Oxytocin in Chronic Neck and Shoulder Pain

Lincoln M Tracy, Nellie Georgiou-Karistianis, Stephen J Gibson, Melita J Giummarra

Abstract

The overall aim of this study is to examine whether oxytocin has a mechanistic effect on pain perception, physical functioning, and physiological and psychological arousal in individuals with chronic musculoskeletal neck and shoulder pain. This study involves receiving a one-off administration of intranasal oxytocin, relative to placebo, to investigate the role of this peptide in pain experience. Oxytocin is a hormone that is produced naturally within the human body. Although it is traditionally known for its role in initiating childbirth and breastfeeding, oxytocin has also been shown to have pain- and anxiety-reducing effects. We expect that oxytocin will help modulate responses to pain-related experiences.

Citation: Lincoln M Tracy, Nellie Georgiou-Karistianis, Stephen J Gibson, Melita J Giummarra Oxytocin in Chronic Neck and Shoulder Pain. **protocols.io**

dx.doi.org/10.17504/protocols.io.hwib7ce

Published: 10 May 2017

Before start

Methods

Study Design

This study employed a randomised, double-blind, placebo-controlled cross-over design adhering to CONSORT guidelines (Moher et al., 2012). Each participant was tested under two acute treatment conditions separated by a washout period of at least 14 days (mean days between testing sessions = 16.3; range = 14 – 36). There are several advantages to this design. First, there is a reduction in the influence of potential extraneous confounding factors (e.g., age and sex of participants), as each participant is tested under both conditions. Second, there is higher statistical power compared to a between-subjects design, so that the required sample size to detect meaningful effects is smaller (Senn, 2002). Potential participants were screened for eligibility and randomised (as per procedures described below) to complete the experimental protocol by author LMT.

The protocol was approved by the Monash University Human Research (CF15/659 – 2015000303) and the Alfred Human Research (111/16) Ethics Committees and followed the Helsinki Declaration of 1975. The study was registered with the Australian Government Therapeutic Goods Administration under the Clinical Trial Notification scheme (protocol number CT-2016-CTN-01313-1) and the Australian New Zealand Clinical Trials Registry (www.anzctr.org.au; registration number ACTRN12616000532404).

We aimed to recruit 25 participants with CNSP, and 25 age- and sex-matched healthy, pain-free,

controls, providing a total sample size of 50. This target was set following power analysis via G*Power, in which we estimated small-medium effects with 80% power and alpha of .05 (Faul et al., 2007). This sample size was consistent with that used in prior research studies on pain (e.g., Rash and Campbell, 2014; Scott et al., 2005), and allowed for participant attrition of 10% and data loss as a result of technical failures and other unforeseen circumstances.

Study Setting

Volunteers with CNSP were recruited via local private physiotherapy clinics, and online advertising between September 2015 and December 2016. Data collection ceased once the a priori defined sample size had been recruited. All participants gave written, informed consent and received $120AU for their participation. Participants were free to withdraw their consent at any point during the study. The two testing sessions were always scheduled to commence between 9:00 am and 10:30 am, to prevent excessive diurnal variation in hormonal levels (Brondino et al., 2017). Participants were greeted by a male experimenter, who guided them through the experimental tasks. Participants completed baseline questionnaires before commencing the experimental tasks. All data were collected in a behavioural research laboratory at Monash University.

Participant Eligibility

Volunteers who self-reported chronic neck and shoulder pain of musculoskeletal origin, aged from 18 to 60 years, were eligible to participate if their pain was constant (i.e., present on a daily basis) and had lasted for at least 12 consecutive months. Initially we planned to recruit individuals with whiplash associated disorders (Grade I-III), however, the initial recruitment rate was slow, despite attempting to recruit through a private physiotherapy clinic that specialised in treating patients with whiplash associated disorders. Therefore, the inclusion criteria for the study were broadened to include individuals with CNSP of more diffuse origin.

Potential CNSP participants were excluded from the study if they were currently receiving any form of active treatment (e.g., physiotherapy, manual therapy); if they reported medical conditions suspected or known to be associated with altered pain sensitivity (e.g., diabetes, dysthyroidism); a current history of cancer or other malignancies, asthma or any other chronic respiratory conditions, psychological or psychiatric conditions, alcohol or substance abuse disorders, hypertension or heart disease, neurodegenerative diseases, colour blindness, moderate-to-severe traumatic brain injury; medications known to affect the autonomic nervous system (e.g., beta blockers, selective serotonin reuptake inhibitors, opioids, benzodiazepines, tricyclic antidepressants, anticholinergics, antiarrhythmics); pregnancy or breastfeeding; any known allergies; or if they were non-ambulant. Participants at risk of anxious and/or depressive disorders, or suicidal ideation (as indicated by scores > 15 on the Beck Anxiety Inventory-I, scores > 20 on the Beck Depression Inventory-II, or a score ≥ two on item nine of the Beck Depression Inventory-II; respectively) were excluded due to the known associations between such disorders and altered pain perception (Dickens et al., 2003; Oktay et al., 2008).

Age- and sex-matched healthy, pain-free individuals were recruited as control participants, provided they had not experienced any musculoskeletal pain in the past 12 months for which they had sought medical treatment (e.g., from a doctor or physiotherapist). Potential control participants were screened against the same inclusion and exclusion criteria as participants with CNSP as described above.

Procedure

Upon arrival, participants administered a nasal spray as per the randomisation schedule (see below), and after a 45 minute waiting period commenced a series of experimental tasks examining the effects of intranasal oxytocin on heart rate variability (HRV), physical functioning, thermal heat pain thresholds, sensitivity to experimentally-induced pain, and the anticipation and experience of thermal heat stimuli. Methods relating specifically to these experimental tasks are described below. Participants were instructed to abstain from alcohol in the 24 hours prior to the experimental testing sessions, and from caffeine and nicotine in the four hours prior to the experimental testing sessions, due to the known effects of these substances on pain perception and HRV (Bigal and Lipton, 2009; Sjaastad and Bakketeig, 2004; Sjoberg and Saint, 2011; Sondermeijer et al., 2002; Spaak et al., 2010).

Randomisation and blinding

A compounding chemist (Dartnell's Pharmacy, Surrey Hills, Australia) generated the randomisation schedule, which was done by means of a computerised method using Microsoft Excel, to ensure the study followed a randomised double-blind placebo-controlled cross-over trial experimental design. Such randomisation prevented any systematic bias from the researcher and the participants due to factors other than the intervention. Due to the short shelf-life of the nasal sprays, nasal sprays were prepared in batches of 10 (i.e., five oxytocin sprays, five placebo sprays). Blocked randomisation of treatment allocation followed the same structure. The randomisation schedules were concealed from the researcher using sequentially labelled and opaque sealed envelopes. To prevent expectation effects and biases, the researcher recruiting participants and conducting the experimental testing sessions, as well as the participants, were unaware of the treatment allocated (i.e., oxytocin or placebo).

The bottles containing oxytocin and placebo, and their contents, were identical in appearance, smell, texture, and taste. The active spray contained 24 international units (IU) of oxytocin (40.32 µg), and a full list of ingredients are listed in Appendix 1. The intranasal placebo spray contained the same ingredients as the oxytocin spray, with the exception of the active ingredient. No serious adverse effects have been reported in previous studies using dosages in excess of 24 IU in healthy humans and clinical samples (MacDonald et al., 2011). Each spray was labelled with the study details (i.e., study name, principal investigator (the primary supervisor of the student researcher), and contact details), storage and administration instructions, and a randomisation code, which corresponded to the randomisation schedule. The randomisation code referred to the relevant administration number for each participant that had been supplied by the compounding chemist (e.g., the first participant received spray number 66 in their first session, and received spray number 45 in their second session).

Nasal spray administration

The nasal sprays were kept in a locked safety deposit box in a PC2 research laboratory, in a temperature controlled fridge at 4 °C. The nasal sprays were kept for a maximum of two months (i.e., the shelf life of the sprays as no preservatives were added), under advice from the compounding

chemist. Prior to administration, the nasal spray bottle was shaken lightly, and four priming puffs were squirted from the spray bottle to ensure normal and even distribution of each spray before use. Participants could blow their nose before they commenced self-administering the nasal sprays, but were instructed to avoid this during the spray administration and for the first 20 minutes after the administration of the last puff. Participants were only allowed to dab off any leaking fluid from their noses with a tissue.

Participants self-administered the intranasal sprays according to verbal instructions from the researcher. Participants were instructed to keep the bottle in an upright but slightly angled position during the administration of the spray. Participants then self-administered three puffs per nostril, with each puff of the active spray administering 4 IU (6.72 µg) of oxytocin. Administration alternated between the nostrils, at intervals of 45 seconds, until three puffs were delivered to each nostril. Following the last puff, participants were instructed to wait for approximately 45 minutes before commencing the first experimental task in the session. This waiting period ensured that the oxytocin concentration had achieved its peak pharmacodynamics and pharmacokinetic effects within the central nervous system (Striepens et al., 2013) and could remain at a consistent level throughout the experimental session. To ensure a standardised experience while waiting for drug absorption, demographic data was collected from participants and the remaining time was spent watching documentaries that had been screened for neutral content for approximately forty minutes. Electrophysiological recording equipment, used in other experimental tasks described elsewhere, was then attached.

Electrophysiological Recording and Processing

This study used an 8/35 PowerLab unit (AD Instruments, Sydney, Australia) for continuous measurement of heart rate (HR) and skin conductance response (SCR). A five-lead electrocardiogram (ECG) system was used with disposable, pre-gelled electrodes (diameter 35mm, Coviden) placed in the following pattern: one beneath the lower right clavicle, one beneath the lower left clavicle, one on the lower right ribs, one on the lower left ribs, and one approximately halfway down the sternum, slightly to the right of the midline. The tasks examining HRV did not commence until a clear and accurate ECG recording was obtained. ECG R-R series was obtained by the identification of a local maximum after crossing the threshold of the derivative series via dedicated software (HRV Module, LabChart, AD Instruments). All relevant segments were visually inspected and corrected for false or undetected R-waves and ectopic beats. Undetected R-waves were manually inserted, and ectopic beats were excluded from analysis. The raw ECG signal was filtered with a 0.3 to 20 Hz band-pass filter, sampled at a rate of 2000 Hz, before being smoothed with a Savitzky-Golay filter (window width 155 samples) and undergoing a Fast Fourier Transformation using Welch's Periodogram (window width 1024s, 50% overlap). For SCR finger electrodes (MLT116F) were placed on the ventral surface of the proximal phalanx of the second and fourth fingers of the participants' non-dominant hand. No isotonic electrode paste was used to record SCR. Prior to the commencement of the paradigm both circuit- and subject-zeroing were performed for SCR to account for inter-individual variability between participants. HR and SCR data was extracted offline using LabChart Pro version 7.3.7 software (AD Instruments, Sydney, Australia).

Self-reported Pain Ratings

Self-reported ratings of current neck and shoulder pain intensity using a 10 cm visual analogue scale (VAS) with anchors of 0 (no pain at all) and 100 (worst pain imaginable) were collected from all participants at three intervals throughout the experimental testing session; immediately prior to the administration of the nasal spray (Time 0), 45 minutes after nasal spray administration (Time 1), and at the completion of the experimental testing session prior to participant debrief (Time 2).

Heart Rate Variability Tasks

Participants were seated upright in a comfortable chair and were instructed on how to affix the electrodes for the ECG recording. Participants then completed a paced breathing (PB) task, where they were prompted to inhale and exhale at regular intervals over a two minute trial (15 cycles per minute). The prompts instructing participants to inhale and exhale (i.e., the words “inhale” and “exhale” alternating on the screen) were presented with SuperLab software (version 4.5) on a laptop computer with a 14” screen. Paced breathing provides a reliable and controlled data for the analysis of HRV at rest (Pinna et al., 2007), while controlling for the influence of respiration on baroreflexes. To induce mild stress, participants then completed the serial sevens (SS) task (Hayman, 1942; Kazmark, 2000). Participants were instructed to count backwards in intervals of seven (e.g., 1,000, 993, 986, etc.) while the experimenter pressured them to “hurry up” and “go faster” for a period of two minutes. If the participant made an error in their counting, they were required to start again. To minimise potential carryover or learning effects, participants starting counted backwards from different starting numbers in the first and second testing sessions. Respiration was not controlled for during the spontaneous breathing of the SS task, as this goes against recent recommendations (Laborde et al., 2017; Thayer et al., 2011).

Both time- and frequency-domain measures of HRV were extracted from the ECG recordings. Time-domain measures focus either on the heart rate at any given time, or on the intervals between successive normal intervals (Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, 1996). With respect to the time-domain measures, the R-R interval (i.e., the interval between successive R peaks) and the root mean square of successive differences (RMSSD) were extracted. The RMSSD is considered to be a stable (Li et al., 2009) and valid (Thayer and Sternberg, 2010) measure of HRV. Frequency-domain measures describe how the variance of heart rate is distributed as a function of frequency. Frequency-domain measures are calculated by applying bands of frequency to the HR recording, and then counting the number of intervals within each frequency band. This distribution is commonly quantified as low-frequency HRV (LF-HRV; 0.04 to 0.15 Hz) and high-frequency HRV (HF-HRV; 0.15 to 0.4 Hz), in accordance with recommendations of the Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (1996).

Physical Functioning Tasks

Hand grip strength

and grip strength was assessed in a standardised way using an isometric hydraulic hand dynamometer “Jamar” (Sammons Preston, Inc, Bolingbrook, IL) in the sitting position with a straight back, shoulder adducted and in a relaxed position, elbow flexed to 90°, and the lower arm and wrist in a neutral position. The participants were instructed to grip the dynamometer in their dominant hand and to squeeze as hard as they could. After maximal squeeze duration of 8 seconds, the peak value was recorded in kilograms. Participants completed three trials, with an inter-trial interval of 60 seconds. The mean of the final two trials was used for statistical and analytic purposes. Reliability of

this procedure has been previously reported as good (ICC > 0.8).

Lifting ability

Lifting ability was assessed via a modification of the progressive isoinertial lifting evaluation (PILE) protocol, utilised to explore the maximal lifting capacity of the participant. The PILE is a test that forms part of the WorkWell FCE, formerly known as the Isernhagen Work System (Isernhagen, 1992). The lifting tasks were executed with a plastic crate and 2.5kg weight plates. The maximum performance was recorded in kg.

The PILE involved lifting the weights in 1) a lumbar test from floor to waist (a height of 0 cm to approximately 75 cm), and 2) a cervical test evaluating shoulder girdle and upper extremity lift capacity from waist to shoulder height (a height of approximately 75 cm to 140 cm). The lumbar and cervical PILE tests were performed separately as the norms and mechanics differ for the two lift heights (Mayer et al., 1988). Participants began with an empty crate and completed two repetitions of the lift to ensure they could follow the correct technique of the lift. One repetition of a lift was defined as the successful transfer from the floor to waist height and then back to the floor, or from waist height to shoulder height and then back to waist height. If a participant successfully completed two repetitions of the lift, an additional weight plate was added to the crate and the lift were repeated until one of the endpoints was triggered.

The PILE was terminated when one of the following endpoints was reached: 1) Biomechanical end point: when the participant altered their lifting technique (e.g., swaying or using momentum to help life the crate), or visibly or audibly exerting larger amounts of effort (e.g., going red in the face, grunting); 2) Pain end point: the participant reported that they were in pain as a consequence of lifting the crate; 3) Fear end point: the participant elected not to continue the task as they feared they would hurt themselves if they continued; and 4) Weight end point: the participant could safely and correctly complete two repetitions of the lift at the highest possible weight load (i.e., 25 kg).

Effort ratings

After completing each paradigm participants were asked to rate the amount of effort they exerted during the completion of the task on an 11-point numerical rating scale (NRS), with anchors of 0 (no effort at all) and 10 (maximum possible effort).

Thermal Heat Pain Threshold

The Medoc Pathway Pain and Sensory Evaluation System (Medoc Advanced Medical Systems Ltd, Ramat Yishay, Israel) with Medoc Main Station software version 6.3.6.18.1, was used to deliver thermal stimulations to participants. The Pathway system allows for exact, controllable delivery of heat stimuli using the Contact Heat Evoked Potentials (CHEPS) thermode. The CHEPS thermode has a round contact area of 573 mm2 (27 mm in diameter) and can produce temperatures between 30 °C and 55 °C, with the ability to increase in temperature at a rate of 70 °C/second.

Thermal heat pain threshold testing was conducted to determine each participant’s individual heat

pain threshold (i.e., the point at which they first perceive the thermal stimulation as painful) on the anterior surface of the dominant forearm, 5 cm proximal from the wrist crease. Three trials were performed, with an interval of 60 seconds was allowed between trials in order to minimise sensitization. The first trial allowed for familiarisation and was discarded, with thermal heat pain threshold calculated as the average of the last two trials. All trials began with the thermode at 32 °C, which increased at a rate of 1 °C/second with a safety limit of 50 °C. If the thermode reached 50 °C the trial was discontinued and heat pain threshold recorded as 50 °C. The participant was instructed to press a button on the Pathway response unit when the thermode reached their subjective pain threshold (i.e., “the point at which the stimulation becomes painful, rather than warm”). After the participant signalled the trial to end, the thermode returned to the baseline temperature at a rate of 8 °C/second.

Magnitude Estimations Task

The magnitude estimation task comprised three blocks, with nine trials per block. Stimuli in each block were delivered to one of three target sites only, with block order pseudorandomised across participants. The exact location of the thermode placement for each stimulus target site was as follows. For the cervical spine site, the thermode was placed over the C4/C5 cervical vertebra. For the deltoid site, the thermode was placed over the point at which the deltoid muscle inserts into the deltoid tuberosity of the humerus. Finally, for the tibialis site, the thermode was placed over the tibialis anterior, halfway between the most superior attachment to the tibia and its tendon in the upper one third of the muscle belly. These stimulus target sites have been used in previous psychophysical studies in patients with chronic neck pain (Sterling, 2010; Sterling et al., 2005; Walton et al., 2011).

As per an earlier study from our research group (Tracy et al., 2016), each trial started with a fixation cross (2 s), followed by the thermal stimulus that was delivered over a four second period (modified from Loggia et al., 2011). During stimulus delivery, the thermode increased from baseline (32 °C) to the target temperature (1 s), remained at the target temperature for two seconds, and then returned to baseline (1 s). The thermode remained at the baseline temperature until the delivery of the stimulus in the next trial (approximately 10 s later). The stimuli were programmed as high pain (45 °C), low pain (42 °C), and innocuous (39 °C). These temperatures were selected based on previous work by our group (Tracy et al., 2016), which confirmed that these temperatures were reliably experienced as high, low, and innocuous in intensity in healthy adults. Within each block, each target temperature was delivered three times in pseudorandomised order. During the trials, the CHEPS thermode was securely held in place by the participant, ensuring the surface of the thermode remained in even contact with the skin of the target area. The experimenter closely supervised the location of the thermode during all trials to ensure it did not become displaced. Having the participant hold the thermode reduced the potential confounding influence of interpersonal touch and social support on stress (Ditzen et al., 2007) and pain reporting (Edwards et al., 2017).

After each trial, participants were asked to rate the perceived intensity and unpleasantness of the stimulus on 11-point computerised numerical rating scales, with anchors of 0 (no pain/not unpleasant), 5 (mildly painful/mildly unpleasant), and 10 (extreme pain/extremely unpleasant). For this study the pain intensity rating was the primary outcome measure, and pain unpleasantness was the secondary outcome measure. The difference between stimulus intensity and unpleasantness was described to participants before the commencement of the task using the analogy of Price et al.

(1983). In brief, the analogy details the concepts of pain intensity and pain unpleasantness like the volume of a radio. As the volume increases, one can perceive how loud (or intense) the radio sounds, as well as how unpleasant it is to hear it at that particular volume. Such a distinction was clarified because we know that pain unpleasantness is only partially related to the intensity of a given stimulus, and that there are a range of other factors that influence the affective experience of pain (e.g., Berna et al., 2010). Participants were given up to 8 s to complete these ratings, and the next trial always started 8 s after the conclusion of the preceding trial.

Pain Anxiety Task

All participants received the following verbal instructions: “In this experiment you will be presented with a series of cues, with thermal stimuli following each cue. Following each stimulus you will be asked a series of questions relating to how you experienced and perceived the cue and its paired stimulus”. Participants first completed five training trials to familiarise them with the process and timing of the trials and the rating procedures. The visual cues used in the training trials were different to those used in the real trials, and thermal stimuli were not delivered, but participants were told when the stimulus would have been delivered. Training trials continued until the participant was comfortable and confident with the experimental procedure.

The pain anxiety task comprised three blocks, with 24 trials per block. Each trial began with a fixation cross (4 s), followed by an anticipatory visual cue (4 s), before the thermal stimulus was delivered over a four second period. The visual cues indicated that the following stimulus would deliver high pain, (45 °C), low pain (41 °C), an innocuous stimulus (32 °C), or that the stimulus intensity was ambiguous. Temperatures were selected based on screening the literature (Thibodeau et al., 2013) and previous work from our lab (Tracy et al., 2016), which confirmed that these temperatures were reliably experienced as different intensities in healthy, pain-free adults.

During stimulus delivery, the thermode increased from baseline (32 °C) to the target temperature (1 s), remained at the target temperature (2 s), and then returned to baseline (1 s). The thermode remained at baseline until the next trial (approximately 20 s). The relationship between the visual cue presented and the intensity of the paired thermal stimulus is described in Table 1. In one session, the relationship between the visual cue and stimulus temperature differed by the colour of the cue, whereas in the second session the relationship differed by the shape of the cue. The order of cue type (i.e., shape or colour) for each participant was determined by a computerised random number generator. The CHEPS thermode was securely attached to the volar surface of the participant’s dominant forearm, 5 cm proximal to the wrist crease with a Velcro strap.

Table 1: Relationship between visual cues and stimulus temperature

Set A Cue	Set B Cue	Stimulus Temperature
Pink	Triangle	32 °C
Orange	Circle	41 °C
Purple	Square	45 °C
Blue	Star	32 °C or 41 °C or 45 32 °C

Note: Each block of the pain anxiety task consisted 24 stimuli. Over the course of each block, each target temperature was delivered eight times in pseudorandomised order, with 50% of these preceded by their partnered unambiguous cue, while the remaining 50% were preceded by the ambiguous cue.

After each trial participants rated their anticipatory anxiety, pain intensity, and pain unpleasantness using computerised 11-point numerical rating scales (NRS), with anchors of 0 (not at all anxious/no pain/not unpleasant), 5 (mildly anxious/mildly painful/mildly unpleasant), and 10 (extremely anxious/worst pain/extremely unpleasant). Participants indicated their responses by pressing the appropriate button on a numerical computer keypad with their finger. The difference between pain intensity and unpleasantness was illustrated by an analogy according to Price et al. (1983). In brief, the analogy relates the concepts of pain intensity and pain unpleasantness to listening to sound (e.g., a radio). As the volume of the radio increases, one can be asked how loud it sounds, or how unpleasant it is to hear it. Pain intensity is similar to the loudness, whereas the unpleasantness of pain depends on the intensity, as well as a variety of other factors that may exert an impact on the experience. Participants were given up to 12 s to complete these ratings, after which the next trial started. Electrophysiological data (SCR and HR) were recorded throughout the experiment, and were extracted for the three phases of each trial (i.e., fixation, anticipation, and stimulus delivery).

Baseline Questionnaires

All participants completed the following questionnaires prior to the administration of nasal sprays in the first testing session. In addition to the array of questionnaires, the following demographic information was also recorded: sex, age, ethnicity, education, employment status, cause of chronic pain (chronic pain participants only), duration of chronic pain (chronic pain participants only), compensation status (chronic pain participants only), and oral contraceptive usage (female participants only), and the phase of menstrual cycle (female participants only).

Brief Pain Inventory (BPI)

The BPI is an 11-item self-report questionnaire designed to evaluate the severity of pain, and the impact of this pain on daily functioning (Cleeland and Ryan, 1994). Participants were asked to rate their worst, least, average, and current pain intensity, as well as the degree to which their pain has interfered with their general activity, mood, walking ability, normal work, relationships with other people, sleep, and enjoyment of life over the past week on an 11-point Likert scale with anchors of 0 (no pain/does not interfere) and 10 (pain as bad as you can imagine/completely interferes). The Pain Severity and Pain Interference Scores were generated by averaging individual item responses. Higher scores indicate greater levels of the construct measured (Cleeland and Ryan, 1994). Previous research has shown that the BPI has good internal consistency, with Cronbach's alpha ranging from .80 to .87 for the pain severity items, and from .89 to .92 for the pain interference items (Cleeland, 2009). For the current sample, high internal consistency was found for both the pain severity items (Cronbach's $\alpha = .95$) and the pain interference items (Cronbach's $\alpha = .95$).

Symptom Intensity Rating Scale (SIRS)
The SIRS is a 12-item patient self-report assessment tool and outcome measure for people with cervical spine dysfunction. This scale provides a measure of neck-specific symptom severity, not just neck related disability (e.g., neck pain, shoulder pain, dizziness, numbness, etc.). Participants rated the intensity of each symptom on an 11-point Likert scale with anchors of 0 (no symptoms) and 10 (extreme symptoms). A total score was summed from all 12 items. Higher scores indicate greater symptom severity (Davidson and De Nardis, 2011). For the current sample, the SIRS displayed excellent internal consistency (Cronbach's $\alpha = .93$).

Beck Depression Inventory-II (BDI-II)
The BDI-II (Beck et al., 1996b) has been recommended for pain research as it has excellent psychometric properties, and has been recommended for use in studies examining the effectiveness of interventions for pain (Dworkin et al., 2008). The BDI-II consists of 21 groups of four statements designed to assess the severity of current symptoms of depression. Participants selected the statement that best described how they had been feeling over the past two weeks. The BDI-II has been shown to have high one-week test-retest reliability (Pearson's $r = .93$), indicating that it is not overly influenced by daily variations in mood (Beck et al., 1996b). The BDI-II also has excellent internal consistency (Cronbach's $\alpha = .91$; Beck et al., 1996a). For the current sample, the BDI-II displayed good internal consistency (Cronbach's $\alpha = .78$).

Beck Anxiety Inventory-I (BAI-I)
The BAI-I is a self-reported measure of anxiety (Beck and Steer, 1990) that consists of 21 groups of four statements designed to assess the severity of current symptoms of anxiety-related disorders. Although the BAI-I was designed to minimize the overlap with the BDI-II, a significant correlation of $r = 0.66$ between the BAI-I and BDI-II has been observed in psychiatric outpatients (Beck et al., 1996a), suggesting that while the constructs measured by the BAI and BDI-II are strongly related, they can equally discriminate between anxiety and depression (Stulz and Crits-Cristoph, 2010). Participants indicated the extent that they felt they had been affected by each symptom (e.g., numbness, dizzy, terrified, etc.) from "not at all" to "severely" over the past week, including the day of the testing session. The BAI-I has been found to have high internal consistency (Cronbach's $\alpha = .94$) and acceptable reliability over a period of 11 days (Pearson's $r = .67$; Fydrich et al., 1992). For the current sample, the BAI-I displayed good internal consistency (Cronbach's $\alpha = .76$).

Treatment Guess
At the end of the first experimental testing session, participants were asked to indicate which treatment (i.e., oxytocin or placebo) they had received at the beginning of that session, and the degree of certainty they had in that judgement from 0 % (not at all certain) to 100 % (completely certain).

Appendix 1 – Ingredients in Nasal Sprays

Oxytocin Spray
1. Oxytocin (0.045 mg)

2. Sorbitol (45.048 mL)
3. Glycerin (0.045 mlL)
4. Preserved water containing parabens (0.51 mL)

Placebo Spray
1. Sorbitol (45.048 mL)
2. Glycerin (0.045 mlL)
3. Preserved water containing parabens (0.51 mL)

References

Beck, A.T., Steer, R.A., 1990. Manual for the Beck anxiety inventory. Psychological Corporation, San Antonio, TX.
Beck, A.T., Steer, R.A., Ball, R., Ranieri, W., 1996a. Comparison of Beck Depression Inventories -IA and -II in psychiatric outpatients. Journal of personality assessment 67, 588-597.
Beck, A.T., Steer, R.A., Brown, G.K., 1996b. Beck Depression Inventory. Psychological Corporation, San Antonio, TX.
Berna, C., Leknes, S., Holmes, E.A., Edwards, R.R., Goodwin, G.M., Tracey, I., 2010. Induction of depressed mood disrupts emotion regulation neurocircuitry and enhances pain unpleasantness. Biological psychiatry 67, 1083-1090.
Bigal, M.E., Lipton, R.B., 2009. Overuse of acute migraine medications and migraine chronification. Current pain and headache reports 13, 301-307.
Brondino, N., Fusar-Poli, L., Politi, P., 2017. Something to talk about: Gossip increases oxytocin levels in a near real-life situation. Psychoneuroendocrinology 77, 218-224.
Cleeland, C.S., 2009. The Brief Pain Inventory User Guide.
Cleeland, C.S., Ryan, K.M., 1994. Pain assessment: Global use of the Brief Pain Inventory. Annals of the Academy of Medicine, Singapore 23, 129-138.
Davidson, M., De Nardis, R., 2011. A twelve-item symptom intensity rating scale for cervical spine dysfunction. Spine 36, E307-E312.
Dickens, C., McGowan, L., Dale, S., 2003. Impact of depression on experimental pain perception: a systematic review of the literature with meta-analysis. Psychosomatic medicine 65, 369-375.
Ditzen, B., Neumann, I.D., Bodenmann, G., von Dawans, B., Turner, R.A., Ehlert, U., Heinrichs, M., 2007. Effects of different kinds of couple interaction on cortisol and heart rate responses to stress in women. Psychoneuroendocrinology 32, 565-574.
Dworkin, R.H., Turk, D.C., Wyrwich, K.W., Beaton, D., Cleeland, C.S., Farrar, J.T., Haythornthwaite, J.A., Jensen, M.P., Kerns, R.D., Ader, D.N., Brandenburg, N., Burke, L.B., Cella, D., Chandler, J., Cowan, P., Dimitrova, R., Dionne, R., Hertz, S., Jadad, A.R., Katz, N.P., Kehlet, H., Kramer, L.D., Manning, D.C., McCormick, C., McDermott, M.P., McQuay, H.J., Patel, S., Porter, L., Quessy, S., Rappaport, B.A., Rauschkolb, C., Revicki, D.A., Rothman, M., Schmader, K.E., Stacey, B.R., Stauffer, J.W., von Stein, T., White, R.E., Witter, J., Zavisic, S., 2008. Interpreting the Clinical Importance of Treatment Outcomes in Chronic Pain Clinical Trials: IMMPACT Recommendations. The Journal of Pain 9, 105-121.
Edwards, R., Eccelston, C., Keogh, E., 2017. Observer influences on pain: An experimental series examining same-sex and opposite-sex friends, strangers, and romantic partners. Pain 158, 846-855.
Faul, F., Erdfelder, E., Lang, A.G., Buchner, A., 2007. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 39, 175-191.
Fydrich, T., Dowdall, D., Chambless, D.L., 1992. Reliability and validity of the Beck Anxiety Inventory.

Journal of Anxiety Disorders 6, 55-61.
Hayman, M., 1942. Two minute clinical test for measurement of intellectual impairment in psychiatric disorders. Archives of Neurology & Psychiatry 47, 454-464.
Isernhagen, S.J., 1992. Functional capacity evaluation: Rationale, procedure, utility of the kinesiophysical approach. Journal of occupational rehabilitation 2, 157-168.
Kazmark, P., 2000. Validity of the serial seven procedure. International Journal of Geriatric Psychiatry 15, 677-679.
Laborde, S., Mosley, E., Thayer, J.F., 2017. Heart Rate Variability and Cardiac Vagal Tone in Psychophysiological Research - Recommendations for Experiment Planning, Data Analysis, and Data Reporting. Front Psychol 8, 213.
Li, Z., Snieder, H., Su, S., Ding, X., Thayer, J.F., Treiber, F.A., Wang, X., 2009. A longitudinal study in youth of heart rate variability at rest and in response to stress. International journal of psychophysiology : official journal of the International Organization of Psychophysiology 73, 212-217.
Loggia, M.L., Juneau, M., Bushnell, M.C., 2011. Autonomic responses to heat pain: Heart rate, skin conductance, and their relation to verbal ratings and stimulus intensity. Pain 152, 592-598.
MacDonald, E., Dadds, M.R., Brennan, J.L., Williams, K., Levy, F., Cauchi, A.J., 2011. A review of safety, side-effects and subjective reactions to intranasal oxytocin in human research. Psychoneuroendocrinology 36, 1114-1126.
Mayer, T.G., Barnes, D., Kishino, N.D., Nichols, G., Gatchel, R.J., Mayer, H., Mooney, V., 1988. Progressive isoinertial lifting evaluation. I. A standardized protocol and normative database. Spine 13, 993-997.
Moher, D., Hopewell, S., Schulz, K.F., Montori, V., Gotzsche, P.C., Devereaux, P.J., Elbourne, D., Egger, M., Altman, D.G., 2012. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. International journal of surgery (London, England) 10, 28-55.
Oktay, C., Eken, C., Ozbek, K., Ankun, G., Eray, O., Avci, A.B., 2008. Pain perception of patients predisposed to anxiety and depressive disorders in emergency department. Pain management nursing : official journal of the American Society of Pain Management Nurses 9, 150-153, 153.e151-153.
Pinna, G., Maestri, R., Torunski, A., Danilowicz-Szymanowicz, L., Szwoch, M., La Rovere, M., Raczak, G., 2007. Heart rate variability measures: A fresh look at reliability. Clinical Science 113, 131-140.
Price, D.D., McGrath, P.A., Rafii, A., Buckingham, B., 1983. The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. Pain 17, 45-56.
Rash, J.A., Campbell, T.S., 2014. The effect of intranasal oxytocin administration on acute cold pressor pain: a placebo-controlled, double-blind, within-participants crossover investigation. Psychosomatic medicine 76, 422-429.
Scott, D., Jull, G., Sterling, M., 2005. Widespread sensory hypersensitivity is a feature of chronic whiplash-associated disorder but not chronic idiopathic neck pain. The Clinical journal of pain 21, 175-181.
Senn, S., 2002. Cross-over Trials in Clinial Research, 2nd ed. John Wiley & Sons Ltd, San Francisco, CA.
Sjaastad, O., Bakketeig, L.S., 2004. Caffeine-withdrawal headache. The Vaga study of headache epidemiology. Cephalalgia : an international journal of headache 24, 241-249.
Sjoberg, N., Saint, D.A., 2011. A single 4 mg dose of nicotine decreases heart rate variability in healthy nonsmokers: implications for smoking cessation programs. Nicotine Tob Res 13, 369-372.
Sondermeijer, H.P., van Marle, A.G., Kamen, P., Krum, H., 2002. Acute effects of caffeine on heart rate variability. Am J Cardiol 90, 906-907.
Spaak, J., Tomlinson, G., McGowan, C.L., Soleas, G.J., Morris, B.L., Picton, P., Notarius, C.F., Floras, J.S., 2010. Dose-related effects of red wine and alcohol on heart rate variability. Am J Physiol Heart Circ Physiol 298, H2226-2231.
Sterling, M., 2010. Differential development of sensory hypersensitivity and a measure of spinal cord

hyperexcitability following whiplash injury. Pain 150, 501-506.
Sterling, M., Jull, G., Vicenzino, B., Kenardy, J., Darnell, R., 2005. Physical and psychological factors predict outcome following whiplash injury. Pain 114, 141-148.
Striepens, N., Kendrick, K.M., Hanking, V., Landgraf, R., Wullner, U., Maier, W., Hurlemann, R., 2013. Elevated cerebrospinal fluid and blood concentrations of oxytocin following its intranasal administration in humans. Scientific Reports 6, 3440.
Stulz, N., Crits-Cristoph, P., 2010. Distinguishing anxiety and depression in self-report: Purification of the Beck Anxiety Inventory and Beck Depression Inventory-II. Journal of Clinical Psychology 66, 927-940.
Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, 1996. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. Circulation 93, 1043-1065.
Thayer, J.F., Loerbroks, A., Sternberg, E.M., 2011. Inflammation and cardiorespiratory control: the role of the vagus nerve. Respiratory physiology & neurobiology 178, 387-394.
Thayer, J.F., Sternberg, E.M., 2010. Neural aspects of immunomodulation: focus on the vagus nerve. Brain, behavior, and immunity 24, 1223-1228.
Thibodeau, M.A., Welch, P.G., Katz, J., Asmundson, G.J.G., 2013. Pain-related anxiety influences pain perception differently in men and women: A quantitative sensory test across thermal pain modalities. Pain 154, 419-426.
Tracy, L.M., Georgiou-Karistianis, N., Gibson, S.J., Giummarra, M.J., 2016. Location, location, location: Variation in sensitivity to pain across the body. European Journal of Pain 20, 1721-1729.
Walton, D.M., Macdermid, J.C., Nielson, W., Teasell, R.W., Reese, H., Levesque, L., 2011. Pressure pain threshold testing demonstrates predictive ability in people with acute whiplash. The Journal of orthopaedic and sports physical therapy 41, 658-665.

Protocol