**Overview of studies reporting on blinding effectiveness and expectation of benefit**

***Table S3: Studies reporting on blinding effectiveness****. Results for correct and incorrect guesses of group allocation are provided as reported in original reports. The third column from the right refers to the authors’ self-reported blinding success, according to their own criteria. Interventions were rated as ‘simple’ if only few procedural steps and little provider-patient interaction were involved. Y, Yes; N, No; NA, Not applicable / not provided. Bang’s blinding index was calculated for each group where information was provided in the required format, and compiled as ratio using Hedge’s g. Ratio values larger than 0 indicate that participants in the active group were more likely to correctly guess their allocation to the active group than those in the control group, and values below 0 that patients in the sham control group were more likely to wrongly guess that they received an active treatment compared to those in the active group (Bang et al., 2010; Colaguiri et al., 2019).*

*Interpretation: Faced with a dichotomous choice, participants are likely to guess when adequately blinded, resulting in a close-to-chance distribution of 50% correct guesses in the placebo control group. As Chaibi et al. (2015) suggest, however, this percentage should be higher in non-pharmacological trials, as these elaborate sham procedures may be more convincing, and the aim is to make patients believe they are receiving a real treatment. The percentage of successfully blinded control participants is highlighted where available. \* James’ Blinding index (BI): 1.0 indicates complete blinding, whereas an index of 0.5 is equivalent to random guessing. A statistically significant amount of blinding beyond chance has occurred if the 95% CI lies completely above 0.50.*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Study** | **Intervention type and complexity**  | **Placebo control**  | **Assessment method** | **Statistical method** | **Interpretation** | **Result** | **Reported ‘successful’ blinding** | **Bang’s blinding index (95% CI)** | **Blinding ratio****(Hedge’s g, SD)** |
| Assefi, 2008 | Spiritual / energetic healing, simple  | Simulated hands-on manoeuvre (actor) | Certainty rating (of real treatment received)Provider skill ratingRating expected treatment efficacy  | Chi-square | No difference between groups as successful blinding | % of participants who believed to be in active group: Active: 77%Sham: 65% *p* = 0.24 | Y | BIa 0.53 (0.29 - 0.77)BIp -0.3 (-0.53 - -0.03) | 1.72 (0.13) |
| Assefi, 2008 | Spiritual / energetic healing, simple  | Simulated hands-off manoeuvre (actor) | As above (blinding results reported together for all 2 sham groups) |  |
| Bennell, 2010 | Other Manual Therapy, complex | Disabled device  | Group guess | James' blinding index (James et al., 1996)  | As per index \* | Correct guesses per group: Active: 32/55 (58%) Sham: 21/61 (34%) Uncertain: Active: 15 (27%) Sham: 27 (44%) Incorrect guesses: Active: 8 (15%)Sham: 13 (21%) “James’ BI = 0.70 (bootstrap 95% confidence interval 0.58 to 0.82), interpreted as a moderate to high degree of blindedness and representing a statistically significant amount of blinding beyond that expected by chance (the value of the blinded index is 0.5 for random guessing).” | Y | BIa 0.44 (0.24-0.63 95% CI)BIp 0.13 (-0.05 - 0.32) | 5.6 (0.1) |
| Bennell, 2014 | Physiotherapy / Rehabilitation, complex | Disabled device  | Group guessCredibility rating  | James' blinding index (James et al., 1996) | As per index \* | BI at last session: 0.53 (0.44-0.62), M, 95% CI“Treatment credibility: 16.8 (6.5) in active group, 14.1 (7.2) in sham group (ratings from 0-24)” | N |  |  |
| Bialosky, 2009 | Other Manual Therapy, complex | Manual, simulated manoeuvre | Group guessRating expected treatment efficacy  | Chi-square | No difference between groups as successful blinding | % of participants who believed to be in active group: Active: 7/19 (37%)Sham: 11/18 (61%)“Frequencies of perceived group assignment did not differ by actual group assignment χ2 (1, n = 37) = 2.10, *p* = 0.15.” | NA | BIa -0.26 (-0.7 - 0.17)BIp -0.22 (-0.67 – 0.23) | -2.15 (0.23) |
| Bialosky, 2014 | Manual Therapy (with spinal manipulation), simple  | Manual, simulated manoeuvre | Group guessRating expected treatment efficacy  | Chi-square | No difference between groups as successful blinding | Correct guesses per group: Active: 18 (66.7%)Sham (“standard”): 17 (63%)Incorrect guesses: Active: 9 (33.3%)Sham (“standard”): 10 (37%) Significant group-related differences were observed in perception of group assignment immediately following the first intervention, χ² (2, n = 81) = 10.02, *p* = .01 “More participants receiving the standard placebo SMT believed they received a placebo than did those receiving spinal manipulation (*p* = .03) or the enhanced placebo SMT (*p* < .01).” | N | BIa 0.33 (-0.02 - 0.68)BIp 0.26 (-0.11- 0.62) | 3.23 (0.18) |
| Bialosky, 2014 (same trial, “enhanced” placebo control) | Manual Therapy (with spinal manipulation), simple  | Manual, simulated manoeuvre | As above | Correct guesses per group: Active: 18 (66.7%)Sham 2 (“enhanced”): 6 (22.2%)Incorrect guesses: Active: 9 (33.3%)Sham 2 (“enhanced”): 21 (77.8%) | Y | BIa 0.33 (-0.02 – 0.68)BIp -0.56 (-0.87 - -0.24) | -1.30 (0.17) |
| Bliddal, 2014 | Spiritual / energetic healing, complex  | Simulated hands-off manoeuvre (actor) | Group guessRating expected treatment efficacy  | Difference between groups (t-test) | No difference between groups as successful blinding | Responses to "Do you think you were treated by the healer (active)? Ratings 0-100 (No-Yes), Median (IQR):Active: 2.0 (0-46)Sham: 15.5 (2-35)“No significant differences between groups (*p* = 0.176)” | Y | NA |  |
| Cerritelli, 2015 | Craniosacral Therapy, simple  | Manual, simulated manoeuvre | Group guess | Descriptive | Unclear | Correct guesses: Sham: 0%(no further detail reported)Incorrect guesses or “Don’t know”:Sham: 100% | Y | NA |  |
| Chaibi, 2017 (migraine trial) | Manual Therapy (with spinal manipulation), simple  | Manual, simulated manoeuvre | Certainty rating (of real treatment received) | Odds ratio | Expecting more than 50% per group to assume having received the active treatment | % of participants who believed to be in active group: Active: >80%Sham: >80%After each of the 12 intervention sessions. “Odds ratio for believing that active treatment was received: >10 at all treatment sessions in both groups (all *p* < 0.001)” | Y | NA  |  |
| Chen, 2008 | Other Manual Therapy, complex  | Manual, simulated manoeuvre | Group guessrating expected treatment efficacy  | Difference between groups (t-test) | No difference between groups as successful blinding | % of participants who believed to be in active group: Active (provider 1): 47% Active (provider 2): 91% Sham (actor): 45% (rest answered “not sure”)“More subjects treated by provider 2 guessed they were being treated by a real healer than other two groups (*p* = 0.04).” | Y | NA |  |
| Costa, 2009 | Physiotherapy / Rehabilitation, complex  | Disabled device  | Group guess | Descriptive | None | % of participants who believed to be in active group: Active: (66) 85%Sham: (65) 84%  | Y | BIa 0.71 (0.56 – 0.87)BIp -0.69 (-0.85 - - 0.53) | .31 (0.08) |
| Craske, 2011 | Cognitive Behavioural & other Psychotherapy, complex  | Multicomponent therapist interaction | Credibility rating(modified 7-item Treatment Credibility Questionnaire; Borkovec & Nau, 1972) | Difference between groups (t-test) | No difference between groups as successful blinding | “On average, participants rated the treatments to be more than moderately credible (M = 7.28, SD = 1.26) with no differences across groups (*p* =.13).” | Y | NA |  |
| Cuñado González, 2021 | Other Manual Therapy, simple | Manual, simulated manoeuvre & technique to different area | Group guess | Difference between groups (t-test) | No difference between groups as successful blinding | % of participants who believed to be in active group: Active: (18) 69%Sham: (14) 56%“Regarding the believability of the sham technique, the frequencies of perceived group assignment showed no differences between groups (P = .329).“ | Y | BIa 0.38 (0.03 – 0.74)BIp -0.12 (-0.51 – 0.27) | 1.39 (0.19) |
| Fisher, 2020 | Manual Therapy (with spinal manipulation, complex  | Manual, simulated manoeuvre | Group guess | Chi-square | No difference between groups as successful blinding | Correct guesses per group: Active: 43 (91.5%)Sham: 16 (37.2%)Incorrect guesses: Active: 4 (8.5%)Sham: 27 (62.8%)“χ2 = 10.70, *p* = .001; odds ratio = 6.62; 95% confidence interval: 2.10, 21.94.” | Y | BIa 0.83 (0.67 – 0.99)BIp -0.26 (-0.54 – 0.03) | 4.8 (0.12) |
| Ganderton, 2018 | Physiotherapy / Rehabilitation, complex | Exercise, nonspecific  | Group guess | Descriptive  | Total number unaware of their group allocation | “No participant was aware of the group allocation and reported being ‘‘unsure’’ if they were in the intervention or sham group.” Incorrect guesses: Sham: 100%  | Y | NA |  |
| Kardouni, 2015 | Manual Therapy (with spinal manipulation, simple  | Manual, simulated manoeuvre | Group guess | Difference between groups (test of proportions | No difference between groups as successful blinding | % of participants who believed to be in active group: Active: (18) 76% Sham: (14) 62%  | Y | BIa 0.5 (0.15 – 0.85)BIp -0.17 (-0.56 – 0.23) | 1.76 (0.19) |
| Krekoukias, 2017 | Other Manual Therapy, simple  | Manual, soft touch | Group guess | Unclear  | Unclear  | "Subjects were unaware of their inclusion in this group and, after being asked following the trial, thought they were receiving actual treatment at the time of the intervention." | Y | NA |  |
| Panagopoulos, 2014 | Other Manual Therapy, complex | Manual, soft touch | Group guess | Unclear  | Unclear  | % of participants who believed to be in active group: Active: (26) 80% Sham: (29) 90% | Y | BIa 0.63 (0.36 – 0.9)BIp -0.8 (-1.0 - -0.61) | -1.54 (0.12) |
| Silva, 2019 | Manual Therapy (with spinal manipulation, simple | Manual, simulated manoeuvre | Group guess | Informal, no detail reported | Unclear  | "After the intervention, the participants were informally questioned about whether they believed they had received the manipulation or placebo, and most believed they had received the manipulation.”  | Y | NA |  |
| Traeger, 2018 | Cognitive-behavioural and other Psychotherapy, complex | Multicomponent therapist interaction | Credibility rating | Difference between groups (t-test) | No difference between groups as successful blinding | "Treatment credibility scores were not different between groups (mean [SD] credibility and expectancy questionnaire score for patient education vs placebo patient education: 36.6 [8.8] vs 35.3 [10.5]; mean difference, –1.3; 95% CI, –4.0 to 1.4)." | Y | NA |  |
| Vernon, 2012 | Manual Therapy (with spinal manipulation, simple | Manual, simulated manoeuvre | Group guess  | Chi-square | No difference between groups as successful blinding (also reference to Bang et al., 50% would signify correct guesses are chance) | Correct guesses per group: Active: 50%Sham: 47.2%Incorrect guesses per group: Active: 50% Sham: 53.8%"There was no statistical difference between the subject perceptions by group (X2 = 0.06, df = 1, p =.80)." | Y | BIa 0.0 (-0.35 – 0.35)BIp 0.06 (-0.41 – 0.28) | -.35 (0.18) |
| Walker, 2013 | Manual Therapy (with spinal manipulation, complex | Disabled device | Group guessTreatment satisfaction | Bang blinding index (Bang et al., 2004) | As per index | Correct guesses per group: Active: 85%Sham: 67%Incorrect guesses: Sham: 33%"Bang Index values showed that 25% of the sham group (95% CI = 10%–40%) and 61% of the usual care group (95% CI = 48%–74%) guessed correctly beyond what would be expected by chance.”  | N | BIa 0.7 (0.55 – 0.84)BIp 0.34 (0.16 – 0.54) | 11.98 (0.09) |
| Young, 2019 | Manual Therapy (with spinal manipulation, simple | Manual, simulated manoeuvre | Group guess | Chi-square | No difference between groups as successful blinding | % of participants who believed to be in active group: Active: 90% Sham: 57% *p* = 0.01 | N | BIa 0.82 (0.58 – 1.06)BIp -0.14 (-0.67 – 0.28) | 3.85 (0.18) |
| **Small trials or validation studies without pain-related outcome measures** |  |  |
| Attali, 2013 | Other Manual Therapy, complex | Manual, simulated manoeuvre | Group guess | Informal, no detail reported | Unclear | "(…) some patients did notice a difference, and some of them expressed a preference for the placebo treatment (noting that the osteopathic treatment caused them more pain)." | Not reported | NA |  |
| Brose, 2013 | Other Manual Therapy, simple | Manual, simulated manoeuvre | Group guessRating expected treatment efficacy Treatment satisfaction | Fisher exact and Chi-square | No difference between groups as successful blinding | % of participants who believed to be in active group: Active: 63.6% Sham: 60% *p* = 0.99 / 0.85, respectively  | Y | BIa 0.27 (-0.3 – 0.84)BIp -0.2 (-0.7 – 0.3) | .27 (0.27) |
| Chaibi, 2015 | Manual Therapy (with spinal manipulation, complex | Manual, simulated manoeuvre | Group guessCertainty rating (of real treatment received) | Descriptive & Odds ratio  | Expecting more than 50% per group to assume having received the active treatment | % of participants who believed to be in active group: Active: >80% (97% as per figure)Sham: >80% (87% as per figure)After each of the 12 intervention sessions. At appointment no. 6 (as per figure): Active: 97% Sham: 87% “Odds ratio for believing that active treatment was received: >10 at all treatment sessions in both groups (all *p* < 0.001)” | Y | BIa 0.94 (0.83 – 1.05)BIp -0.71 (-0.95 - -0.48)(at appointment no. 6) | 2.47 (0.09) |
| Chaibi, 2017 (cervicogenic headache trial) | Manual Therapy (with spinal manipulation, simple | Manual, simulated manoeuvre | Group guessCertainty rating (of real treatment received) | Bang blinding index planned (acc. to protocol) | Unclear  | Group guesses not reported "The four participants who received CSMT believed they received it with a mean 9.1 certainty, whereas the four participants who received placebo believed they received active CSMT with a mean 6.4 certainty." | Y | NA |  |
| Hall, 2008 | Other Manual Therapy, complex | Manual, simulated manoeuvre | Group guessCertainty rating (of real treatment received)Rating expected treatment efficacy  | Descriptive only | Unclear  | Results for first phase of cross-over design: Incorrect guesses: Active: 40%Sham: 75%Second phase: Incorrect guesses: Active: 0%Sham: 100% | Y | BIa 0.2 (-0.66 – 1.06)BIp -0.6 (-1.35 – 0.35)(phase one) | -0.67 (0.44) |
| Kawchuk, 2009 | Manual Therapy (with spinal manipulation, simple | Other, general anaesthesia  | Group guess | Descriptive | Not formalised as study deemed "preliminary" | % of participants who believed to be in active group: Sham: 2/6 (33%) | Not reported | NA |  |
| Lee & Lewis, 2020 | Other Manual Therapy, complex | Manual, simulated manoeuvre | Group guessCertainty rating (of real treatment received) | Difference between groups (t-test) | No difference between groups as successful blinding | % of participants who believed to be in active group: Active: 57% Sham: 47% *p* = 0.06 | Y | BIa 0.14 (-0.38 – 0.66)BIp 0.07 (-0.44 – 0.57) | 0.8 (0.26) |
| Michener, 2015 | Manual Therapy (with spinal manipulation, simple | Manual, simulated manoeuvre | Group guessRating expected treatment efficacy | Chi-square | No difference between groups as successful blinding | % of participants who believed to be in active group: Active: 78.6%Sham: 60.7%*p* = 0.15  | Y | BIa 0.57 (0.27 – 0.88)BIp -0.21 (-0.58 – 0.15) | 2.09 (0.17) |
| Sharpe, 2012 (study 2, same report)  | Cognitive-behavioural and other Psychotherapy, simple | Other, nonspecific cognitive task  | Group guess | Chi-square | No difference between groups as successful blinding | “χ2 = 1.345, *p* = 0.488, with the majority assuming they were allocated to placebo." | Y | NA |  |
| Simcock, 2008  | Other, Music / Sound  | Other, white noise  | asked to recollect distinguishing feature of active intervention (blinding through sedation) | Descriptive | Unclear | "When asked 24 hours after the procedure, patients correctly reported whether they listened to music 60% of the time." | Not reported | NA |  |
| Teys, 2008 | Other Manual Therapy, simple | Manual, simulated manoeuvre | Group guess (cross-over design) | Descriptive | Unclear | "Results of the exit questionnaire showed that three participants (12%) correctly guessed they had only received active treatment and none had correctly guessed that they had received either a sham or control." | Not reported | NA |  |

**Table S4: Overview of studies reporting expectation-related information**, including treatment credibility and satisfaction.The method, timepoint, and result of expectancy-related information are summarized. Criteria for pooling in a meta-analysis of treatment benefit expectation were: Data that were i) clearly attributable to expectation of treatment benefit only (i.e., not presented as a compound measure with questions on treatment credibility, or evaluated as treatment satisfaction), ii) sampled after at least one exposure to the test or control intervention but not after the final of multiple treatment sessions (to avoid confounding with treatment satisfaction), and iii) reported in full and per trial group. Trial reports complying with these criteria are highlighted in green. (n total = 27 trial reports, 29 comparisons between experimental and dedicated control intervention)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Study** | **Intervention type and complexity**  | **Method and measure of expectation assessment**  | **Timepoint of assessment** | **Methods (quoted)** | **Results (quoted)** | **Data pooling of expectation data possible (Y/N)** | **Reason for non-pooling (if applicable)**  |
| Albert, 2012 | Rehabilitation / Physiotherapy, complex | Expectation of outcome:Rating scale (-3 to 3) (Much worse to much better) | Baseline (prior to exposure) | The patients’ expectations of outcome of the 2 treatment regimens were also measured before randomization by patients’ self-report, using the questions (…) “What treatment outcome will you expect to have if you receive ‘Hands off’ exercises = circulation exercises?” Evaluated on a 7-point Likert scale, 3 were worse, 3 were better, 1 and 1 was unchanged. | In % per group and rated score (extracted from figure 2): Much better 44.5% in test, 54.8% in sham; Some better: 42% vs 43.75%; No change: 12.5 % vs 0.9%; Some worse: 0.8% vs. 0%; Much worse: 0% each.  | N | Unexposed expectations only |
| Assefi, 2008 (‘direct’ Reiki vs. ‘direct’ control) | Spiritual / energetic healing, simple | Confidence rating:(in helpfulness of treatment; apparently Y/N answer) | Baseline (prior to exposure) and after final (of eight) treatment session | At week 8, 79 participants rated their certainty about being treated by a Reiki master, 77 reported on their provider’s skill level, and 77 on their confidence in Reiki to relieve symptoms. | (baseline) treatment expectations did not differ across the 4 treatment groups. At week 8 (…) believed that Reiki masters and actors could relieve fibromyalgia symptoms (74% versus 66%, p = 0.38). | N | Unexposed expectations& Lack of construct validity (retrospective assessment after completion of prolonged programme) |
| Assefi, 2008b (‘indirect’ Reiki vs. ‘indirect’ control) | As above | As above | As above  | As above | As above | As above | As above |
| Baldwin, 2013 | Spiritual / energetic healing, complex  | Expectation of benefit: Rating scale (0 to 100) (higher scores indicating higher expectations) | Baseline (prior to exposure) | On arrival (...) Each participant filled out a visual analog scale (VAS) expectancy survey, asking whether they expected the treatment to work. / Secondary outcome variables were (i) expectancy, pretreatment, that the treatment would work / Expectancy that the treatment would work was assessed with a 100mm VAS before the subject entered the treatment room. Each subject was asked to mark a vertical line on the VAS to indicate expectancy. No expectancy at all was represented by 0 mm and definite expectancy by 100 mm. | There was no significant difference between the average self-reported expectation levels of the 5 groups (𝐹 = 0.25, 𝑃 = 0.9). (Mean expectation values were as follows: control: 54.5 Sham: 64.1; PT: 56.5; Reiki: 57.8; and RH: 55.1) | N | Unexposed expectations only |
| Bennell, 2010 | other Manual Therapy, complex | Expectation of benefit:Rating scale (1 to 5) (higher scores indicating higher expectations) | Baseline (prior to exposure) | Baseline demographic information was collected, and participants rated their expectation of a beneficial effect of active physiotherapy treatment on an ordinal scale from 1 to 5, with higher scores indicating higher expectations. | The participants’ expectation of treatment outcomes for active physiotherapy was similar in the two groups (P=0.79); 95/105 (90%) participants who provided this information expected a moderate or large beneficial effect. | N | Unexposed expectations&Insufficient data reporting |
| Bialosky, 2009 | Other Manual Therapy, complex | Expectation of outcome:Rating scale (0 to 100) (no pain to worst imaginable) | At randomization and after "brief exposure to assigned intervention" | Individual expectation for outcome was assessed using the Patient- Centered Outcome Questionnaire (PCOQ). (80) The PCOQ is a 5-item questionnaire that uses individual 101-point numeric rating scales (NRSs) to quantify the usual, desired, and expected levels of pain, fatigue, emotional distress, and interference with daily activities associated with a pain condition. (…) Individual NRSs were anchored with 0 (none) and 100 (the worst imaginable). Participants were asked to answer these questions regarding their CTS-related symptoms. We used ratings for expected level of pain following treatment to quantify expectation for our data analyses. | An interaction of group (NDT versus sham) by time (immediately post-consent versus following brief exposure to assigned intervention) was not observed for expected pain intensity following the study (P = .87), suggesting that the groups did not differ in their expectation of pain relief following brief exposure to the assigned intervention. Additionally, a main treatment effect for time was not observed (P = .13), suggesting that expectation for treatment did not differ at baseline and following brief exposure to the randomly assigned intervention. Table 1: Reports "Expectation of pain" per group | Y | N/A |
| Bialosky, 2014 (vs. ‘enhanced’ control) | As above | As above | As above | As above | As above | As above | N/A |
| Bialosky, 2014 (vs. ‘standard control’) | Manual Therapy (with spinal manipulation), simple | Expectation of outcome:Ordinal rating (More pain, same pain, less pain) | After first treatment | We included satisfaction as a secondary outcome measure to determine whether differing group-related expectations were associated with differences in satisfaction separate from changes in clinical outcomes. We used 2 questions from the North American Spine Society Lumbar Spine Outcome Assessment (22) indicative of satisfaction. (40) Par- ticipants were asked the following questions: 1) ‘‘Would you have the same intervention you received in this study again for low back pain?’’ Possible responses ranged from 1 = definitely not to 5 = definitely yes. 2) ‘‘How would you rate the overall results of the intervention you received in this study for low back pain?’’ Possible responses ranged from 1 = terrible to 6 = excellent. | Raw data per group reported in Table 2; Significant group-related differences were observed in the expected 2-week changes in LBP immediately following the first intervention, c2(6, n = 110) = 20.91, P < .01 (Table 2). A larger percentage of participants receiving the SMT and enhanced placebo SMT expected less pain than those receiving the standard placebo SMT and the no-treatment control group (P < .05). Expected LBP at 2 weeks in response to the intervention did not differ for participants receiving the SMT and enhanced placebo SMT (P = .67) or for participants in the no-treatment control and the standard placebo SMT group (P = .23). These findings suggest that the enhanced placebo SMT was associated with similar expectations for effectiveness as the studied SMT. | YComment: Unclear if expectation or satisfaction were assessed (Authors report that satisfaction was assessed but report expectation-type results) | N/A |
| Bliddal,2014 | Spiritual / energetic healing, complex  | Expectation of benefit:Rating scale (0 to 100) (higher scores indicating higher expectations) | Baseline (prior to exposure) and seven weeks after final (of several) treatment sessions | Before randomisation and at follow-up, participants were asked about their attitudes towards CAM and healing, if they had tried one or both and their expectation concerning the effects. This was done using 0–100 mm visual analogue scales (VAS) with the anchors “0” defined as “no, not at all” or “fully dissatisfied” and “100” as “yes, definitely” or “fully satisfied.” | Neither before nor after therapy did the two groups differ in expectations with respect to the possible effect of healing, and after treatment the participants in the two groups did not differ when asked about the possibility that they had actually received active healing (Table 3). | N | Unexposed expectations& Lack of construct validity (retrospective assessment after completion of prolonged programme |
| Castro-Martín, 2017 | Craniosacral therapy and gentle myofascial release, simple  | Attitudes towards treatment compound evaluation: Attitudes Towards Massage Scale (ATOM): a 9-item measure of overall attitude towards massage comprising the “massage as healthful” and the “massage as pleasant” subscales. | Baseline (unclear if exposed) | The Attitudes Towards Massage Scale (ATOM) is a 9-item measure of overall attitude towards massage comprising the “massage as healthful” and the “massage as pleasant” subscales. Both subscales are reliable (Cronbach a, >.70 and >.80, respectively).43 (Ref: Moyer and Rounds, 2009.) A 2-way repeated-measures analysis of covariance (ANCOVA) with intervention (experimental, placebo) as the between-group variable, time (pre-post) as the within-group variable, and the ATOM subscale scores as covariates was used to examine the effects of the interventions. | The ATOM subscale scores did not influence the results. | N | Unexposed expectations only & Insufficient outcome data reporting |
| Coste, 2021 | Manual Therapy (with spinal manipulation), complex  | Expectation and credibility compound evaluation: Devilly and Borkovec expectancy and credibility questionnaire | After visit 1, 3, 6 | Within 2days after the first, third, and sixth sessions, patients completed the French-language version of the credibility/expectancy questionnaire (CEQ). This instrument includes six items assessing a cognition-based credibility dimension (three items, giving a score of 0–30), and an affect-based expectancy dimension (three items, giving a score of 0–30). Ref: Devilly and Borkovec TD., 2000.  | However, for subjects attending and completing the first ses- sion, there was no difference between groups in terms of credibility and expectancy, in either the scheduled interim assessment on the first 30 patients (2 × 15) performed in October 2016, or in the final samples (Supplemental Table 1). Treatment credibility and expectancy remained stable and similar across groups over the six sessions. At 6weeks (the end of the sessions) and 52weeks, 23 (23%) and 49 (49%) participants, respectively, were lost to follow up. The participants lost to follow up did not differ from the other patients in terms of the principal baseline characteristics (Supplemental Table 2). | PYNote: can be pooled if allowing for a credibility/expectancy compound score. |  |
| Curtis, 2011 | Craniosacral therapy and gentle myofascial release, complex  | Expectation and credibility compound evaluation, individual item responses provided: 4-item modified Borkovec and Nau Scale (mBNS); Each item a scale of 0 to 9, with higher scores indicating higher credibility/expectancy.  | After first treatment | The credibility of the protocols and the expectancy of improvement for each subject were assessed by their com- pletion of a commonly used and adapted self-report instru- ment that had been originally developed by Borkovec and Nau.23,24 Borkovec and Nau used a 0- to 9-point scale for five questions to which the subject could mark his/her de- gree of confidence in treatment rationale and expected out- come, regarding psychologic interventions for a specific psychologic problem: anxiety in public speaking. Others have successfully modified this scale, adapting it for use with other illnesses and interventions.25,26 The first three of the original five items in Borkovec and Nau’s instrument were adapted: (...) The resulting four-item question- naire was completed by each subject at the beginning of the second treatment visit and was placed in an envelope for retrieval by the research staff. The four questions were worded as follows:1. How logical does this type of therapy seem to you for helping people treat or prevent their headache? (Not logical [0]—very logical [9])2. How confident are you that this treatment will be successful in reducing your headache symptoms if you were to have approximately eight such treatments over the course of 2 months? (Not at all confident [0]—very confident [9])3. How confident would you be in recommending this treatment to a friend who has a problem with head- aches? (Not at all confident [0]—very confident [9])4. How competent do you consider the therapist who gave you treatment to be? (Not at all competent [0]—very competent [9]) | Using a 0–9 rating scale, the mean score for perceived logicality of treatment (question 1) was significantly less for LSSM (5.03, SD 2.34) compared to the CST group (6.64, SD 2.19), p = 0.005. Although subjects’ confidence that migraine would improve (question 2) was greater for CST (5.94, SD 2.01) than for LSSM (4.9, SD 2.21), this difference was not statistically significant ( p = 0.06). In response to question 3, significantly more subjects receiving CST (6.08, SD 2.27) would confidently recommend treatment to a friend than those receiving LSSM (4.69, SD 2.49), p=0.03 (Table 3). On the other hand, subjects from both groups were equally and strongly positive about the competence of the therapist (question 4), with no between-group differences (8.41, SD 0.95 versus 8.31, SD 1.09), p = 0.815 on the Wilcoxon sum rank test. The withdrawal/dropout rate was very similar between the groups (11.1% for CST and 15.2% for LSSM [p = 0.62]). (Results also provided in table 4) | Y |  |
| Dougherty, 2014 | Manual Therapy (with spinal manipulation), complex  | Expectation and credibility compound evaluation: 4-item modified Borkovec and Nau Scale, all items averaged (higher the score indicating higher expectations) | After first treatment | (…) we administered the 4-item modified Borkovec and Nau Scale (mBNS)( 28) at the end of the first treatment. The mBNS asks patients to rate on 5-point Likert-type scale the extent to which they were confident in the treatment, confident in recommending the treatment to a friend, felt the treatment was logical, and felt the treatment would be successful in alleviating other complaints. We calculated a mean treatment expectation score by computing the mean of the responses on the 4 items, with the higher the score, the more favorable the expectation.  | Next, we compared the patients’ mBNS treatment expectation scores at the end of the first treatment session using a 1- way analysis of variance with the treatment group as the between-patient factor. We found no significant difference (F1,124 = 0.42, P = .51) in the treatment expectation score between patients who underwent the first sham intervention (mean = 3.88, SE = .08) and patients who underwent the first chiropractic treatment (mean = 3.08, SE = .08). The sham intervention patients’ initial expectations about the treatment were as positive as the expectations of the patients in the SMT group. This pattern of results suggests that we effectively designed our sham intervention as a realistic placebo. | N | Insufficient data reporting |
| Garcia, 2021 | Cognitive behavioural & other psychotherapy, complex | Satisfaction compound score: Four items with various ranges.  | After final treatment session | Satisfaction with treatment was assessed with several items. First, using a 6-point scale (0=strongly disagree and 5=strongly agree), participants rated 4 items: ease of use of the VR headset, enjoyment of the headset, whether the headset helped with pain coping, and desire to continue using the VR headset. These 4 items were summed to create a total satisfaction score. Additionally, 1 item assessed likelihood to recommend VR (0=definitely not recommend and 10=definitely would recommend). One item assessed likelihood to continue using VR if they were able to keep their headset using a response scale (0=definitely would not it and 10=definitely would use it). Because of an error with the electronic survey administration, these data were captured at 1 month posttreatment. | For the 4 summed satisfaction items, the EaseVRx group reported greater satisfaction with treatment than the Sham VR group (4.32 versus 3.46 respectively; P<.001). Similarly, the EaseVRx group reported greater likelihood to recommend VR to someone else compared to the Sham VR group (8.72 versus 6.55, respectively; P<.001). Finally, EaseVRx participants reported greater likelihood to continue using VR if they could keep their headset compared to Sham VR (9.18 versus 7.23, respectively; P<.001). | N | Lack of construct validity |
| González, 2021 | Other Manual therapy, simple  | Meeting of expectations (satisfaction?): Rating scale (1 to 5) (strongly disagree to strongly agree with statement that expectations were met) | 1 week after end of treatment | The degree of compliance with the patients’ expectations on the received intervention was measured with a 5-point Likert scale, one week after the intervention. The statement given to patients was: “The treatment I received has met my expectations”. The patients were then instructed to mark the most appropriate response, ranging from “totally agree” to “strongly disagree". | With regard to the degree of compliance with the patients’ expec- tations on the received intervention, no differences between groups were observed (P = .281). In the NM group, 14 patients (53.8%) marked “totally agree” with the treatment received meeting their expectations; 5 patients (19.2%) chose “agree”; and 7 patients (26.9%) selected “normal.” In the sham NM group, 8 patients (32%) marked “totally agree”; 10 patients (40%) chose “agree”; and 7 patients (28%) marked “normal”. No patient selected the answers “disagree” or “totally disagree”. | N | Lack of construct validity |
| Haller, 2016 | Craniosacral therapy and gentle myofascial release, complex  | Crediblity and Expectancy questionnaire: Devilly and Borkovec (2000) - Rating scale of 1 (not at all) to 9 (very much) | Baseline (prior to exposure) | Furthermore, treatment expectancy was assessed as part of the Credibility/Expectancy Questionnaire on a 9- point rating scale from 1 (not at all) to 9 (very much). (44,45) | Means and SD per group provided (Table 2); Craniosacral therapy: 7.0 (1.1), Sham: 6.5 (1.4), p = 0.136 | N  | Unexposed expectations only |
| Holmgren, 2012 | Rehabilitation / Physiotherapy, complex  | Expectation of benefit: Categoric rating (4 options from no to complete relief) | Baseline (prior to exposure) | At the inclusion visit we documented (...) expectations of effects of treatment (completely restored, quite improved, not improved but some relief of the symptoms, or no expectations of being restored). | All patients included in the study expected a moderate or large treatment effect, with no significant differences between the two groups. |  | Unexposed expectations&Insufficient data reporting |
| Kutner, 2008 | Other Manual Therapy, complex  | Expectation of benefit: Rating scale (1 to 5) (1=not at all helpful, 5=very helpful) | Baseline and at each visit | (…) expected helpfulness of massage for pain (...) were collected at baseline (within 72 hours of study enrollment) and at 3 subsequent weekly visits over the 3-4 week period of participation (sustained outcomes). Final data collection occurred approximately 1 week following the final treatment. | "Perceived helpfulness of massage therapy, mean (SD)" reported per group at BL in Table 1: Massage: 4.0 (1.0), Control: 3.9 (1.1). In text: (…) there was no association between expected helpfulness of massage and study outcomes. | N | Insufficient data reporting |
| Kwekkeboom, 2018 | Cognitive behavioural & other psychotherapy, complex  | Expectation of outcome: Outcome expectancy scale (unreferenced) | Weeks 3, 6, and 9 | We evaluated four potential mediators of intervention effects, collected at baseline, 3, 6, and 9 weeks. The Outcome Expectancy Scale (24,32) was used to measure one's belief in the efficacy of CBS (Cronbach's α = .88‐.91). Higher scores indicate more positive expectancy; From the authors’ own paper (ref 32): "The Outcome Expectancy Scale was used to assess anticipated outcomes of using guided imagery and relaxation interventions, that is, the specific expectation that guided imagery or relaxation could relieve the individual's pain. Beliefs that one's pain can be controlled with guided imagery and relaxation were assessed with six items for each strategy and rated on a seven-point scale. Means were calculated across the six items, with higher scores indicating more positive outcome expectancy." (Ref 24: Also own paper, cannot access; abstract only?) | Change in outcome expectancy differed by group at weeks 6 [F(1,130) = 5.46, P = .01] and 9 [F(1,130) = 3.45, P = .03], with larger improvements in the treatment versus control group. | N | Insufficient data reporting |
| Lehtola, 2010 | Manual Therapy (with spinal manipulation), simple  | Satisfaction rating: Categoric rating (4 options from 'helped remarkably' to 'did not help') | After final (of several) treatment session | The satisfaction with the treatment was examined in the final measurement with three-point patient global impression of change, a modification of the standard seven-point global impression of change (PGIC). This modification was done because satisfaction with the treatment was not the primary outcome measure. The categories were: treatment ‘helped remarkably’, ‘helped to some extent’ and ‘did not help’. | The satisfaction with the treatment was more positive in the manipulation and acupuncture groups than in the placebo group, revealing statistically significance in the comparison of manipulation with placebo (P < 0.00). | N | Lack of construct validity (retrospective satisfaction assessment) |
| Michener, 2015 | Manual Therapy (with spinal manipulation), simple  | Expectation of benefit: Categoric rating (Yes/No answers for 3 expectancy-related questions) | Baseline (prior to exposure) and after first (and only) treatment session | Prior to the treatment, participants answered 3 questions about their perception of the assigned treatment based on the labels of the assigned treatment. The 3 questions were: “Would you expect the effects of the treatment you will receive to: 1- decrease shoulder pain, 2- increase shoulder motion, 3- improve the use of the shoulder”, with yes or no response options to each question. Post-treatment, participants completed the same questions about the perceived effects of the randomly assigned treatment they just received. Each answer was assigned a point value; 0 = no, 1 = yes, with a maximum of 3 points indicating maximum positive perceived effects of treatment. | Differences or lack thereof in perceived effects between treatment groups were assessed in 2 ways. First, there were no significant differences for perceptions of effects between SMT and sham- SMT groups at pre-treatment (t = 0.12, p = 0.90) and at post-treatment (t = 0.40, p = 0.69). We also performed equivalency testing, to test the hypothesis that the groups had the same perceived effects, specifically using a t distribution at 95% confidence indicated with a sample size of n < 30 per group. At pre-test, the mean difference in perceived effects between the SMT and sham-SMT was -0.03 [(95%CI = - 0.60, 0.54), pooled SD = 1.07]. At post-test, the mean difference was 0.11 [(95%CI = -0.41, 0.62), pooled SD = 0.96]. The 95%CI of the perceived effect scores fell entirely within the a priori hypothesized equivalence range of p = 0.73, indicating statistically equivalent means between groups. (Data also in Table 2) | Y |  |
| Moraska, 2017 | Other Manual Therapy, complex  | Expectation of benefit: Details unclear  | Baseline (prior to exposure) | Potential covariates examined at baseline were: (…) massage expectation, Ultrasound expectation (sham) | No results reported. | N | Insufficient data reporting |
| Nguyen, 2021 | Manual Therapy (with spinal manipulation), complex | Expectation and credibility compound evaluation, individual item responses provided: Devilly and Borkovec expectancy and credibility questionnaire | After final treatment session | To ensure no imbalance in the credibility of the interventions and expectancies of the participants, the Credibility/Expectancy Questionnaire was self- administered at 3 months. Ref: 22. Devill and Borkovec, 2000; Borkovec and Nau, 1972. | Scores for all 6 items of the Credibility/Expectancy Questionnaire were numerically higher in the standard OMT group than the sham OMT group (item-specific data provided in eAppendix 8 in Supplement 2).  | N | Lack of construct validity (retrospective expectancy assessment, likely studying satisfaction or even blinding effectiveness) |
| Nicholas, 2013 | Cognitive behavioural & other psychotherapy, complex  | Treatment credibility: Rating scale (1 to 10) (higher scores indicating higher credibility) | After first treatment | At the end of the first session, participants in both intervention groups (PSM and EAC) completed a treatment credibility rating form [8]. These ratings related to expectations of pain relief, of pain management, and how logical treatment appeared, all rated from 0 (not at all confident/logical) to 10 (very confident/logical), and these were compared for equivalence across the 2 treatment groups (PSM and EAC). | T-test analyses revealed no significant differences on treatment credibility and satisfaction for the 2 treatment groups (PSM and EAC). Means (SD) (out of 10) for treatment credibility were 5.62 (1.83) and 5.83 (1.76) (P = 0.59) for the 2 groups, respectively. Means (SD) (out of 10) for treatment satisfaction were 9.24 (1.18) and 9.44 (0.84) (P = 0.34) for the 2 groups, respectively. Sim- ilarly, no significant difference between these groups was found in relation to the question of whether or not they would recommend this treatment for a friend or relative who had a similar pain prob- lem (89% for PSM and 94% for EAC answered ‘‘Yes’’; v2 = 0.080, df = 1, P = 0.44). These results indicate that the 2 interventions had moderate credibility and high satisfaction ratings, and that any differences in outcomes should not be related to differences on these dimensions. | Probably Y (if assuming that treatment credibility and expectancy are comparable) | N/A  |
| Thomas, 2020 | Manual Therapy (with spinal manipulation), complex  | Expectation and credibility compound evaluation: Devilly and Borkovec expectancy and credibility questionnaire | After first (of several) treatment session | Prior to the second treatment intervention, participant expectations for treatment success were assessed using a credibility and expectancy questionnaire.20 The credibility and expectancy instrument consists of 7 questions with a 9-point Likert scale to assess how confident the participant is that the treatment they have received has the potential to reduce their symptoms.  | A comparison of treatment credibility and expectancy ratings across groups was not statistically significant (F2,151 = 1.70, P = .19), indicating that, on average, participants in each group had similar expectations regarding the likely benefit of their assigned treatment. Next, we examined the relationship between these individual treatment expectancy ratings (Table 2), obtained before the second treatment session, and changes in current NPRS ratings (ie, 72 hours posttreatment − baseline). For the sample as a whole, treatment expectancy scores were inversely correlated with the change in pain ratings (r = −0.396; P < .01), indicating that those who had higher expectations of treatment success reported larger decreases in pain with treatment. When these expectancy effects were examined within each treatment group, significant inverse correlations were observed for placebo (r = −0.569; P < .001) and spinal manipulation (r = −0.423; P = .002) but not spinal mobilization (r = −0.188; P = .18). The changes in Roland-Morris Disability Questionnaire ratings (ie, session 9 − session 2) were not significantly related to treatment expectancy. (Note: Table 2 reports correlation analysis between pain outcomes and expectancy); no raw data.  | N | Insufficient data reporting |
| Traeger, 2018 | Cognitive behavioural & other psychotherapy, complex | Expectation and credibility compound evaluation: Devilly and Borkovec expectancy and credibility questionnaire | Baseline (after explanation of trial rationale, before exposure) | Participants completed the credibility and expectancy questionnaire (17) in paper format immediately after the trial clinician explained the rationale for the study and before randomization. (Ref 17: Devilly & Borkovec, 2000) | Treatment credibility scores were not different between groups (mean [SD] credibility and expectancy questionnaire score for patient education vs placebo patient education: 36.6 [8.8] vs 35.3 [10.5]; mean difference, –1.3; 95% CI, –4.0 to 1.4). | N | Unexposed expectations only |
| Vranceanu, 2016 | Cognitive behavioural & other psychotherapy, complex, complex  | Satisfaction rating: Rating scale (1 to 5) (higher scores indicating higher satisfaction) | After final treatment session | At posttest and 6-month follow-up, all participants answered an additional question about overall satisfaction with the program (1–5 Likert scale), with higher scores depicting higher satisfaction. | Satisfaction with participation in the program was high and similar at both posttest and 6- month follow-up between the 3RP-NF (4.19 and 3.92 of 5) and the HEP-NF (4.10 and 3.79 of 5) groups. | N | Lack of construct validity |
| Wang, 2009 | Rehabilitation / Physiotherapy, complex | Expectation of benefit: Rating scale (1 to 5) (higher scores indicating higher expectations) | Baseline (prior to exposure) | Additional measures included (…) outcome expectations for exercise (score range 1–5, where 1 = no outcome expectations) | Table 1: Mean and SD per group; Tai Chi (n = 20): 4.1 ± 0.6, Attention control (20): 4.3 ± 0.4 | N | Unexposed expectations only |
| Wang, 2010 | Rehabilitation / Physiotherapy, complex | Expectation of benefit: Rating scale (1 to 5) (higher scores indicating higher expectations) | Baseline (prior to exposure) | Scores on the Outcome Expectations for Exercise Scale range from 1 to 5, with higher scores indicating high outcomeexpectations. (Table 1 subscript) Ref: 26. Steinhardt MA, Dishman RK. Reliability and validity of expected outcomes and barriers for habitual physical activity. J Occup Med 1989;31:536-46. | Notably, the baseline outcome expectations of benefit from an exercise intervention were similar in the tai chi and control groups (3.7±0.8 and 3.9±0.7, respectively), indicating that our neutral presentation of the interventions may have been successful. Table 1: Tai Chi group (n =33): 3.7±0.8; Control group (33): 3.9±0.7 | N | Unexposed expectations only |