links between theory and practice and will adapt the implementation of the humanistic approach to the needs of individual patients. These competences are more abstract than those in other domains; however, they will guide the procedures and practice used by counsellors within the trial.
3. *Specific humanistic psychological therapies competences*. These are technical interventions employed by some (although not all) forms of humanistic psychological therapies. These include, for example, competences related to approaches to working with emotions and emotional meaning.

# Baseline characteristics for patients with and without missing data at 3 months

| Characteristic | Patients with | | Total (*n* = 755) |
| --- | --- | --- | --- |
| | Available data at 3 months (*n* = 616) | Missing data at 3 months (*n* = 139) | |
| Age (years) | | | |
| Mean (SD) | 44.6 (13.01) | 38.2 (13.42) | 43.5 (13.37) |
| Median (min., max.) | 43.5 (18, 89) | 35.0 (18, 93) | 43 (18, 93) |
| Interquartile range | 35–54 | 27–46 | 33–53 |
| Missing, *n* (%) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Sex, *n* (%) | | | |
| Male | 164 (26.6) | 37 (26.6) | 201 (26.6) |
| Female | 452 (73.4) | 102 (73.4) | 554 (73.4) |
| Missing | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Depression, *n* (%) | | | |
| In last 2 weeks | 457 (74.2) | 117 (84.2) | 574 (76.0) |
| Missing | 11 (1.8) | 1 (0.7) | 12 (1.6) |
| Not first major episode | 408 (89.4) | 105 (89.7) | 513 (89.4) |
| Missing | 8 (1.8) | 1 (0.9) | 9 (1.6) |
| Four or more previous episodes | 309 (75.7) | 80 (76.2) | 389 (75.8) |
| Missing | 4 (1.0) | 0 (0.0) | 4 (0.8) |

| Characteristic | Patients with | | |
| --- | --- | --- | --- |
| | Available data at 3 months (*n* = 616) | Missing data at 3 months (*n* = 139) | Total (*n* = 755) |
| Age at first major depressive episode (years) | | | |
|    Mean (SD) | 25.7 (12.5) | 22.9 (10.8) | 25.2 (12.28) |
|    Median (min., max.) | 22 (0, 79) | 20 (3, 78) | 22 (0, 79) |
|    Interquartile range | 16.5–32 | 15–30 | 16–31 |
|    Missing, *n* (%) | 12 (1.9) | 4 (2.9) | 16 (2.1) |
| Medication, *n* (%) | | | |
|    Depression medication in last 3 months | 426 (69.2) | 93 (66.9) | 519 (68.7) |
|    Missing | 0 (0.0) | 0 (0.0) | 0 (0.0) |
|    Analgesic medication in last 3 months | 301 (48.9) | 58 (41.7) | 359 (47.5) |
|    Missing | 4 (0.7) | 2 (1.4) | 6 (0.8) |
| EQ-5D anxiety/depression, *n* (%) | | | |
|    Not anxious/depressed | 17 (2.8) | 4 (2.9) | 21 (2.8) |
|    Moderately anxious/depressed | 465 (75.7) | 89 (64.0) | 554 (73.6) |
|    Extremely anxious/depressed | 132 (21.5) | 46 (33.1) | 178 (23.6) |
|    Missing | 2 (0.3) | 0 (0.0) | 2 (0.3) |
| PHQ-9 score | | | |
|    Mean (SD) | 15.7 (5.32) | 17.3 (4.94) | 16.0 (5.29) |
|    Median (min., max.) | 16 (3, 27) | 17 (6, 27) | 16 (3, 27) |
|    Interquartile range | 12–20 | 14–21 | 12–20 |
|    Missing, *n* (%) | 1 (0.2) | 0 (0.0) | 1 (0.1) |
| PHQ-9 group, *n* (%) | | | |
|    None (0–4) | 6 (1.0) | 0 (0.0) | 6 (0.8) |
|    Mild (5–9) | 78 (12.7) | 9 (6.5) | 87 (11.5) |
|    Moderate (10–14) | 182 (29.6) | 35 (25.2) | 217 (28.8) |
|    Moderately severe (15–19) | 184 (29.9) | 47 (33.8) | 231 (30.6) |
|    Severe (20–27) | 165 (26.8) | 48 (34.5) | 213 (28.2) |
|    Missing | 1 (0.2) | 0 (0.0) | 1 (0.1) |
| BDI-II score | | | |
|    Mean (SD) | 32.0 (8.62) | 34.5 (8.91) | 32.5 (8.72) |
|    Median (min., max.) | 31 (20, 60) | 34 (20, 53) | 31 (20, 60) |
|    Interquartile range | 25–37.5 | 28–41 | 26–38 |
|    Missing, *n* (%) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| BDI-II group, *n* (%) | | | |
|    Moderate (20–28) | 248 (40.3) | 36 (25.9) | 284 (37.6) |
|    Severe (29–63) | 368 (59.7) | 103 (74.1) | 471 (62.4) |
|    Missing | 0 (0.0) | 0 (0.0) | 0 (0.0) |

| Characteristic | Patients with | | Total (*n* = 755) |
| --- | --- | --- | --- |
| | Available data at 3 months (*n* = 616) | Missing data at 3 months (*n* = 139) | |
| SF-36 bodily pain | | | |
|     Mean (SD) | 57.2 (28.54) | 59.6 (28.01) | 57.6 (28.44) |
|     Median (min., max.) | 56.5 (0, 100) | 52 (0, 100) | 52 (0, 100) |
|     Interquartile range | 31–82 | 41–84 | 32–84 |
|     Missing, *n* (%) | 4 (0.6) | 0 (0.0) | 4 (0.5) |
| Treatment preference, *n* (%) | | | |
|     Acupuncture | 343 (55.7) | 87 (62.6) | 430 (57.0) |
|     Counselling | 137 (22.2) | 27 (19.4) | 164 (21.7) |
|     Usual care | 10 (1.6) | 0 (0.0) | 10 (1.3) |
|     No preference | 122 (19.8) | 22 (15.8) | 144 (19.1) |
|     Missing | 4 (0.7) | 3 (2.2) | 7 (0.9) |

Max., maximum; min., minimum.

## Details of the usual care provided to patients in all three arms

(Except medication, which is set out in *Table 38*.)

*Number of patients seeing a health professional and average number of visits (total and specifically for depression) for these patients in the preceding 3 months*

| Health professional | Time point (months) | Intervention Acupuncture + usual care | | | | Counselling + usual care | | | | Usual care | | | | Total | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Patients | | Mean number of visits | | Patients | | Mean number of visits | | Patients | | Mean number of visits | | Patients | | Mean number of visits | |
| | | n | % | Total | Depression[a] | n | % | Total | Depression[a] | n | % | Total | Depression[a] | n | % | Total | Depression[a] |
| GP | 3[b] | 106 | 45.1 | – | 2.2 | 108 | 50.2 | – | 2.3 | 70 | 58.8 | – | 2.2 | 284 | 49.9 | – | 2.2 |
| | 6 | 153 | 73.6 | 2.8 | 0.8 | 134 | 68.0 | 2.5 | 0.8 | 73 | 74.5 | 2.6 | 1.1 | 360 | 71.6 | 2.7 | 0.9 |
| | 9 | 135 | 67.2 | 2.2 | 0.7 | 121 | 66.5 | 2.2 | 0.8 | 81 | 75.7 | 2.3 | 0.9 | 337 | 68.8 | 2.3 | 0.8 |
| | 12 | 127 | 63.2 | 2.4 | 1.0 | 131 | 68.2 | 2.3 | 0.7 | 73 | 69.5 | 2.4 | 0.7 | 331 | 66.5 | 2.4 | 0.8 |
| Practice nurse | 3[b] | 7 | 3.7 | – | 1.4 | 12 | 6.9 | – | 2.0 | 8 | 9.1 | – | 1.8 | 27 | 6.3 | – | 1.8 |
| | 6 | 58 | 26.1 | 1.7 | 0.0 | 63 | 29.9 | 1.5 | 0.1 | 36 | 33.0 | 1.4 | 0.0 | 157 | 29.0 | 1.6 | 0.0 |
| | 9 | 55 | 25.0 | 1.7 | 0.1 | 50 | 25.1 | 1.9 | 0.0 | 28 | 25.5 | 1.8 | 0.3 | 133 | 25.1 | 1.8 | 0.1 |
| | 12 | 54 | 25.0 | 1.6 | 0.0 | 63 | 30.7 | 1.4 | 0.0 | 39 | 35.5 | 1.8 | 0.0 | 156 | 29.4 | 1.6 | 0.0 |
| Outpatient clinic | 3[b] | 11 | 4.6 | – | 2.0 | 13 | 5.9 | – | 2.1 | 9 | 7.4 | – | 3.3 | 33 | 5.7 | – | 2.6 |
| | 6 | 46 | 20.6 | 2.3 | 0.1 | 60 | 28.4 | 2.2 | 0.1 | 25 | 23.1 | 1.3 | 0.0 | 131 | 24.2 | 2.1 | 0.1 |
| | 9 | 50 | 22.4 | 1.8 | 0.2 | 44 | 22.3 | 2.0 | 0.2 | 27 | 25.2 | 1.5 | 0.1 | 121 | 23.0 | 1.8 | 0.2 |
| | 12 | 61 | 28.1 | 1.7 | 0.3 | 57 | 28.4 | 1.8 | 0.2 | 35 | 32.4 | 1.7 | 0.1 | 153 | 29.1 | 1.8 | 0.2 |
| Mental health nurse | 3[b] | 5 | 2.2 | – | 1.6 | 1 | 0.5 | – | 5.0 | 4 | 3.5 | – | 4.5 | 10 | 1.8 | – | 3.1 |
| | 6 | 4 | 1.8 | 4.0 | 2.3 | 1 | 0.5 | 8.0 | – | 2 | 1.8 | 3.5 | 3.5 | 7 | 1.3 | 4.4 | 2.8 |
| | 9 | 7 | 3.2 | 4.1 | 2.9 | 3 | 1.5 | 5.3 | 6.5 | 4 | 3.8 | 3.0 | 1.8 | 14 | 2.7 | 4.1 | 3.1 |
| | 12 | 11 | 5.1 | 3.5 | 2.3 | 4 | 2.1 | 2.3 | 1.8 | 3 | 2.9 | 4.0 | 2.3 | 18 | 3.5 | 3.3 | 2.2 |
| Psychologist or psychiatrist | 3[b] | 4 | 1.8 | – | 1.8 | 5 | 2.4 | – | 6.2 | 6 | 5.4 | – | 2.7 | 15 | 2.7 | – | 3.6 |
| | 6 | 10 | 4.5 | 1.7 | 1.2 | 9 | 4.5 | 4.7 | 1.7 | 7 | 6.5 | 2.9 | 1.6 | 26 | 4.9 | 3.0 | 1.5 |
| | 9 | 7 | 3.3 | 2.3 | 2.1 | 7 | 3.6 | 1.9 | 1.3 | 4 | 3.8 | 5.5 | 5.3 | 18 | 3.5 | 2.8 | 2.5 |
| | 12 | 11 | 5.1 | 3.5 | 3.5 | 7 | 3.7 | 2.9 | 2.0 | 5 | 4.8 | 2.2 | 1.4 | 23 | 4.5 | 3.0 | 2.6 |

| Health professional | Time point (months) | Intervention | | | | | | | | | | | | Total | | | |
| | | Acupuncture + usual care | | | | Counselling + usual care | | | | Usual care | | | | | | | |
| | | Patients | | Mean number of visits | | Patients | | Mean number of visits | | Patients | | Mean number of visits | | Patients | | Mean number of visits | |
| | | n | % | Total | Depression[a] | n | % | Total | Depression[a] | n | % | Total | Depression[a] | n | % | Total | Depression[a] |
| NHS counsellor | 3[b] | 5 | 2.2 | – | 3.0 | 4 | 1.9 | – | 3.0 | 5 | 4.4 | – | 2.4 | 14 | 2.5 | – | 2.8 |
| | 6 | 9 | 4.2 | 2.7 | 1.0 | 5 | 2.5 | 3.2 | 3.8 | 5 | 4.8 | 1.4 | 0.5 | 19 | 3.6 | 2.5 | 1.6 |
| | 9 | 5 | 2.3 | 3.0 | 1.8 | 2 | 1.0 | 3.5 | 3.5 | 4 | 3.9 | 3.8 | 3.0 | 11 | 2.1 | 3.4 | 2.6 |
| | 12 | 10 | 4.7 | 3.8 | 3.1 | 3 | 1.5 | 3.3 | 4.0 | 5 | 5.0 | 2.2 | 2.2 | 18 | 3.5 | 3.3 | 2.9 |
| Other NHS health professional | 3[b] | 16 | 8.9 | – | 3.9 | 10 | 6.1 | – | 2.5 | 15 | 19.0 | – | 3.1 | 41 | 9.7 | – | 3.2 |
| | 6 | 20 | 9.1 | 2.3 | 0.3 | 26 | 12.6 | 2.7 | 0.0 | 14 | 13.3 | 2.1 | 0.0 | 60 | 11.3 | 2.4 | 0.1 |
| | 9 | 24 | 11.2 | 2.8 | 0.6 | 26 | 13.3 | 2.5 | 0.3 | 16 | 14.5 | 2.2 | 0.5 | 66 | 12.7 | 2.5 | 0.4 |
| | 12 | 25 | 11.5 | 2.3 | 0.5 | 32 | 16.2 | 2.7 | 0.0 | 20 | 18.7 | 3.1 | 0.2 | 77 | 14.8 | 2.7 | 0.2 |

a Specifically for depression (out of total).
b Questions at 3 months were phrased differently from those at following time points, asking for responses in relation to depression only.

## Number of patients attending accident and emergency departments and number of attendances in the preceding 3 months

| Time point (months) | Intervention | | | | | | | | | | | | Total | | | |
| | Acupuncture + usual care | | | | Counselling + usual care | | | | Usual care | | | | | | | |
| | Patients | | Number of attendances | | Patients | | Number of attendances | | Patients | | Number of attendances | | Patients | | Number of attendances | |
| | n | % | Mean | SD | n | % | Mean | SD | n | % | Mean | SD | n | % | Mean | SD |
| 3 | 1 | 0.4 | 3.0 | – | 3 | 1.3 | 1.7 | 1.15 | 3 | 2.4 | 1.3 | 0.58 | 7 | 1.2 | 1.7 | 0.95 |
| 6 | 17 | 7.5 | 1.1 | 0.33 | 20 | 8.8 | 1.8 | 2.51 | 11 | 9.7 | 1.1 | 0.30 | 48 | 8.5 | 1.4 | 1.65 |
| 9 | 11 | 4.8 | 1.3 | 0.47 | 13 | 6.3 | 1.4 | 0.96 | 12 | 10.7 | 1.5 | 1.00 | 36 | 6.6 | 1.4 | 0.84 |
| 12 | 21 | 9.2 | 1.2 | 0.40 | 11 | 5.3 | 1.3 | 0.47 | 12 | 10.5 | 1.1 | 0.29 | 44 | 8.0 | 1.2 | 0.39 |

*Number of patients admitted to hospital and number of overnight stays in the preceding 3 months*

| Time point (months) | Intervention | | | | | | | | | | | | Total | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | Acupuncture + usual care | | | | Counselling + usual care | | | | Usual care | | | | | | | |
| | Patients | | Number of overnight stays | | Patients | | Number of overnight stays | | Patients | | Number of overnight stays | | Patients | | Number of overnight stays | |
| | n | % | Mean | SD | n | % | Mean | SD | n | % | Mean | SD | n | % | Mean | SD |
| **Hospital ward** | | | | | | | | | | | | | | | | |
| 3 | 2 | 0.8 | 3.5 | 0.71 | 0 | 0.0 | – | – | 0 | 0.0 | – | – | 2 | 0.3 | 3.5 | 0.71 |
| 6 | 6 | 2.8 | 2.0 | 0.89 | 10 | 4.8 | 5.6 | 6.59 | 2 | 1.8 | 6.0 | 1.41 | 18 | 3.4 | 4.4 | 5.15 |
| 9 | 4 | 1.9 | 3.8 | 3.20 | 4 | 2.0 | 6.8 | 4.79 | 7 | 6.1 | 3.1 | 3.76 | 15 | 2.9 | 4.3 | 3.95 |
| 12 | 6 | 2.7 | 2.0 | 0.89 | 7 | 3.4 | 1.9 | 1.46 | 8 | 7.1 | 2.5 | 2.00 | 21 | 3.9 | 2.1 | 1.53 |
| **Intensive care** | | | | | | | | | | | | | | | | |
| 3 | 0 | 0.0 | – | – | 0 | 0.0 | – | – | 0 | 0.0 | – | – | 0 | 0.0 | – | – |
| 6 | 0 | 0.0 | – | – | 1 | 0.5 | 2.0 | – | 0 | 0.0 | – | – | 1 | 0.2 | 2.0 | – |
| 9 | 0 | 0.0 | – | – | 1 | 0.5 | 4.0 | – | 0 | 0.0 | – | – | 1 | 0.2 | 4.0 | – |
| 12 | 1 | 0.5 | 6.0 | – | 0 | 0.0 | – | – | 0 | 0.0 | – | – | 1 | 0.2 | 6.0 | – |
| **Mental health unit** | | | | | | | | | | | | | | | | |
| 3 | 0 | 0.0 | – | – | 0 | 0.0 | – | – | 1 | 0.8 | 1.0 | – | 1 | 0.2 | 1.0 | – |
| 6 | 1 | 0.5 | 7.0 | – | 1 | 0.5 | 3.0 | – | 0 | 0.0 | – | – | 2 | 0.4 | 5.0 | 2.82 |
| 9 | 0 | 0.0 | – | – | 1 | 0.5 | 3.0 | – | 1 | 0.9 | 18.0 | – | 2 | 0.4 | 10.5 | 10.61 |
| 12 | 0 | 0.0 | – | – | 1 | 0.5 | 2.0 | – | 1 | 0.9 | 21.0 | – | 2 | 0.4 | 11.5 | 13.44 |
| **Other** | | | | | | | | | | | | | | | | |
| 3 | 0 | 0.0 | – | – | 1 | 0.5 | 3.0 | – | 1 | 0.8 | 1.0 | – | 2 | 0.3 | 2.0 | 1.41 |
| 6 | 0 | 0.0 | – | – | 1 | 0.5 | 1.0 | – | 0 | 0.0 | – | – | 1 | 0.2 | 1.0 | – |
| 9 | 2 | 0.9 | 3.5 | 3.54 | 3 | 1.5 | 1.0 | 0.00 | 0 | 0.0 | – | – | 5 | 1.0 | 2.0 | 2.24 |
| 12 | 1 | 0.5 | 7.0 | – | 1 | 0.5 | 1.0 | – | 0 | 0.0 | – | – | 2 | 0.4 | 4.0 | 4.24 |

## Number of patients paying for private health care and money spent in the preceding 3 months

| Time point (months) | Intervention | | | | | | | | Usual care | | | | Total | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | Acupuncture + usual care | | | | Counselling + usual care | | | | Patients | | Money spent (£) | | Patients | | Money spent (£) | |
| | Patients | | Money spent (£) | | Patients | | Money spent (£) | | | | | | | | | |
| | n | % | Mean | SD | n | % | Mean | SD | n | % | Mean | SD | n | % | Mean | SD |
| **Private acupuncture** | | | | | | | | | | | | | | | | |
| 3 | 3 | 1.2 | 60.0 | 21.79 | 1 | 0.4 | 300.0 | – | 2 | 1.6 | 117.5 | 116.67 | 6 | 1.0 | 119.2 | 107.49 |
| 6 | 27 | 11.5 | 96.8 | 73.29 | 3 | 1.3 | 146.7 | 80.83 | 3 | 2.5 | 63.3 | 49.33 | 33 | 5.7 | 98.4 | 72.48 |
| 9 | 27 | 11.6 | 91.5 | 68.32 | 2 | 0.9 | 130.0 | 70.71 | 2 | 1.7 | 750 | 63.64 | 31 | 5.5 | 93.0 | 66.68 |
| 12 | 22 | 9.6 | 108.1 | 68.75 | 2 | 0.9 | 129.5 | 43.13 | 2 | 1.7 | 75.0 | 63.64 | 26 | 4.6 | 107.2 | 65.80 |
| **Private counselling** | | | | | | | | | | | | | | | | |
| 3 | 2 | 0.8 | 162.5 | 53.03 | 3 | 1.3 | 453.3 | 646.71 | 2 | 1.6 | 152.5 | 137.89 | 7 | 1.2 | 284.3 | 409.96 |
| 6 | 4 | 1.7 | 116.7 | 57.74 | 13 | 5.8 | 198.5 | 103.35 | 1 | 0.8 | 130.0 | – | 18 | 3.1 | 180.0 | 98.04 |
| 9 | 0 | 0.0 | – | – | 10 | 4.7 | 162.5 | 112.18 | 0 | 0.0 | – | – | 10 | 1.8 | 162.5 | 112.18 |
| 12 | 3 | 1.3 | 175.0 | 238.48 | 10 | 4.6 | 170.8 | 107.65 | 1 | 0.9 | 65.0 | – | 14 | 2.5 | 164.1 | 132.63 |
| **Psychotherapy** | | | | | | | | | | | | | | | | |
| 3 | 0 | 0.0 | – | – | 7 | 3.2 | 175.7 | 218.26 | 2 | 1.9 | 40.0 | – | 9 | 1.6 | 158.8 | 207.69 |
| 6 | 0 | 0.0 | – | – | 2 | 0.9 | 152.5 | 67.18 | 1 | 0.8 | 160.0 | – | 3 | 0.5 | 155.0 | 47.70 |
| 9 | 0 | 0.0 | – | – | 2 | 10 | 291.5 | 362.75 | 0 | 0.0 | – | – | 2 | 0.4 | 291.5 | 362.75 |
| 12 | 1 | 0.4 | 450.0 | – | 2 | 0.9 | 182.5 | 166.17 | 1 | 0.9 | 65.0 | – | 4 | 0.7 | 220.0 | 189.16 |

## Effect of trial arm on PHQ-9 depression score at 3 months and BDI-II depression score at 12 months (non-imputed data)

| | | Group | | | | | | | |
| | | 1 | | 2 | | | | | |
| | | Mean | SE | Mean | SE | Group difference | | | |
| Analysis | n | Acupuncture | | Usual care | | Mean | SE | 95% CI | p-value |
|---|---|---|---|---|---|---|---|---|---|
| PHQ-9[a] | 376 | 9.8 | 0.35 | 12.4 | 0.48 | −2.58 | 0.598 | −3.76 to −1.40 | < 0.001 |
| BDI-II[b] | 335 | 21.3 | 1.25 | 24.6 | 1.70 | −3.27 | 1.462 | −6.14 to −0.39 | 0.026 |
| | | Counselling | | Usual care | | | | | |
| PHQ-9[a] | 365 | 11.0 | 0.38 | 12.9 | 0.52 | −1.97 | 0.639 | −3.22 to −0.71 | 0.002 |
| BDI-II[b] | 321 | 23.9 | 1.37 | 27.2 | 1.62 | −3.37 | 1.445 | −6.21 to −0.52 | 0.021 |
| | | Acupuncture | | Counselling | | | | | |
| PHQ-9[a] | 485 | 10.0 | 0.36 | 10.7 | 0.37 | −0.65 | 0.516 | −1.67 to 0.36 | 0.207 |
| BDI-II[b] | 354 | 21.7 | 1.21 | 21.1 | 1.37 | 0.62 | 1.300 | −1.95 to 3.18 | 0.636 |

a Effect of trial arm on non-imputed PHQ-9 score at 3 months (ANCOVA, adjusting for baseline PHQ-9 score).
b Effect of trial arm on non-imputed BDI-II score at 12 months (ANCOVA, adjusting for baseline BDI-II score, baseline expectation of counselling and baseline expectation of allocated treatment; for the comparison between acupuncture and counselling, treatment time*3 months and practitioner's empathy were additionally included).
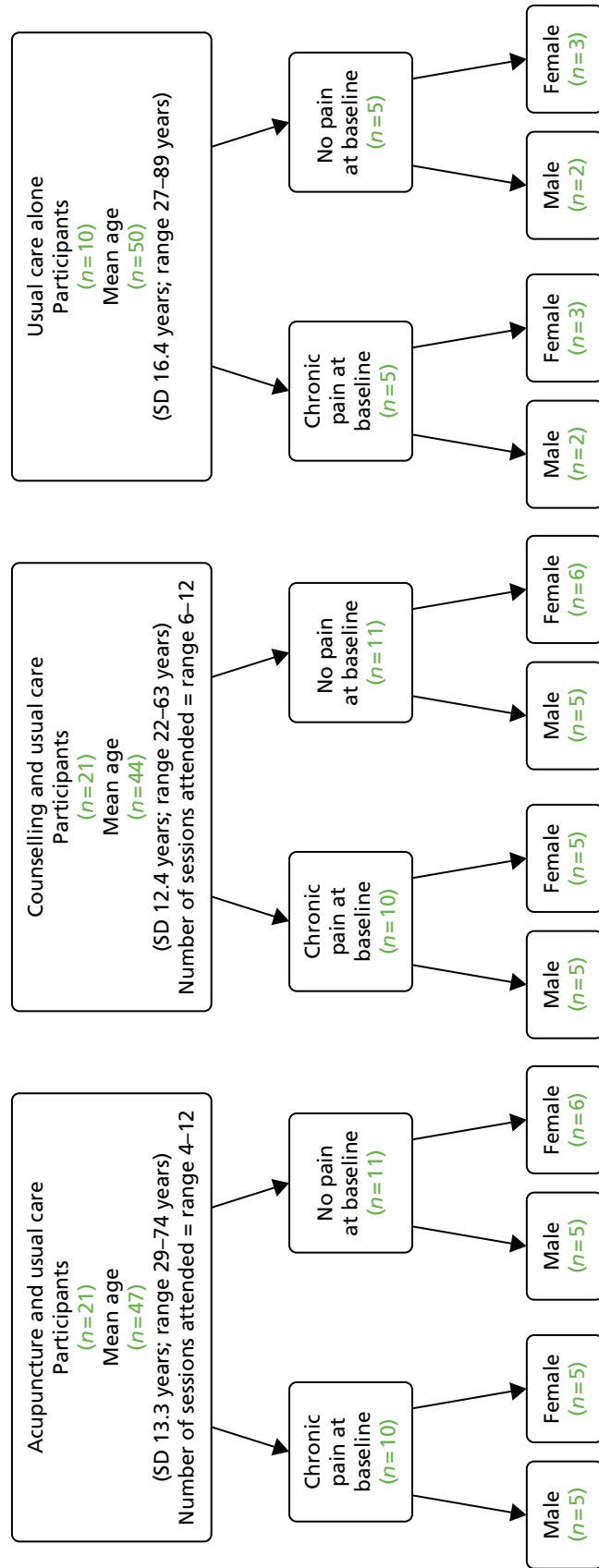
**FIGURE 31** Sampling frame of the patients interviewed in nested substudy 1 (experience of treatment from the patient perspective).

## Demographics and variables of interest at baseline comparing the pain and the no-pain groups in substudy 2

| Characteristic | Group | | Total (n = 755) (100%) |
|---|---|---|---|
| | No pain (n = 371) (49%) | Moderate to extreme pain (n = 384) (51%) | |
| Age (years) | | | |
| Mean (SD) | 39.9 (11.58) | 46.83 (14.11) | 43.5 (13.37) |
| Median (min., max.) | 39 (18, 75) | 46 (18, 93) | 43 (18, 93) |
| Interquartile range (25%, 75%) | 31, 48 | 38, 56 | 33, 53 |
| Missing, n (%) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Sex, n (%) | | | |
| Male | 88 (23.7) | 113 (29.4) | 201 (26.6) |
| Female | 283 (76.3) | 271 (70.6) | 554 (73.4) |
| Missing | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Age left education (years) | | | |
| Mean (SD) | 18.3 (3.90) | 17.53 (4.61) | 18.0 (4.37) |
| Median (min., max.) | 17.0 (13, 52) | 16 (14, 54) | 16 (13-54) |
| Interquartile range (25%, 75%) | 16, 21 | 16, 18 | 16, 19 |
| Missing, n (%) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Employment, n (%) | | | |
| Full-time education | 13 (3.5) | 10 (2.6) | 23 (3.1) |
| Working full-time | 167 (45.0) | 114 (29.7) | 281 (37.2) |
| Working part-time | 83 (22.4) | 61 (15.9) | 144 (19.5) |
| Unable to work | 19 (5.1) | 76 (19.8) | 95 (12.9) |
| Looking after home | 41 (11.1) | 42 (10.9) | 83 (11.0) |
| Retired | 14 (3.8) | 51 (13.3) | 65 (8.8) |
| Other | 26 (7.0) | 22 (5.7) | 48 (6.5) |
| Missing | 8 (2.2) | 8 (2.1) | 16 (2.1) |
| Painful health or medical condition, n (%) | | | |
| Current painful health or medical condition | 41 (12.7) | 299 (77.9) | 346 (45.8) |
| Onset before depression | 32 (8.6) | 216 (56.2) | 248 (32.8) |
| Missing | 2 (0.5) | 4 (1) | 6 (0.8) |
| Type of health problem, n (%) | | | |
| Musculoskeletal | 21 (6) | 221 (58) | 242 (32) |
| Other | 27 (7) | 81 (21) | 108 (14) |
| No health problem | 321 (87) | 81 (21) | 402 (53) |
| SF-36 bodily pain | | | |
| Mean (SD) | 76.97 (22.53) | 39.03 (22.5) | 57.6 (28.44) |
| Median (min., max.) | 77.5 (12, 100) | 41 (0, 100) | 52 (0, 100) |
| Interquartile range (25%, 75%) | 62, 100 | 22, 51 | 32, 84 |
| Missing, n (%) | 3 (0.39) | 1 (1.3) | 4 (0.5) |

| Characteristic | Group | | |
| --- | --- | --- | --- |
| | No pain (*n* = 371) (49%) | Moderate to extreme pain (*n* = 384) (51%) | Total (*n* = 755) (100%) |
| Depression, *n* (%) | | | |
| In last 2 weeks | 264 (71.2) | 310 (80.7) | 574 (76.1) |
| Missing | 2 (0.5) | 10 (2.6) | 12 (1.6) |
| Not first major episode | 240 (64.7) | 273 (71.1) | 513 (67.9) |
| Missing | 108 (29.1) | 82 (21.3) | 190 (25.2) |
| Four or more previous episodes | 162 (43.7) | 227 (59.1) | 389 (51.5) |
| Missing | 133 (35.8) | 113 (29.4) | 246 (32.6) |
| Age at first major depressive episode | | | |
| Mean (SD) | 24.89 (11.56) | 25.47 (12.9) | 25.2 (12-28) |
| Median (min., max.) | 21 (6, 72) | 23 (0, 79) | 22 (0, 79) |
| Interquartile range (25%, 75%) | 16, 32 | 16, 31 | 16, 31 |
| Missing, *n* (%) | 7 (0.9) | 9 (1.2) | 16 (2.1) |
| Medication, *n* (%) | | | |
| Depression medication in last 3 months | 250 (67.3) | 269 (70) | 519 (68.7) |
| Missing | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Analgesic medication in last 3 months | 114 (30.7) | 245 (63.8) | 359 (47.5) |
| Missing | 4 (1) | 2 (0.52) | 6 (0.8) |
| EQ-5D pain, *n* (%) | | | |
| No pain | 371 (49.1) | 0 (0.0) | 371 (49.1) |
| Moderate pain or discomfort | 0 (0.0) | 298 (77.6) | 298 (39.5) |
| Extreme pain or discomfort | 0 (0.0) | 86 (22.4) | 86 (11.4) |
| Missing | 3 | 4 | 7 (0.9) |
| EQ-5D anxiety/depression, *n* (%) | | | |
| Not anxious/depressed | 13 (3.5) | 8 (2.1) | 21 (2.8) |
| Moderately anxious/depressed | 282 (76.0) | 272 (70.8) | 554 (73.4) |
| Extremely anxious/depressed | 74 (19.9) | 104 (27.1) | 178 (23.6) |
| Missing | 2 (0.5) | 0 (0.0) | 2 (0.3) |
| EQ-5D mobility, *n* (%) | | | |
| No problems walking about | 361 (31) | 234 (31) | 595 (79.0) |
| Some problems walking about | 9 (1.2) | 148 (19.7) | 157 (20.8) |
| Confined to bed | 0 (0) | 1 (0.1) | 1 (0.1) |
| Missing | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| EQ-5D self-care, *n* (%) | | | |
| No problems with self-care | 358 (47.5) | 314 (41.7) | 672 (89.2) |
| Some problems with self-care | 11 (1.5) | 68 (9) | 79 (10.5) |
| Unable to wash or dress | 1 (0.1) | 1 (0.1) | 2 (0.3) |
| Missing | 1 (0.1) | 1 (0.1) | 2 (0.3) |

| Characteristic | Group | | Total (*n* = 755) (100%) |
|---|---|---|---|
| | No pain (*n* = 371) (49%) | Moderate to extreme pain (*n* = 384) (51%) | |
| EQ-5D usual activities, *n* (%) | | | |
| No problems performing usual activities | 251 (33.2) | 111 (14.7) | 362 (48.1) |
| Some problems performing usual activities | 115 (15.3) | 247 (32.8) | 115 (48.1) |
| Unable to perform usual activities | 5 (0.7) | 24 (3.2) | 29 (3.9) |
| Missing | 2 (0.3) | 0 (0) | 2 (0.3) |
| PHQ-9 score | | | |
| Mean (SD) | 14.97 (5.15) | 17.0 (5.23) | 16.0 (5.29) |
| Median (min., max.) | 15 (3, 27) | 17 (3, 27) | 16 (3-27) |
| Interquartile range (25%, 75%) | 11, 19 | 13, 21 | 12, 20 |
| Missing, *n* (%) | 0 (0.0) | 1 (0.2) | 1 (0.1) |
| BDI-II score | | | |
| Mean (SD) | 31.3 (8.3) | 33.6 (8.9) | 32.5 (8.7) |
| BDI-II group, *n* (%) | | | |
| Moderate (20–28) | 155 (41.8) | 129 (33.6) | 284 (37.6) |
| Severe (29–63) | 216 (52.8) | 255 (66.4) | 471 (62.4) |
| Trial arm allocation, *n* (%) | | | |
| Acupuncture | 156 (20.7) | 146 (19.3) | 302 (40) |
| Counselling | 151 (20) | 151 (20) | 302 (40) |
| Usual care | 64 (8.5) | 87 (11.5) | 151 (20) |
| Missing | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Expectation of treatment allocated, *n* (%) | | | |
| Very ineffective | 13 (3.5) | 21 (9.6) | 34 (4.5) |
| Fairly ineffective | 39 (10.5) | 57 (14.8) | 96 (12.8) |
| Can't decide | 159 (42.9) | 178 (46.4) | 337(44.9) |
| Fairly effective | 122 (32.9) | 87 (22.7) | 209 (27.8) |
| Very effective | 38 (10.2) | 37 (9.6) | 75 (10.0) |
| Missing | 0 (0.0) | 4 (0.5) | 4 (0.5) |
| Treatment preference, *n* (%) | | | |
| Acupuncture | 196 (52.8) | 234 (60.9) | 430 (57.5) |
| Counselling | 94 (25.3) | 70 (18.2) | 164(21.9) |
| Usual care | 3 (0.8) | 7 (1.8) | 10 (1.3) |
| No preference | 76 (20.5) | 68 (17.7) | 144 (19.3) |
| Missing | 2 (0.5) | 5 (1.3) | 7 (0.9) |

Max., maximum; min., minimum.

## Details of acupuncturists and counsellors participating in an interview or focus group in nested substudy 3

| Group | Number involved in | | | Total number participating |
| --- | --- | --- | --- | --- |
| | Main trial | One-to-one interviews | Focus group | |
| Acupuncturists | 23 | 13 | 4 | 17 |
| Counsellors | 41 | 15 | 4 | 19 |
| Total | 64 | 28 | 8 | 36 |

## Unit costs 2012

| Resource | Unit cost (£) | Description | Source |
| --- | --- | --- | --- |
| Depression medication | 0.047 | Per dose of fluoxetine | BNF[485] |
| Pain medication | 0.330 | Per dose of co-codamol | BNF[485] |
| GP | 43 | Appointment of 11.7 minutes | Curtis 2012[291] |
| Practice nurse | 22 | Band 5, appointment of 15.5 minutes | Curtis 2012[291] |
| Other health professional | 43 | Appointment of 11.7 minutes | Curtis 2012[291] |
| NHS hospital outpatient clinic | 131 | Community setting mental health services | Curtis 2012[291] |
| Hospital ward | 254 | Non-elective inpatient (long stay) excess bed day HRG data | NHS Reference Costs 2011–2012[486] |
| Hospital ICU | 654 | Mental health services inpatient attendances, intensive care – adult | Curtis 2012[291] |
| Hospital mental health unit | 338 | Weighted average of all adult mental health inpatient bed-days | Curtis 2012[291] |
| Accident and emergency | 122 | Weighted average of 'leading to admission' and 'not leading to admission' | NHS Reference Costs 2011–2012[486] |
| Community mental health nurse | 85 | Band 5, per hour of client contact | Curtis 2012[291] |
| Psychologist or psychiatrist | 136 | Band 8a, per hour of client contact | Curtis 2012[291] |
| NHS counsellor | 65 | Band 7, per hour of client contact | Curtis 2012[291] |
| Acupuncturist (initial session) | 47.5 | Average of initial sessions costs of £35 and £60 | MacPherson et al.[326] |
| Acupuncturist (further session) | 37.5 | Average of further session costs of £25 and £50 | MacPherson et al.[326] |

BNF, *British National Formulary*; HRG, Healthcare Resource Group; ICU, intensive care unit.

**EME
HS&DR
HTA
PGfAR
PHR**

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

**Published by the NIHR Journals Library**