**Supplementary Information**

***Lack of Evidence for the Effect of Oxytocin on Placebo Analgesia and Nocebo Hyperalgesia***

**Methods**

*Participants*

Participants were recruited from the university. They reported no history of neurological or psychiatric disorders (including substance abuse and obesity). They were required to not drinking caffeine, alcohol, or nicotine within 2 hours before the experiment. No female participant reported pregnancy or used hormonal contraception in the last month. Prior to the experiment, participants provided demographic information and completed a set of questionnaires: Pain Catastrophizing Scale (1), Eysenck Personality Questionnaire-Neuroticism (2), Interpersonal Trust Scale (3), and State Anxiety Inventory (4). After the formal task, participants conducted the State Anxiety Inventory again and they were asked to guess whether they took OT or a placebo. **Table S1 and S2** show that there were no significant differences between the 4 groups of participants regarding demographic information and psychological traits.

The Institutional Review Board at South China Normal University approved all study procedures. The experiment was carried out in accordance with the approved guidelines. Participants gave written informed consent before beginning any study procedures. Protocols were pre-registered at Chinese Clinical Trial Registry (ChiCTR1800015647, <http://www.chictr.org.cn/showproj.aspx?proj=22981>). Please note the deviation of the final study from the trial preregistration in terms of sample size. The initial preregistration also did not provide important details, such as planned statistical analysis and main outcomes. The CONSORT Flow diagram and checklist are presented in **Figures S1 and S2**.

*Materials*

Electric stimulations were square pulses delivered to the right volar forearm by a Grass SD9 stimulator (Warwick, U.S.A.) with two 0.75cm diameter electrodes. The experiment was conducted in a quiet room with a temperature of 24℃. All stimuli were presented using the Eprime 2.0 software (Version 2.0.8.22, http://www.pstnet.com).

*Experimental procedures*

Both studies consisted of three stages: (1) calibration, (2) intranasal administration, (3) conditioning paradigm (Experiment 1) or verbal suggestion reinforced paradigm (Experiment 2).

*Stage 1: Calibration*

Since individuals differ in their pain sensitivity, the threshold and tolerance to electric stimuli for each participant was calibrated at the beginning of the experiment. Calibrations were manipulated via ascending voltage of the electric currents with a fixed delivering duration of 80ms. Participants were asked to rate their pain intensity on a 9-point self-report Numeric Rating Scale (NRS, 1 = a little pain, 5 = moderate pain, and 9 = unbearable pain). After finding the physical voltage that participants rated around 3 (low pain), this parameter kept constant in further procedures. The next step was to find electrical parameters that would elicit low pain at ≈3 rating, high pain at ≈7 rating, and moderate pain at ≈5 rating on the NRS for each participant. With a previously determined constant voltage, we increased the stimulation time of electric currents, starting from 80ms to 800ms (increasing in sequence at multiples of 80ms), to increase participants’ feeling of pain. Participants were given 2s to rate on the NRS by pressing the corresponding number buttons on the keyboard. Once the low, moderate, and high pain levels for each participant were determined, the participants were tested for rating response consistency. A random sequence of three low- and three high-intensity pain stimuli was administered. If the participants could reliably rate the high stimuli as more intense than the low stimuli, they proceeded to the next step of the experiment. The calibration lasted for 5 to 10 minutes.

*Stage 2: Intranasal administration*

In a randomized double-blind study, participants received either 24 IU oxytocin, 24 IU saline, 40 IU oxytocin, or 40 IU saline intranasally. 24 IU was administered with three puffs per nostril and 40 IU was administered with five puffs per nostril. After the drug administration, participants rested for 40 minutes and then proceeded to the next procedures.

*Stage 3 for Experiment 1: Conditioning paradigm*

In Experiment 1, participants performed the conditioning procedure (5). Two abstract images were used as cues during the conditioning stage. In total, there were 40 trials: one cue was coupled with a high pain level and the other cue was coupled with a low pain level for 20 trials each. The specific assignment of cues to a given trial type was fully counterbalanced across participants. Each trial started with the presentation of a fixation cross. The abstract cue was presented for 2s, followed by an interval of 2s, then the electric current stimulus was delivered to the right volar forearm. Participants rated how much pain they felt, using the same 9-point NRS. There was a 7-9s blank interval between trials to allow the feeling of pain dissipating. The conditioning sequence lasted for 15 minutes, with a break of 1 minute during the procedure. An experimental assistant was seated in a chair near the desk that the monitor was on, facing the side of the participant to make sure that participants were engaged in the experiment.

There were three cues in test stage: the cue previously associated with high pain (high-pain cue), the cue previously associated with low pain (low-pain cue), and a new cue that participants had not seen before (neutral cue). Unknown to participants, the cues were all paired with identical moderate electric shocks for 20 trials each. The streamline of each trial was the same as in the conditioning stage. This testing stage lasted for 20 minutes, with a break of 1 minute during the procedure. In addition, similar to previous studies, another six “booster trials” were added to the test stage (6). Specifically, three high-pain cues and three low-pain cues were respectively paired with their original electric currents. As all electric current stimuli were at the same level of intensity in the test stage, the booster trials served to prevent habituation and extinction and to ensure participants remain vigilant. These booster trials were not included in the statistical analysis of analgesic and hyperalgesic effects.

*Stage 3 for Experiment 2: Verbal suggestion reinforced paradigm*

Experiment 2 followed a well-established placebo analgesia paradigm including both expectation and conditioning components (7). We informed participants that the aim of the study was to investigate the effect of ointments on pain perception. Three identical inert ointments were applied to three sites on each participant’s forearm, with the sites randomized across participants. A female experimenter introduced ointments as creams that increase pain (nocebo, red label), reduce pain (placebo, green label), and have no effect on pain (control, blue label), respectively. Participants were then told to wait for 10 minutes for the creams to take effect. Of note, we did not randomize cream colors across participants given that the color itself impacts on placebo/nocebo responses with red color associating with hazard and green color associating with safety.

Next, verbal suggestions were reinforced by a conditioning procedure to convince participants that creams applied are effective in reducing or increasing pain. Participants were told that they could be stimulated on three skin areas with moderate intensity of shock (i.e., pain level 5). Unknown to them, however, the shock intensity was lowered to pain level 3 during the placebo condition and was heightened to pain level 7 during the nocebo condition. This conditioning stage consisted of 3 sessions with 10 stimulations for each cream session, lasting for 10 minutes in total. Each session started with a word reminder such as high pain, low pain or control pain condition. After the reminder, each trial in the session started with a cross fixation with the font color consisted with the cream label for 2s. Then the electric current stimulus was delivered to the corresponding skin area that administered creams. Participants rated how much pain they felt after each shock, using the same 9-point NRS. There was a 7-9s blank interval between trials. After the reinforced conditioning, participants were asked to rate how much did them expect the ointments to increase pain / reduce pain on a rating scale ranging from 0 = ‘no effect at all’ to 4 = ‘very effective’. Participants (n=7) who reported ‘no effect at all’ were excluded from the data analysis.

In the subsequent testing stage, all three sites were paired with moderate shocks for 20 times each. The testing stage lasted for 20 minutes with short breaks during the procedure. The procedure for each trial was the same as in the reinforced conditioning stage.

**Power calculation**

The main purpose of our study was to examine the effects of oxytocin and its dosage on placebo and nocebo responses. Based on the initial effect size of d = 0.495 reported by Kessner, Sprenger (8) that showed the effect of oxytocin on placebo analgesia in males, a sample size calculation using G\*Power for a repeated-measures between-factors ANOVA with 4 groups (group: oxytocin/saline X dosage: 24 IU/40 IU) and 2 measurements (placebo/nocebo responses) revealed that 60 male participants in total (15 in each group) would be needed to obtain a power of 0.95 at an alpha level of 0.05 (9). To further explore the effects of oxytocin on females, we applied the same sample size of 15 in each group to female participants. This led to a sample size of 30 male and female participants altogether in each group. Considering the exclusion of subjects, we recruited 40 subjects in each group.

**Results**

*Successful induction of placebo and nocebo effects independent of treatment.*

In experiment 1, a 2 (group: oxytocin/saline) X 2 (dosage: 24 IU/40 IU) X 3 (cue: high/control/low) repeated measures ANOVA showed that there was a significant effect of the cue on pain ratings during the test stage (F (1, 156) = 233.78, p < 0.001). Post hoc tests using Bonferroni corrections indicated that participants rated the stimuli following the high-pain cues (mean ± SE: 5.02 ± 0.10) more painful than stimuli following control cues (mean ± SE: 4.44 ± 0.09) and low-pain cues (mean ± SE: 3.65 ± 0.09) which indicated a nocebo effect, and low-pain cues stimuli as significantly less painful than control cues stimuli which indicated a placebo effect (all Ps < 0.001). There were no other significant main effects and interactions (all Ps > 0.05).

In experiment 2, a similar 2 X 2 X 3 repeated measures ANOVA showed that there was a significant effect of the cue on pain ratings during the test stage (F (1, 142) = 40.33, p < 0.001). Post hoc tests using Bonferroni corrections indicated that participants rated the stimuli following the high-pain cues (mean ± SE: 4.82 ± 0.15) more painful than stimuli following control cues (mean ± SE: 4.01 ± 0.11) and low-pain cues (mean ± SE: 3.37 ± 0.12), and low-pain cues stimuli as significantly less painful than control cues stimuli (all Ps < 0.001), indicating a successful induction of placebo and nocebo effects. Unexpectedly, there was a main effect of dosage (F (1, 142) = 5.50, p = 0.020), with 40 IU (mean ± SE: 4.26 ± 0.13) showing higher pain ratings than 24 IU (mean ± SE: 3.86 ± 0.12). There were no other significant main effects and interactions (all Ps > 0.05).

Altogether, these results demonstrated that the conditioning and verbal suggestions successfully induced placebo and nocebo effects independent of treatment.

*Effects of oxytocin on the extinction of placebo and nocebo responses.*

To test whether oxytocin affects the extinction process of placebo analgesia and nocebo hyperalgesia responses in the test stage, we divided the test stage into early and late sessions. Using group (oxytocin versus saline) and dosage (24 IU versus 40 IU) as between-subject factors and time (early/late: first versus second half trials in the test) as within-subject factor, we tested how these factors affected placebo responses by treating control minus low cue rating difference as the dependent variable. The same analysis was conducted for nocebo responses with high minus control cue rating differences as the dependent variable.

In Experiment 1, results showed no main effect of time on placebo responses (time, F (1, 156) = 3.15, p = 0.078) as well as no interactions with other factors (time X group, F (1, 156) = 0.02, p = 0.902; time X dosage, F (1, 156) = 0.55, p = 0.460; time X group X dosage, F (1, 156) = 0.24, p = 0.627). The analysis on nocebo responses showed a significant effect of time (time, F (1, 156) = 23.91, p < 0.001), with the first half of trials (mean ± SE: 0.73 ± 0.07) revealing greater nocebo responses than the second half of trials (mean ± SE: 0.42 ± 0.06). There were no interactions between time and other factors (time X group, F (1, 156) = 0.19, p = 0.666; time X dosage, F (1, 156) = 0.40, p = 0.527; time X group X dosage, F (1, 156) = 0.30, p = 0.587).

In Experiment 2, results showed no main effect of time on placebo responses (time, F (1, 142) = 0.001, p = 0.981) and no interactions between time and other factors (time X group, F (1, 142) = 0.96, p = 0.328; time X dosage, F (1, 142) = 0.01, p = 0.942; time X group X dosage, F (1, 142) = 2.83, p = 0.095). The analysis on nocebo responses also showed no main effect of time (time, F (1, 142) = 0.51, p = 0.478), and no interactions between time and other factors (time X group, F (1, 142) = 0.05, p = 0.819; time X dosage, F (1, 142) = 0.29, p = 0.590; time X group X dosage, F (1, 142) = 0.08, p = 0.784).

Taken together, these results demonstrated that oxytocin did not influence the extinction of placebo analgesia and nocebo hyperalgesia effects, in consistent with Skvortsova et al.’ findings (10).

*Equivalence test and Bayesian hypothesis test to assess the null findings of oxytocin effects.*

Given that all main effects and the interaction effect of the group X dosage ANOVA results were non-significant in two experiments, we followed these null hypothesis significance tests with equivalence testing and Bayesian hypothesis testing to assess the sensitivity of our null findings of oxytocin on placebo and nocebo responses compared to the saline treatment.

Equivalence testing was conducted in RStudio using the TOSTER package (11).To determine the smallest effect size of interest (SESOI) for setting equivalence bounds, we calculated the mean of SESOI based on a recent meta-analysis report that synthesized 32 intranasal oxytocin studies as well as a set of unpublished data (12). The calculated SESOI average was 0.58. Therefore, we set the lower and upper equivalence bounds to −0.58 and 0.58. In Experiment 1, equivalence tests in comparing oxytocin and saline treatment was significant on both placebo (t(131.24) = 3.20, p < 0.001) and nocebo responses (t(155.6) = -3.04, p = 0.001). Equivalence tests for Experiment 2 also revealed significant results (placebo: t(143.79) = 3.09, p = 0.001; nocebo: t(143.99) = 3.20, p < 0.001).

We also used Bayesian hypothesis testing in JASP (version 0.11.1.0) to assess the null effects of oxytocin on placebo and nocebo response (13). Bayesian testing is particularly beneficial for providing information on the relative degree of evidence that the data provide in favor of either the alternative or null hypotheses (14). In Bayesian testing, Bayes Factor is calculated to infer the ratio of the posterior odds of the alternative and null hypothesis to its prior odds (15). Bayesian hypothesis tests in comparing oxytocin and saline groups found no support for the alternative hypothesis in both experiments (Experiment 1: placebo BF10 = 0.187, nocebo BF10 = 0.204; Experiment 2: placebo BF10 = 0.212; nocebo BF10 = 0.212).

To sum up, findings from equivalence testing and Bayesian testing demonstrated that the oxytocin and saline treatment showed equivalent effect on placebo and nocebo responses.

*Meta-analysis of oxytocin effects on placebo effects.*

To improve estimates of the size of the effect, we combined the results from multiple studies and did a meta-analysis. Four studies and our two experiments were included **(Table S3)**. The results of the individual studies were transformed into common metric of the standardized difference (Cohen’s d) between oxytocin and control conditions. The meta-analysis was conducted by using the random-effects model of the “metafor” R package (16). Our analysis of six studies showed that the combined effect size of oxytocin on placebo effect was small and not significantly different from zero (Cohen’s d = 0.02, 95% CI [-0.14, 0.18], z = 0.25, p = 0.80 **(Figure S3)**. Influence measures showed that Kessner, Sprenger et. al.’ study (8) with the largest effect size was identified as a potential outlier and influential case (17). Compared with the other five cases, this study incurred the largest change in the Cook’s distance (0.62) and standardized residuals (2.16). Although meta-analyses possess more power to detect effects than individual studies, our meta-analysis is based on a small number of individual studies and has limited power. The results of our exploratory data analysis should be interpreted with great care.

**References**

1. Sullivan MJ, Bishop SR, Pivik J. The pain catastrophizing scale: development and validation. Psychological assessment. 1995;7(4):524.

2. Eysenck HJ, Eysenck SBG. Manual of the Eysenck Personality Scales (EPS Adult): Comprising the EPQ-Revised (EPQ-R), EPQ-R Short Scale, Impulsiveness (IVE) Questionnaire: Hodder & Stoughton; 1991.

3. Rotter JB. A new scale for the measurement of interpersonal trust 1. Journal of personality. 1967;35(4):651-65.

4. Spielberger CD. State‐Trait anxiety inventory. The Corsini encyclopedia of psychology. 2010:1.

5. Egorova N, Yu R, Kaur N, Vangel M, Gollub RL, Dougherty DD, et al. Neuromodulation of conditioned placebo/nocebo in heat pain: anodal vs cathodal transcranial direct current stimulation to the right dorsolateral prefrontal cortex. Pain. 2015;156(7):1342-7.

6. Jensen K, Kirsch I, Odmalm S, Kaptchuk TJ, Ingvar M. Classical conditioning of analgesic and hyperalgesic pain responses without conscious awareness. Proceedings of the National Academy of Sciences. 2015;112(25):7863-7.

7. Freeman S, Yu R, Egorova N, Chen X, Kirsch I, Claggett B, et al. Distinct neural representations of placebo and nocebo effects. Neuroimage. 2015;112:197-207.

8. Kessner S, Sprenger C, Wrobel N, Wiech K, Bingel U. Effect of oxytocin on placebo analgesia: a randomized study. JAMA. 2013;310(16):1733-5.

9. Faul F, Erdfelder E, Lang A-G, Buchner A. G\* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior research methods. 2007;39(2):175-91.

10. Skvortsova A, Veldhuijzen DS, Van Middendorp H, Colloca L, Evers AW. Effects of oxytocin on placebo and nocebo effects in a pain conditioning paradigm: a randomized controlled trial. The Journal of Pain. 2019.

11. Lakens D. Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. Social psychological and personality science. 2017;8(4):355-62.

12. Quintana DS. Revisiting non-significant effects of intranasal oxytocin using equivalence testing. Psychoneuroendocrinology. 2018;87:127-30.

13. Marsman M, Wagenmakers E-J. Bayesian benefits with JASP. European Journal of Developmental Psychology. 2017;14(5):545-55.

14. Tabak BA, Teed AR, Castle E, Dutcher JM, Meyer ML, Bryan R, et al. Null results of oxytocin and vasopressin administration across a range of social cognitive and behavioral paradigms: Evidence from a randomized controlled trial. Psychoneuroendocrinology. 2019;107:124-32.

15. Jeffreys H. The theory of probability: OUP Oxford; 1998.

16. Viechtbauer W. Conducting meta-analyses in R with the metafor package. Journal of statistical software. 2010;36(3):1-48.

17. Viechtbauer W, Cheung MWL. Outlier and influence diagnostics for meta‐analysis. Research synthesis methods. 2010;1(2):112-25.

**Table S1.** Participants’ characteristics and experimental results (mean ± SE) in Experiment 1.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **24 IU** | | **40 IU** | | ***F/χ2*** | ***P*** |
|  | **Oxytocin** | **Saline** | **Oxytocin** | **Saline** |
|  | **(N=46)** | **(N=41)** | **(N=37)** | **(N=36)** |
| **Characteristics** |  |  |  |  |  |  |
| Age (years) | 20.43±0.30 | 19.54±0.22 | 19.78±0.26 | 20.03±0.24 | 2.33 | 0.076 |
| BMI | 20.59±0.54 | 20.85±0.61 | 20.07±0.36 | 19.76±0.32 | 1.01 | 0.392 |
| PCS | 19.67±1.55 | 18.22±1.34 | 17.73±1.34 | 18.74±1.75 | 0.32 | 0.810 |
| NEO | 4.89±0.52 | 4.88±0.40 | 4.62±0.55 | 4.63±0.49 | 0.09 | 0.965 |
| ITS | 79.00±1.30 | 78.12±1.56 | 79.97±1.70 | 78.82±2.01 | 0.214 | 0.887 |
| S-AI before | 35.72±1.20 | 37.40±1.25 | 35.83±1.15 | 36.06±1.34 | 0.41 | 0.747 |
| S-AI after | 37.17±1.26 | 37.90±1.29 | 37.08±1.06 | 37.18±1.44 | 0.09 | 0.965 |
| Guess oxytocin | 23 | 18 | 9 | 16 | 6.10 | 0.107 |
| Guess saline | 23 | 23 | 28 | 20 |
| Shock intensity | 2.19±0.13 | 2.64±0.25 | 2.44±0.11 | 2.31±0.15 | 1.36 | 0.258 |
| **Results** |  |  |  |  |  |  |
| **Pain rating differences in conditioning stage** | | | | | | |
| High-low | 3.62±0.19 | 3.46±0.16 | 3.52±0.20 | 3.79±0.21 | 0.55 | 0.652 |
| **Pain rating differences in testing stage** | | | | | | |
| Control-low | 0.75±0.09 | 0.86±0.14 | 0.79±0.09 | 0.77±0.14 | 0.18 | 0.911 |
| High-control | 0.63±0.10 | 0.44±0.11 | 0.60±0.12 | 0.66±0.13 | 0.67 | 0.554 |

Abbreviations: BMI, body mass index; PCS, pain catastrophizing scale; NEO, Eysenck personality questionnaire-neuroticism; ITS, interpersonal trust scale; S-AI before, state anxiety inventory conducted before drug administration; S-AI after, state anxiety inventory conducted after the whole experimental procedures.

**Table S2.** Participants’ characteristics and experimental results (mean ± SEM) in Experiment 2.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **24 IU** | | **40 IU** | | ***F/χ2*** | ***P*** |
|  | **Oxytocin** | **Saline** | **Oxytocin** | **Saline** |
|  | **(N=38)** | **(N=39)** | **(N=30)** | **(N=39)** |
| **Characteristics** |  |  |  |  |  |  |
| Age (years) | 20.21±0.30 | 19.90±0.28 | 19.97±0.32 | 19.97±0.24 | 0.24 | 0.869 |
| BMI | 21.15±0.61 | 20.35±0.45 | 19.90±0.37 | 19.96±0.33 | 1.55 | 0.205 |
| PCS | 21.45±1.48 | 17.38±1.28 | 17.57±1.53 | 18.68±1.48 | 1.74 | 0.162 |
| NEO | 4.55±0.56 | 4.64±0.40 | 4.57±0.55 | 4.81±0.50 | 0.06 | 0.982 |
| ITS | 80.89±1.52 | 77.05±1.49 | 81.70±1.82 | 78.08±1.91 | 1.71 | 0.162 |
| S-AI before | 36.34±1.61 | 37.42±1.22 | 37.46±1.09 | 35.77±1.32 | 0.38 | 0.767 |
| S-AI after | 34.97±1.60 | 36.61±1.38 | 34.92±1.46 | 36.26±1.57 | 0.33 | 0.806 |
| Guess oxytocin | 20 | 11 | 11 | 19 | 5.89 | 0.117 |
| Guess saline | 18 | 28 | 19 | 20 |
| Shock intensity | 1.92±0.10 | 1.97±0.10 | 2.34±0.12 | 2.15±0.14 | 2.492 | 0.063 |
| **Results** |  |  |  |  |  |  |
| **Pain rating differences in conditioning stage** | | | | | | |
| Control-low | 2.18±0.24 | 1.68±0.25 | 2.27±0.24 | 2.20±0.29 | 1.15 | 0.332 |
| High-control | 1.40±0.26 | 2.03±0.31 | 1.51±0.36 | 1.77±0.34 | 0.83 | 0.480 |
| **Pain rating differences in testing stage** | | | | | | |
| Control-low | 0.57±0.21 | 0.76±0.28 | 0.60±0.31 | 0.63±0.29 | 0.10 | 0.959 |
| High-control | 0.53±0.26 | 0.84±0.33 | 1.09±0.36 | 0.77±0.32 | 0.48 | 0.694 |

Abbreviations: BMI, body mass index; PCS, pain catastrophizing scale; NEO, Eysenck personality questionnaire-neuroticism; ITS, interpersonal trust scale; S-AI before, state anxiety inventory conducted before drug administration; S-AI after, state anxiety inventory conducted after the whole experimental procedures.

**Table S3. A summary of studies examining the effect of oxytocin on placebo analgesia and nocebo hyperalgesia.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Study** | **Groups and sample size** | **Pain stimulation** | **Paradigms inducing**  **placebo/nocebo responses** | **Findings** |
| Kessner, 2013 | 1. 40 IU OT: n=37 male  2. Control: n=38 male | Heat stimuli | Verbal suggestions:  Ointment that reduces pain (placebo)  An inert ointment (control) | OT boosted placebo analgesia. |
| Colloca, 2016 | 1. 24 IU OT: n=17 female+17 male  2. Control: n=12 female+12 male | Electrical stimuli | Verbal suggestions:  Green light indicates no pain/less pain (placebo)  Red light indicates pain (control) | OT had no effect on placebo analgesia. |
| Skvortsova, 2018 | 1. 24 IU OT+suggestion: n=27 female  2. Control+suggestion: n=27 female  3. 24 IU OT: n=27 female  4. Control: n=27 female | Pain:  cold pressor test  Itch:  histamine iontophoresis | Verbal suggestions:  Oxytocin that decreases pain and itch  Placebo that decrease pain and itch  Oxytocin without suggestion  Placebo without suggestion | OT had no effect on placebo analgesia. |
| Skvortsova, 2019 | 1. 40 IU OT: n=39 male  2. Control: n=37 male | Heat stimuli | Conditioning + verbal suggestions:  Low-pain cue (placebo)  High-pain cue (nocebo)  Moderate-pain cue (control) | OT had no effect on placebo analgesia and nocebo hyperalgesia. |
| Our study:  Experiment 1 | 1. 24 IU OT: 28 female+18 male  2. 24 IU control: 24 female+17 male  3. 40 IU OT: 19 female+18 male  4. 40 IU control: 20 female+16 male | Electrical stimuli | Conditioning:  Low-pain cue (placebo)  High-pain cue (nocebo) | OT had no effect on placebo analgesia and nocebo hyperalgesia. |
| Our study:  Experiment 2 | 1. 24 IU OT: 19 female+19 male  2. 24 IU control: 23 female+16 male  3. 40 IU OT: 19 female+11 male  4. 40 IU control: 23 female +16 male | Electrical stimuli | Verbal suggestions:  Ointment that decreases pain (placebo)  Ointment that increases pain (nocebo)  An inert ointment (control) | OT had no effect on placebo analgesia and nocebo hyperalgesia. |

**Figure S1.** CONSORT Flow Diagram for Experiment 1.



**CONSORT 2010 Flow Diagram**

Analysed (n=77)  
 Excluded from analysis (give reasons) (n=2)

Failure of distinguishing different pain intensities during calibration

Allocated to intervention (oxytocin group, n=85)

 Received allocated intervention (n=85)

 Did not receive allocated intervention (give reasons) (n=0)

## Follow-Up

## Analysis

Lost to follow-up (give reasons) (n=0)

Discontinued intervention (give reasons) (n=0)

Lost to follow-up (give reasons) (n=0)

Discontinued intervention (give reasons) (n=0)

## Enrollment

## Allocation

Allocated to intervention (placebo group n=79)

 Received allocated intervention (n=79)

 Did not receive allocated intervention (give reasons) (n=0)

Randomized (n=164)

Excluded (n=0)

  Not meeting inclusion criteria (n=0)

  Declined to participate (n=0)

  Other reasons (n=0)

Assessed for eligibility (n=164)

**Experiment 1**

**Figure S2.** CONSORT Flow Diagram for Experiment 2.

Analysed (n=83)  
 Excluded from analysis (give reasons) (n=2)

Failure of distinguishing different pain intensities during calibration



**CONSORT 2010 Flow Diagram**

Analysed (n=68)  
 Excluded from analysis (give reasons) (n=5)

Doubt of ointments’ effects

Analysed (n=78)  
 Excluded from analysis (give reasons) (n=2)

Doubt of ointments’ effects

Allocated to intervention (oxytocin group, n=73)

 Received allocated intervention (n=73)

 Did not receive allocated intervention (give reasons) (n=0)

## Follow-Up

## Analysis

Lost to follow-up (give reasons) (n=0)

Discontinued intervention (give reasons) (n=0)

Lost to follow-up (give reasons) (n=0)

Discontinued intervention (give reasons) (n=0)

## Enrollment

## Allocation

Allocated to intervention (placebo group n=80)

 Received allocated intervention (n=80)

 Did not receive allocated intervention (give reasons) (n=0)

Randomized (n=153)

Excluded (n=1)

  Not meeting inclusion criteria (n=0)

  Declined to participate (n=1)

  Other reasons (n=0)

Assessed for eligibility (n=154)

**Experiment 2**

**Figure S3.** Forest plot of effect sizes for oxytocin on placebo effect studies. Square sizes represent study weights. Filled diamonds represent summary effect sizes.

