



## Practice, practitioner, or placebo? A multifactorial, mixed-methods randomized controlled trial of acupuncture

Peter White<sup>a,\*</sup>, Felicity L. Bishop<sup>b</sup>, Phil Prescott<sup>c</sup>, Clare Scott<sup>b</sup>, Paul Little<sup>b</sup>, George Lewith<sup>b</sup>

<sup>a</sup> Faculty of Health Sciences, University of Southampton, Highfield, Southampton, UK

<sup>b</sup> School of Medicine, Primary Medical Care, Aldermoor Health Centre, University of Southampton, Southampton, UK

<sup>c</sup> School of Mathematics, University of Southampton, Highfield, Southampton, UK

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

### ARTICLE INFO

#### Article history:

Received 1 July 2011

Received in revised form 26 October 2011

Accepted 4 November 2011

Available online xxx

#### Keywords:

Acupuncture

Mixed methods

Nonspecific effects

Placebo

RCT

### ABSTRACT

The nonspecific effects of acupuncture are well documented; we wished to quantify these factors in osteoarthritic (OA) pain, examining needling, the consultation, and the practitioner. In a prospective randomised, single-blind, placebo-controlled, multifactorial, mixed-methods trial, 221 patients with OA awaiting joint replacement surgery were recruited. Interventions were acupuncture, Streitberger placebo acupuncture, and mock electrical stimulation, each with empathic or nonempathic consultations. Interventions involved eight 30-minute treatments over 4 weeks. The primary outcome was pain (VAS) at 1 week posttreatment. Face-to-face qualitative interviews were conducted (purposive sample, 27 participants). Improvements occurred from baseline for all interventions with no significant differences between real and placebo acupuncture (mean difference  $-2.7$  mm, 95% confidence intervals  $-9.0$  to  $3.6$ ;  $P = .40$ ) or mock stimulation ( $-3.9$ ,  $-10.4$  to  $2.7$ ;  $P = .25$ ). Empathic consultations did not affect pain ( $3.0$  mm,  $-2.2$  to  $8.2$ ;  $P = .26$ ) but practitioner 3 achieved greater analgesia than practitioner 2 ( $10.9$ ,  $3.9$  to  $18.0$ ;  $P = .002$ ). Qualitative analysis indicated that patients' beliefs about treatment veracity and confidence in outcomes were reciprocally linked. The supportive nature of the trial attenuated differences between the different consultation styles. Improvements occurred from baseline, but acupuncture has no specific efficacy over either placebo. The individual practitioner and the patient's belief had a significant effect on outcome. The 2 placebos were equally as effective and credible as acupuncture. Needle and nonneedle placebos are equivalent. An unknown characteristic of the treating practitioner predicts outcome, as does the patient's belief (independently). Beliefs about treatment veracity shape how patients self-report outcome, complicating and confounding study interpretation.

© 2011 International Association for the Study of Pain. Published by Elsevier B.V. All rights reserved.

#### Trial Registration:

UKCRN study ID: 4881

ISRCTN78434638

### 1. Introduction

All complex therapeutic interventions achieve their outcomes through a combination of specific and nonspecific mechanisms, including the patient–therapist relationship, treatment process and rituals, expectation, suggestibility, conditioning, and practitioner bias [1,11]. Acupuncture studies show very large improvements from pretreatment to posttreatment [2,28], with mean improvements of 80% for symptom severity (across painful pathologies) and 62% for pain [6]. There is a significantly greater benefit

from acupuncture than from standard conventional care [5], with acupuncture “placebos” routinely outperforming conventional care [3,12,25]. This suggests that the effectiveness of acupuncture is not point- or therapy-specific.

There is considerable interest in maximising treatment effects through understanding the placebo more generically [10,16]. Acupuncture's effectiveness coupled with marginal proven efficacy implies that contextual factors may have a profound influence on outcome. Kaptchuk et al. suggest that acupuncture needles create enhanced placebo effects [15], implying that different placebo rituals may demonstrate different relative benefits.

The aims of this study were to investigate whether there is an enhanced nonspecific effect associated with needling, to ascertain the effects of the consultation process and the practitioner, to evaluate the efficacy of acupuncture on severe osteoarthritic (OA) pain, and to enhance interpretation of the quantitative study through nested qualitative work.

\* Corresponding author. Address: Faculty of Health Sciences, University of Southampton, Highfield, Southampton SO17 1BJ, UK. Tel.: +44 (0) 238059 8308.

E-mail address: [pjw1@soton.ac.uk](mailto:pjw1@soton.ac.uk) (P. White).

## 2. Materials and methods

A randomised, single-blind (patient), multifactorial trial was conducted involving 3 interventions, 2 consultation types, and 3 practitioners with a nested qualitative study to elucidate the quantitative findings. Recruitment was via joint replacement waiting lists at Southampton General and Salisbury District Hospitals (Table 1).

### 2.1. Randomisation

After informed consent and 1-week baseline pain recording, a 2-stage randomisation occurred via an independent third party using a computer-generated list. Patients were first randomised to treatment type and consultation type, and then to specific practitioners.

### 2.2. Interventions

#### 2.2.1. Real acupuncture

A Western acupuncture approach was used involving a flexible but prescribed range of points. Practitioners were free to pick appropriately from the list as clinically indicated. A mean of 6 points was used at each treatment, with deep needling, which lasted for 20 minutes during the 30-minute appointment twice per week for 4 weeks. Deqi (needle sensation) was elicited for each needle through needle rotation. The acupuncture treatment was designed to provide high-quality Western acupuncture of the type commonly used in UK clinics.

#### 2.2.2. Placebo acupuncture

The format was exactly the same as those for real acupuncture (RA), but using Streitberger needle (SN) nonpenetrating needles instead of RA needles. These work rather like a stage dagger, have been validated, and patients cannot distinguish them from RA needles [27].

#### 2.2.3. Mock electrical stimulation

An electroacupuncture stimulator was used (NOMA Ltd., Southampton, UK) to provide mock transcutaneous electrical stimulation to acupoints via electrodes fixed to the surface of the patient's skin. The cables were disconnected inside the output plug. All aspects of the intervention were exactly the same as with RA and SN, including time of treatment and checking on patients at regular intervals. This control has previously been well validated [22].

### 2.3. Consultation types

#### 2.3.1. Empathic

Empathic (EMP) consultations were deemed to be normal pragmatic treatment sessions. Patients were greeted in a friendly, warm manner and were free to enter into conversation with their

practitioner, who in turn would willingly do so. Practitioners did their utmost to comply with participants' wishes, providing detailed answers to questions and emphasising patient comfort and well-being.

#### 2.3.2. Non-EMP

This encounter was more "clinical" in nature. Patients were greeted in an efficient manner and quietly shown to the treatment cubicle. Practitioners would only discuss matters directly relating to the treatment to enable them to effectively carry out that treatment, e.g., pattern of pain and side effects. Necessary explanations were kept as short as possible, and if patients attempted to enter into any discussion, the practitioner would respond using the words "I'm sorry but because this is a trial I am not allowed to discuss this with you." Between needle stimulations, patients were left on their own in a curtained cubicle.

#### 2.3.3. Acupuncturists

Three qualified, experienced (range 3 to 10 years) practitioners were available and funded to provide treatments (physiotherapist, nurse, and licensed acupuncturist). They met frequently throughout the trial to ensure that treatments, including the acupoints used, and consultation types were as comparable and equivalent as possible.

### 2.4. Protocol/procedure

Patients were assessed and gave consent (including consent to differing consultation types). Ethics approval was gained from the Southampton and South West Hampshire and the Salisbury and South Wiltshire Research ethics committees (approval number 170/03/t).

The information sheet explained that "a proportion of patients will receive placebo (dummy) treatment and so the acupuncture points will not be stimulated; i.e., the needle will not penetrate your skin or the machine will not deliver any current."

Patients were given a daily pain diary (100 mm visual analogue scale VAS) to complete for 7 (pretreatment) days, which included recording their analgesia. A minimum mean weekly score of 30 of 100 was an inclusion requirement. Patients were then randomised and treated. Throughout treatment, patients recorded their weekly pain and analgesia with a final extra weeks' diary after treatment completion.

Face-to-face open-ended narrative qualitative interviews were conducted with 27 randomized controlled trial (RCT) participants purposively sampled to obtain representation from all treatment groups (EMP and non-EMP, all 3 interventions and the 2 practitioners not conducting the qualitative study), and both responders and nonresponders. Interviews took place 4 to 8 weeks after treatment completion. There was no topic guide, apart from the opening question in which participants were asked to talk about their experiences of taking part in the trial. Interviews lasted between 30 and 120 minutes (typically 60 minutes), and were audio-taped, transcribed, and annotated with the interviewer's field notes. Summaries were sent to interviewees to provide a member check on the analysis. Patients were not interviewed by their treating acupuncturist (Table 2).

### 2.5. Outcomes

Pain measured on a 100-mm VAS in the 7 days immediately after treatment completion. The primary outcome was the differences between groups in percentage change in pain. Differences of >30% improvement from baseline were defined as clinically important [8,9]. Secondary outcomes were Nottingham Health Profile (part 1) and Western Ontario and McMaster University Osteoarthritis Index and pain over time (weekly VAS).

**Table 1**  
Inclusion/exclusion criteria.

Inclusion criteria	Exclusion criteria
Age 18 to 80 years	Pregnancy
Suffering chronic osteoarthritic pain from a single joint (hip or knee)	Serious comorbidity (including severe back pain)
Awaiting joint replacement surgery	History of prolonged or current steroid use
Having a mean score of $\geq 30$ mm during the baseline week (7 daily recordings) on a 100-mm visual analogue pain scale	Awaiting hip/knee revision (i.e., current prosthesis)
Not on any current physical treatment (e.g., physiotherapy)	Needle phobia
	Allergy to sticking plaster

## 2.6. Potential confounders

To assess potential confounding, we recorded treatment credibility [7], attitudes to complementary medicine holistic complementary and alternative medicine questionnaire (HCAMQ) [13], empathy consultation and relational empathy (CARE) questionnaire [19], analgesic intake (tablet count), and needling sensation [20]. Patients were asked at treatment completion, “Do you think the treatment you had was real” and required to give a yes or no answer.

## 2.7. Sample size and analyses

Assuming a 13% dropout, 96 subjects were required per arm for 80% power to detect a 20% difference between each group on VAS (assuming SD = 38 mm) for our primary outcome. To avoid an accumulated type I error from multiple testing, we required a *P* value of .01 or less, creating a sample size of 288.

An analysis of covariance (ANCOVA) for pain (VAS) estimated main effects and interactions between factors and controlled for baseline pain, stratifiers, and confounders. If interactions were not found, the main effects with confidence intervals were reported. Missing data were dealt with by either taking a mean of the values on either side of the missing value or by carrying forward the last entry as appropriate. Our per-protocol analysis examined needling sensation, belief in treatment, and attitudes to complementary medicine, as well as treatment type, the empathetic nature of the consultation, and practitioner.

Framework analysis [24] incorporating deductive and inductive elements was used to relate the qualitative data to the quantitative trial. An index was used by 2 authors (F.B., G.L.) to identify talk that related to the 4 main predetermined elements of the framework: outcomes, intervention characteristics, practitioners, and empathy. Talk that related to each of these themes was explored inductively to identify subthemes. Predetermined comparisons were then made across intervention groups by examining how each subtheme manifested in the different participants' accounts. These comparisons suggested consistent explanations for the main RCT findings. Table 3 shows how the RCT findings were used to determine the 4 key themes and how data mapped to themes, and were explored inductively to identify subthemes. The findings are presented with illustrative quotes (using pseudonyms) selected as typical and/or particularly clear examples.

## 3. Results

Patients (279) were assessed and completed baseline pain recordings (01/2004 to 08/2007) with 221 patients recruited (Fig. 1). Eleven patients (4.9%) dropped out (Table 4); 3 completed some pain scores, and their data were carried forward per intention-to-treat analysis and included in the final analysis. Eight did not complete any outcomes and were excluded from analysis.

### 3.1. Demographics and baseline measurements

Age (mean = 66.75 years, SD = 8.29) and sex were balanced across all treatments, consultation types, and practitioners; 57.5%

were female; 59.7% knee joints were treated. Table 5 shows baseline data and mean posttreatment scores.

The ANCOVA showed that treatment credibility and HCAMQ scores had no effect on outcome, implying that equipoise between groups was achieved. The final empathy questionnaire demonstrated statistically significant differences between consultation types, with mean scores of 45.73 (95% confidence interval [CI] 44.73 to 46.73) (empathic) and 36.38 (95% CI 34.34 to 38.42, Mann–Whitney *U*, *P* < .001) (nonempathic) demonstrating that this factor was delivered effectively. This difference in empathy for the CARE score, based on available data, also demonstrates a clinically significant effect for the EMP and non-EMP consultation types [18]. Strength of needle sensation was not correlated with pain relief (Spearman correlation coefficient .121, *P* = .083).

### 3.2. Primary outcome—pain

Pain diminished for all groups, with only RA achieving clinically relevant improvements at 29.5% (Table 5). SN reduced pain by 23.0% and mock electrical stimulation by 16.6%. Those allocated to nonempathic consultations achieved slightly (nonsignificant) greater analgesia than empathic consultations (25.2% vs 21.1%; mean VAS difference 3.00 mm, 95% CI −2.2 to 8.2, *P* = .26).

ANCOVA at week 5 showed no significant effect on immediate posttreatment pain of the factors practitioner, treatment, and consultation type, and their first-order interactions with all of the possible baseline confounders as covariates (age, sex, joint, previous knowledge of acupuncture, credibility, CARE score, and HCAMQ scores). The ANCOVA was repeated with the insignificant covariates removed. Group differences are shown in Table 6. Patients who thought that their treatment was real recorded pain scores 11.5 mm (95% CI 3.4 to 19.5, *P* = .005) lower than those who did not. Treatment type and consultation type had no significant effect on posttreatment pain (Table 6). There was a significant practitioner effect. Practitioner 3 achieved greater analgesia than practitioner 2 (mean difference 10.9 mm, 95% CI 3.9 to 18.0, *P* = .002) across all treatments and consultation types. This also neared significance against practitioner 1 (mean difference 5.8 mm, 95% CI −.1 to 11.7, *P* = .05).

### 3.3. Secondary outcomes

Pain over time decreased for all treatments (Fig. 2). This was assessed using a repeated-measures ANCOVA and showed that there was a highly significant downward trend in pain scores over time, with an average reduction of 3.4 mm in pain scores per week on the VAS scales (*P* < .001), but there was not a significant difference between the downward trends for the different treatments. All scores for Western Ontario and McMaster University Osteoarthritis Index and Nottingham Health Profile improved in all groups, with no significant differences between treatment or consultation types. Practitioner 3 achieved the largest improvements.

### 3.4. Blinding

A high percentage of patients believed that their treatment had been real; 96% (RA), 93% (SN), and 75% (mock electrical stimula-

**Table 2**  
Interviewees from each condition in the randomized controlled trial.

	Real acupuncture	Streitberger needles	Mock electrical stimulation	Total
Practitioner 1 empathic	Hester, Emily, Colin	Helen, Sidney	David	6
Practitioner 1 non-empathic	Ursula, Norman	Susan	Dorothy, Betty, Stella, Elizabeth, Beatrice, Arthur	9
Practitioner 3 empathic	Lillian, John	Linda, Melanie, Cynthia, Roy	Agnes, Peter	8
Practitioner 3 non-empathic	Martin	Wendy	Marion, Margaret	4
Total	8	8	11	27

**Table 3**  
The relationship between RCT findings, framework analysis, and inductive thematic analysis.

Framework analysis		Inductive thematic analysis	
RCT finding	Deductive framework	Inductive themes	Comparisons
No effect of intervention condition, and effect of perceiving treatment to be real	Outcomes	Nature of outcomes Confidence in outcomes Reporting outcomes	Treatment Condition (MES vs RA vs SN)
	Treatment characteristics	Nature of treatment Past experience of treatment Confidence in treatment veracity	Treatment Condition (MES vs RA vs SN)
Effect of practitioner	Practitioners	Personal characteristics Interpersonal skills Technical skills Name	Practitioner (1 vs 2 vs 3)
No effect of consultation type	Empathy	Feeling cared for and/or respected within the consultation Feeling cared for and/or respected within the trial	Consultation Type (EMP vs non-EMP)

EMP, empathic consultation; MES, mock electrical stimulation; RA, real acupuncture; RCT, randomized controlled trial; SN, Streitberger needle.

tion), which suggests that blinding was successful. These differences between treatments were significant ( $\chi^2 = 15.486$ ,  $P = .001$ ). Three adverse events were recorded, none of which were related to treatment. There were 28 cases of minor side effects (Table 7).

### 3.5. Qualitative findings

The qualitative analysis provides insight into how perceptions of treatment veracity influenced outcomes, why practitioner 3 was more effective, and why consultation type had no effect on posttreatment pain.

#### 3.5.1. Treatment veracity

Interviewees expressed uncertainty regarding the veracity of their treatment, which was accompanied by uncertainty regarding subjective outcomes. Beliefs about treatment veracity and confidence in outcomes were reciprocally linked. As people became more confident that they were experiencing benefits, such as decreased pain, they also became more confident that they were receiving real acupuncture. As they became more confident that they were receiving real acupuncture, they also became more confident that they were experiencing positive health changes. Confidence in health changes in turn impacted how participants completed the quantitative self-report outcome measures.

*The only thing that you do think about is—you're having your treatment and you're not sure if you are having something that is supposed to do something or whether you're just having something that's just sticking needles in you. . . And you're always a bit hesitant to say 'well I think it's doing me good' because you've got that feeling that perhaps what you're having done isn't. . . you know."* (Martin, receiving RA)

#### 3.5.2. The practitioner effect

Interviewees described practitioners in predominantly positive ways, related to personality (friendly, nice, understanding) and technical skills (being very good at giving treatments, and as offering clear explanations). One difference was identified: terminology. Interviewees referred to Practitioner 1 by her first name and as a "girl" and a "young lady," and some described her using affectionate terms such as "sweet." Practitioner 3 was referred to as "Doctor," was never referred to by his first name, and was typically described in more respectful than affectionate terms, including "courteous" and "formal but friendly." Participants seem to have seen practitioner 3 as more authoritative than practitioner 1.

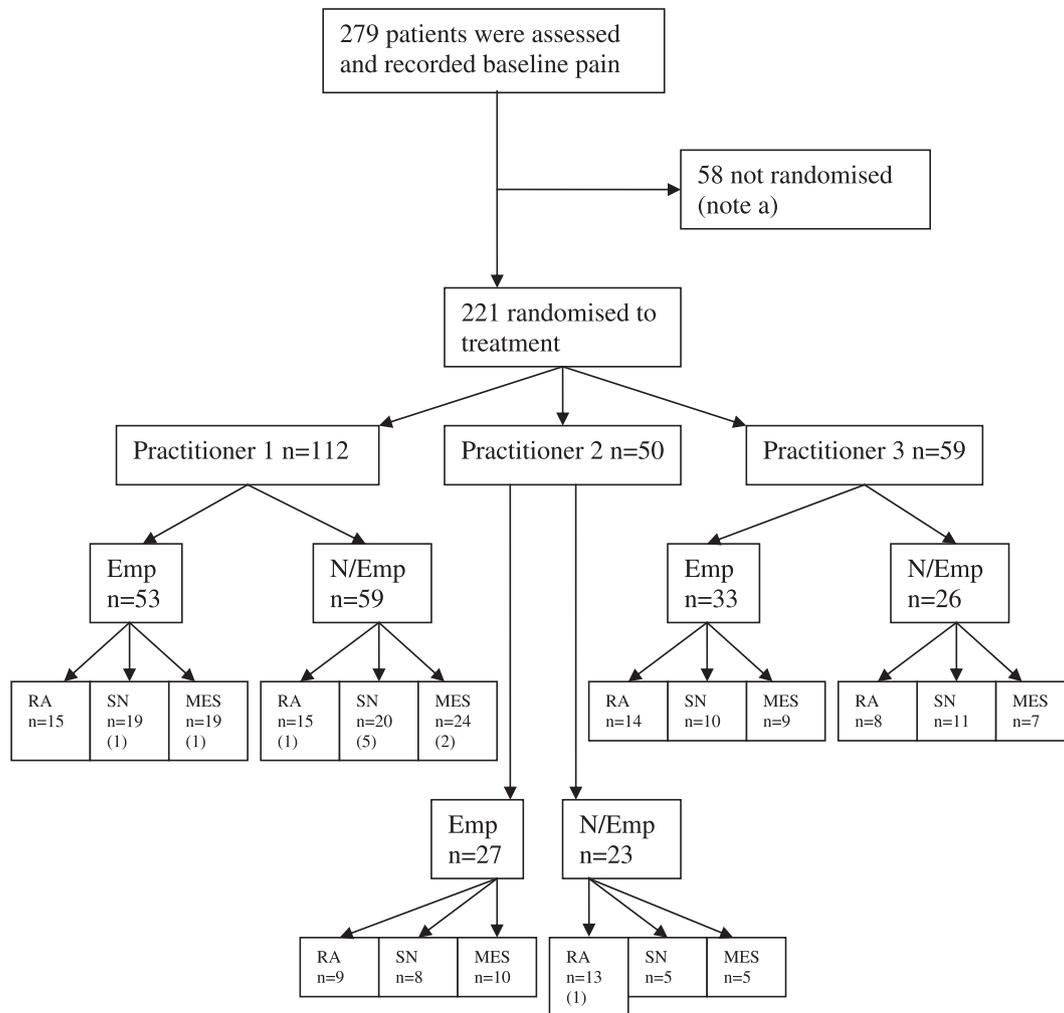
#### 3.5.3. Consultation type

Participants in empathic consultations described practitioners as caring, friendly, and communicative. Those in nonempathic consultations undertook a little more work to explain their similarly positive views of their practitioners. Thus interviewees who had received nonempathic consultations talked about how they colluded with the practitioner to obey the study rules and have limited personal interactions. They suggested that the practitioners were not really nonempathic, they were just acting that way for the sake of the trial. For example, "I had the feeling that she sort of felt that you know, not being able to converse properly, that she felt a bit awkward about it" (Betty, nonempathic). Other aspects of the trial beyond the consultation (friendly and polite reception staff, convenient appointments) were also described positively as evidence that the researchers and practitioners did actually care about the interviewees and so the interviewees described feeling "cared for" by all those involved in the study.

Dorothy was the only interviewee to describe her experiences of being in the trial in general and her experiences of her (nonempathic) practitioner in negative terms: "in the end I decided that she couldn't be a practitioner, she was just being told where to put these things and was bored stiff with the job but needed the job." Dorothy described not knowing that her practitioner's behaviour was part of the trial (despite being fully informed at the recruitment interview) and was also the interviewee for whom the nonempathic consultation could be described as most successful. If she had recalled her recruitment interview, the nonempathic consultation may have lost its potency: "if I had this young lady and I'd known that some of you were going to be just businesslike and some are going to be friendly, I would have thought I had got one of the businesslike ones and brushed it off. But not to know was I think, wrong."

## 4. Discussion

All groups improved from baseline, with many achieving clinically significant improvements. There were no differences between verum and placebo treatments. The empathic nature of the consultation process had no effect on outcome, but believing in the treatment and a powerful practitioner effect significantly affected pain at 5 weeks. All primary and secondary outcomes showed similar trends. The qualitative study helps to explain these findings, suggesting that beliefs about treatment veracity and confidence in outcomes were reciprocally linked and the supportive nature of the study and consent process may have attenuated differences in outcomes between the different consultation styles. In this



**Fig. 1.** consolidated standards of reporting trials (CONSORT) diagram. Key: Figures in brackets indicate patients withdrawn from treatment. Emp = Empathic, N/Emp = non empathic, RA = real acupuncture, SN = Streitberger needle, MES = mock electrical stimulation. Note a 58: 21 – too little pain; 9 – called for operation; 6 – flare up of comorbidity; 5 – failed to complete diary; 4 – did not attend; 3 – awaiting test results; 3 – changed mind; 7 – ‘other’ from ms.

**Table 4**  
Reasons for withdrawals.

	Treatment	Empathic (n)	Nonempathic (n)	Total	Reason for withdrawal (n)
Practitioner 1	Acupuncture	0	1	1	Work commitment (1)
	Streitberger needle	1	5	6	Not improving (1) Transport problem (2) Pain exacerbation (1) Time commitment (1)
	Mock electrical stimulation	1	2	3	Called for surgery (1) Called for surgery (2) Flare up of comorbidity (1)
Practitioner 2	Acupuncture	0	1	1	Pain exacerbation (1)
Total		2	9	11	

population, an authoritative practitioner seemed to be more effective across all outcomes.

The methodology is rigorous, with particular focus on the analysis of potential confounders. All groups were comparable at entry, and treatment credibility ratings confirm that each treatment was considered equally likely to be effective (i.e., there was no a priori bias). Outcome was not confounded by pre-existing beliefs or previous acupuncture treatment. Most patients in all groups believed their treatment had been real. The nested qualitative study offers important and unique insights into the mechanisms involved in delivering and interpreting an RCT as well as novel explanatory hypotheses for our main quantitative findings.

A type II error is unlikely; there was underrecruitment but dropouts were approximately half that anticipated, so the main-effect comparisons were adequately powered. Using the smallest groups, the absolute difference between the mean pain improvement for practitioners 2 and 3 was 13 mm, which is less than our predefined clinically important difference of 20 mm, and corresponds to 80% power at 1% significance. Other main comparisons were based on much larger group sizes. Type I error is possible for the practitioner effect, but the level of significance and the supportive findings from the qualitative study make this unlikely.

This study did not employ a no-treatment arm, and therefore it is possible that improvements recorded might simply be due to the

**Table 5**  
Mean pretreatment and posttreatment scores for prime outcome by group.

	RA	SN	MES	Practitioner 1	Practitioner 2	Practitioner 3	EMP	Non-EMP
Baseline pain VAS (SD)	60.5 (14.2)	58.6 (14.6)	58.3 (16.0)	60.3 (15.6)	59.6 (13.3)	56.6 (14.9)	59.1 (14.9)	59.2 (15.0)
Posttreatment pain VAS (SD)	43.5 (25.5)	44.0 (21.7)	49.2 (25.7)	47.0 (24.8)	53.2 (23.0)	37.0 (22.7)	47.4 (25.3)	43.6 (23.4)
Mean pain improvement (% of patients with clinically relevant improvement in pain)	29.5 (61.9)	23.0 (62.3)	16.6 (41.6)	22.2 (55.2)	10.3 (45.8)	34.9 (62.7)	21.1 (52.3)	25.2 (58.4)
Baseline analgesia (SD)	4.6 (3.9)	4.4 (3.6)	4.0 (3.74)	4.6 (4.0)	4.8 (3.7)	3.4 (3.2)	4.6 (4.3)	4.1 (3.0)
Posttreatment analgesia (SD)	3.8 (4.6)	4.2 (6.9)	3.7 (4.1)	4.4 (6.6)	4.2 (3.9)	2.7 (3.3)	3.8 (4.6)	4.0 (6.0)

EMP, empathic consultation; MES, mock electrical stimulation; RA, real acupuncture; SN, Streitberger needle; VAS, visual analogue score.

**Table 6**  
Adjusted means and differences for pain at week 5 from the analysis of covariance.

Treatments	Mean	Standard error	95% confidence interval		P value
			Lower bound	Upper bound	
RA	43.5	2.2	39.2	47.8	
SN	46.2	2.3	41.6	50.8	
SE	47.3	2.4	42.6	52.1	
Differences	Mean	Standard error	Lower bound	Upper bound	P value
RA–SN	–2.7	3.2	–9.0	3.6	.40
RA–SE	–3.9	3.3	–10.4	2.7	.25
SN–SE	–1.2	3.4	–7.8	5.5	.73
Practitioner	Mean	Standard error	95% confidence interval		P value
Practitioner			Lower bound	Upper bound	
1	45.9	1.8	42.4	49.430	
2	51.0	2.6	45.8	56.201	
3	40.1	2.4	35.4	44.745	
Differences	Mean	Standard error	Lower bound	Upper bound	P value
1–2	–5.1	3.2	–11.4	1.2	.11
1–3	5.8	3.0	–0.1	11.7	.05
2–3	10.9	3.6	3.9	18.0	.002
Consultation	Mean	Standard error	95% confidence interval		P value
Consultation			Lower bound	Upper bound	
EM	47.2	1.8	43.7	50.6	
ME	44.2	2.0	40.3	48.0	
Difference	Mean	Standard error	Lower bound	Upper bound	P value
EM–ME	3.0	2.6	–2.2	8.2	.26

Covariates appearing in the model are evaluated at the following values: mean baseline pain = 58.86, baseline mean daily tablet count = 4.32, real = 0.88, posttreatment mean daily tablet count = 3.64.

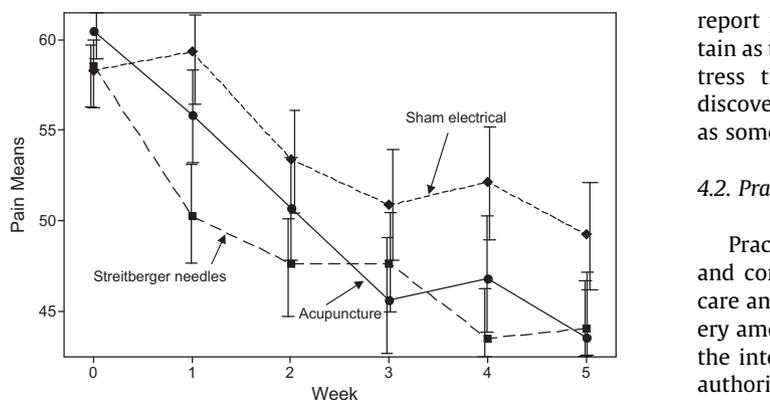
EM, empathic; ME, minimal empathy; RA, real acupuncture; SE, sham electrical; SN, Streitberger needle.

natural history of the condition, i.e., leading to spontaneous remission (a regression to the mean). The effect of this cannot be measured from this study, and we cannot discount this possibility. However, this possibility might be attenuated by the population chosen for this study, i.e., all were chronic end-stage OA sufferers such that all were awaiting joint replacement. Conservative treatment options for this group had been exhausted and pain was still persistent and consistently high enough to warrant surgical intervention.

#### 4.1. The effect of the consultation process and treatment context

Manipulating the consultation process was effective quantitatively and resulted in significantly different CARE scores across the groups. Mercer [18] suggest that small differences in CARE represent large differences in real terms because patients are reluctant to give very low scores to their doctors. The 10-point difference found in this trial is therefore substantial, reflecting clinically important differences that may be much larger differences in ob-

server rated scores [18]. This did not result in a different therapeutic outcome in spite of the acknowledged importance of a patient-centred approach [17]. It is possible that CARE is failing to adequately capture patient centeredness [17], but the qualitative study suggests that this is not the case. It seems that patients appear to have made allowances and colluded when receiving non-empathic consultations, which might have subverted the trial process. They drew empathy from other sources (reception staff, the initial prerandomisation consent/screening session), thus implying it may be very difficult to deliver a nonempathic process to patients in a trial in which patients have received prior informed consent. Researchers and ethics committees should consider this carefully in relation to consent and subsequent trial debriefing. The other more fundamental issue is that we do not yet fully understand the relationship between verbal and nonverbal communication in the consultation with respect to patient centeredness and therapeutic outcome [17]. Interpretations of outcome related to empathy should therefore be approached with great caution.



**Fig. 2.** Mean pain scores from pretreatment to week 5 by treatment group. NB: Error bars are based on the means plus and minus the standard error of the mean (they are not confidence intervals)

**Table 7**  
Side effects.

Nature of side effect	n	Group allocation
Temporary increase in pain	1	RA/EMP
	4	RA/non-EMP
	1	SN/non-EMP
Bleed or bruise at needle site	4	RA/EMP
	11	RA/non-EMP
	2	SN/EMP
Reddening over patella	1	MES/non-EMP
Headache posttreatment	1	MES/non-EMP
Tired posttreatment	1	RA/EMP
	1	SN/EMP
Became tearful posttreatment	1	RA/non-EMP

EMP, empathic consultation; MES, mock electrical stimulation; non-EMP, nonempathic consultation; RA, real acupuncture; SN, Streitberger needle.

Kaptchuk et al. suggest that enhanced consultations produced better outcomes [14]. The differences between our studies could relate to the diagnoses (OA pain vs irritable bowel syndrome) or to the differences in the initial consent process (Kaptchuk et al. did not require consent to different consultation types, whereas we did). We only manipulated empathy, and Kaptchuk et al. manipulated the acupuncture treatment (palpation), consultation time, and therapists' confidence in treatment. There may also be an interaction between different components within acupuncture treatment (talking, listening, and diagnosis) that produce a clinical effect that is greater than the sum of its individual elements. Pater-son and Dieppe [21] suggest that these factors, in contrast to drug trials, are integral to complex nonpharmaceutical interventions.

Belief in treatment veracity was significantly correlated to outcome, suggesting that improving the management of expectation could enhance outcome. This is supported by the qualitative analysis, which suggests that patients link positive outcomes and recorded improved symptoms when they believed that the treatment had been real and vice versa. Where patients were unsure as to the "reality" of their treatment, this was manifest in a reluctance to score symptom improvement, and therefore they were less willing to commit themselves when completing the outcome questionnaires. This has implications for all placebo-controlled trials inasmuch as the real effects of an unconvincing verum treatment might be underreported (a type II error). One solution might be to encourage participants to understand and believe (correctly) that placebos are powerful interventions that act through a variety of complex sociocultural, psychological, and neurological mechanisms [23]. They have a proven range of beneficial effects, including pain relief. This might encourage participants to

report perceived improvements confidently despite being uncertain as to their treatment allocation. It might also minimise the distress that some participants experience on unblinding when discovering they have responded to a placebo [4] and perceive this as something deceitful.

#### 4.2. Practitioner effect

Practitioner 3 produced better outcomes across all treatment and consultation types. This occurred in spite of the meticulous care and planning taken to ensure consistency of treatment delivery among the 3 practitioners. The qualitative data suggested that the interviewees perceived practitioner 3 as a paternalistic male authority figure. Practitioner 3, as the primary investigator, might have been seen by patients as the expert, consequently establishing higher expectations of success, which in turn influenced outcome. Although this is consistent with previous research [10,29] a larger explanatory study involving many practitioners is needed.

#### 4.3. Control type

It has been suggested that that needling carries an enhanced placebo effect [14,15]. We are the first to compare 2 acupuncture placebo controls demonstrating no difference between needle and nonneedle placebos. The magnitude of the improvements shown in the placebo arms in this and other trials indicates that placebo acupuncture has powerful effects outperforming standard care [12,25]. This supports the hypothesis that "exotic" processes or rituals may have powerful therapeutic effects.

#### 4.4. Efficacy of acupuncture

Efficacy studies of acupuncture for OA pain have not reported consistent results. Although our results concur those of with Scharf et al. [25], other large studies demonstrate a statistically significant effect for verum acupuncture in knee pain [2,26,28]. The reasons for these differences are unclear. We included patients with hip pain, and this might not respond as well to acupuncture; it is also possible that acupuncture is much less effective in this group of severely affected OA patients.

#### 4.5. Conclusions

This trial uniquely reports a mixed-methods approach in a single report to enable a contextual interpretation of the quantitative data. Patients receiving acupuncture demonstrated clinically important improvements from baseline (i.e., a 29.5% reduction in pain), but despite this, acupuncture has no specific efficacy over placebo for this group of patients. The clinical effect of acupuncture treatment and associated controls is not related to the use of an acupuncture needle, nor mediated by empathy, but is practitioner related and may be linked to the perceived authority of the practitioner. Qualitative analysis indicated that beliefs about treatment veracity and confidence in outcomes were reciprocally linked, and this appears to affect how patients self-report treatment outcomes, leading to a greater risk of demonstrating type II errors. This suggests that the process of interpreting RCT data is complex and contextual.

#### Acknowledgements

P.W. and this study was funded by a Department of Health Post-doctoral Research Award. C.S. was also funded by the same award. G.L.'s post is partially funded by the Rufford Maurice Laing Foundation. The Southampton Complementary Medicine Research Trust contributed funding for this trial. The funders played no role in

study design, interpretation, or writing of this paper. All authors have no conflicts of interest to declare and no financial interests related to this study.

## References

- [1] Araujo M. Does the choice of placebo determine the results of clinical studies on acupuncture. *Res Complement Med* 1998;5:8–11.
- [2] Berman BM, Lao L, Langenberg P, Lee W, Gilpin A, Hochberg M. Effectiveness of acupuncture as adjunctive therapy in osteoarthritis of the knee. *Ann Intern Med* 2004;141:901–11.
- [3] Cherkin D, Sherman K, Avins A, Erro J, Ichikawa L, Barlow W, Delaney K, Hawkes R, Hamilton L, Pressman A, Khalsa P, Deyo R. A randomized trial comparing acupuncture, simulated acupuncture, and usual care for chronic low back pain. *Arch Intern Med* 2009;169:858–66.
- [4] Chung S, Price D, Verne G, Robinson M. Revelation of a personal placebo response: its effects on mood, attitudes and future placebo responding. *Pain* 2007;132:281–8.
- [5] Cummings M. Modellvorhaben Akupunktur—a summary of the ART, ARC and GERAC trials. *Acupunct Med* 2009;27:28–30.
- [6] Department of Health SSaPS, Social and market research. Complementary and Alternative Medicines Pilot Project, 2008. Available from: [http://www/dhsspsni.gov.uk/final\\_report\\_from\\_smr\\_on\\_the\\_cam\\_pilot\\_project\\_may\\_2008.pdf](http://www/dhsspsni.gov.uk/final_report_from_smr_on_the_cam_pilot_project_may_2008.pdf).
- [7] Devilly G, Borkovec T. Psychometric properties of the credibility/expectancy questionnaire. *J Behav Ther* 2000;31:73–86.
- [8] Dworkin R, Turk D, Farrar J, Haythornthwaite J, Jensen M, Katz N, Kerns R, Stucki G, Allen R, Bellamy N, Carr D, Chandler J, Cowan P, Dionne R, Galer B, Hertz S, Jadad A, Kramer L, Manning D, Martin S, McCormick C, McDermott M, McGrath P, Quessy S, Rappaport B, Robbins W, Robinson J, Rothman M, Royal M, Simon L, Stauffer J, Stein W, Tollett J, Wernicke J, Witter J. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain* 2005;113:9–19.
- [9] Farrar J, Young J, LaMoreaux L, Werth J, Poole R. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain* 2001;94:149–58.
- [10] Finniss D, Kaptchuk T, Miller F, Benedetti F. Biological, clinical and ethical advances of placebo effects. *Lancet* 2010;375:686–95.
- [11] Gracely RH. Clinician's expectations influence placebo analgesia. *Lancet* 1985;1:43.
- [12] Haake M, Muller H, Schade-Brittinger C, Basler H, Schafer H, Maier C, Endres H, Tampisch H, Molsberger A. German acupuncture trials (GERAC) for chronic low back pain—randomised, multicenter, blinded, parallel-group trial with 3 groups. *Arch Intern Med* 2007;167:1892–8.
- [13] Hyland M, Lewith G, Westoby C. Developing a measure of attitudes: the holistic complementary and alternative medicine questionnaire. *Complement Ther Med* 2003;11:33–8.
- [14] Kaptchuk T, Stason W, Davis R, Legedza A, Schnyer R, Kerr C, Stone D, Nam B, Kirsch I, Goldman R. Sham device vs inert pill: randomised controlled trial of two placebo treatments. *BMJ* 2006;332:391.
- [15] Kaptchuk TJ, Goldman P, Stone D, Stason W. Do medical devices have enhanced placebo effects? *J Clin Epidemiol* 2000;53:786–92.
- [16] Laurance J. On reflection; magic is acceptable. *Lancet* 2010;375:885.
- [17] Little P, Everitt H, Williamson I, Warner G, Moore M, Gould C, Ferrier K, Payne S. Observational study of effect of patient centredness and positive approach on outcomes of general practice consultations. *BMJ* 2001;323:908–11.
- [18] Mercer S. Empathy 'CARE' measure, differences in score, 2009; personal communication.
- [19] Mercer S, Maxwell M, Heaney D, Watt G. The consultation and relational empathy (CARE) measure: development and preliminary validation and reliability of an empathy based consultation process measure. *Fam Pract* 2004;21:699–705.
- [20] Park H, Park J, Lee H, Lee H. Does Deqi (needle sensation) exist? *Am J Chin Med* 2002;30:45–50.
- [21] Paterson C, Dieppe P. Characteristic and incidental (placebo) effects in complex interventions such as acupuncture. *BMJ* 2005;330:1202–5.
- [22] Petrie J, Hazleman B. Credibility of placebo transcutaneous nerve stimulation and acupuncture. *Clin Exp Rheumatol* 1985;3:151–3.
- [23] Price D, Finniss D, Benedetti F. A comprehensive review of the placebo effect: recent advances and current thought. *Annu Rev Psychol* 2008;59:565–90.
- [24] Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Bryman A, Burgess R, editors. *Analyzing qualitative data*. London: Routledge; 1994. p. 173–94.
- [25] Scharf H, Mansmann U, Streitberger K, Witte S, Kramer J, Maier C, Trampisch H, Victor N. Acupuncture and knee osteoarthritis: a three-armed randomized trial. *Ann Intern Med* 2006;145:12–20.
- [26] Vas J, Mendez C, Perea-Milla E, Vega E, Panadero MD, Leon JM, Borge MA, Gaspar O, Sanchez-Rodriguez F, Aguilar I, Jurado R. Acupuncture as a complementary therapy to the pharmacological treatment of osteoarthritis of the knee: randomised controlled trial. *BMJ* 2004;329:1216.
- [27] White P, Lewith G, Hopwood V, Prescott P. The placebo needle, is it a valid and convincing placebo for use in acupuncture trials. A randomised, single blind, cross-over trial. *Pain* 2003;106:401–9.
- [28] Witt C, Brinkhaus B, Jena S, Linde K, Streng A, Wagenpfeil S, Hummelsberger J, Walther H, Melchart D, Willich S. Acupuncture in patients with osteoarthritis of the knee: a randomised trial. *Lancet* 2005;366:136–43.
- [29] Wright E, Holcombe C, Salmon P. Doctors communication of trust, care, and respect in breast cancer: qualitative study. *BMJ* 2004;328:864–7.