

Discrimination accuracy between real and sham needles using the Park sham device in the upper and lower limbs

Chee-Wee Tan, Padraig Sheehan, Derek Santos

► Additional data (online appendix) are published online only. To view these files please visit the journal online at (<http://aim.bmj.com>).

School of Health Sciences, Queen Margaret University Edinburgh, Edinburgh, UK

Correspondence to

Chee-Wee Tan, Queen Margaret University Edinburgh, Queen Margaret University Drive, Edinburgh EH21 6UU, UK; ctan@qmu.ac.uk

Accepted 26 April 2011
Published Online First
7 June 2011

Abstract

Objective To evaluate the blinding effectiveness of the Park sham acupuncture device using participants' ability to discriminate between the real and sham acupuncture needles in the (1) upper limb (TE points) compared with pure guessing and (2) lower limb (BL points) compared with pure guessing.

Methods 20 healthy acupuncture-naïve university students and staff were recruited through convenience sampling. Participants made Yes–No judgements on whether the real or sham needle was administered to four TE acupoints on the dominant upper limb, and four acupoints along the BL meridian on the dominant lower limb. The proportion of correct judgements, P(C), for each participant was calculated to indicate the discrimination accuracy of participants in distinguishing between the real and sham needles. Separate P(C) were computed for the upper limb acupoints and lower limb acupoints. The data were also pooled to calculate a P(C) for a combination of both body regions.

Results The participants' discrimination accuracy between the real and sham needles was not statistically significant from P(C)=0.5 (chance level) for the lower limb alone and combined body regions' acupoint comparisons (lower limb: $t_{19}=0.00$, unadjusted $p=1.00$; combined: $t_{19}=1.75$, unadjusted $p=0.10$). However, the participants' discrimination accuracy was statistically significant from P(C)=0.5 for the upper limb acupoints alone comparison ($t_{19}=2.36$, unadjusted $p=0.03$).

Conclusions This study showed that the Park sham device is more likely to blind participants in differentiating between the real and sham needles in the lower limb (BL meridian) acupoints than in the upper limb (TE meridian). However, the participants' ability to differentiate between the needle types for the upper limb acupoints was significantly different from chance levels.

INTRODUCTION

One important principle for designing a randomised controlled trial is the inclusion of a control group in which no relevant active intervention is administered. For research studies examining the effects of acupuncture, one strategy is to include a viable and credible sham acupuncture device mimicking

the real acupuncture intervention. In recent years, blunt acupuncture needles that retract into the handles were invented as a possible solution for blinding research participants.^{1,2} Randomised controlled trials examining the effects of acupuncture have started to use these sham devices and needles.^{3–5} Credibility studies had generally found that the participant-blinding effectiveness of these sham devices were fair or good.^{6–8}

Studies examining the blinding effectiveness of the sham devices usually use only one or a small number of acupoints. Tsukayama *et al*⁷ suggested that the blinding effectiveness of the sham needle may depend on the locations of the acupoints in the body regions. One contributing factor may be the apparent variations in sensation thresholds in different regions of the body. In a classic study by Weinstein,⁹ the pressure threshold on various body sites were mapped using von Frey hairs, and it was found that the face, upper limb and trunk had lower pressure thresholds than the lower limbs. In a large multicentre study, Rolke *et al*¹⁰ collected normative values for quantitative sensory testing parameters (of which pressure and thermal pain thresholds are two examples), and also found that the face region is more sensitive than the hand, followed by the foot.¹⁰ In relation to acupuncture, Tsukayama *et al*⁷ showed that higher proportions of participants could correctly differentiate between the sham and real needles in the hand (LI4) compared with the lower back (BL23). These findings imply that participants may be better able to discriminate between the sham and real needle in more sensitive body regions compared with the less sensitive ones. This could have a biased effect on the perceived effectiveness of the acupuncture intervention by the participants when acupoints chosen for a clinical study span across more than one acupoint or body region. It is therefore important to ascertain the participants' discrimination accuracy for sham and real needles at different body regions for the purpose of informing the design and implementing future clinical studies.

It is worth noting that we have operationalised blinding effectiveness as the (in)ability of participants to differentiate between the

real and sham needles administered in an experimental or clinical study, measured in terms of proportion of correct identification between the needle types administered. In other words, the participants' discrimination accuracy between the needle types is used as the outcome measure for blinding effectiveness. Blinding effectiveness is seen as a continuous variable because it is stated in terms of proportions, percentages and probability. That is, we do not view blinding effectiveness as an all-or-none phenomenon. We have used proportions as the units of measurement for blinding effectiveness in this study. This is also a commonly used definition for one aspect of the credibility of sham devices in other studies.^{6,11,12} We estimated, based on a previous study by Tan *et al*,¹² that a proportion difference of 0.2 for correct judgements from chance levels was a reasonable allowance for participants to make errors.

This study will therefore observe participants' discrimination accuracy between real and sham needles for two different peripheral body regions (upper and lower limbs). This study will also examine if the participants' discrimination accuracy within each limb is significantly different compared with correctly guessing the needle types simply by chance.

METHODS

Participants

A favourable ethical opinion for this study was obtained from the Queen Margaret University Research Ethics Committee. Students and staff of the University were recruited as participants using convenience sampling. The inclusion criteria for this study were as follows: (A) age of 18 years or more, (B) naïve to acupuncture intervention and (C) able to provide informed consent. The exclusion criteria were as follows: (A) the presence of medical conditions that caused anaesthesia to the dominant upper and lower limbs, (B) any wounds or injury to the limbs, (C) needle phobia, (D) consumption of potentially analgesic medications 24 h before the study procedures and (E) pregnancy. All the participants provided written, informed consent for this study. Participants were allowed to withdraw at any juncture of the study without providing reasons for doing so.

Experimenters

One author, in the role of the assistant, recruited and inducted participants to the study, prepared the equipment and assisted the practitioner during the procedure. Another author took the role of the practitioner and administered the needles to all the participants. The practitioner is a trained physiotherapist from the Acupuncture Association of Chartered Physiotherapists and has 5 years of experience in acupuncture administration.

The assistant provided the participant the following information, in both verbal and written form, before the experiment for consent and information-giving purposes:

This study will investigate whether people can tell the difference between a real acupuncture needle and a fake acupuncture needle with a blunt tip, using a small device known as a Park sham device. As a participant, you will be asked to expose your arm and the back of your leg (A screen will be present if you wish to change into shorts). At 8 specific acupuncture points either a real or fake needle will be inserted, and you will be asked which one you think it is.

MATERIALS

The Park sham acupuncture device was used for administering the sham and real needles (figure 1).² The device consists of a ring-base unit and a special oversized tube (Park tube). The ring-base of the device is kept in place on the participant's skin using double-sided tape. The internal circumference of the ring-base fits tightly around the Park tube. A guide tube that is included with the device makes a sliding fit into the Park tube. This arrangement allowed different penetration depths of the acupuncture needle by telescopically adjusting the guide tube within the Park tube. Both types of needles were of the same dimensions (0.25×40 mm) and manufactured by Dong Bang Acupuncture (Kyunggi-do, Korea). However, the real needle had a shorter steel handle compared with the sham needle. This issue was resolved by positioning the participants in such a way that they could not see the needles.

Acupuncture points

A total of eight (four upper limb and four lower limb) acupoints on the dominant limbs of the participants were chosen for this study. The upper limb acupoints along the TE meridian and their designated needle types were (1) TE11 (sham needle), (2) TE12 (real needle), (3) TE13 (sham needle) and (4) TE14 (real needle). The lower limb acupoints along the BL meridian were (1) BL37 (real needle),

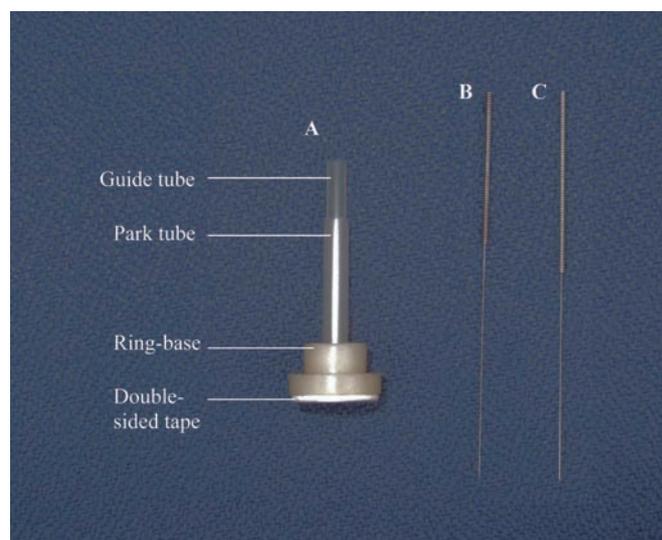


Figure 1 Photograph showing (A) the Park sham device; (B) a real acupuncture needle and (C) a sham acupuncture needle.

(2) BL55 (sham needle), (3) BL56 (real needle) and (4) BL57 (sham needle). The allocation of the designated needle type to the acupoints on both limbs was randomised. The depths of needle penetration were 0.5 cun for the TE acupoints and 1 cun for the BL acupoints. The cun system was adopted for this study so that the insertion depths are relative to the body proportions of the individual participant. This ensured safety in terms of needle insertion depths. The acupoints were chosen for ease of access when the participant is lying prone.

Procedures

Before the participant entered the laboratory, all the sham and real needles were removed from their packaging by the assistant and were placed on a sterile tray. The sequence of needle administrations for the eight acupoints was randomised using an online randomisation generator (<http://www.randomization.com> (accessed 17 February 2008)). Based on this randomisation scheme, 10 participants had their first acupoint located on the upper limb and 10 participants on the lower limb. Upon arrival, the participant was in a prone position lying on a plinth with the dominant arm rested comfortably beside the participant's body. The participants rested their head in rotation facing away from the tested side. The practitioner then attached the Park sham devices over the eight chosen acupoints on the dominant upper and lower limbs.

At the start of the needle insertion procedure, the assistant handed the appropriate needle (real or sham) to the practitioner for administration. The needle was placed inside the sham device. The real needle insertion into the skin occurred in two stages. In the first stage, the practitioner slid the guide tube downwards to approximately halfway down the acupoint penetration depth (0.25 cun for the upper limb and 0.5 cun for the lower limb acupoints). The practitioner then held the interface between the guide tube and the Park tube firmly and slightly depressed the device down onto the skin. This procedure simulated the clinical practice of using the needle guide tube to stretch the skin over the acupoints before needle insertion. A quick gentle tap was applied to the proximal end of the needle handle so that needle penetration into the skin was achieved. In the second stage, the guide tube was moved down further to the chosen acupoint penetration depth (0.5 cun for the upper limb and 1.0 cun for the lower limb acupoints). The practitioner then gripped the needle handle and slid the needle to the appropriate tissue depth without twirling the needle. The procedure for the administration of sham needles is similar to that of real needles. However, the sham needle did not penetrate the skin but retracted into the handle thereby providing the illusion of actual skin insertion.

The assistant asked the participants: "Do you think the real acupuncture needle has been administered?" The participants answered 'yes' or 'no'. The research assistant recorded the participant's judgement on a score sheet. This denoted the end of one administration. Each participant received eight administrations, providing a total of eight

judgements. The entire procedure for each participant lasted approximately 30 min. This methodology is called a Yes–No experiment^{13 14} and was performed for all eight acupoints. It is an established methodology in experimental psychology, physiology and decision-making science. The function of the Yes–No experiment is to repeatedly expose the participant to the stimuli for differentiation (in our case, needle type). Similar studies using this methodology have looked at how clinicians make decisions about radiographic evidence,¹⁵ how engineers determine structural defect using ultrasonography¹⁶ and the accuracy of HIV testing.¹⁷ Therefore, the methodology has been used to investigate real world decision-making problems. The sham acupuncture credibility issue is similar in many ways to the decision-making problems noted above. When the Yes–No experiment is used in the real–sham needle differentiation context, the methodology may make it more likely for participants to be able to differentiate between the needles because they have been afforded the luxury of a comparison between the needle types. This means that if the participant is still unable to differentiate between the needle types for a Yes–No experiment, then it may be very unlikely that he/she is able to differentiate between the needle types for a real trial setting.

Analysis

The outcome measure $P(C)$ (proportion of correct judgements) was computed for the upper limb, lower limb and a combination of both body region acupoints for each participant using the same data set. In order for $P(C)$ to be calculated, each participant's true positive rate or $P(TP)$ (the proportion of times the participant correctly judges the real needle to be the real needle) and the true negative rate or $P(TN)$ (the proportion of times the participant correctly judges the sham needle to be the sham needle) were tallied. The $P(TP)$ and $P(TN)$ are analogous to the outcome measures of sensitivity and specificity, respectively, in the study of diagnostic accuracy.¹⁸ The outcome measure $P(C)$ was calculated by adding both $P(TP)$ and $P(TN)$ before dividing by 2.¹³ For an example of a $P(C)$ calculation, see online appendix 1. When $P(C)$ is 0.50, this means that the real and sham needles are indistinguishable by the participant. Perfect participant ability to distinguish between needle types yields a $P(C)$ of 1.00.

The $P(C)$ values were transformed to z , the inverse of the normal distribution function for analysis using parametric statistics. One-sample t tests at $\alpha=0.05$ were performed to compare the z scores generated for the upper limb alone, lower limb alone and both body region acupoints with $z=0.00$. The z score of zero is the transformed value of $P(C)=0.50$, where the participants were unable to differentiate between needle types completely. All z score results were reverse transformed back to $P(C)$ for reporting and ease of interpretation.

Correction methods for multiple testing (eg, Bonferroni method) were not used in this study. The main rationale for using correction methods is to maintain the family-wise error rate thereby preventing the inflation of Type I

error (the probability of rejecting a null hypothesis when it should not have been rejected). This comes at the cost of increasing Type II error (the probability of not rejecting a null hypothesis when it should have been rejected). Correction methods may be appropriately applied when it is desirable to reject the null hypothesis. However, for our study, it is more desirable not to reject the null hypothesis (ie, participants are unable to differentiate between real and sham acupuncture needles). This implies that an inflated Type II error (due to multiple testing correction) for our study may lead to a higher probability of failing to reject the null hypothesis when it should have been rejected. We have therefore reported the unadjusted p and CI values.

We also explored the data generated for each acupoint by breaking down the number of correct judgements (expressed as percentage of correct judgements) and summarised the data by each acupoint. In addition, we examined the presence of fatigue or learning effects by tallying the number of correct judgements (expressed as percentage of correct judgements) according to the sequence of needle administrations regardless of acupoint. These analysis were purely descriptive and no inferential tests were applied. It is important to note that the percentage of correct judgements for the acupoints and needle sequence breakdown are different to $P(C)$. Percentage of correct judgements is the number of correct guesses expressed as a percentage of the total number of judgements made for that particular acupoint or needle sequence. In contrast, $P(C)$ is the weighted average of $P(TP)$ and $P(TN)$ (online appendix 1).

We judged, based on the study by Tan *et al.*¹² that a $P(C)$ difference of 0.2 from chance level was a reasonable error allowance for participants. The SD $P(C)=0.29$ was calculated using raw data from the same study.¹² For an a priori determined power=0.8 and $\alpha=0.05$ (two-tailed), the sample size required for this study was 19 participants. Twenty participants were recruited in the event of participant drop-out or occurrence of unusable data.

RESULTS

Participants

Twenty healthy volunteers (13 women and 7 men) took part in the experiment with no participant drop outs. The participants' median age was 24 years (range 21–28 years).

Discrimination accuracy of participants

To obtain a general overview of the participants' ability to differentiate between the needle types, the judgements for all participants were pooled and categorised into the four judgement categories. Table 1 shows the needle type–response matrix of all judgements and the matrices of all participants' judgements separated by the body regions (both limbs, upper limb alone or lower limb alone).

Each participant's $P(C)$ was also computed. The mean $P(C)$ of participants for acupoints on both limbs was 0.56

(SD 0.16, 95% CI 0.49 to 0.64). The mean $P(C)$ for the upper limb acupoints alone and the lower limb acupoints alone were 0.63 (SD 0.24, 95% CI 0.51 to 0.74) and 0.50 (SD 0.26, 95% CI 0.38 to 0.62), respectively.

The participants' $P(C)$ for differentiating acupoints on both limbs and lower limb acupoints alone were not significantly larger than $P(C)=0.50$ (both: $t_{19}=1.75$, unadjusted $p=0.10$; lower limb: $t_{19}=0.00$, unadjusted $p=1.00$). The $P(C)$ for upper limb acupoints alone was significantly larger than $P(C)=0.50$ ($t_{19}=2.36$, unadjusted $p=0.03$).

The number of correct judgements for each acupoint was tallied and summarised as percentages (figure 2). Out of the eight acupoints, two had a percentage of less than 50 for correct judgements. These two acupoints were chosen for the administration of sham needles. To ascertain if fatigue or learning effects were present, the number of correct judgements according to the sequence of needle administrations was also tallied and summarised as percentages (figure 3). The results showed an increase in percentage of correct judgements from the third to the fifth needle inserted. There were slightly more correct judgements from the real needles contributing to the total percentage of correct judgements compared with the sham needles. At the insertion of the sixth needle, there was a drastic drop in correct judgements. At this point, there was no contribution from judgements for the real needle to the total number of correct judgements. Then, this increased back again to 65% by the eighth needle.

DISCUSSION

This study found that the participants appeared to be able to differentiate between real and sham acupuncture needles for the TE acupoints within the upper limbs. However, the participants did not appear to be able to differentiate between the needle types for the BL acupoints within the lower limb and when the acupoints on both limbs were considered together.

Our study is very similar in research design to the study by Tan *et al.*¹² However, it is important to note that there are differences in the aims and methodologies between the two studies. The first difference is in the aims of the studies. Tan *et al.*¹² investigated the participants' ability to discriminate between real and sham needles for traditional

Table 1 Needle type–response matrix for upper and lower acupoints

Needle type administered	Participants' responses (proportions)					
	Upper limb*		Lower limb*		Both†	
	Real	Sham	Real	Sham	Real	Sham
Real	0.60	0.40	0.62	0.38	0.61	0.39
Sham	0.35	0.65	0.62	0.38	0.49	0.51

*The proportions are expressed as a total of 40 real and 40 sham needles pooled from 20 participants' judgements.

†The proportions are expressed as a total of 80 real and 80 sham needles pooled from 20 participants' judgements.

and non-traditional acupoints as a secondary hypothesis, whereas we investigated participants' needle-type discrimination ability for the upper and lower limb traditional acupoints as the primary hypothesis. Tan *et al*¹² found that participants were able to discriminate between real and sham needles for the traditional acupoints but not for the non-traditional acupoints in the upper limb. Our study has followed this up by first attempting to gather evidence for disconfirmation of the hypothesis of Tan *et al*¹². We have also investigated if body regions could affect the partici-

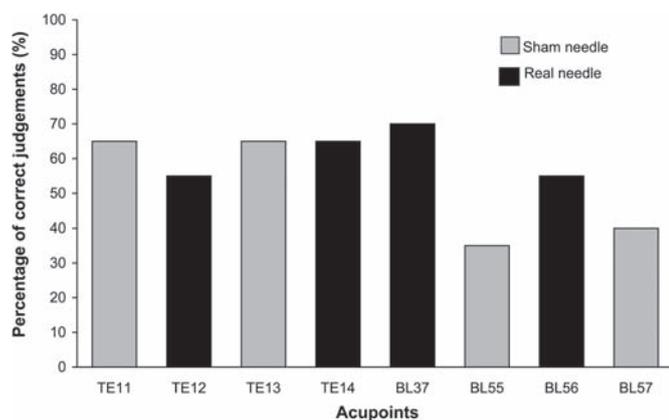


Figure 2 Percentage of correct judgements made by the participants for each acupoint is represented by bars. The BL acupoints that received sham needles generally have a lower percentage of correct judgements compared with the acupoints that received real needles.

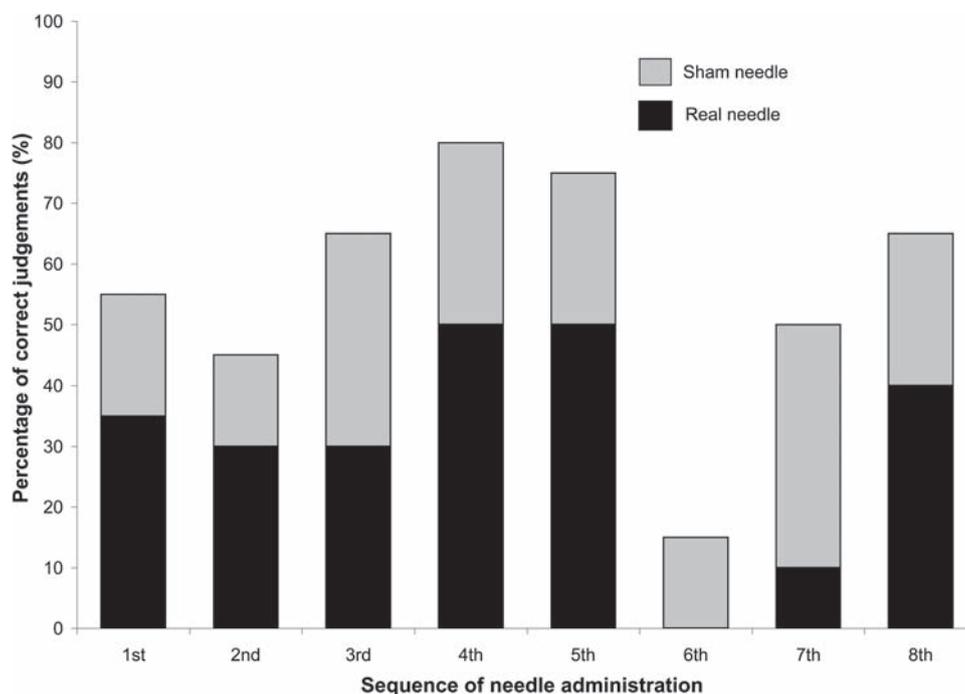


Figure 3 Cumulated percentage of correct judgements from real and sham needles based on the sequence of needle administration. The bars represent the cumulated percentage of correct judgements made by the participants. The grey and black sub-bars represent the percentage of correct judgements contributed by the sham and real needles, respectively. There appears to be no clear pattern of learning effect or task fatigue during the study.

pants' discrimination ability. The second difference is in the methodologies. Because of the insertion of needles on the lower limb at a greater penetration depth, we felt that the 'tap and treat' used by Tan *et al*¹² may not be appropriate or safe in the context of our study. We have therefore introduced a two-stage needle insertion, which, to some extent, mimics the clinicians' insertion technique when using guide tubes.

In our study, the difference in statistical results for discrimination accuracy in the upper limb and in the lower limb acupoints seems to suggest that different body regions may have dissimilar responses to real–sham needle differentiation. This pattern is also observed in a study by Tsukayama *et al*,⁷ which found that certain body regions appeared to be more sensitive than others. This may have implications for clinical trials that use the upper limbs, specifically TE11 to TE14. In such situations, our results suggest that most participants may potentially be able to guess which treatment condition of the clinical trial they belong to. However, when the data for upper limbs and lower limbs are pooled together, the discrimination accuracy is no longer significantly better than pure guessing (ie, $P(C)=0.5$). Although, this suggests that participants appeared unable to differentiate between the needle types when the body regions are analysed together, we propose that this is simply an averaging effect of pooling data from both limbs.

The averaging effect of the pooled data may not be a true reflection of the way by which participants make their decisions in a clinical trial. It is unknown how participants aggregate sensations and determine an overall

judgement of whether the sham or real treatment has been provided in clinical trials. For example, some clinical trials require participants at the end of the study to state whether they believe the real or sham treatment has been provided.^{19 20} If each body region has a different sensation threshold, how did the participants arrive at the final decision of which treatment arm of the trial they belong to despite potentially facing conflicting sensory evidence from different body regions? Unfortunately, our study was not designed to investigate participants' decision processes through which a final aggregate judgement is made on the needle type administered.

Due to the repeated administration of needles for the same participant, there is a potential for learning effects and task fatigue to influence the discrimination accuracy. The percentage of correct judgements was broken down according to the sequence of administration (figure 3). This analysis showed that there may be some learning effect happening from the first to the fifth needle. However, on the sixth needle, there is a drastic reduction in correct judgements mainly contributed by the absence of correct guesses made for the real needles. This data pattern did not fit neatly into a learning effect (displaying an asymptotic function) or task fatigue (displaying a suppressed percentage for correct judgements). Another plausible interpretation for the observed pattern may be that these are simply due to random fluctuation. Learning effects and task fatigue are important factors for study designs that involved repeated administration of needles. Further investigation is needed to establish the number of needles for minimisation of both learning effects and task fatigue.

It is also important to ascertain that the participants have actually been blinded in the sham or placebo condition. There are indices available that help clinical researchers confirm the extent to which blinding has been achieved for the study.²¹ We recommend that these blinding indices be used for post-trial evaluation of blinding effectiveness because the assessment protocol integrates well into a clinical trial. We also recommend that our study design and index be used for pretrial and sham device evaluation because it is more conservative, that is, more likely to conclude that a sham device may not provide adequate blinding. The reason for this apparent conservativeness is attributed to the repeated administration and comparison of multiple real and sham needles that provides more cues for the participants to make their judgements. Consider the situation when the participants' ability to judge the needle type is no better than chance levels despite the repeated needle administration. Then it may be argued that for the investigated independent variable, it is more probable that participants are less able to guess the needle type when they do not have the luxury of needle-type comparisons.

Based on our study's preliminary results, it may be necessary to examine the potential differences in discrimination accuracy to sham needles of different body regions or commonly selected acupoints. This precautionary pro-

cedure may prevent the potential contamination of participants' judgement by acupoints, which has been shown to be more difficult in instigating a blinding effect. Our result suggests that the blinding effect of sham devices may not be generalisable when acupoints investigated are located on different regions of the body. If it is shown that a class of acupoints is consistently resistant to the blinding attempts of investigators, it may be necessary to either implement more rigorous blinding measures or adopt a different type of acupuncture sham device.

Limitations

It is interesting to note the lower limb acupoints that received the sham needle had lower percentage of correct judgements compared with the acupoints that received the real needles (figure 2). Tan *et al*¹² also observed a similar finding. An explanation is that the sham needle may provide cues that make it difficult for the participant to discriminate it from the real needle, especially for the lower limb. However, a participant's discrimination accuracy may not be based solely on whether the sham needle is convincing, it may also encompass other factors such as the comparability of cues between the sham and real needle, the sensitivity of the particular bodily region and procedural cues from the practitioner. For example, during the needle administration protocol, the slight depression of the Park sham device onto the skin and the pressure exerted by the blunt tip of the sham needle may partly contribute to the discrimination accuracy of the participations. However, it is difficult to ascertain the magnitude of this effect on the discrimination accuracy because no study, to our knowledge, that used the Park sham device has investigated this particular issue in detail.

A limitation of our study is that the needle sensations by both the real and sham needles were not recorded for estimating their weighted influence on the participants' judgements. Needle sensations encompass a variety of sensory experiences, for example, heavy, stinging, radiating and electric.²² Participants may have used the sensations generated by the real and sham needles as cues for making their judgements. Benham *et al*²³ found that deep needle insertion produced a higher intensity of needle sensation compared with superficial insertion. In our study, the needle insertion depth for the upper limb acupoints was more superficial than the lower limb acupoints. If needle depth was indeed a contributing factor to discrimination accuracy, a likely outcome would be that the lower limb would have better discrimination accuracy than the upper limb. Our study found the opposite result. However, this comparison may be inappropriate because the upper limb and lower limbs have different cutaneous and subcutaneous tissue thicknesses. We recommend that future studies should establish the relationship between insertion depths, needle sensations and discrimination accuracy for real and sham needles.

It is important to note that the practitioner in our study was not blinded to the needle type. This may have introduced some form of bias to the participants' discrimination accuracy through the practitioner's behaviour, both consciously

and unconsciously. In the studies by Kim²⁴ and Takakura and Yajima,²⁵ where both the practitioners and participants were blinded, the P(C) made by the participants were 0.48–0.67 and 0.56, respectively. In our single-blind study, the P(C) were 0.63 (upper limb) and 0.50 (lower limb), which is very similar in range to those in the previous studies. However, it is inappropriate to conclude from this post-hoc comparison that practitioner blinding may be unnecessary. The main reason for this is because the methodologies for counting P(C) were different between our study and the previous studies. We used the mean of participants' P(C) compared with that from the studies by Kim²⁴ and Takakura and Yajima,²⁵ which used pooled P(C). Mean P(C) adds each individual participant's P(C) score, which is then divided by the participant number. Pooled P(C) cumulates responses from all participants, adds up the P(TP) and P(TN), which is then further divided by 2. Unknown statistical bias from these different calculation methods may be introduced into the proportion of correct judgement estimates, which may render the comparison of results difficult.

Another limitation of this study is that each acupoint in this study was allocated to be administered either a real or sham needle. Ideally, there should be both real and sham needles administered for each acupoint. An advantage of the 'both needles to one acupoint' approach is that it allows intra-acupoint discrimination accuracy to be determined in addition to the general limb region discrimination accuracy. However, due to the complexity of the randomisation scheme, we have chosen not to randomise the needle type within each acupoint.

Our study assumed that different sensitivity of body regions will lead to differences in discrimination accuracy for the needle types compared with simply guessing by chance. However, our study did not investigate or verify the relationship between discrimination accuracy and pre-needling sensation threshold, procedural cues from the practitioner, participants' decision processes, learning effects, task fatigue or the comparability of cues between real and sham needle. Future studies could test these relationships.

CONCLUSION

This study showed that within the context of the Yes–No experiment, the Park sham device is more likely to blind participants in differentiating between the real and sham needles in the lower limb (BL meridian) acupoints than in the upper limb (TE meridian). However, the participants' ability to differentiate between the needle types for the upper limb acupoints was significantly different from chance levels. It is recommended that future studies address limitations inherent in this study in order to further the design, study and clinical trial implementation of acupuncture sham devices.

Competing interests None.

Ethics approval This study was conducted with the approval of the Queen Margaret University Research Ethics Committee.

Summary points

- ▶ Blunt sham needles are used as controls for acupuncture
- ▶ We tested whether healthy volunteers could discriminate them from standard needles
- ▶ They are likely to be identifiable in the arm but not in the leg

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Streitberger K, Kleinhenz J. Introducing a placebo needle into acupuncture research. *Lancet* 1998;352:364–5.
2. Park JB, White A, Lee HJ, *et al.* Development of a new sham needle. *Acupunct Med* 1999;17:110–12.
3. Goldman RH, Stason WB, Park SK, *et al.* Acupuncture for treatment of persistent arm pain due to repetitive use: a randomized controlled clinical trial. *Clin J Pain* 2008;24:211–18.
4. Jubb RW, Tukmachi ES, Jones PW, *et al.* 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.
5. Kennedy S, Baxter GD, Kerr DP, *et al.* Acupuncture for acute non-specific low back pain: a pilot randomised non-penetrating sham controlled trial. *Complement Ther Med* 2008;16:139–46.
6. Park J, White A, Stevinson C, *et al.* Validating a new non-penetrating sham acupuncture device: two randomised controlled trials. *Acupunct Med* 2002;20:168–74.
7. Tsukayama H, Yamashita H, Kimura T, *et al.* Factors that influence the applicability of sham needle in acupuncture trials: two randomized, single-blind, crossover trials with acupuncture-experienced subjects. *Clin J Pain* 2006;22:346–9.
8. White P, Lewith G, Hopwood V, *et al.* The placebo needle, is it a valid and convincing placebo for use in acupuncture trials? A randomised, single-blind, cross-over pilot trial. *Pain* 2003;106:401–9.
9. Weinstein S. Intensive and extensive aspects of tactile sensitivity as a function of body part, sex, and laterality. In: Kenshalo DR, ed. *The Skin Senses*. Springfield, MA: Thomas 1968:195–222.
10. Rolke R, Baron R, Maier C, *et al.* Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): standardized protocol and reference values. *Pain* 2006;123:231–43.
11. Takakura N, Yajima H. A double-blind placebo needle for acupuncture research. *BMC Complement Altern Med* 2007;7:31.
12. Tan CW, Christie L, St-Georges V, *et al.* Discrimination of real and sham acupuncture needles using the park sham device: a preliminary study. *Arch Phys Med Rehabil* 2009;90:2141–5.
13. Gescheider GA. *Psychophysics: The Fundamentals*. 3rd edition. Mahwah, NJ: Lawrence Erlbaum Associates 1997.
14. Macmillan NA, Creelman CD. *Detection Theory: A User's Guide*. 2nd edition. Mahwah, NJ: Lawrence Erlbaum Associates 2005.
15. Metz CE. ROC methodology in radiologic imaging. *Invest Radiol* 1986;21:720–33.
16. Swets JA. Assessment of NDT systems: the relationship of true and false detections and indices of performance. *Mater Eval* 1983;41:1294–8.
17. Schwartz JS, Dans PE, Kinosian BP. Human immunodeficiency virus test evaluation, performance, and use. Proposals to make good tests better. *JAMA* 1988;259:2574–9.
18. Griner PF, Mayewski RJ, Mushlin AI, *et al.* Selection and interpretation of diagnostic tests and procedures. Principles and applications. *Ann Intern Med* 1981;94:557–92.
19. Alecrim-Andrade J, Maciel-Júnior JA, Carnè X, *et al.* Acupuncture in migraine prevention: a randomized sham controlled study with 6-months posttreatment follow-up. *Clin J Pain* 2008;24:98–105.
20. Diener HC, Kronfeld K, Boewing G, *et al.*; GERAC Migraine Study Group. Efficacy of acupuncture for the prophylaxis of migraine: a multicentre randomised controlled clinical trial. *Lancet Neurol* 2006;5:310–16.
21. Bang H, Flaherty SP, Kolahi J, *et al.* Blinding assessment in clinical trials: a review of statistical methods and a proposal of blinding assessment protocol. *Clin Res Regul Aff* 2010;27:42–51.
22. Park H, Park J, Lee H, *et al.* Does Deqi (needle sensation) exist? *Am J Chin Med* 2002;30:45–50.
23. Benham A, Phillips G, Johnson MI. An experimental study on the self-report of acupuncture needle sensation during deep needling with bi-directional rotation. *Acupunct Med* 2010;28:16–20.
24. Kim S. Creating an instrument for a successful double-blind acupuncture placebo. *J Acupunct Meridian Stud* 2008;1:36–41.
25. Takakura N, Yajima H. A placebo acupuncture needle with potential for double blinding – a validation study. *Acupunct Med* 2008;26:224–30.