Effectiveness and adequacy of blinding in the moderation of pain outcomes: Systematic review and meta-analyses of dry needling trials

Felicity A Braithwaite [Corresp., 1, 2], Julie L Walters [1], Lok Sze Katrina Li [1], G. Lorimer Moseley [1, 2], Marie T Williams [1, 3], Maureen P McEvoy [1]

[1] School of Health Sciences, University of South Australia, Adelaide, Australia

[2] Body in Mind research group, University of South Australia, Adelaide, Australia

[3] Alliance for Research in Exercise, Nutrition and Activity (ARENA), University of South Australia, Adelaide, Australia

Corresponding Author: Felicity A Braithwaite
Email address: felicity.braithwaite@gmail.com

Background: Blinding is critical to clinical trials because it allows for separation of specific intervention effects from bias, by equalising all factors between groups except for the proposed mechanism of action. Absent or inadequate blinding in clinical trials has consistently been shown in large meta-analyses to result in overestimation of intervention effects. Blinding in dry needling trials, particularly blinding of participants and therapists, is a practical challenge; therefore, specific effects of dry needling have yet to be determined. Despite this, dry needling is widely used by health practitioners internationally for the treatment of pain. This review presents the first empirical account of the influence of blinding on intervention effect estimates in dry needling trials. The aim of this systematic review was to determine whether participant beliefs about group allocation relative to actual allocation (blinding effectiveness), and/or adequacy of blinding procedures, moderated pain outcomes in dry needling trials. **Methods:** Twelve databases (MEDLINE, EMBASE, AMED, Scopus, CINAHL, PEDro, The Cochrane Library, Trove, ProQuest, trial registries) were searched from inception to February 2016. Trials that compared active dry needling with a sham that simulated dry needling were included. Two independent reviewers performed screening, data extraction, and critical appraisal. Available blinding effectiveness data were converted to a blinding index, a quantitative measurement of blinding, and meta-regression was used to investigate the influence of the blinding index on pain. Adequacy of blinding procedures was based on critical appraisal, and subgroup meta-analyses were used to investigate the influence of blinding adequacy on pain. Meta-analytical techniques used inverse-variance random-effects models. **Results:** The search identified 4894 individual publications with 24 eligible for inclusion in the quantitative syntheses. In 19 trials risk of methodological bias was high or unclear. Five trials were adequately blinded, and blinding was assessed and sufficiently reported to compute the

blinding index in 10 trials. There was no evidence of a moderating effect of blinding index on pain. For short-term and long-term pain assessments pooled effects for inadequately blinded trials were statistically significant in favour of active dry needling, whereas there was no evidence of a difference between active and sham groups for adequately blinded trials. **Discussion:** The small number and size of included trials meant there was insufficient evidence to conclusively determine if a moderating effect of blinding effectiveness or adequacy existed. However, with the caveats of small sample size, generally unclear risk of bias, statistical heterogeneity, potential publication bias, and the limitations of subgroup analyses, the available evidence suggests that inadequate blinding procedures could lead to exaggerated intervention effects in dry needling trials.

1 **Effectiveness and adequacy of blinding in the moderation of pain outcomes: systematic**

2 **review and meta-analyses of dry needling trials**

4 Felicity A. Braithwaite[1,2]; Dr Julie L. Walters[1]; Lok Sze Katrina Li[1]; Professor G. Lorimer

5 Moseley[1,2]; Associate Professor Marie T. Williams[1,3]; Dr Maureen P. McEvoy[1]

7 [1]School of Health Sciences, University of South Australia, Adelaide, Australia; [2]Body in Mind

8 research group, University of South Australia, Adelaide, Australia; [3]Alliance for Research in

9 Exercise, Nutrition and Activity (ARENA), University of South Australia, Adelaide, Australia.

11 **Corresponding author:**

12 Felicity A. Braithwaite

13 Email: felicity.braithwaite@gmail.com

15 **Previous presentation of the research:** The findings of this systematic review were presented

16 at the Australian Physiotherapy Association Conference (19-21 October, 2017).

17 **Abstract**

18 **Background:** Blinding is critical to clinical trials because it allows for separation of specific

19 intervention effects from bias, by equalising all factors between groups except for the proposed

20 mechanism of action. Absent or inadequate blinding in clinical trials has consistently been shown

21 in large meta-analyses to result in overestimation of intervention effects. Blinding in dry

22 needling trials, particularly blinding of participants and therapists, is a practical challenge;

23 therefore, specific effects of dry needling have yet to be determined. Despite this, dry needling is

24 widely used by health practitioners internationally for the treatment of pain. This review presents

25 the first empirical account of the influence of blinding on intervention effect estimates in dry

26 needling trials. The aim of this systematic review was to determine whether participant beliefs

27 about group allocation relative to actual allocation (blinding effectiveness), and/or adequacy of

28 blinding procedures, moderated pain outcomes in dry needling trials.

29 **Methods:** Twelve databases (MEDLINE, EMBASE, AMED, Scopus, CINAHL, PEDro, The

30 Cochrane Library, Trove, ProQuest, three trial registries) were searched from inception to

31 February 2016. Trials that compared active dry needling with a sham that simulated dry needling

32 were included. Two independent reviewers performed screening, data extraction, and critical

33 appraisal. Available blinding effectiveness data were converted to a blinding index, a

34 quantitative measurement of blinding, and meta-regression was used to investigate the influence

35 of the blinding index on pain. Adequacy of blinding procedures was based on critical appraisal,

36 and subgroup meta-analyses were used to investigate the influence of blinding adequacy on pain.

37 Meta-analytical techniques used inverse-variance random-effects models.

38 **Results:** The search identified 4894 individual publications with 24 eligible for inclusion in the

39 quantitative syntheses. In 19 trials, risk of methodological bias was high or unclear. Five trials

40 were adequately blinded, and blinding was assessed and sufficiently reported to compute the

41 blinding index in 10 trials. There was no evidence of a moderating effect of blinding index on

42 pain. For short-term and long-term pain assessments, pooled effects for inadequately blinded

43 trials were statistically significant in favour of active dry needling, whereas there was no

44 evidence of a difference between active and sham groups for adequately blinded trials.

45 **Discussion:** The small number and size of included trials meant there was insufficient evidence

46 to conclusively determine if a moderating effect of blinding effectiveness or adequacy existed.

47 However, with the caveats of small sample size, generally unclear risk of bias, statistical

48   heterogeneity, potential publication bias, and the limitations of subgroup analyses, the available

49   evidence suggests that inadequate blinding procedures could lead to exaggerated intervention

50   effects in dry needling trials.

51   **PROSPERO registration number:** 42016029340.

Background

Blinding is widely considered critical to the internal validity of clinical trials because it allows separation of specific intervention effects from effects due to bias. This separation is possible because blinding equalises all factors between groups except for the proposed mechanism of action of the intervention under investigation (Hróbjartsson et al. 2014).

Blinding adequacy relates to procedures in the design of a trial to blind relevant parties (i.e. trial staff, therapists, recipients, outcome assessors, data analysts). In the absence of adequate procedures, the inclination of these parties to favour a particular result can lead to distorted findings that most commonly manifest as exaggerated intervention effects (Hróbjartsson et al. 2014; Hróbjartsson et al. 2012; Hróbjartsson et al. 2013; Jüni et al. 2001; Moher et al. 2010; Nüesch et al. 2009; Savović et al. 2012; Schulz et al. 1995; Wood et al. 2008). For example, a review of trials involving head-to-head comparisons of blinded versus non-blinded participants demonstrated pronounced bias in non-blinded groups for complementary/alterative interventions (N=12 trials, 11 of which were acupuncture trials) (Hróbjartsson et al. 2014). Self-reported outcomes such as pain, which is often used to evaluate physical interventions, are particularly susceptible to the effects of inadequate blinding procedures (Hróbjartsson et al. 2014; Hróbjartsson et al. 2013; Moher et al. 2010; Savović et al. 2012; Wood et al. 2008). The complex nature of physical interventions means that blinding of relevant parties, particularly participants and therapists, is often extremely difficult (Boutron et al. 2004). As a result, blinding procedures for these types of interventions have been generally inadequate or omitted completely (Armijo-Olivo et al. 2017; Boutron et al. 2007; Boutron et al. 2004; Machado et al. 2008; Moseley et al. 2011).

Inclusion of adequate blinding procedures is recognised as crucial to robust trial design (Moher et al. 2010), but evaluation of the actual effectiveness of blinding procedures and its influence on clinical trial outcomes has been poorly addressed (Bang et al. 2010; Fergusson et al. 2004; Hróbjartsson et al. 2007). Needling therapies [acupuncture and dry needling (DN)] provide a unique intervention on which to focus evaluation of blinding effectiveness, because unlike trials of many other physical interventions (Armijo-Olivo et al. 2017; Hróbjartsson et al. 2007; Machado et al. 2008; Villamar et al. 2013), blinding assessments are becoming common practice

86   in needling therapy trials (Moroz et al. 2013). In addition, needling therapies are growing

87   globally in popularity to manage pain (Cagnie et al. 2013; Carlesso et al. 2014; Dommerholt

88   2011; Legge 2014) and encompass a range of factors known to predict large non-specific

89   responses; needling therapies are ritualistic, invasive, involve a medical device, are highly

90   credible to patients, and are often held in high regard by the person delivering them (Benedetti

91   2013; Finniss et al. 2010; Kaptchuk 2002; Kaptchuk et al. 2008; Kaptchuk & Miller 2015;

92   Kaptchuk et al. 2006). Exaggeration of intervention effects in acupuncture is associated with

93   expectation of intervention outcomes (Colagiuri & Smith 2011; Linde et al. 2007), and in

94   acupuncture trials, beliefs about group allocation have been shown to bear a stronger relationship

95   to pain than actual allocation (Bausell et al. 2005; Vase et al. 2013; White et al. 2012). These

96   findings suggest that failed blinding could be a significant confounder of trial outcomes, and

97   confirms that well-blinded trials will be required to determine the mechanisms of needling

98   therapies. However, a recent systematic review of acupuncture and dry needling trials (N=54

99   trials) reported that only 61% of trials *might* have had effective participant blinding based on

100   empirical data (i.e. where participant beliefs about the intervention to which they were allocated

101   were approximately balanced between active and sham groups) (Moroz et al. 2013). Ineffective

102   participant blinding, coupled with potentially inadequate or omitted blinding procedures for other

103   relevant parties (particularly therapists), calls into question any specific intervention effect of

104   needling therapy reported to date.

105

106   This systematic review presents the first empirical account of the influence of blinding on

107   intervention effect estimates in dry needling trials. Dry needling differs from acupuncture

108   because while acupuncture needles are used, they are inserted into clinically identified locations

109   in muscles (such as tender areas, palpable nodules or bands) rather than the largely pre-

110   determined insertion sites based on traditional Chinese medicine used in acupuncture. As such,

111   dry needling aims at local effects whereas acupuncture aims at systemic effects. The aim of this

112   review was to determine the influence of blinding effectiveness and blinding adequacy on pain in

113   sham-controlled dry needling trials. Blinding effectiveness was determined by participant beliefs

114   about group allocation relative to actual allocation, and blinding adequacy was determined by

115   critical appraisal. This review posed two questions: (1) 'Does blinding effectiveness moderate

116   intervention effect on pain?' and (2) 'Does blinding adequacy moderate intervention effect on

117   pain?'

Methods

The methods complied with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist (Moher et al. 2009). The protocol was prospectively registered with the International Prospective Register of Systematic Reviews (PROSPERO) (registration number: 42016029340; URL:

www.crd.york.ac.uk/prospero/display_record.asp?ID=CRD42016029340).

Disclosure of deviations from prospectively registered protocol

Following original registration on March 2, 2016, two changes were made to the protocol of this review: (1) the data extraction template was pilot tested using an iterative process rather than a sample of 10 included trials, and percentage agreement was used to determine agreement, rather than an intra-class correlation coefficient (ICC); and (2) the time-point that was used to investigate the influence of blinding effectiveness on pain outcomes was the time-point at which blinding was assessed (instead of the pre-defined time-points of immediate, short-term, and long-term pain assessments), because the authors agreed that this time-point would most accurately reflect intervention beliefs (i.e. blinding effectiveness) as beliefs can change over time (Bang et al. 2010; Rees et al. 2005). The second change was updated in PROSPERO prior to data analyses (revision posted on February 5, 2017). This review presents only review questions 1 and 2 of the protocol; review questions 3 and 4 will be reported elsewhere.

Information sources and search strategy

One reviewer (FAB) executed the search strategy. Databases (MEDLINE, EMBASE, AMED, Scopus, CINAHL, PEDro, The Cochrane Library) were searched from inception to February 2016. The general search terms were (needl* OR acupuncture OR intramuscular stimulation) AND (sham OR placebo*), and Medical Subject Headings (MeSH) were used where possible. The full electronic search strategy for MEDLINE is presented in Table 1. Searches were modified to suit the functionality of each database. Thesis databases (Trove, ProQuest) and clinical trial registries [Australian New Zealand Clinical Trials Registry (ANZCTR), Clinicaltrials.gov, World Health Organization International Clinical Trials Registry Platform (WHO ICTRP)] were crosschecked with database searches to identify further potential trials.

148 The reference lists of systematic reviews identified by the search were examined to locate

149 additional or unpublished trials. There were no limits on year, language, or publication status.

150

151 INSERT TABLE 1 ABOUT HERE

152

153 *Eligibility and study selection*

154 Trials were eligible for inclusion in this review if they (1) were prospective experimental designs

155 (e.g. randomised, non/quasi-randomised trials, pre-post, n-of-1) of any duration, which included

156 a 'real' dry needling intervention (referred to as 'active' dry needling in this review) and a

157 placebo/sham dry needling intervention; (2) included human adults (≥18 years of age) who were

158 asymptomatic or with symptomatic health conditions; (3) involved a recognised dry needling

159 approach with needle insertion sites based on anatomical or clinical rationales; (4) assessed and

160 reported an outcome for pain [visual analogue scale (VAS) or numeric rating scale (NRS)].

161 Trials were also eligible for inclusion if they reported blinding assessment data, without

162 reporting on pain, but the results from these trials are not presented in this review. Trials were

163 ineligible for inclusion if the needling therapy involved pre-designated needle insertion sites (e.g.

164 traditional acupuncture points) or involved injection of a substance (wet needling).

165

166 Records identified from the search strategy were exported to Endnote, duplicates were removed,

167 and the remaining records were imported into the online screening tool 'Covidence systematic

168 review software' (2018). Titles and abstracts were screened against the eligibility criteria by

169 three independent reviewers in teams of two (FAB and MPM or JLW), and trials potentially

170 meeting the criteria were progressed to full text review. The same three reviewers independently

171 screened the full-text articles in teams of two. Discrepancies were resolved through discussion,

172 with an independent third reviewer (MPM, JLW, or LSKL) consulted where necessary. Where

173 full-text was unavailable, authors were contacted to clarify eligibility and/or to provide full-texts.

174 Non-English publications were translated using Google Translate; the extracted data were then

175 checked with fluent speakers of each language.

176

177 *Data extraction and Risk of Bias (RoB) assessment*

178   A prospectively designed data extraction template was developed based on the Standards for

179   Reporting Interventions in Controlled Trials of Acupuncture (STRICTA) (MacPherson et al.

180   2010) and the Cochrane Handbook 'Checklist of items to consider in data collection or data

181   extraction' (Higgins & Green 2011). The domains of data extraction were: source details, trial

182   demographics, trial design, participant details, therapist details, intervention details, outcomes

183   (pain and blinding assessment), blinding strategies, sample size and dropouts, results (pain and

184   blinding assessment), and key conclusions of the authors.

185

186   The provisional data extraction template was pilot tested for inter-rater agreement by two

187   reviewers (FAB and JLW) using an iterative process (two randomly selected included trials in

188   each iteration). Once the pre-specified level of inter-rater agreement was established ($\geq$75%

189   agreement of items within an individual trial), two independent reviewers performed the

190   remaining data extraction (FAB and LSKL, JLW, or MPM), with a third reviewer consulted to

191   resolve disagreements as required.

192

193   Only data from the first phase of crossover trials were extracted due to the risk of carry-over

194   intervention effects. Where necessary (i.e. where no text or table data were provided), graphical

195   data were extracted using a ruler; if there were differences in these values between the two

196   extracting reviewers, the average value was calculated. Pain intensity data were converted to a

197   100-point continuous scale where required (e.g. if collected using a 10cm VAS or an NRS).

198

199   Risk of Bias (RoB) of individual trials was assessed using the Cochrane RoB assessment tool for

200   randomised trials (because all included studies were randomised trials) (Higgins et al. 2011).

201   Three key domains (allocation concealment, performance bias, detection bias) were determined *a*

202   *priori* based on relevance to the review questions. The key domains were informed by empirical

203   evidence for the likelihood and magnitude of these biases influencing trial outcomes (Higgins et

204   al. 2011; Hróbjartsson et al. 2014; Hróbjartsson et al. 2013; Savović et al. 2012; Wood et al.

205   2008). The overall RoB for individual trials was determined using the three key domains (low =

206   low RoB for all key domains, unclear = low or unclear RoB for all key domains, high = high

207   RoB for one or more key domains) (Higgins et al. 2011). Two independent reviewers appraised

208 each trial (FAB and MPM, JLW, or LSKL), with a third reviewer consulted to resolve

209 disagreements as required.

210

211 Publication bias for each meta-analysis was assessed by visual inspection of asymmetry of

212 funnel plots, which were contour-enhanced to allow consideration of the potential influence of

213 the statistical significance of trial outcomes on publication bias (Peters et al. 2008). A statistical

214 test for asymmetry was also computed for funnel plots containing ≥10 trials using the method

215 specified in Egger et al. (1997) at a significance level of p<0.10 (Higgins & Green 2011; Sterne

216 et al. 2011).

217

218 ***Data syntheses***

219 For both review questions, meta-analyses used generic inverse variance and random-effects

220 models. Restricted Maximum Likelihood (REML) was used to estimate between-trial variance.

221 Stata statistical software (version 15.1) (StataCorp. 2017) was used to compute inferential

222 statistics and create plots. The x^2 test and I^2 statistic were used to assess statistical heterogeneity;

223 p<0.10 was interpreted as statistically significant heterogeneity and I^2 >50% was interpreted as

224 substantial heterogeneity (Higgins & Green 2011). Intervention effects were interpreted as

225 statistically significant when p<0.05, and between-group effect sizes [Standardised Mean

226 Difference (SMD)] were considered large if >0.80, moderate if between 0.20 and 0.80, and small

227 if <0.20 as defined by Cohen (1988).

228

229 A blinding index (BI) (Bang et al. 2004) was used to quantify the effectiveness of blinding

230 (participant belief about group allocation relative to actual group allocation), where blinding

231 assessments were sufficiently reported. The BI estimates the degree of unblinding (i.e. correct

232 identification of group allocation) beyond random chance (Bang et al. 2004). To assist with

233 interpretation of blinding effectiveness, groups within included trials were classified based on the

234 BI cut-offs proposed by Moroz et al. (2013) (Table 2). Trials were then categorised based on

235 paired classifications for the active and sham groups, termed a 'blinding scenario' (e.g.

236 'Correct/Incorrect', which means that the active group was classified as 'Correct' and the sham

237 group was classified as 'Incorrect') (Bang et al. 2010). Using this classification method, a total of

238 nine blinding scenarios were possible (Bang et al. 2010). The 'R' software package (version

239 3.4.3) (R Core Team 2017) was used to compute BIs and their 95% Confidence Intervals (CIs).

240

241 INSERT TABLE 2 ABOUT HERE

242

243 *Review question 1: Does blinding effectiveness moderate intervention effect on pain?*

244 It was hypothesised that if the proportion of participants who believed they had the active or

245 sham intervention differed between active and sham groups (i.e. unbalanced intervention

246 beliefs), this would have a moderating effect on between-group pain outcomes (i.e. increase or

247 decrease between-group differences). To interrogate the hypothesis, a summary value for

248 blinding effectiveness for each trial was calculated by adding the BI scores from each group (i.e.

249 BI active group + BI sham group) [adapted from Freed et al. (2014)], and a meta-regression of

250 the influence of the summary BI (blinding effectiveness) on between-group effect size (pain) was

251 computed for the time-point closest to which blinding was assessed (as this is likely to most

252 accurately reflect intervention beliefs at that moment).

253

254 *Review question 2: Does blinding adequacy moderate intervention effect on pain?*

255 The Cochrane RoB tool (Higgins et al. 2011) was also used to assess blinding adequacy of trials.

256 Adequacy was based on the four RoB domains that relate to blinding (allocation concealment,

257 participant blinding, therapist blinding, and outcome assessor blinding) (Higgins et al. 2011)

258 [adapted from Feys et al. (2014)]. Trials were conservatively categorised as either '*adequately*

259 *blinded*' or '*inadequately blinded*' based on the following rules:

260 • Adequately blinded: low RoB across all four domains, or low RoB in the three domains

261   excluding 'therapist blinding' if no trials attempted therapist blinding.

262 • Inadequately blinded: high or unclear RoB in at least one domain.

263

264 Meta-analyses were used to assess differences in between-group effect sizes based on adequacy

265 of blinding. It was hypothesised that inadequate blinding would favour active dry needling.

266 Separate meta-analyses were completed for three time-periods: immediately after the first/only

267 intervention (<24 hours); short-term (24 hours to one month from baseline, using closest

268 assessment to one week); long-term (one to six months from baseline, using closest assessment

269 to three months).

270

271

272 **Results**

273 ***Outcome of search strategy***

274 The outcome of the search strategy is presented in Fig. 1. The search identified 11835 records.

275 Four additional publications were identified by searching personal records (Itoh et al. 2004) and

276 through hand searching reference lists of 199 systematic reviews (Itoh & Katsumi 2005; Itoh et

277 al. 2006b; Katsumi et al. 2004). Following removal of duplicates, 4894 potentially relevant

278 publications were screened. Title and abstract screening excluded 4280 publications. Of the

279 remaining 614 publications, 588 were excluded following full-text review, leaving 26

280 publications (Fig. 1). The exclusion of two research questions from the current review resulted in

281 the exclusion of three trials (within one publication) from this review because they did not report

282 a pain outcome (Braithwaite 2014) (this publication is included in Fig. 1 because the two omitted

283 review questions that did include results from this publication are reported elsewhere). The 25

284 relevant publications included one trial that presented results over two publications (Tough et al.

285 2010; Tough et al. 2009), and two single publications with two eligible sham groups (Itoh &

286 Katsumi 2005; Itoh et al. 2007); therefore, 25 publications (with 26 group comparisons from 24

287 trials) are presented in the current review. Of these 25 publications, 24 publications (with 25

288 group comparisons from 23 trials) provided sufficient data for inclusion in the current meta-

289 analyses. For the meta-analyses, in the two trials with two eligible sham groups (Itoh & Katsumi

290 2005; Itoh et al. 2007) the active group data were used twice.

291

292   INSERT FIG. 1 ABOUT HERE

293

294 Five non-English publications were included in the current three Japanese publications (Itoh &

295 Katsumi 2005; Itoh et al. 2006b; Katsumi et al. 2004) and two Spanish publications (Espejo

296 Antúnez et al. 2014; García-Gallego et al. 2011).

297

298 Of 22 authors who were contacted to clarify eligibility and/or to provide full-texts, 12 replied

299 confirming ineligibility and 10 did not reply. For eight further records, author contact details

300 could not be found. Nine authors of included trials were contacted to clarify trial details or

301 request data; one author replied stating they no longer had access to the data, and the remaining

302 eight authors did not reply.

303

304 ***Risk of Bias (RoB) assessment***

305 A summary of results for the RoB assessment is presented in Table 3. Overall RoB was high in

306 one trial, unclear in 18 trials, and low in five trials (Table 3). The areas with least RoB were

307 participant blinding and reporting bias (low RoB in all included trials). The areas with greatest

308 potential for bias were blinding of therapists and research personnel (high or unclear RoB in all

309 included trials), allocation concealment, and attrition bias (Table 3).

310

311 INSERT TABLE 3 ABOUT HERE

312

313 *Assessment of publication bias*

314 Visual inspection of asymmetry of contour-enhanced funnel plots suggested that publication bias

315 may be present (Fig. 2) (Peters et al. 2008). A statistical test for asymmetry was computed for

316 funnel plots containing ≥10 group comparisons (Fig. 2A, 2C, and 2D) and a statistically

317 significant result was found for all three plots (p<0.001, p=0.083, and p=0.061, respectively),

318 which further supports the presence of publication bias (Egger et al. 1997).

319

320 INSERT FIG. 2 ABOUT HERE

321

322

323 ***Description of included trials***

324 Table 4 presents a summary of trial characteristics and results of the 24 trials involving 26 group

325 comparisons. Non-penetrating (NP) shams were used in 16 group comparisons, penetrating (P)

326 shams were used in nine group comparisons, and one group comparison used anaesthesia

327 (general or spinal) to blind participants (Mayoral et al. 2013) (Table 4). The 16 NP devices were

328 guide-tubes alone (N=3 group comparisons), custom-made blunted/retracting needles (N=12

329   group comparisons), and one commercial device [the Park sham (Park et al. 1999)]. Of the nine

330   group comparisons that used penetrating shams, six inserted needles subcutaneously only [i.e.

331   superficial dry needling above trigger points (TrP SDN) or away from trigger points (Non-TrP

332   SDN)], and three inserted needles into muscle but away from trigger points (Non-TrP DN)

333   (Table 4).

334

335   INSERT TABLE 4 ABOUT HERE

336

337   Fourteen trials (16 group comparisons) assessed blinding effectiveness (N=13 trials) or

338   intervention credibility (N=1 trial). To assess blinding effectiveness, participants were asked

339   whether they thought a needle had been inserted (N=9 trials), if they felt a 'needling sensation'

340   (N=1 trial), or which group they thought they were in (N=1 trial). The remaining two trials did

341   not report how blinding was assessed (Huguenin et al. 2005; Sterling et al. 2015). To assess

342   intervention credibility, participants completed the Credibility/Expectancy Questionnaire (CEQ)

343   (Devilly & Borkovec 2000) (N=1 trial). Of the 13 trials that assessed blinding effectiveness, 10

344   trials (12 group comparisons) presented blinding data in a way that the BI could be calculated for

345   active and sham groups (Table 4). To evaluate blinding effectiveness, nine trials used inferential

346   statistics to determine if there was a difference in proportions of guesses, and four trials

347   described, but did not statistically analyse, the blinding data (Table 4). For pain outcome

348   assessments, 22 trials used a VAS (N=11 trials used a 100mm scale and N=11 trials used a 10cm

349   scale) and two trials used an NRS (NRS 0-10).

350

351   ***Data syntheses***

352   *Review question 1: Does blinding effectiveness moderate intervention effect on pain?*

353   Figure 3 presents a bubble plot (meta-regression) of the influence of the summary BI on effect

354   size (pain). There was no evidence of a moderating effect of the summary BI on effect size

355   [meta-regression coefficient -1.87 (95% CI -5.63 to 1.88); p=0.292; N=12 group comparisons;

356   n=248] (Fig. 3). There was evidence of statistically significant and substantial statistical

357   heterogeneity (I^2 =79.0%; p<0.001) (Higgins & Green 2011).

358

359   INSERT FIG. 3 ABOUT HERE

360

361   *Review question 2: Does blinding adequacy moderate intervention effect on pain?*

362   Five of the 24 trials were adequately blinded (Table 3). All trials demonstrated adequate

363   participant blinding and no trials attempted to blind therapists, so by default blinding adequacy

364   was determined based on the remaining two domains (allocation concealment and blinding of

365   outcome assessors).

366

367   *Immediate intervention effect (<24 hours after the first/only intervention)*

368   There were seven group comparisons where immediate pain outcomes were collected (Fig. 4).

369   One group comparison (n=31) met the requirements for adequate blinding (Inoue et al. 2006),

370   and intervention effects were statistically significant in favour of active dry needling [SMD -0.76

371   (95% CI -1.49 to -0.03)]. For inadequately blinded group comparisons (N=6; n=206), there was

372   no evidence of a difference in intervention effects between active and sham groups [pooled SMD

373   -0.47 (95% CI -0.95 to 0.02)]. There was evidence of significant and substantial heterogeneity in

374   the pooled group comparisons (Higgins & Green 2011) (Fig. 4).

375

376   INSERT FIG. 4 ABOUT HERE

377

378   *Short-term intervention effect (24 hours to one month, closest assessment to one week)*

379   There were 20 group comparisons where short-term pain outcomes were collected (Fig. 5). For

380   adequately blinded group comparisons (N=3; n=122) there was no evidence of a difference in

381   intervention effects between active and sham groups [pooled SMD -0.40 (95% CI -0.96 to 0.15)],

382   whereas inadequately blinded group comparisons (N=17; n=504) had statistically significant

383   intervention effects that favoured active dry needling [pooled SMD -0.71 (95% CI -1.05 to -

384   0.38)]. There was evidence of statistically significant and substantial heterogeneity of pooled

385   group comparisons for the inadequately blinded subgroup, whereas the adequately blinded

386   subgroup had moderate heterogeneity that was not significant (Higgins & Green 2011) (Fig. 5).

387

388   INSERT FIG. 5 ABOUT HERE

389

390   *Long-term intervention effect (one to six months, closest assessment to three months)*

391   There were 16 group comparisons where long-term pain outcomes were collected (Fig. 6). For

392   adequately blinded group comparisons (N=4; n=202) there was no evidence of a difference in

393   intervention effects between active and sham groups [pooled SMD -0.30 (95% CI -0.62 to 0.02)],

394   whereas inadequately blinded group comparisons (N=12; n=284) had statistically significant

395   intervention effects that favoured active dry needling [pooled SMD -1.14 (95% CI -1.64 to -

396   0.65)]. There was evidence of statistically significant and substantial heterogeneity of pooled

397   group comparisons for the inadequately blinded subgroup, whereas the adequately blinded

398   subgroup had low heterogeneity that was not significant (Higgins & Green 2011) (Fig. 6).

399

400   INSERT FIG. 6 ABOUT HERE

401

402   **Discussion**

403   ***Key findings***

404   This review aimed to determine whether blinding effectiveness and/or blinding adequacy

405   moderated pain outcomes in dry needling trials. Of the 23 trials included in the meta-analyses,

406   only 10 (43.5%) reported data that were sufficient to calculate the BI, and only five (21.7%)

407   reported adequate blinding procedures. The small number and size of included trials meant that

408   there was insufficient evidence to determine if a moderating effect of blinding effectiveness or

409   adequacy existed (Button et al. 2013; Higgins & Green 2011).

410

411   *Review question 1: Does blinding effectiveness moderate intervention effect on pain?*

412   Blinding effectiveness was determined based on participant beliefs about whether they received

413   active or sham dry needling. Table 5 presents the hypothesised moderation effect of the nine

414   possible blinding scenarios on pain outcomes [adapted from Bang et al. (2010)], and the number

415   of group comparisons in this review that fell into those scenarios. In this hypothesis, within each

416   group (active or sham), intervention benefits would increase as more participants believe they

417   received active dry needling (⬆; Table 5). Theoretically, effective blinding would exist in

418   Scenarios 4, 5, or 6, where intervention beliefs were approximately balanced between groups

419   (shaded in Table 5). In contrast, the imbalance in active and sham groups in Scenarios 1-3 would

420   favour the sham group and in 7-9 would favour the active group.

421

422   INSERT TABLE 5 ABOUT HERE

423

424   It was hypothesised that unbalanced beliefs between active and sham groups would moderate

425   between-group differences in pain. No evidence of a moderating effect of the summary BI on

426   pain was found, but the analysis may have been underpowered to detect it, as too it may have

427   been underpowered to confidently conclude against it. Only 12 group comparisons were included

428   in the analysis, marginally more than the minimum recommended number (N=10) for a meta-

429   regression (Higgins & Green 2011). The current findings are in contrast to previous studies

430   where significant associations between intervention outcomes and beliefs about allocation have

431   been demonstrated (Baethge et al. 2013; Dar et al. 2005; McRae et al. 2004), including pain

432   outcomes in acupuncture trials (Bausell et al. 2005; Vase et al. 2013; White et al. 2012).

433   Intervention effects favoured active dry needling irrespective of whether intervention beliefs

434   were balanced between groups; this finding is consistent with Moroz et al. (2013) who found that

435   the majority of acupuncture and dry needling trials reported positive outcomes, regardless of

436   blinding effectiveness. The findings of the meta-regression are likely to have been threatened by

437   the inclusion of underpowered trials (Button et al. 2013), by other threats to internal validity [e.g.

438   biases associated with therapist expectation (Gracely et al. 1985) and the high likelihood of

439   publication bias (Fig. 2A)], and by substantial statistical heterogeneity and confounding due to

440   the non-randomised nature of the meta-regression analysis (Higgins & Green 2011). Therefore,

441   further research is needed to quantify a moderating effect of blinding effectiveness on pain

442   outcomes.

443

444   Inconsistent techniques and incomplete reporting of blinding assessments make it difficult to

445   draw robust conclusions. Overall, 14 trials (58%) in this review reported some form of blinding

446   effectiveness or intervention credibility data, which is markedly greater than in previous samples

447   [e.g. between 2-8% of random samples of clinical trials reported assessments of blinding

448   (Fergusson et al. 2004; Hróbjartsson et al. 2007)]. However, of these 14 trials, only 10 reported

449   data that were sufficient to calculate the BI. Where reported blinding data were insufficient to

450   calculate the BI, authors were contacted to request the raw data but authors either did not

451   respond or no longer had access to the data. Given the strong motivation to report success, there

452 is a possibility of underreporting when blinding assessments indicate ineffective blinding (i.e.

453 reporting bias) (Hróbjartsson et al. 2007).

454

455 The lack of data to confirm the influence of blinding effectiveness on trial outcomes means that

456 currently blinding 'success' is largely subjective (Bang et al. 2010). This is evidenced by the

457 universal author conclusion of blinding 'success' (where reported), despite varied patterns in the

458 blinding data (Table 4).

459

460 *Review question 2: Does blinding adequacy moderate intervention effect on pain?*

461 It was hypothesised that inadequate blinding procedures would exaggerate intervention effects.

462 Threats to the internal validity of included trials, coupled with the limitations of meta-analytical

463 techniques precluded definitive conclusions. For immediate assessments, there was no evidence

464 of a difference in intervention effects between adequately and inadequately blinded group

465 comparisons (Fig. 4), but drawing inferences from this finding is difficult due to the small

466 sample of group comparisons (N=7; with only N=1 adequately blinded). However, with the

467 caveats of small samples, generally unclear RoB, and the limitations of subgroup analyses, the

468 available evidence suggests that inadequate blinding procedures could lead to exaggerated

469 intervention effects in dry needling trials in the short-term and long-term.

470

471 In the short-term and long-term, there were statistically significant intervention effects in favour

472 of active dry needling for inadequately blinded group comparisons, whereas adequately blinded

473 group comparisons showed no difference between groups (Fig. 5 and 6). Differences in pooled

474 pain outcomes between inadequately and adequately blinded group comparisons were moderate

475 to large (short-term difference in SMD=0.31; long-term difference in SMD=0.84), and in the

476 long-term there was no overlap of pooled 95% CIs (i.e. significance guaranteed at $p<0.05$). In

477 addition, in both the short-term and long-term, the adequately blinded group comparisons had

478 more statistically homogenous results, and in the long-term the 95% CI for adequately blinded

479 comparisons was also more precise despite having fewer group comparisons. These findings

480 together suggest that inadequate blinding might be associated with greater heterogeneity and

481 lower precision in group comparisons.

482

483   The current findings are consistent with the findings of previous meta-analyses investigating

484   moderating effects of inadequate blinding procedures (Hróbjartsson et al. 2014; Savović et al.

485   2012). More specifically, exaggeration of intervention effects has been found in trials with

486   inadequate allocation concealment and/or outcome assessor blinding (Hróbjartsson et al. 2012;

487   Hróbjartsson et al. 2013; Jüni et al. 2001; Nüesch et al. 2009; Schulz et al. 1995; Wood et al.

488   2008), and these two domains were the only determinates of blinding adequacy in the current

489   review because all included trials demonstrated adequate participant blinding and no trials

490   attempted to blind therapists.

491

492   Lack of adequate blinding procedures means that at present, specific effects of dry needling

493   cannot be distinguished from effects due to bias. Blinding of therapists and research personnel

494   was either not attempted or poorly reported by all included trials (Table 3). There are clearly

495   substantial practical challenges with therapist blinding, however, potential effects of non-blinded

496   therapists (Cook et al. 2013; Gracely et al. 1985; Moher et al. 2010; Savović et al. 2012; Vase et

497   al. 2015) warrants research in this direction. Acupuncture studies have attempted therapist

498   blinding using custom-made sham needle devices (Takakura et al. 2010; Takakura & Yajima

499   2007), which may have potential for application in future dry needling trials. Blinding of

500   research personnel should be a relatively simple procedure and needs greater attention and/or

501   clearer reporting. Participant attrition, another major source of potential bias, should be

502   accounted for using statistical methods such as intention-to-treat analysis using multiple

503   imputation, possibly with adjustments for informative missingness, or adjustments based on

504   covariates. In addition, there were more dropouts in sham groups due to 'no response to

505   intervention' (where reported: n=15 in sham groups versus n=4 in active groups; Table 4), which

506   could have contributed to biases favouring active dry needling.

507

508   ***Strengths and limitations***

509   The strengths of this review included prospective peer review and registration of the protocol

510   (PROSPERO), adherence to the PRISMA statement for reporting (Moher et al. 2009), and

511   independent screening for trial eligibility, data extraction, and RoB assessments by two

512   reviewers. The search strategy was comprehensive and trials were not limited to the English

513   language. Despite attempts to limit the impact of publication bias on the current results by

514 searching trial registrations and thesis databases, asymmetry of funnel plots suggests publication

515 bias was present (Fig. 2). The small number of trials in one of the funnel plots (<10; Fig. 2B)

516 meant that it could not be confidently interpreted (Egger et al. 1997; Higgins & Green 2011).

517

518 The current findings should be interpreted with caution. The meta-analytical techniques used in

519 this review are not randomised comparisons and are therefore observational in nature (Higgins &

520 Green 2011). The strength of inferences is therefore limited by potential confounding by

521 uncontrolled covariates, and subgroups may have differed in capacity to detect effects (Higgins

522 & Green 2011). However, for review question 2, the *a priori* hypothesis, the statistical

523 significance of the findings, and the consistency of the difference across comparisons strengthen

524 the validity of the inferences (Oxman & Guyatt 1992). That no studies to date have made head-

525 to-head comparisons of blinded versus non-blinded dry needling interventions precludes any

526 further analysis, aside from indirect comparisons. The meta-analyses used random effects

527 modelling to allow for statistical heterogeneity between trials (Higgins & Green 2011).

528 Homogeneity was improved for meta-analyses investigating blinding adequacy (I^2 values <75%;

529 Fig. 4, 5, 6), in which comparisons were grouped based on four RoB domains, for three pre-

530 defined time periods, so these analyses may be more reliable (Higgins & Green 2011).

531

532 The included trials were methodologically heterogeneous, and many were likely to have had a

533 high risk of null findings in the presence of small to moderate effects due to insufficient power

534 (N=20 group comparisons with n<50 participants, with power clearly achieved for the pain

535 outcome in only three trials, and zero for blinding assessment outcomes) (Button et al. 2013).

536 Trials were also clinically diverse in terms of participant health condition, pain chronicity, age,

537 and intervention dose, which may have confounded results, in particular because the aetiology of

538 pain may influence the specific effects of dry needling (Cagnie et al. 2013; Dommerholt 2011),

539 as well as non-specific effects (Tracey 2010). The limited number of trials precluded

540 investigation of potential covariates (i.e. sensitivity or multivariable meta-regression analyses)

541 (Higgins & Green 2011).

542

543 Contrary to best practice, active group data were used twice in several meta-analyses that

544 included trials with two eligible sham groups (Itoh & Katsumi 2005; Itoh et al. 2007) (Fig. 3, 5,

545 6), which may have caused unit-of-analysis errors due to correlations between the non-

546 independent comparisons (Higgins & Green 2011). However, due to extremely small sample

547 sizes (n=8 to 10 participants in the relevant active groups) and potential differences in

548 physiological effects of the sham interventions (i.e. penetrating versus non-penetrating) (Lund et

549 al. 2009), it was decided that the active group data could not be split, nor could the sham groups

550 be combined as recommended by Higgins and Green (2011).

551

552 To determine whether the current review required updating (original search completed in

553 February 2016), a citation search was undertaken for trials included in the current systematic

554 review (292 citations since January 2016 as at 18th of September 2017, with reference lists of 39

555 potentially relevant systematic reviews also reviewed). This search revealed 47 new prospective

556 primary studies of dry needling; of these, only one was blinded using sham dry needling and this

557 trial did not report an assessment of blinding effectiveness (Mason et al. 2016). Addition of one

558 trial to the current review was unlikely to significantly alter the results, therefore the review was

559 not updated (Elkins 2018).

560

561 **Conclusions**

562 This review found insufficient data to understand moderating effects of blinding effectiveness or

563 adequacy on pain; therefore recommendations about interpreting trial outcomes with reference to

564 blinding are premature. However, consistent with previous reviews, the current review found a

565 bias in favour of active dry needling when trials were inadequately blinded for short-term and

566 long-term pain outcomes. Due to the limitations of subgroup comparative analyses and threats to

567 the validity of the included trials (particularly insufficient power), the findings of this review

568 should be interpreted with caution. We did not aim to determine whether or not dry needling is

569 superior to sham, but we can confidently conclude that should researchers propose further trials

570 in this or related areas, they should be adequately blinded and collect robust blinding data.

571

572

573 **Acknowledgements**

574 The authors would like to thank Dr. Tasha R. Stanton for her valuable assistance in the
575 preparation of this manuscript, and Dr. Beben Benyamin and Dr. Terry Boyle for their statistical
576 support with the meta-analytical techniques.

577

References

579 2018. Covidence systematic review software. Melbourne, Australia: Veritas Health Innovation,
580   <http://www.covidence.org>.

581 Armijo-Olivo S, Fuentes J, da Costa BR, Saltaji H, Ha C, and Cummings GG. 2017. Blinding in
582   physical therapy trials and its association with treatment effects: a meta-epidemiological
583   study. *American Journal of Physical Medicine and Rehabilitation* 96:34-44.

584 Baethge C, Assall OP, and Baldessarini RJ. 2013. Systematic review of blinding assessment in
585   randomized controlled trials in schizophrenia and affective disorders 2000-2010.
586   *Psychotherapy and Psychosomatics* 82:152-160.

587 Bang H, Flaherty SP, Kolahi J, and Park JJ. 2010. Blinding assessment in clinical trials: a review
588   of statistical methods and a proposal of blinding assessment protocol. *Clinical Research*
589   *and Regulatory Affairs* 27:42-51.

590 Bang H, Ni L, and Davis CE. 2004. Assessment of blinding in clinical trials. *Controlled Clinical*
591   *Trials* 25:143-156.

592 Bausell RB, Lao L, Bergman S, Lee W-L, and Berman BM. 2005. Is acupuncture analgesia an
593   expectancy effect? Preliminary evidence based on participants' perceived assignments in
594   two placebo-controlled trials. *Evaluation and the Health Professions* 28:9-26.

595 Benedetti F. 2013. Placebo and the new physiology of the doctor-patient relationship.
596   *Physiological Reviews* 93:1207-1246.

597 Boutron I, Guittet L, Estellat C, Moher D, Hróbjartsson A, and Ravaud P. 2007. Reporting
598   methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS*
599   *Medicine* 4:370-380.

600 Boutron I, Tubach F, Giraudeau B, and Ravaud P. 2004. Blinding was judged more difficult to
601   achieve and maintain in nonpharmacologic than pharmacologic trials. *Journal of Clinical*
602   *Epidemiology* 57:543-550.

603 Braithwaite FA. 2014. Testing a sham dry needle for the cervical spine in healthy adults: a
604   randomised controlled trial. Honours thesis. University of South Australia.

605 Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, and Munafò MR. 2013.
606   Power failure: why small sample size undermines the reliability of neuroscience. *Nature*
607   *Reviews Neuroscience* 14:365-376.

608 Cagnie B, Dewitte V, Barbe T, Timmermans F, Delrue N, and Meeus M. 2013. Physiologic
609   effects of dry needling. *Current Pain and Headache Reports* 17:1-8.

610 Carlesso LC, MacDermid JC, Gross AR, Walton DM, and Santaguida PL. 2014. Treatment
611   preferences amongst physical therapists and chiropractors for the management of neck
612   pain: results of an international survey. *Chiropractic & Manual Therapies* 22:11.

613 Cohen J. 1988. *Statistical power analysis for the behavioral sciences*. Hillsdale, New Jersey:
614   Lawrence Erlbaum Associates.

615 Colagiuri B, and Smith CA. 2011. A systematic review of the effect of expectancy on treatment
616   responses to acupuncture. *Evidence-Based Complementary & Alternative Medicine*
617   2012:12 pages.

618 Cook C, Learman K, Showalter C, Kabbaz V, and O'Halloran B. 2013. Early use of thrust
619   manipulation versus non-thrust manipulation: a randomized clinical trial. *Manual*
620   *Therapy* 18:191-198.

621 Cotchett MP, Munteanu SE, and Landorf KB. 2014. Effectiveness of trigger point dry needling
622   for plantar heel pain: a randomized controlled trial. *Physical Therapy* 94:1083-1094.

623 Dar R, Stronguin F, and Etter J-F. 2005. Assigned versus perceived placebo effects in nicotine
624   replacement therapy for smoking reduction in Swiss smokers. *Journal of Consulting and*
625   *Clinical Psychology* 73:350-353.

626 Devilly GJ, and Borkovec TD. 2000. Psychometric properties of the credibility/expectancy
627   questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry* 31:73-86.

628 Dıraçoğlu D, Vural M, Karan A, and Aksoy C. 2012. Effectiveness of dry needling for the
629   treatment of temporomandibular myofascial pain: a double-blind, randomized, placebo
630   controlled study. *Journal of Back and Musculoskeletal Rehabilitation* 25:285-290.

631 Dommerholt J. 2011. Dry needling - peripheral and central considerations. *Journal of Manual &*
632   *Manipulative Therapy* 19:223-227.

633 Egger M, Smith GD, Schneider M, and Minder C. 1997. Bias in meta-analysis detected by a
634   simple, graphical test. *BMJ* 315:629-634.

635 Elkins M. 2018. Updating systematic reviews. *Journal of Physiotherapy* 64:1-3.

636 Espejo Antúnez L, Gacimartín García A, Pérez Cardeñosa MR, Cardero Durán MA, de la Cruz-
637   Torres B, and Albornoz-Cabello M. 2014. Efectos sobre la tensión neural adversa medida
638   mediante test de slump tras punción seca de punto gatillo miofascial del músculo
639   gastrocnemio [Effects on adverse neural tension by slump test after dry needling of
640   myofascial trigger point of the gastrocnemius muscle]. *Fisioterapia* 36:127-134.
641 Fergusson D, Glass KC, Waring D, and Shapiro S. 2004. Turning a blind eye: the success of
642   blinding reported in a random sample of randomised, placebo controlled trials. *BMJ*
643   328:5 pages.
644 Feys F, Bekkering GE, Singh K, and Devroey D. 2014. Do randomized clinical trials with
645   inadequate blinding report enhanced placebo effects for intervention groups and nocebo
646   effects for placebo groups? *Systematic Reviews* 3:14 pages.
647 Finniss DG, Kaptchuk TJ, Miller F, and Benedetti F. 2010. Biological, clinical, and ethical
648   advances of placebo effects. *Lancet* 375:686-695.
649 Freed B, Assall OP, Panagiotakis G, Bang H, Park JJ, Moroz A, and Baethge C. 2014. Assessing
650   blinding in trials of psychiatric disorders: A meta-analysis based on blinding index.
651   *Psychiatry Research* 219:241-247.
652 García-Gallego R, Tormos-Claramunt L, Vilanova-Salcedo P, Morales-Rodríguez R, Pérez-
653   Villalba A, and Segura-Ortí E. 2011. Efectividad de la punción seca de un punto gatillo
654   miofascial versus manipulación de codo sobre el dolor y fuerza máxima de prensión de la
655   mano [Effectiveness of a myofascial trigger point dry needling versus elbow
656   manipulation on pain and maximum hand grip strength]. *Fisioterapia* 33:248-255.
657 Gracely RH, Dubner R, Deeter WR, and Wolskee PJ. 1985. Clinicians' expectations influence
658   placebo analgesia. *Lancet* 325:43.
659 Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Savović J, Schulz KF,
660   Weeks L, and Sterne JA. 2011. The Cochrane Collaboration's tool for assessing risk of
661   bias in randomised trials. *BMJ* 343:9 pages.
662 Higgins JP, and Green S. 2011. Cochrane handbook for systematic reviews of interventions. In:
663   Higgins JP, Altman DG, and Sterne JA, editors. Cochrane handbook for systematic
664   reviews of interventions. Chichester: The Cochrane Collaboration.

665 Hróbjartsson A, Emanuelsson F, Thomsen ASS, Hilden J, and Brorson S. 2014. Bias due to lack
666   of patient blinding in clinical trials. A systematic review of trials randomizing patients to
667   blind and nonblind sub-studies. *International Journal of Epidemiology* 43:1272-1283.
668 Hróbjartsson A, Forfang E, Haahr M, Als-Nielsen B, and Brorson S. 2007. Blinded trials taken
669   to the test: an analysis of randomized clinical trials that report tests for the success of
670   blinding. *International Journal of Epidemiology* 36:654-663.
671 Hróbjartsson A, Thomsen ASS, Emanuelsson F, Tendal B, Hilden J, Boutron I, Ravaud P, and
672   Brorson S. 2012. Observer bias in randomised clinical trials with binary outcomes:
673   systematic review of trials with both blinded and non-blinded outcome assessors. *BMJ*
674   344:11 pages.
675 Hróbjartsson A, Thomsen ASS, Emanuelsson F, Tendal B, Hilden J, Boutron I, Ravaud P, and
676   Brorson S. 2013. Observer bias in randomized clinical trials with measurement scale
677   outcomes: a systematic review of trials with both blinded and nonblinded assessors.
678   *Canadian Medical Association Journal* 185:e201-e211.
679 Huguenin L, Brukner PD, McCrory P, Smith P, Wajswelner H, and Bennell K. 2005. Effect of
680   dry needling of gluteal muscles on straight leg raise: a randomised, placebo controlled,
681   double blind trial. *British Journal of Sports Medicine* 39:84-90.
682 Inoue M, Kitakoji H, Ishizaki N, Tawa M, Yano T, Katsumi Y, and Kawakita K. 2006. Relief of
683   low back pain immediately after acupuncture treatment - a randomised, placebo
684   controlled trial. *Acupuncture in Medicine* 24:103-108.
685 Itoh K, Asai S, Ohyabu H, Imai K, and Kitakoji H. 2012. Effects of trigger point acupuncture
686   treatment on temporomandibular disorders: a preliminary randomized clinical trial.
687   *Journal of Acupuncture & Meridian Studies* 5:57-62.
688 Itoh K, Hirota S, Katsumi Y, Ochi H, and Kitakoji H. 2008. Trigger point acupuncture for
689   treatment of knee osteoarthritis - a preliminary RCT for a pragmatic trial. *Acupuncture in*
690   *Medicine* 26:17-26.
691 Itoh K, and Katsumi T. 2005. 高齢者の慢性腰下肢痛に対する鍼治療の効果 [Effect of
692   acupuncture on chronic lumbar pain in the elderly: comparative examination on the
693   usefulness of trigger point acupuncture]. *全日本鍼灸学会雑誌* 55:530-537.

694   Itoh K, Katsumi Y, Hirota S, and Kitakoji H. 2006a. Effects of trigger point acupuncture on
695   chronic low back pain in elderly patients - a sham-controlled randomised trial.
696   *Acupuncture in Medicine* 24:5-12.

697   Itoh K, Katsumi Y, Hirota S, and Kitakoji H. 2007. Randomised trial of trigger point
698   acupuncture compared with other acupuncture for treatment of chronic neck pain.
699   *Complementary Therapies in Medicine* 15:172-179.

700   Itoh K, Katsumi Y, and Kitakoji H. 2004. Trigger point acupuncture treatment of chronic low
701   back pain in elderly patients - a blinded RCT. *Acupuncture in Medicine* 22:170-177.

702   Itoh K, Saito S, Sahara S, Naitoh Y, Imai K, and Kitakoji H. 2014. Randomized trial of trigger
703   point acupuncture treatment for chronic shoulder pain: a preliminary study. *Journal of*
704   *acupuncture and meridian studies* 7:59-64.

705   Itoh K, Wave M, Reiyo N, Kawamoto M, Hideki O, and Hiroshi K. 2006b. 大学生の肩こり被
706   験者を対象にしたトリガーポイント鍼治療の試み [Trigger point for college students
707   with shoulder stiffness trial of acupuncture treatment: questionnaire survey on shoulder
708   stiffness and clinical trial on the effect of acupuncture]. *全日本鍼灸学会雑誌*56:150-
709   157.

710   Jüni P, Altman DG, and Egger M. 2001. Systematic reviews in health care: assessing the quality
711   of controlled clinical trials. *BMJ* 323:42-46.

712   Kaptchuk TJ. 2002. The placebo effect in alternative medicine: can the performance of a healing
713   ritual have clinical significance? *Annals of Internal Medicine* 136:817-825.

714   Kaptchuk TJ, Kelley JM, Conboy LA, Davis RB, Kerr CE, Jacobson EE, Kirsch I, Schyner RN,
715   Nam BH, and Nguyen LT. 2008. Components of placebo effect: randomised controlled
716   trial in patients with irritable bowel syndrome. *BMJ* 336:999-1003.

717   Kaptchuk TJ, and Miller FG. 2015. Placebo effects in medicine. *New England Journal of*
718   *Medicine* 373:8-9.

719   Kaptchuk TJ, Stason WB, Davis RB, Legedza AR, Schnyer RN, Kerr CE, Stone DA, Nam BH,
720   Kirsch I, and Goldman RH. 2006. Sham device v inert pill: randomised controlled trial of
721   two placebo treatments. *BMJ* 332:391-397.

722   Katsumi Y, Itoi M, Kojima A, Takatori R, Totani Y, Hirasawa Y, and Itoh K. 2004. 高齢者の慢
723   性腰痛に対する阿是穴鍼療法 (< ワークショップ> 鍼灸とリハビリテーション)(第

724   41 回 日本リハビリテーション学術集会) [Tender point acupuncture of chronic low

725   back pain in aged patients]. *リハビリテーション医学: 日本リハビリテーション医学*

726   *会誌*41:824-829.

727 Legge D. 2014. A history of dry needling. *Journal of Musculoskeletal Pain* 22:301-307.

728 Linde K, Witt CM, Streng A, Weidenhammer W, Wagenpfeil S, Brinkhaus B, Willich SN, and

729   Melchart D. 2007. The impact of patient expectations on outcomes in four randomized

730   controlled trials of acupuncture in patients with chronic pain. *Pain* 128:264-271.

731 Lund I, Näslund J, and Lundeberg T. 2009. Minimal acupuncture is not a valid placebo control in

732   randomised controlled trials of acupuncture: a physiologist's perspective. *Chinese*

733   *Medicine* 4:9 pages.

734 Machado L, Kamper S, Herbert R, Maher C, and McAuley J. 2008. Imperfect placebos are

735   common in low back pain trials: a systematic review of the literature. *European Spine*

736   *Journal* 17:889-904.

737 MacPherson H, Altman DG, Hammerschlag R, Youping L, Taixiang W, White A, and Moher D.

738   2010. Revised standards for reporting interventions in clinical trials of acupuncture

739   (STRICTA): extending the CONSORT statement. *Journal of Evidence-Based Medicine*

740   3:140-155.

741 Mason JS, Crowell M, Dolbeer J, Morris J, Terry A, Koppenhaver S, and Goss DL. 2016. The

742   effectiveness of dry needling and stretching vs. stretching alone on hamstring flexibility

743   in patients with knee pain: a randomized controlled trial. *International Journal of Sports*

744   *Physical Therapy* 11:672-683.

745 Mayoral O, Salvat I, Martin MT, Martin S, Santiago J, Cotarelo J, and Rodriguez C. 2013.

746   Efficacy of myofascial trigger point dry needling in the prevention of pain after total knee

747   arthroplasty: a randomized, double-blinded, placebo-controlled trial. *Evidence-Based*

748   *Complementary & Alternative Medicine* 2013:8 pages.

749 McMillan AS, Nolan A, and Kelly PJ. 1997. The efficacy of dry needling and procaine in the

750   treatment of myofascial pain in the jaw muscles. *Journal of Orofacial Pain* 11:307-314.

751 McRae C, Cherin E, Yamazaki TG, Diem G, Vo AH, Russell D, Ellgring JH, Fahn S, Greene P,

752   and Dillon S. 2004. Effects of perceived treatment on quality of life and medical

753   outcomesin a double-blind placebo surgery trial. *Archives of General Psychiatry* 61:412-

754   420.

755 Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux P, Elbourne D, Egger
756   M, and Altman DG. 2010. CONSORT 2010 explanation and elaboration: updated
757   guidelines for reporting parallel group randomised trials. *BMJ* 340:28 pages.
758 Moher D, Liberati A, Tetzlaff J, and Altman DG. 2009. Preferred reporting items for systematic
759   reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine*
760   151:264-269.
761 Moroz A, Freed B, Tiedemann L, Bang H, Howell M, and Park JJ. 2013. Blinding measured: a
762   systematic review of randomized controlled trials of acupuncture. *Evidence-Based*
763   *Complementary & Alternative Medicine* 2013:12 pages.
764 Moseley AM, Herbert RD, Maher CG, Sherrington C, and Elkins MR. 2011. Reported quality of
765   randomized controlled trials of physiotherapy interventions has improved over time.
766   *Journal of Clinical Epidemiology* 64:594-601.
767 Myburgh C, Hartvigsen J, Aagaard P, and Holsgaard-Larsen A. 2012. Skeletal muscle
768   contractility, self-reported pain and tissue sensitivity in females with neck/shoulder pain
769   and upper Trapezius myofascial trigger points - a randomized intervention study.
770   *Chiropractic & Manual Therapies* 20:10 pages.
771 Nabeta T, and Kawakita K. 2002. Relief of chronic neck and shoulder pain by manual
772   acupuncture to tender points - a sham-controlled randomized trial. *Complementary*
773   *Therapies in Medicine* 10:217-222.
774 Nüesch E, Reichenbach S, Trelle S, Rutjes A, Liewald K, Sterchi R, Altman DG, and Jüni P.
775   2009. The importance of allocation concealment and patient blinding in osteoarthritis
776   trials: a meta-epidemiologic study. *Arthritis and Rheumatism* 61:1633-1641.
777 Oxman AD, and Guyatt GH. 1992. A consumer's guide to subgroup analyses. *Annals of Internal*
778   *Medicine* 116:78-84.
779 Park J, White A, Lee H, and Ernst E. 1999. Development of a new sham needle. *Acupuncture in*
780   *Medicine* 17:110-112.
781 Pecos-Martín D, Montañez-Aguilera FJ, Gallego-Izquierdo T, Urraca-Gesto A, Gómez-Conesa
782   A, Romero-Franco N, and Plaza-Manzano G. 2015. Effectiveness of dry needling on the
783   lower trapezius in patients with mechanical neck pain: a randomized controlled trial.
784   *Archives of Physical Medicine and Rehabilitation* 96:775-781.

785 Peters JL, Sutton AJ, Jones DR, Abrams KR, and Rushton L. 2008. Contour-enhanced meta-
786   analysis funnel plots help distinguish publication bias from other causes of asymmetry.
787   *Journal of Clinical Epidemiology* 61:991-996.

788 R Core Team. 2017. R: A language and environment for statistical computing. Vienna, Austria:
789   R Foundation for Statistical Computing, <http://www.R-project.org/>.

790 Rees JR, Wade TJ, Levy DA, Colford JM, and Hilton JF. 2005. Changes in beliefs identify
791   unblinding in randomized controlled trials: a method to meet CONSORT guidelines.
792   *Contemporary Clinical Trials* 26:25-37.

793 Savović J, Jones HE, Altman DG, Harris RJ, Jüni P, Pildal J, Als-Nielsen B, Balk EM, Gluud C,
794   and Gluud LL. 2012. Influence of reported study design characteristics on intervention
795   effect estimates from randomized, controlled trials. *Annals of Internal Medicine* 157:429-
796   438.

797 Schulz KF, Chalmers I, Hayes RJ, and Altman DG. 1995. Empirical evidence of bias:
798   dimensions of methodological quality associated with estimates of treatment effects in
799   controlled trials. *JAMA* 273:408-412.

800 StataCorp. 2017. Stata Statistical Software: Release 15. College Station, Texas: StataCorp LLC.

801 Sterling M, Vicenzino B, Souvlis T, and Connelly LB. 2015. Dry-needling and exercise for
802   chronic whiplash associated disorders (WAD): a randomised single blind placebo-
803   controlled trial. *Pain* 56:635-643.

804 Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, Carpenter J, Rücker G, Harbord
805   RM, and Schmid CH. 2011. Recommendations for examining and interpreting funnel plot
806   asymmetry in meta-analyses of randomised controlled trials. *BMJ* 343:8 pages.

807 Takakura N, Takayama M, Kawase A, Kaptchuk TJ, and Yajima H. 2010. Double blinding with
808   a new placebo needle: a further validation study. *Acupuncture in Medicine* 28:144-148.

809 Takakura N, and Yajima H. 2007. A double-blind placebo needle for acupuncture research. *BMC*
810   *Complementary and Alternative Medicine* 7:5 pages.

811 Tekin L, Akarsu S, Durmuş O, Çakar E, Dinçer Ü, and Kıralp MZ. 2013. The effect of dry
812   needling in the treatment of myofascial pain syndrome: a randomized double-blinded
813   placebo-controlled trial. *Clinical Rheumatology* 32:309-315.

814 Tough EA, White AR, Richards SH, and Campbell JL. 2010. Myofascial trigger point needling
815   for whiplash associated pain - a feasibility study. *Manual Therapy* 15:529-535.

816   Tough EA, White AR, Richards SH, Lord B, and Campbell JL. 2009. Developing and validating
817   a sham acupuncture needle. *Acupuncture in Medicine* 27:118-122.

818   Tracey I. 2010. Getting the pain you expect: mechanisms of placebo, nocebo and reappraisal
819   effects in humans. *Nature Medicine* 16:1277-1283.

820   Tsai CT, Hsieh LF, Kuan TS, Kao MJ, Chou LW, and Hong CZ. 2010. Remote effects of dry
821   needling on the irritability of the myofascial trigger point in the upper trapezius muscle.
822   *American Journal of Physical Medicine and Rehabilitation* 89:133-140.

823   Vase L, Baram S, Takakura N, Takayama M, Yajima H, Kawase A, Schuster L, Kaptchuk TJ,
824   Schou S, and Jensen TS. 2015. Can acupuncture treatment be double-blinded? An
825   evaluation of double-blind acupuncture treatment of postoperative pain. *PloS One*
826   10:e0119612.

827   Vase L, Baram S, Takakura N, Yajima H, Takayama M, Kaptchuk TJ, Schou S, Jensen TS,
828   Zachariae R, and Svensson P. 2013. Specifying the nonspecific components of
829   acupuncture analgesia. *Pain* 154:1659-1667.

830   Villamar MF, Contreras VS, Kuntz RE, and Fregni F. 2013. The reporting of blinding in physical
831   medicine and rehabilitation randomized controlled trials: a systematic review. *Journal of*
832   *Rehabilitation Medicine* 45:6-13.

833   White P, Bishop FL, Prescott P, Scott C, Little P, and Lewith G. 2012. Practice, practitioner, or
834   placebo? A multifactorial, mixed-methods randomized controlled trial of acupuncture.
835   *Pain* 153:455-462.

836   Wood L, Egger M, Gluud LL, Schulz KF, Jüni P, Altman DG, Gluud C, Martin RM, Wood AJ,
837   and Sterne JA. 2008. Empirical evidence of bias in treatment effect estimates in
838   controlled trials with different interventions and outcomes: meta-epidemiological study.
839   *BMJ* 336:601-605.

840

841

MEDLINE search strategy

1   **Table 1** MEDLINE search strategy

Search terms	Limits applied
1. Needl*.tw	Humans only
2. *Acupuncture therapy/	
3. Acupuncture.tw	Keyword searches limited to
4. Intramuscular stimulation.tw	title/abstract/keyword fields
5. Sham.tw	
6. *Placebo effect/	
7. *Placebos/	
8. Placebo$1.tw	
9. #1 OR #2 OR #3 OR #4	
10. #5 OR #6 OR #7 OR #8	
11. #9 AND #10	

Interpretation of the Blinding Index (BI) and classifications

*Cut-off scores were developed by consensus of authors of Moroz et al. (2013) and should not be interpreted as definitive classifications of blinding effectiveness. **BI** Blinding Index.

1   **Table 2** Interpretation of the Blinding Index (BI) and classifications

BI	Interpretation	BI cut-offs*	Classification
-1.00	All participants mistakenly guess the alternative intervention (incorrect guessing)	BI ≤ -0.20	Incorrect
0.00	Random guessing (ideal blinding)	-0.20 < BI < 0.20	Random
+1.00	All participants correctly guess their allocation (correct guessing)	BI ≥ 0.20	Correct

2   *Cut-off scores were developed by consensus of authors of Moroz et al. (2013) and should not
3   be interpreted as definitive classifications of blinding effectiveness. **BI** Blinding Index.

Risk of bias assessment [Cochrane Risk of Bias tool for randomised trials (Higgins et al. 2011)]

*Key domains (used to determine overall Risk of Bias for individual trials); #Domains used to determine blinding adequacy; **PB** Participant Blinding; **RPB** Research Personnel Blinding; **TB** Therapist Blinding; **OAB** Outcome Assessor Blinding; ✔ Low RoB; **?** Unclear RoB; ✖ High RoB.

Table 3 Risk of bias assessment (N=24 trials) [Cochrane Risk of Bias tool for randomised trials (Higgins et al. 2011)]

1st author & year	Random allocation	**Allocation concealed	Performance bias				**Detection bias (OAB)	Attrition bias	Reporting bias	OVERALL	Adequately blinded?
			#PB	RPB	#TB	*Overall					
Cotchett 2014	✔	✔	✔	?	✗	✔	✔	✔	✔	✔	Adequate
Dıraçoğlu 2012	✔	?	✔	?	✗	✔	✔	✗	✔	?	Inadequate
Espejo Antúnez 2014	✔	?	✔	?	✗	✔	✔	✔	✔	?	Inadequate
García-Gallego 2011	✔	?	✔	?	✗	✔	✔	✔	✔	?	Inadequate
Huguenin 2005	✔	✔	✔	?	✗	✔	✔	✗	✔	✔	Adequate
Inoue 2006	✔	✔	✔	?	✗	✔	✔	✔	✔	✔	Adequate
Itoh 2004	✔	?	✔	?	✗	✔	✔	✗	✔	?	Inadequate
Itoh 2005	✔	✔	✔	?	✗	✔	✔	✗	✔	✔	Adequate
Itoh 2006a	✔	?	✔	?	✗	✔	✔	✗	✔	?	Inadequate
Itoh 2006b	?	?	✔	?	✗	✔	?	✗	✔	?	Inadequate
Itoh 2007	✔	?	✔	?	✗	✔	✔	✗	✔	?	Inadequate
Itoh 2008	✔	?	✔	?	✗	✔	✔	✗	✔	?	Inadequate
Itoh 2012	✔	?	✔	?	✗	✔	✔	✗	✔	?	Inadequate
Itoh 2014	✔	?	✔	?	✗	✔	✔	✔	✔	?	Inadequate
Katsumi 2004	✔	?	✔	?	✗	✔	?	✔	✔	?	Inadequate
Mayoral 2013	✔	?	✔	?	✗	✔	✔	✗	✔	?	Inadequate
McMillan 1997	?	?	✔	?	✗	✔	✔	?	✔	?	Inadequate
Myburgh 2012	✔	?	✔	?	✗	✔	?	✗	✔	?	Inadequate
Nabeta 2002	✔	?	✔	?	✗	✔	?	✔	✔	?	Inadequate
Pecos-Martín 2015	✔	?	✔	?	✗	✔	✔	✔	✔	?	Inadequate
Sterling 2015	✔	✔	✔	?	✗	✔	✔	✔	✔	✔	Adequate
Tekin 2013	✔	?	✔	?	✗	✔	✔	✗	✔	?	Inadequate
Tough 2009/2010	✔	?	✔	✗	✗	✗	?	✔	✔	✗	Inadequate
Tsai 2010	✔	?	✔	?	✗	✔	✔	✔	✔	?	Inadequate

*Key domains (used to determine overall Risk of Bias for individual trials); #Domains used to determine blinding adequacy; **PB** Participant Blinding; **RPB** Research Personnel Blinding; **TB** Therapist Blinding; **OAB** Outcome Assessor Blinding; ✔ Low RoB; **?** Unclear RoB; ✗ High RoB.

Characteristics and results of included group comparisons

n number of participants (analysed for pain outcome); **#**Dropouts for pain outcome; **95% CI** 95% Confidence Interval; **AG** Active Group; **SG** Sham Group; **SMD** Standardised Mean Difference; **-ve** Negative; **Ax** Assessment; **Ix** Intervention; **NP** Non Penetrating; **NSD** No Significant Difference; **CEQ** Credibility/Expectancy Questionnaire; **ST** Short-Term (24 hours to four weeks, closest assessment to one week); **LT** Long-Term (one to six months, closest assessment to three months); **Immed** Immediately post-intervention (<24 hours after first/only intervention); **B** time-point at which Blinding was assessed; **†** Data not reported as mean/SD (could not calculate SMD); **NR** Not Reported; **DNR** Did Not Respond (to intervention); **AE** Adverse Effects; **P** Penetrating; **TrP SDN** Superficial Dry Needling above Trigger Point; **LTFU** Loss To Follow Up; **DK** Don't Know; **Non-TrP SDN** Superficial Dry Needling away from Trigger Point; **Non-TrP DN** Dry Needling away from Trigger Point; **AUC** Area Under Curve; **GA** General Anaesthesia; **SA** Spinal Anaesthesia. **Notes:** Shading represents adequately blinded trials (based on critical appraisal criteria for Review Question 2); *Itoh et al. (2005) and *Itoh et al. (2007) each had two eligible sham groups; ^Itoh et al. (2005) only reported the number of participants from each group who guessed they were in the active group, therefore, to calculate the BI it was conservatively assumed that the remaining participants guessed they were in the sham group (i.e. no DK responses).

1 **Table 4** Characteristics and results of included group comparisons (N=26 group comparisons)

1st author & year	n	Dropouts# [reasons]	Type of sham	Blinding index (95% CI) Blinding scenario (AG/SG)	Reported blinding results	Reported blinding conclusion	Between-group SMD (pain) and reported p-values [-ve values in favour of AG]
Cotchett 2014	84	5 AG: 3 [1 missed Ax; 2 ceased Ix] SG: 2 [1 missed Ax; 1 ceased Ix]	NP: Custom (blunt needle)	Insufficient data	NSD between groups (CEQ) (p>0.05 for all questions)	Success	ST: -0.05 (p=0.026) LT: -0.42 (p=0.007)
Huguenin 2005	52	7 [difficulty attending]	NP: Custom (blunt needle)	Insufficient data	AG only: • Immed: Correct (p=0.001) • ST: NSD between correct and incorrect guesses (p=0.062)	Success	Immed/B: † (NSD) ST: † (NSD)
Inoue 2006	31	0	NP: Custom (guide tube only)	AG: 0.20 (-0.30 − 0.70) SG: 0.25 (-0.22 − 0.72) Correct/Correct	NSD between groups (p NR) AG: 9/15 correct SG: 10/16 correct	Success	Immed/B: 0.76 (p=0.020)
*Itoh 2005 (NP)	19	3 AG: 1 SG: 2 [All groups: 5 DNR; 2 AE]	NP: Custom (guide tube only)	^AG: 0.60 (0.10 − 1.10) ^SG: -0.11 (-0.76 − 0.54) Correct/Random	NSD between groups (p=0.64) AG: 8/10 correct SG: 4/9 correct	Success	ST: -0.82 (p<0.05) B: -2.35 (p<0.01) LT: -0.98 (NSD)
*Itoh 2005 (P)	19	3 AG: 1 SG: 2 [All groups: 5 DNR; 2 AE]	P: TrP SDN	^AG: 0.60 (0.10 − 1.10) ^SG: -0.20 (-0.81 − 0.41) Correct/Incorrect	NSD between groups (p=0.64) AG: 8/10 correct SG: 4/10 correct	Success	ST: -0.77 (NSD) B: -0.67 (p<0.05) LT: -0.13 (NSD)
Sterling 2015	80	7 AG: 3 [LTFU] SG: 4 [LTFU]	NP: Commercial (Park sham)	Insufficient data	Descriptive only SG: 1/36 correct [All remaining participants believed AG or DK]	Success	LT: -0.04 (NSD) B: -0.09 (NSD)
Dıraçoğlu 2012	50	2 AG: 1 [difficulty attending] SG: 1 [DNR]	P: Non-TrP SDN	Did not assess blinding	-	-	ST: 0.06 (p=0.478)
Espejo Antúnez 2014	45	0	NP: Custom (retracting needle)	Did not assess blinding	-	-	Immed: -1.15 (p<0.01)
García-Gallego 2011	33	0	P: Non-TrP DN	Did not assess blinding	-	-	Immed: 0.09 (NSD) ST: 0.17 (NSD)
Itoh 2004	18	4 AG: 1 [AE] SG: 3 [DNR]	P: TrP SDN	Did not assess blinding	-	-	ST: -0.72 (NSD) LT: -0.21 (NSD)
Itoh 2006a	19	7 AG: 3 [2 DNR; 1 AE] SG: 4 [DNR]	NP: Custom (blunt needle)	AG: 0.50 (0.00 − 1.00) SG: -0.11 (-0.68 − 0.46) Correct/Random	NSD between groups (p=0.38) AG: 7/10 correct; 1/10 DK SG: 3/9 correct; 2/9 DK	Success	ST: -1.38 (NSD) B: -3.43 (p<0.001) LT: -1.19 (NSD)
Itoh 2006b	18	5	NP: Custom (blunt	AG: NR	Descriptive only (SG only)	NR	ST/B: -1.11 (p<0.05)

		AG: 2 SG: 3 [All groups: 4 DNR; 2 drugs]	needle)	SG: -0.56 (-1.00 – -0.11) NR/Incorrect	SG: 1/9 correct; 2/9 DK		LT: -0.35 (NSD)
*Itoh 2007 (NP)	15	5 AG: 2 [1 DNR; 1 AE] SG: 3 [2 DNR; 1 AE]	NP: Custom (blunt needle)	AG: 0.38 (-0.11 – 0.86) SG: -0.29 (-0.94 – 0.37) Correct/Incorrect	NSD between groups (p=0.89) AG: 4/8 correct; 3/8 DK SG: 2/7 correct; 1/7 DK	Success	ST: -0.71 (NSD) B: -1.87 (NSD) LT: -2.52 (NSD)
*Itoh 2007 (P)	16	4 AG: 2 [1 DNR; 1 AE] SG: 2 [1 DNR; 1 AE]	P: Non-TrP DN	AG: 0.38 (-0.11 – 0.86) SG: -0.38 (-0.86 – 0.11) Correct/Incorrect	NSD between groups (p=0.89) AG: 4/8 correct; 3/8 DK SG: 1/8 correct; 3/8 DK	Success	ST: -1.32 (NSD) B: -2.25 (NSD) LT: -3.25 (NSD)
Itoh 2008	15	5 AG: 2 [1 DNR; 1 AE] SG: 3 [DNR]	NP: Custom (blunt needle)	AG: 0.75 (0.29 – 1.21) SG: -0.43 (-1.10 – 0.24) Correct/Incorrect	NSD between groups (p=0.74) AG: 7/8 correct SG: 2/7 correct	Success	ST: -1.95 B: -2.67 LT: -0.81 (AUC p=0.025)
Itoh 2012	15	1 AG: 1 [AE]	NP: Custom (blunt needle)	AG: 1.00 (1.00 – 1.00) SG: -1.00 (-1.00 – -1.00) Correct/Incorrect	Descriptive only [All participants believed they were in AG]	Success	ST: -0.46 B: -1.83 LT: -1.65 (AUC p=0.003)
Itoh 2014	15	1 SG: 1 [DNR]	NP: Custom (blunt needle)	AG: 0.56 (0.01 – 1.10) SG: -0.50 (-1.10 – 0.10) Correct/Incorrect	NSD between groups (p=0.89) AG: 7/9 correct SG: 2/8 correct	Success	ST: -0.96 B: -1.29 LT: -1.44 (AUC p=0.024)
Katsumi 2004	9	0	NP: Custom (guide tube only)	AG: 1.00 (1.00 – 1.00) SG: -0.60 (-1.30 – 0.10) Correct/Incorrect	Descriptive only AG: 4/4 correct SG: 1/5 correct	NR	ST: -0.64 (NR) B: -4.36 (NR) LT: -0.73 (NR)
Mayoral 2013	31	9 AG: 4 [LTFU] SG: 5 [LTFU]	No needle: GA/SA	Did not assess blinding	-	-	ST: -0.34 (p=0.294) LT: -0.23 (p=0.516)
McMillan 1997	20	NR	P: Non-TrP SDN	Did not assess blinding	-	-	Immed: 0.35 (NSD) ST: 0.26 (NSD)
Myburgh 2012	77	4 AG: 4 [2 non-compliant; 1 AE; 1 NR]	P: TrP SDN	Did not assess blinding	-	-	ST: -0.37 (NSD)
Nabeta 2002	34	7 AG: 2 [difficulty attending] SG: 5 [difficulty attending]	NP: Custom (blunt needle)	AG: 0.41 (0.01 – 0.81) SG: -0.18 (-0.62 – 0.26) Correct/Random	NSD between groups (p=0.74) AG: 11/17 correct; 2/17 DK SG: 6/17 correct; 2/17 DK	Success	Immed: -0.12 (NSD) ST: -0.31 (NSD) B: -0.25 (NSD)
Pecos-Martín 2015	72	0	P: Non-TrP DN	Did not assess blinding	-	-	ST: -1.59 (p<0.001) LT: -1.93 (p<0.001)
Tekin 2013	39	7 AG: 1 [ceased Ix] SG: 6 [ceased Ix]	NP: Custom (blunt needle)	Did not assess blinding	-	-	Immed: -0.88 (p=0.034) ST: -1.62 (p=0.000)
Tough	41	7	NP: Custom (blunt	AG: 0.53 (0.30 – 0.75)	NSD between groups (p>0.2)	Success	ST/B: 0.11 (NR)

2009/2010	AG: 3 [LTFU] SG: 4 [LTFU]	needle)	SG: -0.67 (-0.93 – -0.40) Correct/Incorrect	AG: 10/19 correct; 9/19 DK SG: 1/18 correct; 4/18 DK		LT: -0.61 (p=0.67)
Tsai 2010	35	0	P: TrP SDN	Did not assess blinding	-	-	Immed: -0.91 (p<0.05)

2 **n** number of participants (analysed for pain outcome); **#**Dropouts for pain outcome; **95% CI** 95% Confidence Interval; **AG** Active Group; **SG** Sham

3 Group; **SMD** Standardised Mean Difference; **-ve** Negative; **Ax** Assessment; **Ix** Intervention; **NP** Non Penetrating; **NSD** No Significant Difference;

4 **CEQ** Credibility/Expectancy Questionnaire; **ST** Short-Term (24 hours to four weeks, closest assessment to one week); **LT** Long-Term (one to six

5 months, closest assessment to three months); **Immed** Immediately post-intervention (<24 hours after first/only intervention); **B** time-point at

6 which Blinding was assessed; **†** Data not reported as mean/SD (could not calculate SMD); **NR** Not Reported; **DNR** Did Not Respond (to

7 intervention); **AE** Adverse Effects; **P** Penetrating; **TrP SDN** Superficial Dry Needling above Trigger Point; **LTFU** Loss To Follow Up; **DK** Don't Know;

8 **Non-TrP SDN** Superficial Dry Needling away from Trigger Point; **Non-TrP DN** Dry Needling away from Trigger Point; **AUC** Area Under Curve; **GA**

9 General Anaesthesia; **SA** Spinal Anaesthesia.

10 **Notes:** Shading represents adequately blinded trials (based on critical appraisal criteria for Review Question 2); *Itoh & Katsumi (2005) and *Itoh

11 et al. (2007) each had two eligible sham groups; in both of these trials one group had a non-penetrating (NP) sham and the other had a

12 penetrating (P) sham (labelled accordingly in the first column of the table). ^Itoh & Katsumi (2005) only reported the number of participants

13 from each group who guessed they were in the active group, therefore, to calculate the BI it was conservatively assumed that the remaining

14 participants guessed they were in the sham group (i.e. no DK responses).

Hypothesised effects of intervention belief on pain for group comparisons where the Blinding Index (BI) could be calculated (N=12) [adapted from Bang et al. (2010)]

No. Scenario Number; **AG** Active Group; **SG** Sham Group; **N** Number of group comparisons; **n** number of participants.

Table 5 Hypothesised effects of intervention belief on pain for group comparisons where the
Blinding Index (BI) could be calculated (N=12 group comparisons) [adapted from Bang et al. (2010)]

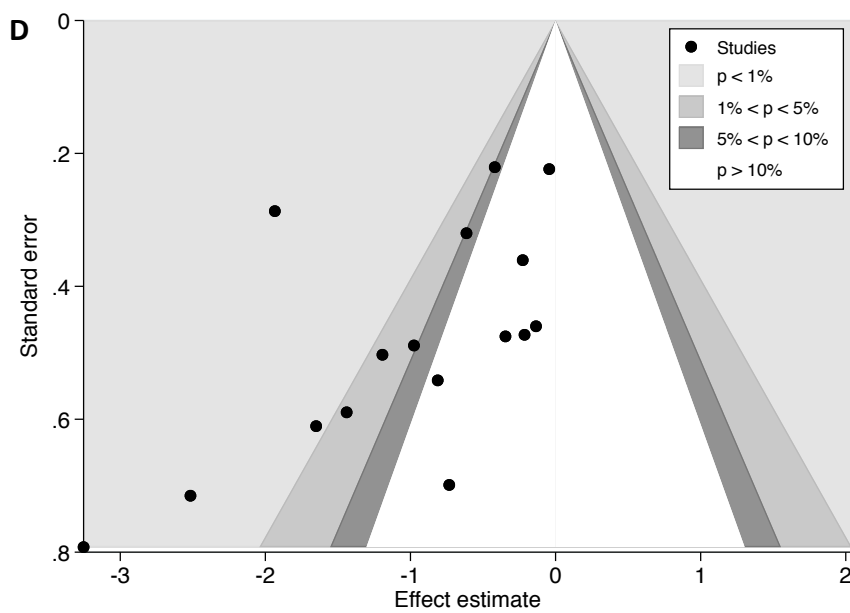
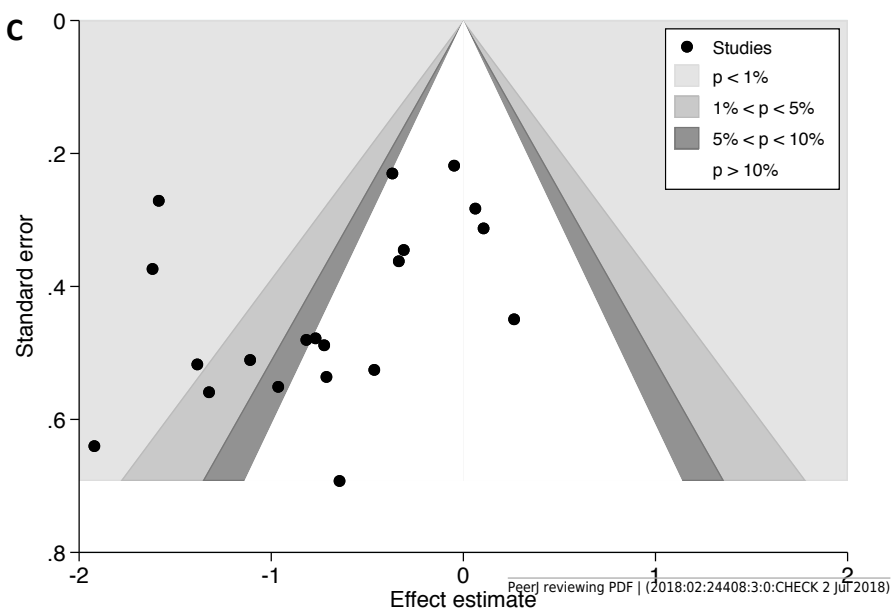
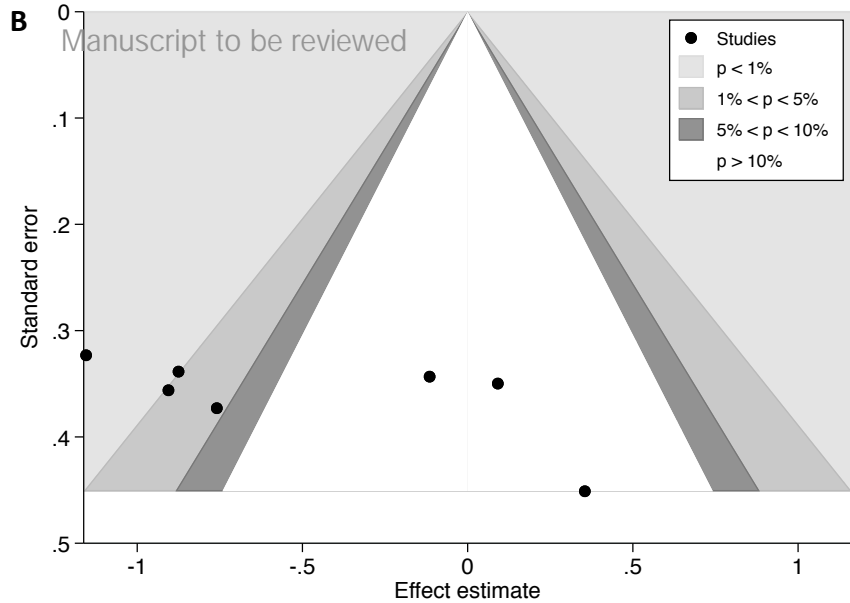
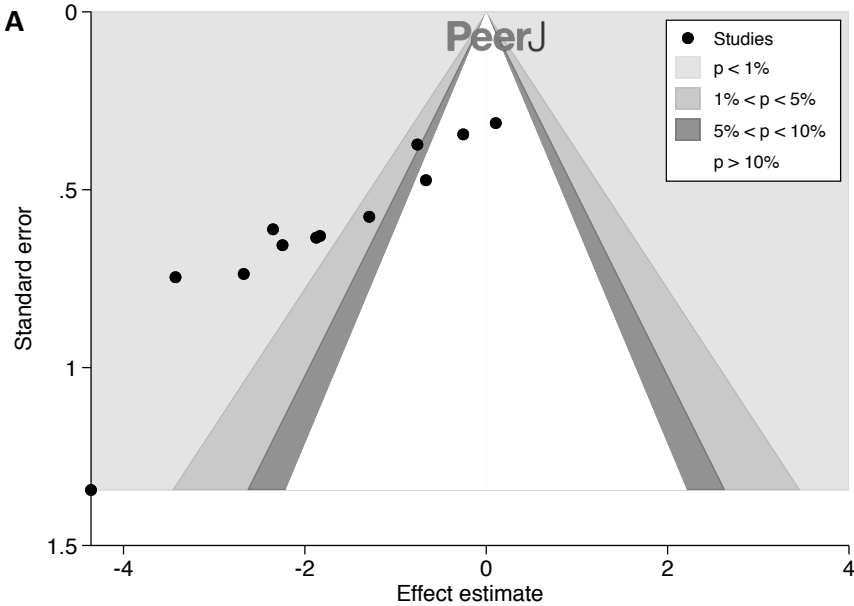
No.	AG beliefs	SG beliefs	Hypothesised moderation effect of intervention belief on pain outcomes			N (%)	n (%)
			AG	SG	Between group		
1	Incorrect (sham)	Incorrect (active)	↓	↑	Large; in favour of SG	0 (0)	0 (0)
2	Random	Incorrect (active)	-	↑	Small; in favour of SG	0 (0)	0 (0)
3	Incorrect (sham)	Random	↓	-	Small; in favour of SG	0 (0)	0 (0)
4	Incorrect (sham)	Correct (sham)	↓	↓	None (reduced in both groups)	0 (0)	0 (0)
5	Random	Random	-	-	None	0 (0)	0 (0)
6	Correct (active)	Incorrect (active)	↑	↑	None (inflated in both groups)	8 (67)	145 (58)
7	Correct (active)	Random	↑	-	Small; in favour of AG	3 (25)	72 (29)
8	Random	Correct (sham)	-	↓	Small; in favour of AG	0 (0)	0 (0)
9	Correct (active)	Correct (sham)	↑	↓	Large; in favour of AG	1 (8)	31 (13)

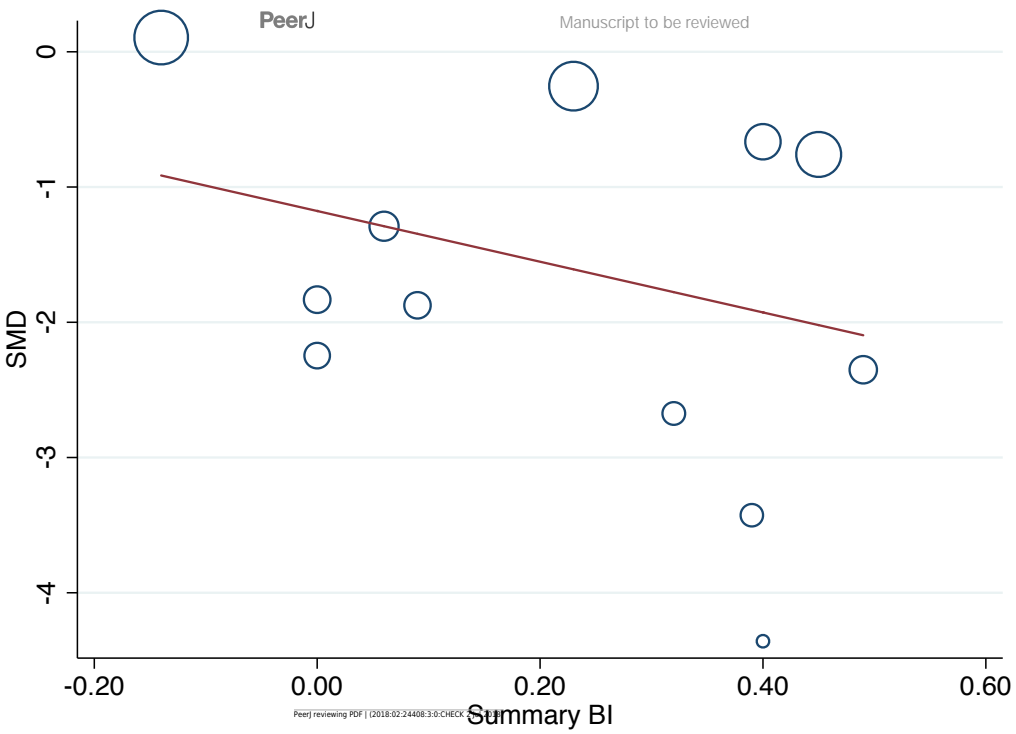
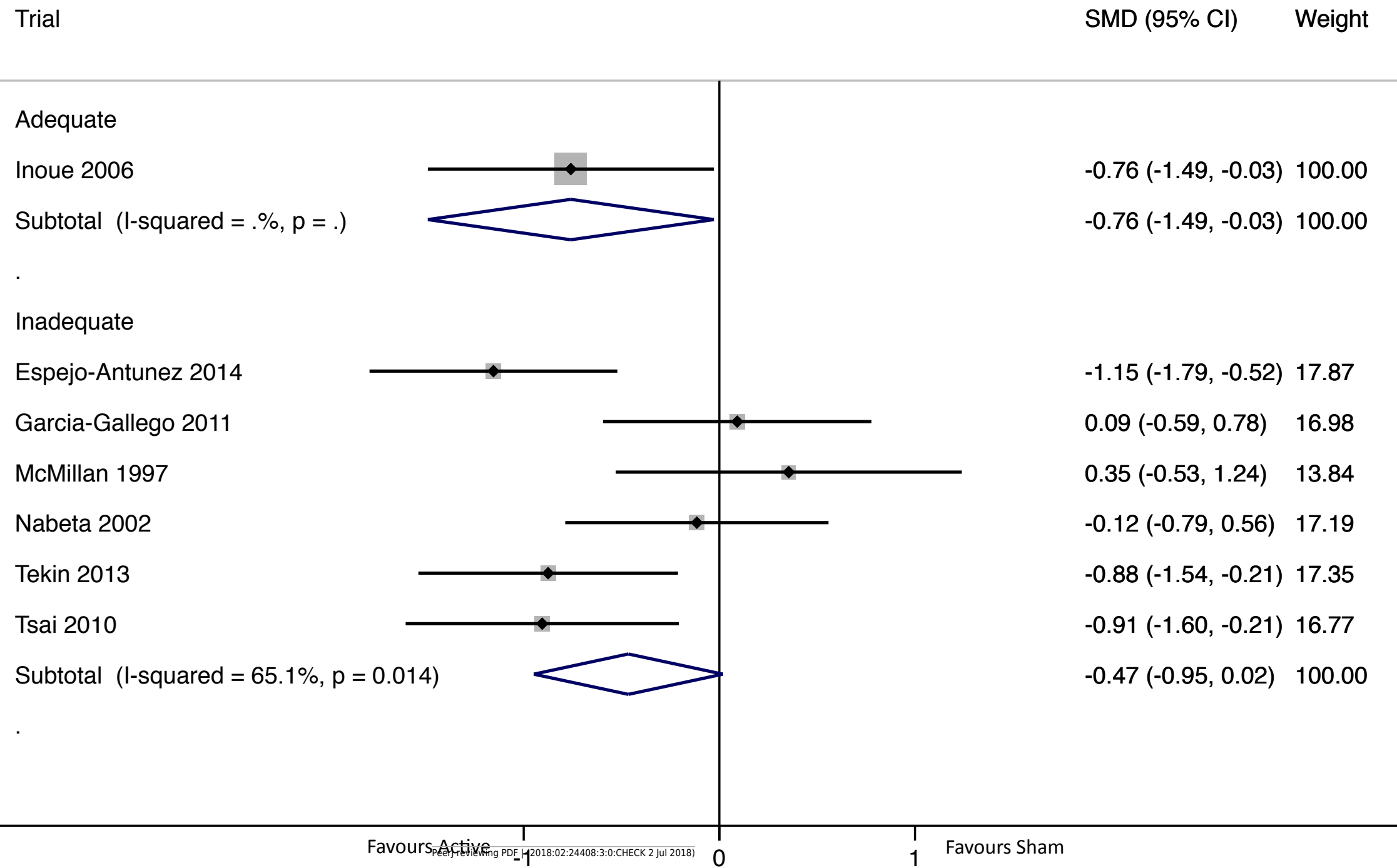
No. Scenario Number; **AG** Active Group; **SG** Sham Group; **N** Number of group comparisons; **n**
number of participants.

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram (www.prisma-statement.org)

*Processes performed by two independent reviewers.

Contour-enhanced funnel plots for pain outcomes in dry needling trials.

(A) Funnel plot for Review Question 1 (blinding effectiveness): time-point closest to when blinding was assessed. (B, C, D) Funnel plots for Review Question 2 (blinding adequacy) [(B) Pain assessments immediately after first/only intervention; (C) Short-term pain assessments; (D) Long-term pain assessments].

Bubble plot (meta-regression) of the influence of the summary BI (blinding effectiveness) on between-group effect size (pain) for pain assessments closest to the time point blinding was assessed (N=12 group comparisons).

Each bubble represents one group comparison, and the size of each bubble is proportional to weight (inverse variance). Negative values for SMD are in favour of active dry needling. **SMD** Standardised Mean Difference (effect size); **BI** Blinding Index.

Forest plot of pooled between-group effect sizes (pain) based on blinding adequacy, for pain assessments immediately after the first/only intervention (<24 hours; N=7 group comparisons).

Trial		SMD (95% CI)	% Weight

Adequate

Inoue 2006 — -0.76 (-1.49, -0.03) 100.00

Subtotal (I-squared = .%, p = .) — -0.76 (-1.49, -0.03) 100.00

.

Inadequate

Espejo-Antunez 2014 — -1.15 (-1.79, -0.52) 17.87

Garcia-Gallego 2011 — 0.09 (-0.59, 0.78) 16.98

McMillan 1997 — 0.35 (-0.53, 1.24) 13.84

Nabeta 2002 — -0.12 (-0.79, 0.56) 17.19

Tekin 2013 — -0.88 (-1.54, -0.21) 17.35

Tsai 2010 — -0.91 (-1.60, -0.21) 16.77

Subtotal (I-squared = 65.1%, p = 0.014) — -0.47 (-0.95, 0.02) 100.00

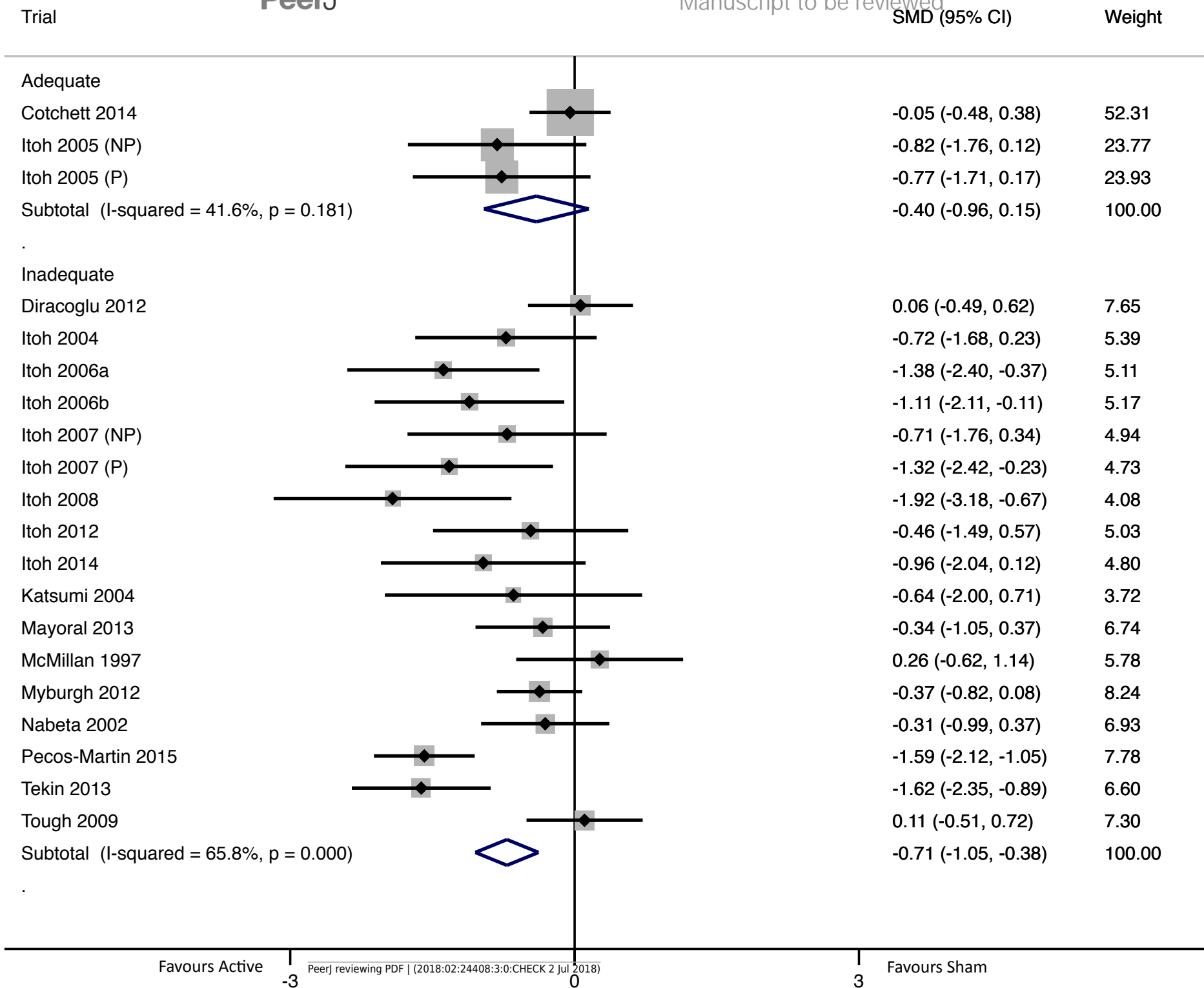
.

Favours Active   -1   0   1   Favours Sham

Forest plot of pooled between-group effect sizes (pain) based on blinding adequacy, for pain assessments in the short-term (24 hours to one month; N=20 group comparisons).

Note: Itoh & Katsumi (2005) and Itoh et al. (2007) each had two eligible sham groups; in both of these trials one group had a non-penetrating (**NP**) sham and the other had a penetrating (**P**) sham (labelled accordingly in the figure).

Trial		SMD (95% CI)	% Weight
Adequate			
Cotchett 2014		-0.05 (-0.48, 0.38)	52.31
Itoh 2005 (NP)		-0.82 (-1.76, 0.12)	23.77
Itoh 2005 (P)		-0.77 (-1.71, 0.17)	23.93
Subtotal (I-squared = 41.6%, p = 0.181)		-0.40 (-0.96, 0.15)	100.00
.			
Inadequate			
Diracoglu 2012		0.06 (-0.49, 0.62)	7.65
Itoh 2004		-0.72 (-1.68, 0.23)	5.39
Itoh 2006a		-1.38 (-2.40, -0.37)	5.11
Itoh 2006b		-1.11 (-2.11, -0.11)	5.17
Itoh 2007 (NP)		-0.71 (-1.76, 0.34)	4.94
Itoh 2007 (P)		-1.32 (-2.42, -0.23)	4.73
Itoh 2008		-1.92 (-3.18, -0.67)	4.08
Itoh 2012		-0.46 (-1.49, 0.57)	5.03
Itoh 2014		-0.96 (-2.04, 0.12)	4.80
Katsumi 2004		-0.64 (-2.00, 0.71)	3.72
Mayoral 2013		-0.34 (-1.05, 0.37)	6.74
McMillan 1997		0.26 (-0.62, 1.14)	5.78
Myburgh 2012		-0.37 (-0.82, 0.08)	8.24
Nabeta 2002		-0.31 (-0.99, 0.37)	6.93
Pecos-Martin 2015		-1.59 (-2.12, -1.05)	7.78
Tekin 2013		-1.62 (-2.35, -0.89)	6.60
Tough 2009		0.11 (-0.51, 0.72)	7.30
Subtotal (I-squared = 65.8%, p = 0.000)		-0.71 (-1.05, -0.38)	100.00
.			

Favours Active   -3   0   3   Favours Sham

Forest plot of pooled between-group effect sizes (pain) based on blinding adequacy, for pain assessments in the long-term (one to six months; N=16 group comparisons).

Note: Itoh & Katsumi (2005) and Itoh et al. (2007) each had two eligible sham groups; in both of these trials one group had a non-penetrating (**NP**) sham and the other had a penetrating (**P**) sham (labelled accordingly in the figure).

Trial	SMD (95% CI)	% Weight
Adequate		
Cotchett 2014	-0.42 (-0.85, 0.01)	39.33
Itoh 2005 (NP)	-0.98 (-1.93, -0.02)	10.42
Itoh 2005 (P)	-0.13 (-1.04, 0.77)	11.65
Sterling 2015	-0.04 (-0.48, 0.39)	38.60
Subtotal (I-squared = 17.5%, p = 0.303)	-0.30 (-0.62, 0.02)	100.00
.		
Inadequate		
Itoh 2004	-0.21 (-1.14, 0.71)	8.90
Itoh 2006a	-1.19 (-2.18, -0.21)	8.55
Itoh 2006b	-0.35 (-1.28, 0.59)	8.87
Itoh 2007 (NP)	-2.52 (-3.92, -1.11)	6.35
Itoh 2007 (P)	-3.25 (-4.81, -1.70)	5.69
Itoh 2008	-0.81 (-1.87, 0.25)	8.11
Itoh 2012	-1.65 (-2.85, -0.45)	7.37
Itoh 2014	-1.44 (-2.60, -0.28)	7.59
Katsumi 2004	-0.73 (-2.10, 0.64)	6.50
Mayoral 2013	-0.23 (-0.93, 0.48)	10.24
Pecos-Martin 2015	-1.93 (-2.50, -1.37)	11.09
Tough 2009	-0.61 (-1.24, 0.01)	10.72
Subtotal (I-squared = 69.2%, p = 0.000)	-1.14 (-1.64, -0.65)	100.00
.		

Favours Active   Favours Sham

-3   0   3