Learned nocebo effects on cutaneous sensations of pain and itch: a systematic review and meta-analysis of experimental behavioral studies on healthy humans

Thomaïdou, A.M.; Blythe, J.S.; Peerdeman, K.J.; Laarhoven, A.I.M. van; Schothorst, M.M.E. van; Veldhuijzen, D.S.; Evers, A.W.M.

OPEN

Psychosomatic Medicine

Author's Accepted Manuscript

Article Title: Learned nocebo effects on cutaneous sensations of pain and itch: A systematic review and meta-analysis of experimental behavioral studies on healthy humans

Authors: Mia A. Thomaidou, Joseph S. Blythe, Kaya J. Peerdeman, Antoinette I. M. van Laarhoven, Myrthe M. E. Van Schothorst, Dieuwke S. Veldhuijzen, and Andrea W. M. Evers

DOI: 10.1097/PSY.0000000000001194

Psychosomatic Medicine Publish Ahead of Print
DOI: 10.1097/PSY.0000000000001194

Learned nocebo effects on cutaneous sensations of pain and itch: A systematic review and meta-analysis of experimental behavioral studies on healthy humans

Mia A. Thomaidou, PhD [a,b], Joseph S. Blythe, MSc [a,b], Kaya J. Peerdeman, PhD [a,b], Antoinette I. M. van Laarhoven, PhD [a,b], Myrthe M. E. Van Schothorst, MSc [a], Dieuwke S. Veldhuijzen, PhD [a,b], Andrea W. M. Evers, PhD [a,b,c,d]

[a] Health, Medical & Neuropsychology unit, Leiden University, Leiden, the Netherlands, [b] Leiden Institute for Brain and Cognition, Leiden, the Netherlands, [c] Medical Delta Healthy Society Leiden University, TU Delft, and Erasmus University, [d] Leiden University Medical Centre, Department of Psychiatry, Leiden, the Netherlands

Author emails: a.m.thomaidou@fsw.leidenuniv.nl, j.s.blythe@fsw.leidenuniv.nl, k.j.peerdeman@fsw.leidenuniv.nl, a.vanlaarhoven@fsw.leidenuniv.nl, d.s.veldhuijzen@fsw.leidenuniv.nl, a.evers@fsw.leidenuniv.nl.

Corresponding author: Mia A. Thomaidou. Leiden University, Health, Medical and Neuropsychology Unit, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands.

Conflicts of Interest and Source of Funding: The authors have no conflicts of interest to declare. This work was supported by a VICI grant awarded to Andrea Evers by the Netherlands Organization for Scientific Research (NWO; 45316004).

Article Editor: Harald Guendel

Abstract

Objective: In past decades, the field of nocebo research has focused on studying how sensory perception can be shaped by learning. Nocebo effects refer to aggravated sensory experiences or increased sensitivity to sensations such as pain and itch resulting from treatment-related negative experiences. Behavioral conditioning as well as verbal suggestions of a negative treatment outcome may aggravate pain and itch perception. Gaining a comprehensive view of the magnitude of nocebo effects and contributing factors will help steer nocebo research towards fruitful directions for understanding complex sensory phenomena. **Methods:** We conducted a systematic review and meta-analysis of a total of 37 distinct experimental nocebo studies on healthy participants (all published in English between 2008-2021), with four separate meta-analyses for nocebo effects on pain or itch. We conducted subgroup analyses and meta-regression on factors such as type and intensity of sensory stimuli, and length of conditioning paradigms. **Results:** This meta-analysis showed that on average, effect sizes of nocebo effects were moderate to large (Hedges *g* between 0.26-0.71 for the four primary outcomes). The combination of conditioning and verbal suggestions yielded stronger nocebo responses on pain in particular. Subgroup analyses, including factors such as the type of sensory stimulation, did not explain the moderate heterogeneity in nocebo magnitudes between different studies. Risk of bias was generally low and was not related to nocebo magnitudes either. **Conclusions:** We discuss these results in relation to the role of conditioning as well as aversive learning, and we recommend more consistency in designing and reporting nocebo experiments.

Keywords

Nocebo, Conditioning, Learning, Pain, Itch, Hyperalgesia

1. Introduction

Negative expectations regarding the effects of a treatment can result in the aggravation of cutaneous sensations such as pain and itch (1–3). Such learned responses can be induced experimentally, allowing for the study of processes by which nocebo effects lead to symptom aggravation (4–10). In experimental studies, nocebo responses are defined as a significant increase in a sensation after a nocebo treatment, relative to no-treatment or a control treatment. To date, studies show that nocebo responses are able to aggravate sensations such as pain or itch –but may not necessarily elicit sensations in the absence of a baseline stimulus (11,12). Research on conditioned allodynia (pain or itch that persists in the absence on a sensory stimulus) has produced mixed results (13–16), but research has indicated that conditioned effects could transfer from one cutaneous sensation to another (17) and itch sensations can arise from social observation (18). Negative expectations leading to aggravated sensations of pain or itch are typically induced through classical conditioning, verbal suggestions, or their combination (4,5,10,19–21). Classical conditioning induces nocebo effects by building implicit associations between an (inert) treatment and the aggravation of sensations such as pain or itch (22–24). Verbal suggestions explicitly provide negative information regarding the pain- or itch-aggravating effects of a treatment (7). Because nocebo studies employ diverse methods, to better understand their potential impact on nocebo outcomes these methodological features warrant a systematic investigation.

Learning has consistently been shown to underlie induced nocebo effects (5–7,9,25), and verbal suggestions seem to induce stronger nocebo responses when combined with conditioning (26). The positive counterpart to nocebo, placebo effects, also appear to be stronger when induced through a combination of conditioning with verbal suggestions, compared to conditioning alone, both on pain (27) and itch (4,12). One meta-analysis included results from ten nocebo experiments published up to 2013 and reports that the overall magnitude of the nocebo effect was moderate to large and effects were generally larger when verbal suggestions were used in combination with conditioning (26). That early meta-analysis had a limited sample of studies available, and an up-to-date review is needed to examine how different types of learning may induce nocebo effects of different magnitudes. Other recent relevant reviews of the nocebo literature found that nocebo effects can be induced across many different sensations, including pain and itch, as a result of instructional learning, such as verbal suggestions, and associative learning, through conditioning mechanisms (28,29). Importantly, such mechanisms of induction of negative associations may be especially potent in settings with poor patient-clinician communication (30).

At the same time, other variables, such as the type of sensation (i.e., pain or itch), stimulus modality (e.g., thermal, electrical), the intensity of pain or itch stimulations, and the length of conditioning (learning) phases in different behavioral paradigms, also require a systematic examination across studies. For example, in experimental nocebo research, some nocebo conditioning paradigms include as few as four associative learning trials (5), while others employ much longer paradigms (6,8,31). A diverse set of cutaneous (pain/itch) sensory induction methods are also used, such as thermal (25), electrical (6,12), or laser pain stimulations (32). Such methodological choices, often meant to target specific underlying processes in nocebo experiments, can potentially influence nocebo responding and thus merit further investigation.

Given the recent growth of nocebo research, we conducted a systematic review and meta-analysis of experimental nocebo studies in healthy participants to provide novel insights into distinct contributions of methodological factors in the induction of nocebo responses. We focused on the cutaneous sensations of pain and itch, aiming to examine nocebo responses induced with comparable sensory inductions externally on the skin. We also focused on methodological and design choices for experimental models, as well as on the types of learning mechanisms involved. This meta-analysis did not delve into potential effects of demographic characteristics on nocebo responses, as demographic variables have not systematically been studied in the nocebo literature and are often reported as secondary, if at all, and we did not have a meaningful rationale for why, for example, small variations in age would impact nocebo responding. First, we examined nocebo magnitudes between pain and itch and based on the learning method used. Then, we conducted subset analyses and meta-regression to assess how the type and intensity of stimulations, the length of conditioning, the timing of measurement of nocebo magnitudes, and risk of bias in studies may impact nocebo magnitudes.

2. Methods

2.1. Protocol and registration

The protocol for this study was pre-registered on ClinicalTrials.gov (ID: NCTxxxxx851) and conducted based on the PRISMA statement 2009 (see checklist, Supplemental Digital Content 1, http://links.lww.com/PSYMED/A917) and Cochrane recommendations (2020). The protocol was registered based on a single search strategy for both nocebo and placebo studies, which, due to the volume of the studies returned, is now divided in two separate papers. Here, we report only the nocebo (arms of) experimental studies.

2.2. Databases and selection criteria

PubMed, PsycINFO, EMBASE, and the Cochrane CENTRAL Methodology Library were searched to identify studies. Languages were a-priori restricted to English, Dutch, and German and the publication period was not restricted. Searches were initially conducted on March 18th, 2019. Repeated searches for studies published after this time were conducted in June 2020 and July 2021. See Supplemental Digital Content 2, http://links.lww.com/PSYMED/A918, for a detailed key-word strategy.

We searched for original, peer-reviewed, controlled experimental studies (or study arms) on healthy human participants that aimed to experimentally induce placebo and/or nocebo effects. Patient samples were not included, to improve the homogeneity of the results, and for the same reason we focused on cutaneous sensations (i.e., pain and/or itch stimulations that were administered on the skin), excluding for example visceral pain studies. We considered as nocebo studies in our inclusion process only those studies that employed a verbal suggestion that specifically informed participants that their pain/itch would be worsened as a result of a (sham) treatment. We therefore did not include studies that, for example, simply instructed participants that they will experience increased pain when viewing a particular cue on a screen, or studies that did not instruct participants on why in some trials they experience increased pain, as these studies were considered to be pain conditioning (but not *nocebo* conditioning studies explicitly inducing expectations regarding a treatment). For the purposes of in- and exclusion, studies were considered to have induced a placebo or nocebo effect if a learning paradigm was used to induce positive or negative outcome expectations about an inert treatment. We considered as *nocebo* learning paradigms only those that aimed to induce negative expectations regarding an intervention, such as sham electrical stimulation or an inert cream. This meant that most conditioning without verbal suggestion studies were excluded from this review, as they did not include treatment associations, and were considered to be pain-conditioning, not nocebo-conditioning studies (albeit explicit mention of the terms *nocebo* and *placebo* was not a specific inclusion criterion). Additionally, we only included studies that had a control group or a control condition within-subjects, so that nocebo effects could be calculated as the difference between nocebo and control/no treatment on self-reported scores. We excluded studies that excluded or did not report data from nocebo non-responders. Post-hoc, we excluded observational learning studies as they were too few for a meaningful analysis. Studies that did not fulfill one or more of the criteria mentioned above were excluded from the meta-analysis, see **Figure 1**.

2.3. Study selection

Eligibility assessment for the inclusion of studies was performed independently by two authors in each of the following steps. Titles and abstracts of articles retrieved using the search strategy were screened by two authors independently (M.M.E.v.S. and J.S.B.). The full text of articles to be included and articles about which doubts existed were then retrieved and assessed for eligibility by two authors independently (M.A.T. and J.S.B.). The reference lists of all included articles were also screened for study inclusion by two authors (M.A.T. and J.S.B.) and included articles were also entered in Web of Science to identify articles that have cited them and should potentially be included in the meta-analysis. When necessary, authors of studies were contacted in order to provide full-text articles that were not accessible online. Any disagreements regarding study inclusion were resolved by consultation with a third author (K.J.P.).

2.4. Data extraction

One author (J.S.B.) used a standardized form to independently extract data from the included studies to derive data for analyses. Another author (M.A.T.) checked 25% of extracted values for accuracy. Extracted information included details of the intervention such as the learning method used, the control condition, study population, sensation type, pain/itch rating data, type of cutaneous stimulation (e.g., heat pain, pressure pain), type of outcome expected (i.e., placebo or nocebo), information for quality assessment, and outcome data for meta-analysis (e.g., sample size, pain/itch rating means and standard deviations). Doubts regarding data-extraction were resolved through discussion with a third review author (K.J.P.). Missing data were requested directly from the study authors. When there was no response from authors, but data could be extracted from published figures, this was done using WebPlotDigitizer version 4.4 (Rohatgi, 2020).

2.5. Risk of bias

Risk of bias (RoB) was assessed and checked by two authors (M.A.T. and J.S.B.) using the method developed by Marcuzzi and colleagues specifically for quantitative sensory testing studies (33). This method assesses whether the sample was clearly described and was representative of the population, whether the somatosensory assessment methods are standardized, validated, and well described, if potential confounders were considered, and adequate blinding. Each category was scored as being satisfied (0 points), not satisfied (2 points), partially satisfied (unclear; 1 point), or not applicable. Scores were selected based on criteria described in Marcuzzi and colleagues (33). We additionally concocted numerical scores (0-34) for each study, by summing each item score, with higher scores indicating higher risk of bias (see Supplemental Digital Content 3, http://links.lww.com/PSYMED/A919 for an example of the RoB scoring).

2.6. Statistical analyses and results synthesis

All analyses were conducted and checked by two reviewers (J.S.B. and M.A.T), using the Comprehensive Meta-Analysis software (version 3.3.070; Biostat, Englewood, USA) and R programming software for visualizations (34). Funnel plots were inspected for outliers (i.e., studies falling outside the funnel of expected results), and to assess publication bias across studies we checked for number of imputed missing studies with Duval and Tweedie's trim and fill method (35). Heterogeneity between studies was assessed with the I^2 statistic and visual inspection of the forest plot. I^2 is a measure of the proportion of observed variance reflecting real differences in effect sizes (36) with values of 25%, 50%, and 75% considered as low, moderate, and high degrees of heterogeneity, respectively (37). For forests plots, we calculated study weights in *R*, by inversing the variance of each effect size.

Given the heterogeneity of study designs, random effects models were used for all meta-analyses. Effect sizes were calculated using means and standard deviations for each group (between subjects) or trial type (within subjects). (36). Magnitude of nocebo responses was the main outcome variable, with *nocebo magnitude* commonly defined in the nocebo literature as the size of nocebo responses on a standardized pain/itch scale as a function of difference scores (26). Nocebo *magnitudes* thus represent the size of the difference in participants' pain/itch ratings during nocebo evocation trials as compared to control trials. When standardized pain/itch scales were not ranging from 0 to 10, we transformed the difference scores for evocation vs control trials by dividing the difference rating by the highest possible value on the scale used and multiplying by 10 to convert to a rating on a 0-10 scale. We selected nocebo and control conditions based on what was reported in studies: some reported nocebo magnitudes between groups, other within groups in the first pair of evocation trials, and others reported nocebo magnitudes as the mean difference of all control and evocation trials. When only nocebo/control difference scores were reported, these were used instead. When only standard errors were reported, they were converted to standard deviations by multiplying the standard error by the square root of the group size (n). For each study, an effect size Hedges's *g*, weighted to the sample size (N), was computed as the mean pain or itch response in the nocebo condition minus the mean response for the control condition of the evocation phase of experiments. Positive *g* values indicate a nocebo response, with values around .2 considered small, .5 medium, and .8 large.

For studies that used within-subjects comparisons, the nocebo-control condition correlation coefficient could not be derived, therefore an average r of .5 was imputed (38). Meta-analysis was only conducted when the data of at least 4 studies were available in total. Specifically for subset analyses, we ran statistical tests when 2 or more studies were available per subgroup. Studies with multiple eligible conditions were treated as separate subgroups and averaged across in CMA (e.g., when cheap vs. expensive inert treatments were used as nocebo, we averaged the results and treated this as one group (see **Table 1** for results synthesis per study).

2.7. Primary outcome measures and subset analyses

Our primary outcomes were the overall magnitude of nocebo responses (i.e., the difference in self-reported pain/itch between a nocebo and a control trial in the evocation phase) separately for pain and itch studies employing verbal suggestions with or without classical conditioning. We thus computed 4 pooled effect sizes: verbal suggestions in pain, conditioning with verbal suggestions in pain, verbal suggestions in itch, conditioning with verbal suggestions in itch. Whenever possible, the mean of pain or itch ratings across the entire evocation phase was used. If only values from the first trial(s) were reported, these were used instead, and sensitivity analyses tested for differences in magnitudes between studies reporting the mean versus the first trials.

We also did subset analyses to compare Hedge's *g* between nocebo responses based on the type of learning (verbal suggestion or combination with conditioning) and type of sensory stimulation (e.g., thermal, electric) and the timing of nocebo measurement (as the mean of evocation or only the first evocation trials, by trial type). Meta-regression assessed the impact of the length of conditioning, (quantified as the number of learning trials during induction, while we also separately examined number of trials evocation), the timing of the measurement of nocebo hyperalgesia in the evocation phase (first trials versus mean of evocation trials), the stimulus intensity (calculated as the calibrated difference in pain intensity for control vs. nocebo trials) and the Risk of Bias score on nocebo magnitudes for the included studies.

3. Results

3.1. Study selection

Figure 1 shows the flow of the study selection process including the reasons for exclusion at each stage. A total of 17546 nocebo and placebo papers were initially identified through the database searches. We searched for more eligible studies through reviewing the reference lists as well as web of science for each included study, as well as conducting repeat database searches in June 2020 and July 2021. At each stage of study inclusion, duplicates were removed, and remaining articles were considered based on title and abstract, or full text. In total, we identified 24814 articles through our searches, of which 24687 were excluded.

We did not follow a strict hierarchical approach in marking exclusion criteria, but selected criteria based on what was deemed to be the major exclusion reason, for example when screening abstracts where limited information is available, therefore the following exclusion numbers provide less than precise estimates of exclusion reasons. We excluded articles for the following reasons: 8302 articles for not aiming to study nocebo or placebo effects or not using a learning paradigm to induce placebo or nocebo effects (explicit use of the terms *nocebo* or *placebo* was not an inclusion criterion), 4328 for not reporting original data or (full length, peer reviewed) experimental studies, 1229 studies for not being conducted in humans, 10440 because they were duplicates or already screened during a previous round, 101 articles for not studying (placebo/nocebo on) cutaneous sensations, 242 articles for not studying (placebo/nocebo in) healthy human participants, 20 articles because they did not report self-reported pain/itch intensity ratings, 13 for not being in English, Dutch, or German, 2 studies for not using a within- or between-subjects controlled design, 5 studies for not responding to requests for data, and 5 for excluding data from participants that were considered placebo/nocebo non-responders. A total of 127 articles were selected of which 108 included placebo conditions and 39 nocebo conditions. Of these articles, we excluded 2 observational learning studies as they were too few for a meaningful analysis. Thus, in total, **37** studies were included in this meta-analysis on nocebo effects. The references for the included studies are available in Supplemental Digital Content 4.

3.1. Study characteristics

Table 1 displays study characteristics. We included 37 distinct nocebo studies, published between 2008 and 2021. Including additional experimental conditions in a number of studies (see **Table 1**) in total we analyzed 40 study arms (30 pain and 10 itch). Thermal pain inductions were used in 19 arms, electrical pain was used in 6, pressure pain was used in 1, and mechanical, cold pressor, hot water bath, and histamine methods were each used in 1 study arm. Only 7 studies (10 arms) induced nocebo effects on itch, one of which also included pain (this study, van Laarhoven et al., 2011, is listed under *Pain* in **Table 1**). Electrical itch was used in 3 studies, one of which (van Laarhoven et al., 2011) used additional mechanical and histamine inductions in both the pain and itch groups (see **Table 1**). Histamine was used in 3 more itch studies and cowhage was used in 1 study.

For nocebo induction, most studies (18 pain and 4 itch studies) used a combination of classical conditioning and negative verbal suggestions, and for 3 we included additional study arms that employed verbal suggestions alone (**Table 1**). Verbal suggestions alone were used as the main manipulation in 10 pain studies (in total 12 arms) and 3 itch studies (in total 6 arms). Risk of bias was low within all studies, with most studies showing low risk of bias (max. 5/34) and only one study scoring in the low-moderate range with a score of 6/34 (**Table 1**). The funnel plots as well as a trim and fill method that suggested a small number of imputed studies (**Figure 2**) indicated that overall, there was a low degree of potential publication bias across all studies, with a total estimated 7 studies missing.

3.2. Magnitude of nocebo responses

See **Figures 2 and 3** for forest and funnel plots, respectively, that display effect sizes per study and pooled effects. For pain **(Figures 2A and 2B),** the magnitude of nocebo responses on a standardized scale of 0-10 (with higher scores indicating larger nocebo magnitudes) across studies using classical conditioning with verbal suggestions ranged from 0.28 to 1.42, with the mean standardized response being *M* = **0.79** (*SE* = 0.24). Verbal suggestions alone induced effects on pain ranging from 0.00 to 1.27 (*M* = **0.70**, *SE* = 0.30). For itch, the magnitude of nocebo responses in studies that used conditioning with verbal suggestions ranged from 0.21 to 0.47 (*M* = **0.35**, *SE* = 0.24). Verbal suggestions alone induced effects on itch ranging from 0.41 to 0.75 (*M* = **0.58**, *SE* = 0.26). Based on these results, on average our meta-analysis indicated medium effects of the nocebo manipulations (Hedges *g* between 0.26-0.71 for each of the four pooled effects), a moderate degree of heterogeneity (I^2 average 41% across the four pooled effects), with the study effect sizes ranging between *g* = 0.00 and *g* = 1.34.

3.3. Classical conditioning and verbal suggestions in pain and itch

A range of different verbal suggestions were used to induce nocebo responses on pain and itch. Most studies used either an inert cream or inactive electrodes as the nocebo stimulus that would supposedly increase pain/itch sensitivity. For example, studies suggested to participants that their pain will be increased upon the activation of electrodes on their skin because these electrodes "enhance the conductivity of the pain signal being sent to the brain" (39) or "the cream that will be applied to your arm increases the effect of the heat pain and you will feel more pain after the application." (25). Most such suggestions were delivered orally by a researcher, with few studies providing such information in writing.

For **pain**, a somewhat larger pooled nocebo effect of the combination of **conditioning with verbal suggestions** (k = 21, g = **0.71**, 95% CI 0.60 – 0.82, I^2 = 50.71%; **Figure 3A**) was observed than of **verbal suggestions** alone (k = 12, g = **0.63**, 95% CI 0.40 – 0.86, I^2 = 55.59%; **Figure 3B**). In **itch**, however, **conditioning with verbal suggestions** yielded a smaller pooled effect on the magnitude of nocebo responses (k = 4, g = **0.26**, 95% CI 0.09 – 0.43, I^2 = 0%; **Figure 3C**) compared to a medium pooled effect of **verbal suggestions** alone (k = 4, g = **0.53**, 95% CI 0.23 – 0.82, I^2 = 53.81%; **Figure 3D**) on nocebo responses. Overall, nocebo responses (see Table 1 for the relevant studies) were thus associated with medium pooled effects in pain, while in itch they were associated with slightly smaller pooled effects overall.

3.4. Magnitude of nocebo responses based on the type of stimulation

For **pain** studies that used **conditioning with verbal suggestions**, we compared effects of different pain administration methods (k = 13 thermal, k = 7 electrical) excluding the single study using laser. Thermal pain yielded a somewhat larger pooled effect on the magnitude of nocebo responses (k = 13, g = 0.75, 95% CI 0.59 – 0.91) compared to medium pooled effects of electrical pain (k = 7, g = 0.65, 95% CI 0.51 – 0.79) on nocebo responses. For **pain** studies that used only **verbal suggestions**, we examined effects of different pain administration methods (k =

4 thermal, *k* = 5 electrical, *k* = 2 mechanical) excluding the single studies using laser, cold pressor, hot water bath, pressure, and histamine. Electrical pain yielded slightly larger pooled effect on the magnitude of nocebo responses (*k* = 5, *g* = 0.91, 95% CI 0.65 – 1.17) compared to medium effects of thermal (*k* = 4, *g* = 0.69, 95% CI 0.21 – 1.16) and mechanical (*k* = 2, *g* = 0.60, 95% CI 0.14 – 1.06).

For **itch** studies that used **conditioning with verbal suggestions**, there were too few studies to analyze (cowhage *k* = 1, electrical itch *k* = 2, and histamine *k* = 1). For **itch** studies that used only **verbal suggestions**, there were again too few studies (*k* = 2 electrical, *k* = 3 histamine, *k* = 1 mechanical).

3.5. Magnitude of nocebo hyperalgesia based on the pain stimulus intensity

For **pain** studies that employed **classical conditioning with verbal suggestions** we had a sufficient sample to examine any relationship between differences in intensity of pain stimulations in the learning phase and the magnitude of nocebo responses, but a meta-regression found no significant association (Q = 0.89, p = 0.35).

3.6. Magnitude of nocebo hyperalgesia based on the number of trials

Studies that employed **classical conditioning** used varying numbers of learning and evocation trials. For **pain only**, there were sufficient studies to examine the effects of different lengths of conditioning and different lengths of evocation (i.e., the length of extinction) on nocebo magnitudes. The shortest pain learning paradigm used 6 nocebo and 6 control trials, while the longest paradigms used up to 30 nocebo and 30 control trials. Evocation phases ranged from 3 nocebo and 3 control trials to 30 nocebo and 30 control trials. A meta-regression of different lengths of conditioning showed no association with the magnitude of nocebo responses (Q = 0.81, p = 0.37). Similarly, there was no association between the length of evocation and nocebo magnitudes (Q = 0.19, p = 0.67).

3.7. Magnitude of nocebo hyperalgesia based the timing of measurement

All itch conditioning studies measured the nocebo effect as the mean of all evocation trials. Among **pain** studies that employed a combination of **conditioning with verbal suggestion**, however, 13 paradigms measured nocebo responses as the mean of all evocation (testing) trials, 6 measured the magnitude of responses in the first pair of evocation trials, and 2 studies specified different timing such as pre-post measures. Studies in which first evocation trials were used yielded a large pooled effect on the magnitude of nocebo responses (k = 6, g = 0.82, 95% CI 0.57 – 1.07) compared to medium pooled effects of measuring the effect as the mean of all evocation trials (k = 13, g = 0.66, 95% CI 0.54 – 0.79) and non-specified (k = 2, g = 0.67, 95% CI 0.23 – 1.11).

3.8. Magnitude of nocebo responses based on the Risk of Bias score

Lastly, we examined how RoB scores (see **Table 1**) may be related to nocebo magnitudes. A meta-regression showed no significant relationship between RoB scores and the magnitude of nocebo responses for **pain** studies that used **conditioning and verbal suggestions** (Q = 0.75, p = 0.39), for **pain** studies that used only **verbal suggestions** (Q = 0.00, p = 0.95), for **itch** studies that used **conditioning and verbal suggestions** (Q = 0.08, p = 0.77), or for **itch** studies that used **verbal suggestions** alone (Q = 1.9, p = 0.05).

4. Discussion

We conducted a systematic review and meta-analysis of a total of 37 distinct nocebo studies on healthy participants. This meta-analysis showed that on average, nocebo effects were moderate to large in magnitude. The combination of verbal suggestions with classical conditioning yielded stronger nocebo responses on pain, but this may not necessarily be the case in the small number of itch studies. Measures of the type or intensity of pain or itch, and length of conditioning, did not explain the moderate heterogeneity in nocebo magnitudes between different studies. Timing of nocebo measurement in the first evocation trials yielded slightly larger nocebo magnitudes. Risk of bias was generally low and was not related to nocebo magnitudes either. We discuss these results in relation to the role of conditioning as well as aversive learning, and we speculate of the reasons why none of the factors collected in the nocebo literature appear to consistently explain variations in the magnitudes of learned nocebo effects on pain and itch.

Generally, studies that aimed to experimentally induce nocebo responses yielded average responses (across the included sample) of moderate to large magnitudes, ranging from 0 (a magnitude of zero indicating no nocebo response) to high response magnitudes up to 4 points on a 0-10 scale. Few studies reported that their experimental manipulations did not induce a nocebo response on average across participants, but interindividual variations are prevalent in nocebo responding. It should be noted that little attention has been devoted in the literature included here regarding the *prevalence* of nocebo responses, i.e., the difference between the absence and presence of a nocebo response.

We found that nocebo magnitudes had a moderate heterogeneity) and were moderated only by the timing of measurement. This is unsurprising, as measuring nocebo magnitudes in the first trials of the attenuation phase yields larger nocebo effects before extinction has had a chance to take place. It is important to note that this result may show that nocebo effects start becoming extinct shortly after negative learning is discontinued, even if some studies indicate that nocebo effects are more persistent than placebo effects (40).

Often conceptualized as the counterpart of nocebo responses, placebo effects appear to be comparable in magnitude to the overall nocebo magnitude found in the current meta-analysis, but heterogeneity in placebo responses may be higher (27). In a more recent meta-analysis on experimental placebo studies, placebo responses were found to yield small to moderate effects, with moderate to large heterogeneity in results (41). We speculate that this may indicate that the negativity of suggestions and experiences in nocebo paradigms may result in stronger learned effects, as compared to the positive expectations induced in placebo excrements. Indeed, aversive learning has consistently be shown to be prioritized over the learning of neutral or positive information in the brain (42–45), something that is thought to have an evolutionary basis (46).

Magnitudes of nocebo responses were found to be moderate to large in pain studies when looking at both verbal suggestions and combination with conditioning. As expected, in pain experiments the addition of classical conditioning yielded somewhat larger nocebo responses, suggesting that learning by experience during behavioral conditioning may be more potent than mere negative suggestions regarding pain outcomes. For itch, however, verbal suggestions alone yielded moderate effects whereas combination with conditioning resulted in small effects across studies. The number of studies included in each of the two itch conditions (k = 4 in each) may be insufficient to allow for further conclusions to be drawn regarding this apparent distinction between learned pain and itch effects.

While the number of itch studies included in this meta-analysis was small (8) compared to pain (30), overall effects on pain appear to be larger than those on itch across both learning methods, based on the present findings. Itch has been shown to be prone to suggestions and can be influenced by expectations (4) with one study that compared placebo effects induced with verbal suggestions for either pain or itch indicating that itch might be more prone to suggestions (47). Our finding that pain resulted in larger nocebo magnitudes across the studies included here, could suggest that compared to itch, the learning of pain associations may be facilitated to a larger degree. In other words, we speculate that, as pain is perhaps more threatening and aversive than itch, it may signal a more vital threat to the person and thereby, from an evolutionary perspective (46), result in stronger learning. Further research into nocebo effects is needed, however, to reach a sufficient sample size for reliable comparison results between pain and itch.

The variability found in nocebo response magnitudes was not explained by differences between the type or intensity of pain or itch stimulation, or the length of conditioning. That the length of conditioning paradigms, during which learning of a negative association took place, is not related to the nocebo response, is somewhat surprising, but could be explained by learning reaching a ceiling effect in early conditioning trials or reaching a ceiling effect due to the strong role of verbal suggestions in such experimental studies. It is also possible that the length of conditioning becomes secondary due to other variations in paradigms, for example, shorter conditioning phases using longer exposure to conditioning cues and pain stimulations (9,48–50). It would be valuable for future research to explore whether a learning curve possibly takes place in nocebo conditioning, with the number of conditioning trials mattering up to a point, but after a certain threshold, or as a result of strong preceding verbal suggestions, the number of trials could matter less over time.

A moderate dispersion of effect sizes across the studies analyzed is important to note, especially when the measures that are systematically reported in studies, such as the duration of learning or the intensity of pain, are unable to explain such variability in nocebo response magnitudes. The large differences in applied experimental models of nocebo effects (e.g., different types of verbal suggestions, whether the experiment was conducted in a hospital or university setting, or types of nocebo and conditioned stimuli presented) may explain some of this variability in results (11). Similarly to the efforts for aligning experimental paradigms in animal models of disease (51,52), it is essential for the field of nocebo to focus on replicating experimental paradigms and aligning paradigms according to ecologically valid models that yield comparable results across studies. For example, in the field of fear conditioning, one study reviewed the literature and summarized the methods and analyses commonly used for experimental fear induction and extinction, identifying the state-of-the-art in this domain and proposing methodological considerations for the design and analysis of such studies, aiming to set a methodological standard for experimental fear models and address the replicability crisis and inconsistency in methodological designs (53). Such an endeavor could be very valuable in the field of nocebo research, as this meta-analysis shows that methodological variations are all too common and compromise the comparability of results.

One of the most consistent differences between experimental nocebo studies seems to be the type of verbal suggestion delivered to participants. No two studies administered the same verbal suggestion. Different verbal suggestions could contain distinct emotional loads and be perceived as more or less threatening, which may in turn influence nocebo responses (21,25). While beyond the scope and reach of the current meta-analysis, a future systematic review of distinct verbal suggestions, for example using content analysis approaches borrowed from linguistics (54,55), could shed a light on how different verbal suggestions could impact nocebo responses. A method such as Natural Language Processing could be implemented by future research to help understand the specific valence of language included in verbal suggestions that leads to stronger nocebo responses, and identify the linguistic content of negative suggestions that are most potent and result in stronger and/or more persistent negative pain associations.

There are other variables that could explain variability of induced nocebo responses, such as sampling, demographics, and the inclusion criteria for participation, but a limitation is that these factors are not consistently reported in papers and could not be investigated in the current meta-analysis. Additionally, studies do not systematically measure fear, which is shown repeatedly to be involved in nocebo responses (21,25,56–58). Other variables relevant to the emotional context of studies, such as the demeanor of the experimenter (59) or whether the experiment is set in an academic building or hospital, are also often not clearly documented, and could not be analyzed here. Finally, risk of bias was low across all studies and showed no relationship to nocebo magnitudes. However, the assessment tool used for this meta-analysis is designed for quantitative sensory testing studies (33) but could have missed bias aspects, such as potential publication bias for significant results, which meta-analyses studies should consider addressing.

The growing field of nocebo research has begun to shed light on biobehavioral mechanisms that support the involvement of learning and expectations in the processing of sensory inputs such as pain and itch. A number of studies show that brain areas that are responsible for integrating prior experiences and the expectations formed regarding a particular treatment into sensory processing are involved in the aggravation of sensations such as pain (56–58). Particularly, results that implicated regions such as the ACC and insula in learning nocebo associations suggest that a prominent difference between the perception of nocebo and control cues can be seen in brain areas that are thought to synthesize sensory perception based on beliefs and expectations (60,61). Past pain experiences, leading to negative expectations, have been shown to form differential expectations that influence pain processing (6,10,62,63). Similar mechanisms are thought to underly the perception of pain in light of learned negative expectations (29). It is notable, however, that while the present meta-analysis focused on studies that set out to induce explicit negative expectations regarding a particular (nocebo) treatment, there is also extensive evidence that aggravated pain responses can result from subliminal conditioning on an unconscious level, in a phenomenon similar to conscious expectation of negative pain outcomes, but that is likely mediated by distinct subconscious mechanisms (64–66). In the studies included in this meta-analysis, there is little focus on explicitly measuring self-reported expectations: five studies measured participants' expectations of pain increase at the start of the experiment (47,67–70), three studies measured expectations of overall pain increase at the end of the experiment (9,62,71), and three studies measured expectations within the experimental paradigm (6,39,72). In the field of nocebo effects on cutaneous sensations, it is now generally accepted that whether consciously or subconsciously, nocebo responses are acquired as a result of prior negative experiences, leading to negative expectations. In line with a long line of research into learning that results from conditioning, as well as based on our current understanding of predictive processing –which is also a well-established mechanism by which past experiences and resulting expectations shape the way incoming stimuli are processed and perceived (73,74)– it appears that nocebo responses may be induced by conscious expectations, but could also be induced on a subconscious level that could not be captured through the measurement of self-reported expectations.

To date, the literature remains mixed and uncertain regarding the precise variables that may make particular nocebo induction methods, contexts, situations, or learning modes more potent, or the types of people that may be more susceptible to presenting nocebo effects on pain and itch. Future research should consider investigating individual differences in nocebo responding in a targeted manner, using sufficiently large sample sizes, and endeavoring to systematically compare experimental nocebo effects between different demographic groups, as well as between healthy participants and patient populations.

This systematic review and meta-analysis quantified magnitudes of nocebo responses on cutaneous sensations (pain and itch) for distinct learning paradigms in experimental studies (classical conditioning with verbal suggestion, or verbal suggestion alone). We replicated previous findings that classical conditioning combined with negative verbal suggestions was strongest for inducing nocebo responses on pain. Meta-analyzing nocebo effects on itch is new and obtained small to moderate effects overall. Subset analyses indicated that factors related to the length of conditioning paradigms or intensity and type of sensory stimuli did not explain the moderate heterogeneity in nocebo effect sizes. This review provides a comprehensive summary of current findings in the field of nocebo research. We have ruled out some factors that were consistently reported in papers and could not explain the variability in results across studies, and we recommended some important directions for the field, such as increased consistency between study designs for inducing nocebo effects, as well as a systematic examination of the effects of different verbal suggestions on magnitudes of learned nocebo effects.

Acknowledgements

The authors would like to thank Barclay P. Wohlstetter and Elif S. Uyanik for their valuable assistance.

References

1. Barsky AJ. Nonspecific Medication Side Effects and the Nocebo Phenomenon. JAMA. 2002;287(5):622-627.
2. Benedetti F, Lanotte M, Lopiano L, Colloca L. When words are painful: Unraveling the mechanisms of the nocebo effect. Neuroscience. 2007 Jun;147(2):260–71.
3. Evers AWM, Peerdeman KJ, van Laarhoven AIM. What is new in the psychology of chronic itch? Exp Dermatol. 2019 Dec;28(12):1442–7.
4. Bartels DJP, van Laarhoven AIM, Haverkamp EA, Wilder-Smith OH, Donders ART, van Middendorp H, et al. Role of Conditioning and Verbal Suggestion in Placebo and Nocebo Effects on Itch. Sakakibara M, editor. PLoS ONE. 2014 Mar 19;9(3):e91727.
5. Blythe J, Peerdeman K, Veldhuijzen D, van Schothorst M, Thomaidou M, Laarhoven A, et al. Nocebo Effects on Cowhage-evoked Itch: A Randomized Controlled Trial of Classical Conditioning and Observational Learning. Acta Derm Venereol. 2021;101(1):adv00370.
6. Colagiuri B, Quinn VF, Colloca L. Nocebo Hyperalgesia, Partial Reinforcement, and Extinction. The Journal of Pain. 2015 Oct 1;16(10):995–1004.
7. Colloca L, Sigaudo M, Benedetti F. The role of learning in nocebo and placebo effects. Pain. 2008 May 1;136(1):211–8.
8. Colloca L, Petrovic P, Wager TD, Ingvar M, Benedetti F. How the number of learning trials affects placebo and nocebo responses. Pain. 2010;151(2):430–9.
9. Thomaidou MA, Veldhuijzen DS, Peerdeman KJ, Wiebing NZS, Blythe JS, Evers AWM. Learning mechanisms in nocebo hyperalgesia: the role of conditioning and extinction processes. Pain. 2020 Jul;161(7):1597–608.

10. Tu Y, Park J, Ahlfors SP, Khan S, Egorova N, Lang C, et al. A neural mechanism of direct and observational conditioning for placebo and nocebo responses. NeuroImage. 2019;
11. Thomaidou MA, Peerdeman KJ, Koppeschaar MI, Evers AWM, Veldhuijzen DS. How Negative Experience Influences the Brain: A Comprehensive Review of the Neurobiological Underpinnings of Nocebo Hyperalgesia. Frontiers in Neuroscience [Internet]. 2021 [cited 2021 Oct 20];15. Available from: https://www.frontiersin.org/article/10.3389/fnins.2021.652552
12. Blythe JS, Peerdeman KJ, Veldhuijzen DS, van Laarhoven AIM, Evers AWM. Placebo and nocebo effects on itch: A review of experimental methods. Itch. 2019 Jul;4(3):e27.
13. Taub CJ, Sturgeon JA, Johnson KA, Mackey SC, Darnall BD. Effects of a Pain Catastrophizing Induction on Sensory Testing in Women with Chronic Low Back Pain: A Pilot Study. Pain Research and Management. 2017;2017:1–10.
14. Madden VJ, Russek LN, Harvie DS, Vlaeyen JWS, Moseley GL. Classical Conditioning Fails to Elicit Allodynia in an Experimental Study with Healthy Humans. Pain Med. 2016 Sep 28;pnw221.
15. Crettaz B, Marziniak M, Willeke P, Young P, Hellhammer D, Stumpf A, et al. Stress-Induced Allodynia – Evidence of Increased Pain Sensitivity in Healthy Humans and Patients with Chronic Pain after Experimentally Induced Psychosocial Stress. Paul F, editor. PLoS ONE. 2013 Aug 7;8(8):e69460.
16. Adamczyk WM, Buglewicz E, Szikszay TM, Luedtke K, Bąbel P. Reward for Pain: Hyperalgesia and Allodynia Induced by Operant Conditioning: Systematic Review and Meta-Analysis. The Journal of Pain. 2019 Aug;20(8):861–75.

17. Weng L, van Laarhoven AIM, Peerdeman KJ, Evers AWM. Induction and generalization of nocebo effects on itch. Experimental Dermatology. 2022 Jun;31(6):878–89.
18. Papoiu ADP, Wang H, Coghill RC, Chan YH, Yosipovitch G. Contagious itch in humans: a study of visual "transmission" of itch in atopic dermatitis and healthy subjects. Br J Dermatol. 2011 Jun;164(6):1299–303.
19. Benedetti F, Durando J, Vighetti S. Nocebo and placebo modulation of hypobaric hypoxia headache involves the cyclooxygenase-prostaglandins pathway. Pain. 2014;155(5):921–8.
20. Hird EJ, Jones AKP, Talmi D, El-Deredy W. A comparison between the neural correlates of laser and electric pain stimulation and their modulation by expectation. Journal of Neuroscience Methods. 2018;293:117–27.
21. Thomaidou MA, Veldhuijzen DS, Meulders A, Evers AWM. An experimental investigation into the mediating role of pain-related fear in boosting nocebo hyperalgesia. Pain. 2021 Jan;162(1):287–99.
22. Stockhorst U. Classical conditioning of endocrine effects : Current Opinion in Psychiatry. Current Opinion in Psychiatry. 2005;18(2).
23. Reicherts P, Gerdes ABM, Pauli P, Wieser MJ. Psychological Placebo and Nocebo Effects on Pain Rely on Expectation and Previous Experience. 2016;
24. Bräscher AK, Witthöft M, Becker S. The Underestimated Significance of Conditioning in Placebo Hypoalgesia and Nocebo Hyperalgesia. Pain Research and Management. 2018 Jan 28;2018:1–8.
25. Aslaksen PM, Lyby PS. Fear of pain potentiates nocebo hyperalgesia. Journal of pain research. 2015 Oct;8:703–10.

26. Petersen GL, Finnerup NB, Colloca L, Amanzio M, Price DD, Jensen TS, et al. The magnitude of nocebo effects in pain: A meta-analysis. PAIN®. 2014 Aug;155(8):1426–34.
27. Vase L, Riley JL, Price DD. A comparison of placebo effects in clinical analgesic trials versus studies of placebo analgesia. Pain. 2002 Oct 1;99(3):443–52.
28. Bagarić B, Jokić-Begić N, Sangster Jokić C. The Nocebo Effect: A Review of Contemporary Experimental Research. Int J Behav Med. 2022 Jun;29(3):255–65.
29. Meeuwis SH, van Middendorp H, van Laarhoven AIM, van Leijenhorst C, Pacheco-Lopez G, Lavrijsen APM, et al. Placebo and nocebo effects for itch and itch-related immune outcomes: A systematic review of animal and human studies. Neuroscience & Biobehavioral Reviews. 2020 Jun 1;113:325–37.
30. Yetman HE, Cox N, Adler SR, Hall KT, Stone VE. What Do Placebo and Nocebo Effects Have to Do With Health Equity? The Hidden Toll of Nocebo Effects on Racial and Ethnic Minority Patients in Clinical Care. Frontiers in Psychology [Internet]. 2021 [cited 2022 Sep 22];12. Available from: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.788230
31. Kong J, Gollub RL, Polich G, Kirsch I, LaViolette P, Vangel M, et al. A Functional Magnetic Resonance Imaging Study on the Neural Mechanisms of Hyperalgesic Nocebo Effect. Journal of Neuroscience. 2008;28(49):13354-13362-13354–62.
32. Pazzaglia C, Testani E, Giordano R, Padua L, Valeriani M. Expectation to feel more pain disrupts the habituation of laser-pain rating and laser-evoked potential amplitudes. Neuroscience. 2016;333:244-251-244–51.
33. Marcuzzi A, Dean CM, Hush JM. Early changes in somatosensory function in spinal pain: protocol for a systematic review. Systematic reviews. 2013;2:90.

34. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2019; Available from: http://www.r-project.org/

35. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics. 2000 Jun;56(2):455–63.

36. Hak T, van Rhee H, Suurmond R. How to interpret results of meta-analysis. Rotterdam, The Netherlands: Erasmus Rotterdam Institute of Management [Internet]. 2016;SSRN 3241367. Available from: www.erim.eur.nl/research- support/meta-essentials/downloads

37. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003 Sep 6;327(7414):557–60.

38. Peerdeman KJ, van Laarhoven AIM, Keij SM, Vase L, Rovers MM, Peters ML, et al. Relieving patients' pain with expectation interventions: a meta-analysis. Pain. 2016 Jun;157(6):1179–91.

39. Colagiuri B, Quinn VF. Autonomic Arousal as a Mechanism of the Persistence of Nocebo Hyperalgesia. The journal of pain : official journal of the American Pain Society. 2018 May;19(5):476–86.

40. Badzińska J, Rubanets D, Wasylewski M, Honcharova S, Bajcar A, Bąbel P. The use of learning processes to change placebo effects in pain: a systematic review. 19th World Congress on Pain. 2022 Sep 20;

41. Hróbjartsson A, Gøtzsche PC. Placebo interventions for all clinical conditions. Cochrane Database Syst Rev. 2010 Jan 20;2010(1):CD003974.

42. Fullana MA, Harrison BJ, Soriano-Mas C, Vervliet B, Cardoner N, Àvila-Parcet A, et al. Neural signatures of human fear conditioning: An updated and extended meta-analysis of fMRI studies. Molecular Psychiatry. 2016 Apr;21(4):500–8.

43. Boddez Y, Moors A, Mertens G, De Houwer J. Tackling fear: Beyond associative memory activation as the only determinant of fear responding. Neuroscience & Biobehavioral Reviews. 2020 May;112:410–9.

44. Mechias ML, Etkin A, Kalisch R. A meta-analysis of instructed fear studies: Implications for conscious appraisal of threat. NeuroImage. 2010 Jan;49(2):1760–8.

45. Kahneman D, Tversky A. The Psychology of Preferences. Scientific American. 1982;246(1):160–73.

46. Ohman A, Mineka S. Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. Psychological review. 2001 Jul;108(3):483–522.

47. Van Laarhoven AIM, Vogelaar ML, Wilder-Smith O, Van Riel P, Van De Kerkhof P, Kraaimaat FW, et al. Induction of nocebo and placebo effects on itch and pain by verbal suggestions. Pain. 2011 Jul;152(7):1486–94.

48. Weng L, Peerdeman K, Della Porta D, Laarhoven AIM, Evers A. Can placebo and nocebo effects generalize within pain modalities and across somatosensory sensations? Pain. 2021 Jul 2;Publish Ahead of Print.

49. Feldhaus MH, Horing B, Sprenger C, Büchel C. Association of nocebo hyperalgesia and basic somatosensory characteristics in a large cohort. Sci Rep. 2021 Dec;11(1):762.

50. Geuter S, Buchel C. Facilitation of Pain in the Human Spinal Cord by Nocebo Treatment. Journal of Neuroscience. 2013;33(34):13784-13790-13784–90.

51. Jones C, Watson D, Fone K. Animal models of schizophrenia. Br J Pharmacol. 2011 Oct;164(4):1162–94.

52. Lu Y, Yin DM, Xiong WC, Mei L. Modeling Schizophrenia in Neuregulin 1 and ErbB4 Mutant Mice. In: O'Donnell P, editor. Animal Models of Schizophrenia and Related Disorders [Internet]. Totowa, NJ: Humana Press; 2011 [cited 2022 Jan 21]. p. 261–77. (Neuromethods). Available from: https://doi.org/10.1007/978-1-61779-157-4_12
53. Lonsdorf TB, Menz MM, Andreatta M, Fullana MA, Golkar A, Haaker J, et al. Don't fear 'fear conditioning': Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. Neuroscience & Biobehavioral Reviews. 2017;77:247–85.
54. Ben-David BM, Moral MI, Namasivayam AK, Erel H, van Lieshout PHHM. Linguistic and emotional-valence characteristics of reading passages for clinical use and research. Journal of Fluency Disorders. 2016 Sep 1;49:1–12.
55. Shaikh MAM, Prendinger H, Ishizuka M. Sentiment Assessment of Text by Analyzing Linguistic Features and Contextual Valence Assignment. Applied Artificial Intelligence. 2008 Jul 18;22(6):558–601.
56. Jensen K, Kaptchuk TJ, Chen X, Kirsch I, Ingvar M, Gollub RL, et al. A neural mechanism for nonconscious activation of conditioned placebo and nocebo responses. Cerebral Cortex. 2015;25(10):3903–10.
57. Schmid J, Bingel U, Ritter C, Benson S, Schedlowski M, Gramsch C, et al. Neural underpinnings of nocebo hyperalgesia in visceral pain: A fMRI study in healthy volunteers. NeuroImage. 2015;120:114-122-114–22.
58. Tinnermann A, Geuter S, Sprenger C, Finsterbusch J, Büchel C. Interactions between brain and spinal cord mediate value effects in nocebo hyperalgesia. Science. 2017;358:105–8.

59. Howe LC, Goyer JP, Crum AJ. Harnessing the Placebo Effect: Exploring the Influence of Physician Characteristics on Placebo Response. Health Psychol. 2017 Nov;36(11):1074–82.

60. Lobanov OV, Zeidan F, McHaffie JG, Kraft RA, Coghill RC. From cue to meaning: Brain mechanisms supporting the construction of expectations of pain. PAIN®. 2014 Jan 1;155(1):129–36.

61. Zeidan F, Lobanov OV, Kraft RA, Coghill RC. Brain mechanisms supporting violated expectations of pain. Pain. 2015;156(9):1772-1785-1772–85.

62. Albu S, Meagher MW. Expectation of nocebo hyperalgesia affects EEG alpha-activity. International Journal of Psychophysiology. 2016;109:147–52.

63. Benedetti F, Amanzio M, Vighetti S, Asteggiano G. The Biochemical and Neuroendocrine Bases of the Hyperalgesic Nocebo Effect. Journal of Neuroscience. 2006 Nov;26(46):12014–22.

64. Benedetti F, Pollo A, Lopiano L, Lanotte M, Vighetti S, Rainero I. Conscious Expectation and Unconscious Conditioning in Analgesic, Motor, and Hormonal Placebo/Nocebo Responses. J Neurosci. 2003 May 15;23(10):4315–23.

65. 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):1342-1347-1342–7.

66. Linnman C, Rougemont-Bücking A, Beucke JC, Zeffiro TA, Milad MR. Unconditioned responses and functional fear networks in human classical conditioning. Behavioural Brain Research. 2011 Aug;221(1):237–45.

67. Skvortsova A, Veldhuijzen DS, Van Middendorp H, Van den Bergh O, Evers AWM. Enhancing Placebo Effects in Somatic Symptoms Through Oxytocin. Psychosom Med. 2018 May;80(4):353–60.
68. Aslaksen PM, Åsli O, Øvervoll M, Bjørkedal E. Nocebo hyperalgesia and the startle response. Neuroscience. 2016;339:599–607.
69. 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.
70. Meeuwis SH, van Middendorp H, Lavrijsen APM, Veldhuijzen DS, Evers AWM. Open- and Closed-Label Placebo and Nocebo Suggestions About a Sham Transdermal Patch. Psychosom Med. 2021 Jan;83(1):33–42.
71. Camerone EM, Piedimonte A, Testa M, Wiech K, Vase L, Zamfira DA, et al. The Effect of Temporal Information on Placebo Analgesia and Nocebo Hyperalgesia. Psychosom Med. 2021 Jan;83(1):43–50.
72. Corsi N, Colloca L. Placebo and Nocebo Effects: The Advantage of Measuring Expectations and Psychological Factors. Frontiers in Psychology [Internet]. 2017 [cited 2022 Feb 11];8. Available from: https://www.frontiersin.org/article/10.3389/fpsyg.2017.00308
73. Atlas LY, Bolger N, Lindquist MA, Wager TD. Brain Mediators of Predictive Cue Effects on Perceived Pain. Journal of Neuroscience. 2010 Sep;30(39):12964–77.
74. Walsh KS, McGovern DP, Clark A, O'Connell RG. Evaluating the neurophysiological evidence for predictive processing as a model of perception. Annals of the New York Academy of Sciences. 2020;1464(1):242–68.

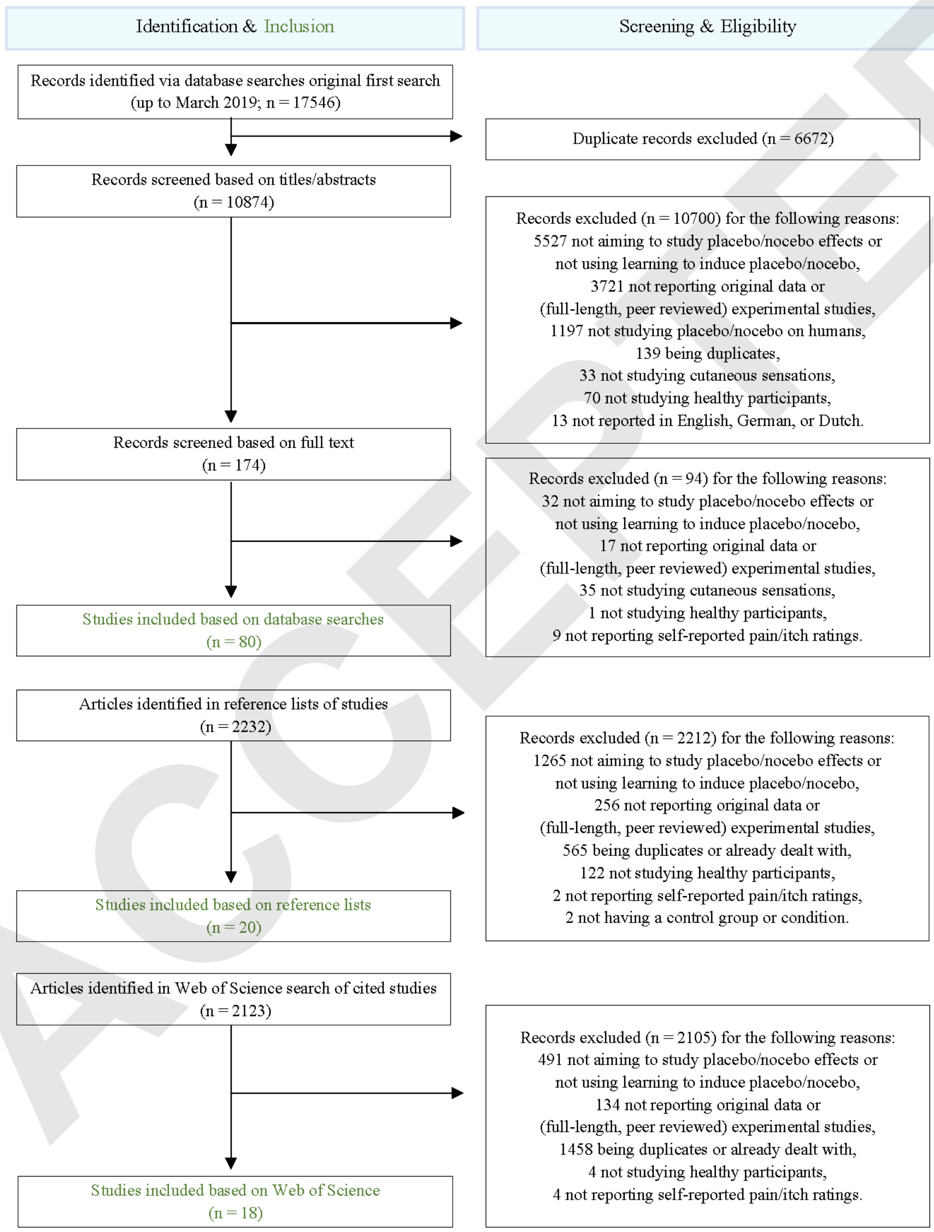
Figure Captions:

Figure 1. Flow diagram detailing the inclusion and exclusion of studies. The final sample included 127 articles, of which 106 investigated placebo effects, and 37 investigated nocebo effects (i.e., 16 studies overlapped as they investigated both placebo and nocebo). Color image is available online only at the Psychosomatic Medicine web site.

Figure 2. Forest plot of the meta-analysis indicating the magnitudes of nocebo responses following a combination of classical conditioning and verbal suggestions (CC+VS) or verbal suggestions alone (VS) on pain (A, B) and itch (C, D). Sample sizes marked with (c) indicate the combined sample from different study arms.

Figure 3. Funnel plots displaying studies within and outside of 95% (dotted line) and 99% (dashed line) CI, for pain verbal suggestions with (A) and without (B) conditioning, and for itch verbal suggestions with (A) and without (B) conditioning.

Figure 1


Identification & Inclusion
Screening & Eligibility
Records identified via database searches original first search (up to March 2019; n = 17546)
Duplicate records excluded (n = 6672)
Records screened based on titles/abstracts (n = 10874)
Records excluded (n = 10700) for the following reasons:
5527 not aiming to study placebo/nocebo effects or not using learning to induce placebo/nocebo,
3721 not reporting original data or (full-length, peer reviewed) experimental studies,
1197 not studying placebo/nocebo on humans,
139 being duplicates,
33 not studying cutaneous sensations,
70 not studying healthy participants,
13 not reported in English, German, or Dutch.
Records screened based on full text (n = 174)
Records excluded (n = 94) for the following reasons:
32 not aiming to study placebo/nocebo effects or not using learning to induce placebo/nocebo,
17 not reporting original data or (full-length, peer reviewed) experimental studies,
35 not studying cutaneous sensations,
1 not studying healthy participants,
9 not reporting self-reported pain/itch ratings.
Studies included based on database searches (n = 80)
Articles identified in reference lists of studies (n = 2232)
Records excluded (n = 2212) for the following reasons:
1265 not aiming to study placebo/nocebo effects or not using learning to induce placebo/nocebo,
256 not reporting original data or (full-length, peer reviewed) experimental studies,
565 being duplicates or already dealt with,
122 not studying healthy participants,
2 not reporting self-reported pain/itch ratings,
2 not having a control group or condition.
Studies included based on reference lists (n = 20)
Articles identified in Web of Science search of cited studies (n = 2123)
Records excluded (n = 2105) for the following reasons:
491 not aiming to study placebo/nocebo effects or not using learning to induce placebo/nocebo,
134 not reporting original data or (full-length, peer reviewed) experimental studies,
1458 being duplicates or already dealt with,
4 not studying healthy participants,
4 not reporting self-reported pain/itch ratings.
Studies included based on Web of Science (n = 18)

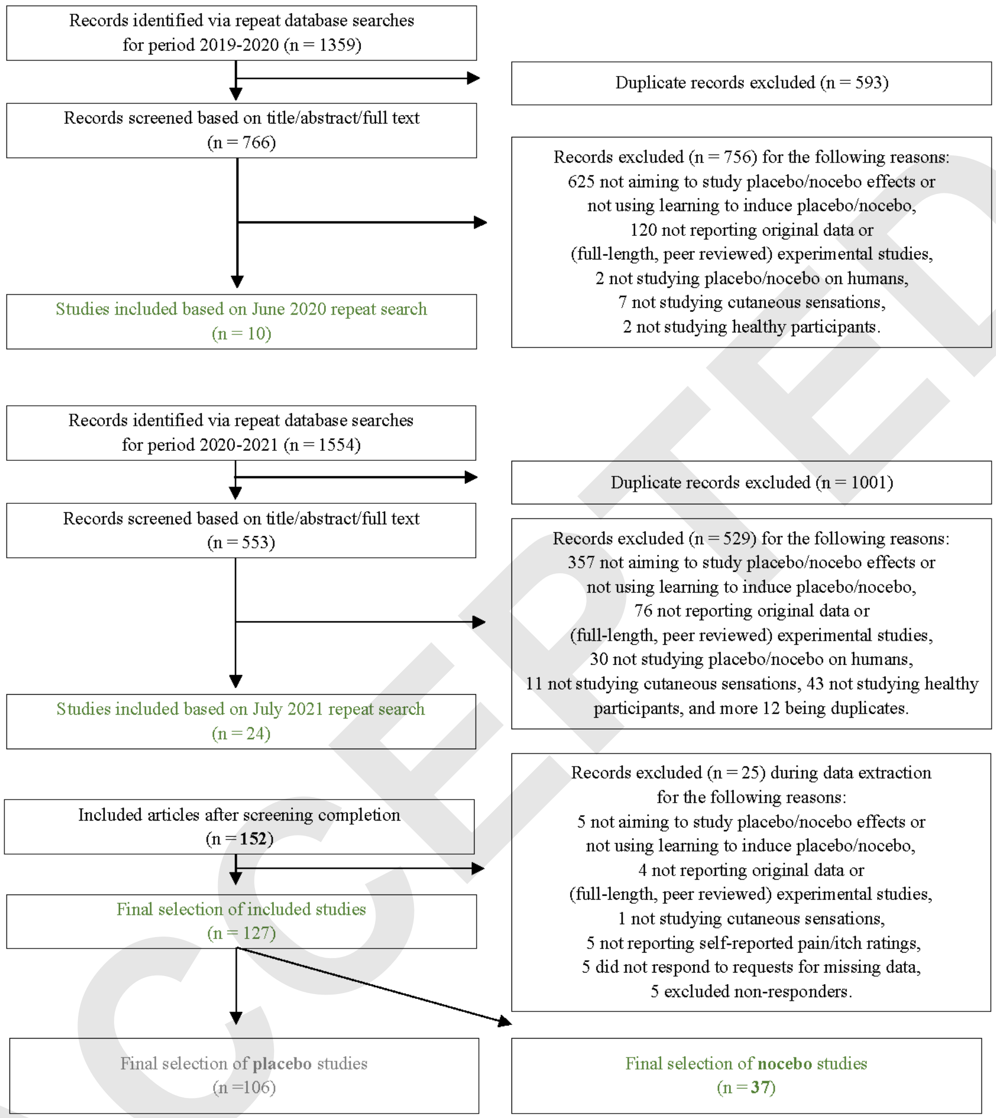

Records identified via repeat database searches
for period 2019-2020 (n = 1359)
Duplicate records excluded (n = 593)
Records screened based on title/abstract/full text
(n = 766)
Records excluded (n = 756) for the following reasons:
625 not aiming to study placebo/nocebo effects or
not using learning to induce placebo/nocebo,
120 not reporting original data or
(full-length, peer reviewed) experimental studies,
2 not studying placebo/nocebo on humans,
7 not studying cutaneous sensations,
2 not studying healthy participants.
Studies included based on June 2020 repeat search
(n = 10)
Records identified via repeat database searches
for period 2020-2021 (n = 1554)
Duplicate records excluded (n = 1001)
Records screened based on title/abstract/full text
(n = 553)
Records excluded (n = 529) for the following reasons:
357 not aiming to study placebo/nocebo effects or
not using learning to induce placebo/nocebo,
76 not reporting original data or
(full-length, peer reviewed) experimental studies,
30 not studying placebo/nocebo on humans,
11 not studying cutaneous sensations, 43 not studying healthy
participants, and more 12 being duplicates.
Studies included based on July 2021 repeat search
(n = 24)
Records excluded (n = 25) during data extraction
for the following reasons:
5 not aiming to study placebo/nocebo effects or
not using learning to induce placebo/nocebo,
4 not reporting original data or
(full-length, peer reviewed) experimental studies,
1 not studying cutaneous sensations,
5 not reporting self-reported pain/itch ratings,
5 did not respond to requests for missing data,
5 excluded non-responders.
Included articles after screening completion
(n = 152)
Final selection of included studies
(n = 127)
Final selection of placebo studies
(n =106)
Final selection of nocebo studies
(n = 37)

Figure 2

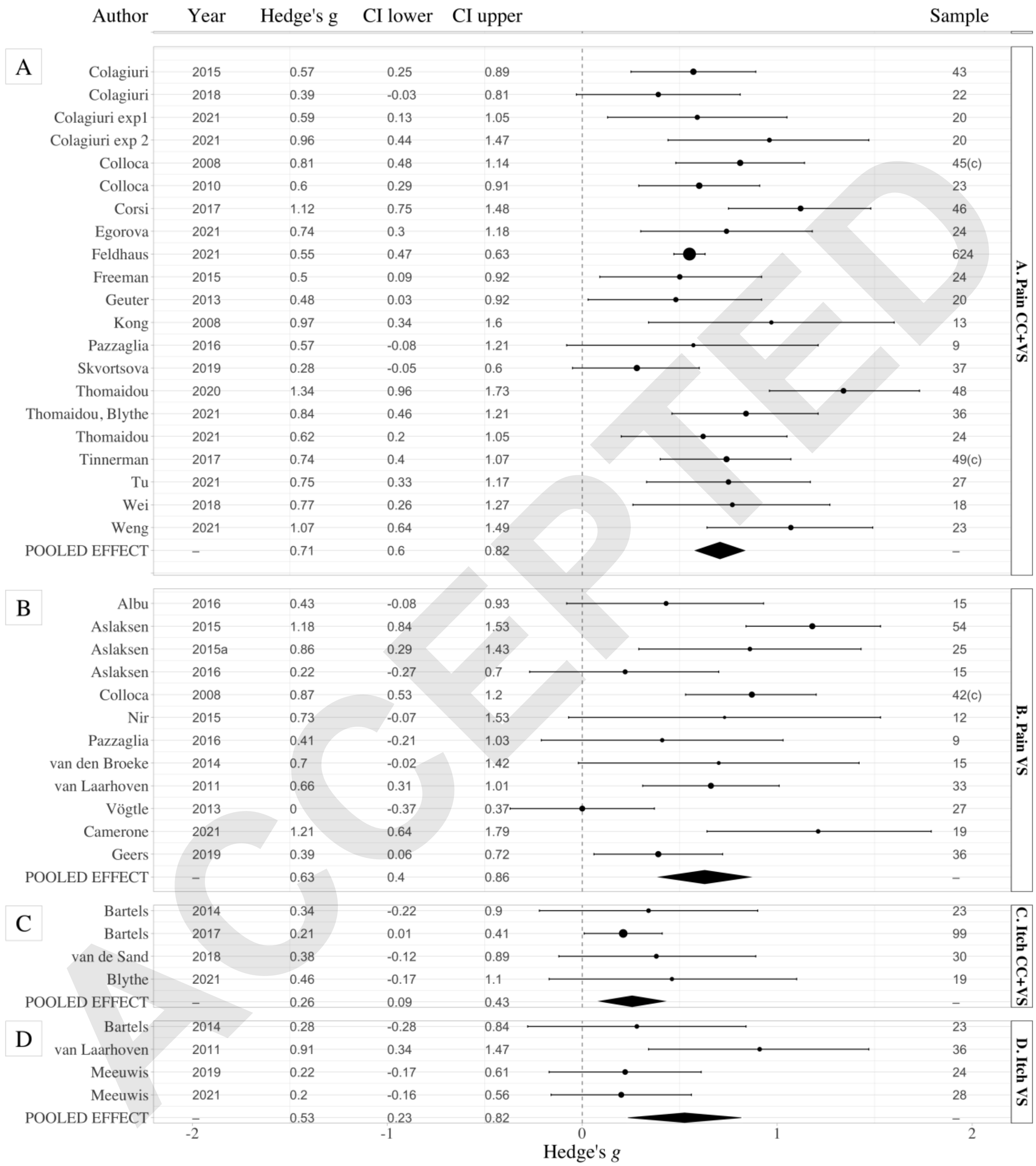

Author Year Hedge's g CI lower CI upper Sample
A
Colagiuri 2015 0.57 0.25 0.89 43
Colagiuri 2018 0.39 -0.03 0.81 22
Colagiuri exp1 2021 0.59 0.13 1.05 20
Colagiuri exp 2 2021 0.96 0.44 1.47 20
Colloca 2008 0.81 0.48 1.14 45(c)
Colloca 2010 0.6 0.29 0.91 23
Corsi 2017 1.12 0.75 1.48 46
Egorova 2021 0.74 0.3 1.18 24
Feldhaus 2021 0.55 0.47 0.63 624
Freeman 2015 0.5 0.09 0.92 24
Geuter 2013 0.48 0.03 0.92 20
Kong 2008 0.97 0.34 1.6 13
Pazzaglia 2016 0.57 -0.08 1.21 9
Skvortsova 2019 0.28 -0.05 0.6 37
Thomaidou 2020 1.34 0.96 1.73 48
Thomaidou, Blythe 2021 0.84 0.46 1.21 36
Thomaidou 2021 0.62 0.2 1.05 24
Tinnerman 2017 0.74 0.4 1.07 49(c)
Tu 2021 0.75 0.33 1.17 27
Wei 2018 0.77 0.26 1.27 18
Weng 2021 1.07 0.64 1.49 23
POOLED EFFECT – 0.71 0.6 0.82 –
A. Pain CC+VS
B
Albu 2016 0.43 -0.08 0.93 15
Aslaksen 2015 1.18 0.84 1.53 54
Aslaksen 2015a 0.86 0.29 1.43 25
Aslaksen 2016 0.22 -0.27 0.7 15
Colloca 2008 0.87 0.53 1.2 42(c)
Nir 2015 0.73 -0.07 1.53 12
Pazzaglia 2016 0.41 -0.21 1.03 9
van den Broeke 2014 0.7 -0.02 1.42 15
van Laarhoven 2011 0.66 0.31 1.01 33
Vögtle 2013 0 -0.37 0.37 27
Camerone 2021 1.21 0.64 1.79 19
Geers 2019 0.39 0.06 0.72 36
POOLED EFFECT – 0.63 0.4 0.86 –
B. Pain VS
C
Bartels 2014 0.34 -0.22 0.9 23
Bartels 2017 0.21 0.01 0.41 99
van de Sand 2018 0.38 -0.12 0.89 30
Blythe 2021 0.46 -0.17 1.1 19
POOLED EFFECT – 0.26 0.09 0.43 –
C. Itch CC+VS
D
Bartels 2014 0.28 -0.28 0.84 23
van Laarhoven 2011 0.91 0.34 1.47 36
Meeuwis 2019 0.22 -0.17 0.61 24
Meeuwis 2021 0.2 -0.16 0.56 28
POOLED EFFECT – 0.53 0.23 0.82 –
D. Itch VS
-2
-1
0
1
2
Hedge's g

Figure 3

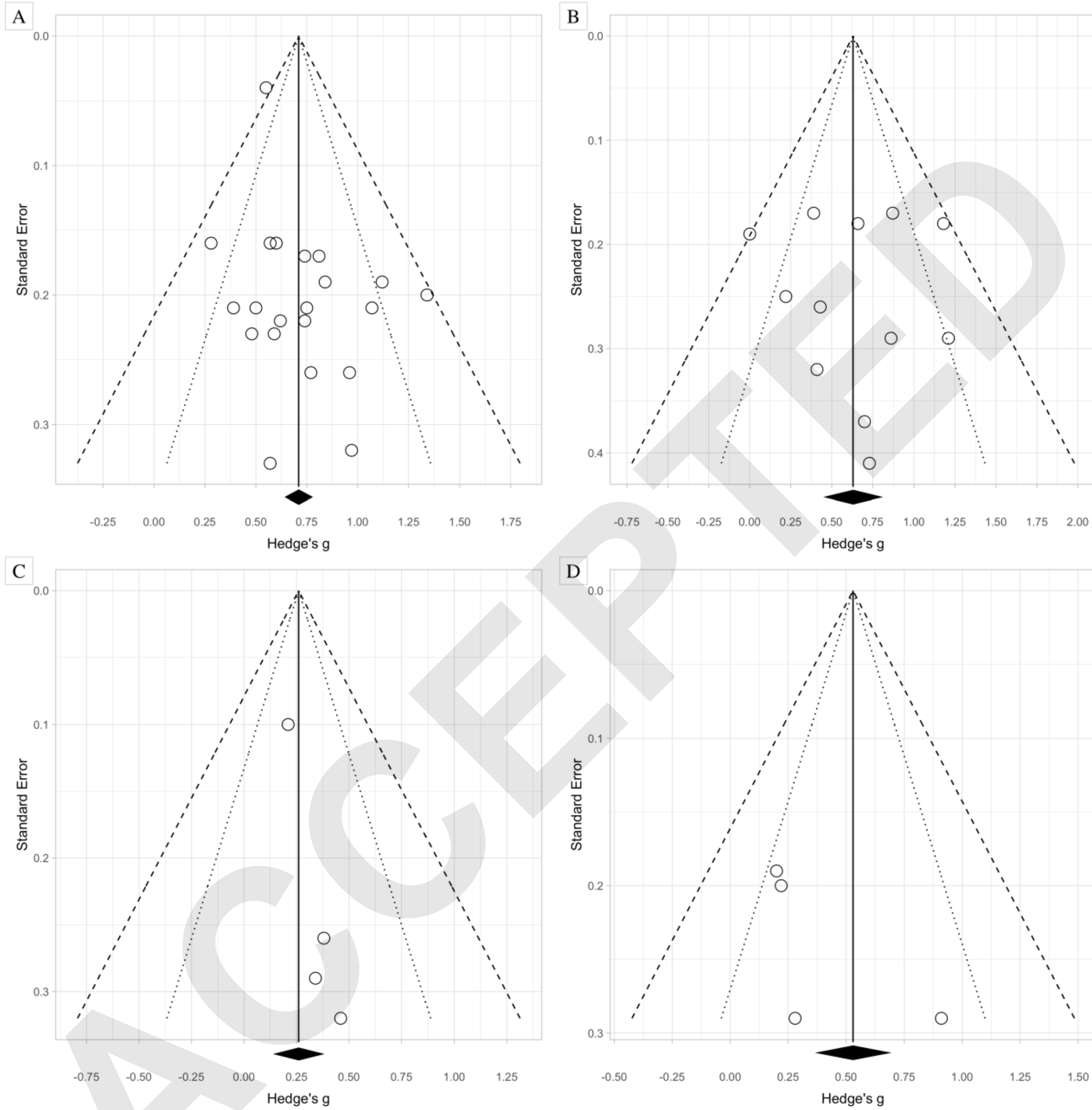

A
Standard Error
0.0
0.1
0.2
0.3
-0.25
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
Hedge's g
B
Standard Error
0.0
0.1
0.2
0.3
0.4
-0.75
-0.50
-0.25
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
Hedge's g
C
Standard Error
0.0
0.1
0.2
0.3
-0.75
-0.50
-0.25
0.00
0.25
0.50
0.75
1.00
1.25
Hedge's g
D
Standard Error
0.0
0.1
0.2
0.3
-0.50
-0.25
0.00
0.25
0.50
0.75
1.00
1.25
1.50
Hedge's g

Table 1. Study characteristics for all included articles.

	Authors	Year	Sample size nocebo group	Sample size control group	Total sample size (Male/ Female)	Mean age (SD)	Stimulation type	Learning method	Results synthesis where applicable	Number of conditioning trials (N/C) (length of conditioning)	Outcome measure: evocation first trials or mean of trials by trial type	Control condition for nocebo outcome (for VS only studies: between-subjects always)	Language of assessment	Risk of Bias score (0-34)
PAIN														
1	Colagiuri, Quinn, et al	**2015**	37	42	46 (22M/24F)	20.3 (4.0)	Electrical	CC+VS		32 (16/16)	First	Within subjects	English (native)	3
2	Colagiuri & Quinn	**2018**	20	20	135 (62M/73F)	20.2 (4.0)	Electrical	CC+VS		32 (16/16)	Mean	Within subjects	English (native)	5
3	Colagiuri, Park, et al.	**2021**	20 + 20	21 + 20	65 (19M/46F)	20.7 (3.6)	Electrical	CC+VS	Lengthier learning condition treated and analyzed as a separate study arm	32 (16/16)	Mean	Within subjects	English (native)	3
4	Colloca, Petrovic, et al.	**2010**	23 + 23	n/a	61 (26M/35F) + 80 (17M/63F)	22.8 (3.4)	Electrical	CC+VS	Four vs. of one learning sessions averaged together	20 (10/10) or 80 (40/40)	Mean	Within subjects	Not reported	3
5	Colloca, Sigaudo, et al	**2008**	42 VS & 45 CC+VS	n/a	46 (16M/30F)	22.3 (2.4)	Electrical	CC+VS & VS	Three pain intensities averaged across VS and CC+VS conditions and analyzed as two separate study arms	24 (12/12)	Mean	Within subjects	Not reported	3
6	Corsi & Colloca	**2017**	46	n/a	116 (0M/116F)	27.4 (1.1)	Thermal	CC+VS		12 (6/6)	Mean	Within subjects	English (native)	3
7	Egorova, Benedetti, et al	**2020**	24	n/a	24 (12M/12F)	n/a	Thermal	CC+VS		48 (24/24)	Mean	Within subjects	English (native)	5
8	Feldhaus, Horing, et al.	**2021**	624	n/a	624 (251M/373F)	24.6 (3.6)	Thermal	CC+VS		16 (8/8)	Mean	Within subjects	German (native)	3

9	Freeman, Yu, et al.	**2015**	24	n/a	24 (12M/12F)	21 to 49	Thermal	CC+VS		18 (9/9)	First	Within subjects	English (native)	5
10	Geuter & Büchel	**2013**	20	n/a	20 (12M/8F)	26.4	Thermal	CC+VS		24 (12/12)	Mean	Within subjects	Not reported	3
11	Kong, Gollub, et al.	**2008**	13	n/a	13 (5M/8F)	26.3 (3.6)	Thermal	CC+VS		48 (24/24)	First	Within subjects	English (native)	5
12	Pazzaglia, Testani, et al.	**2016**	9 + 9	n/a	18 (10M/8F)	29 (5.0)	Laser	CC+VS &VS	VS condition treated and analyzed as a separate study arm	60 (30/30)	Mean	Within subjects	Not reported	5
13	Skvortsova, Veldhuijzen, et al.	**2019**	37	n/a	37 (37M/0F)	23.1 (2.9)	Thermal	CC+VS		24 (12/12)	Mean	Within subjects	Dutch (native)	0
14	Thomaidou, Blythe, et al.	**2021 b**	36	n/a	36 (11M/25F)	22.9 (2.2)	Thermal	CC+VS		32 (16/16)	First	Within subjects	English (mixed)	5
15	Thomaidou, Veldhuijzen, et al.	**2020**	48	25	122 (20M/102F)	21.8 (2.1)	Thermal	CC+VS		30 (15/15)	First	Within subjects	Dutch (native)	5
16	Thomaidou, Veldhuijzen, et al.	**2021 a**	24	n/a	72 (18M/54F)	22.2 (1.9)	Thermal	CC+VS		24 (12/12)	First	Within subjects	English (mixed)	5
17	Tinnermann, Geuter, et al.	**2017**	25 + 24	n/a	49 (27M/22F)	25.4 (3.8)	Thermal	CC+VS	Cheap vs. expensive conditions were averaged together	16 (8/8)	Mean	Within subjects	Not reported	6
18	Tu, Wilson, et al.	**2021**	27	n/a	81 (44M/37F)	27.4 (6.4)	Thermal	CC+VS		48 (24/24)	Mean	Within subjects	English (native)	3
19	Wei, Zhou, et al.	**2018**	18	n/a	76 (0M/76F)	20.9 (1.4)	Electrical	CC+VS		40 (20/20)	Mean	Within subjects	Chinese (native)	3
20	Weng, Peerdeman, et al.	**2021**	33	n/a	33 (8M/25F)	21.6 (3.0)	Thermal	CC+VS		30 (15/15)	Mean	Within subjects	English (mixed)	1
21	Albu & Meagher	**2016**	15	15	30 (11M/19F)	19.1 (1.2)	Thermal	VS		n/a	Mean	Within subjects	English (native)	3
22	Aslaksen & Lyby	**2015**	57	54	111 (35M/76F)	22.2 (3.1)	Thermal	VS		n/a	First	Between subjects	Norwegian (native)	3
23	Aslaksen, Åsli, et al.	**2016**	15	16	61 (28M/33F)	21.6 (3.3)	Thermal	VS		n/a	First	Between subjects	Norwegian (native)	0
24	Aslaksen, Zwarg, et al.	**2015**	25	25	142 (69M/73F)	23.4 (4.1)	Thermal	VS		n/a	First	Between subjects	Norwegian (native)	0
25	Camerone, Piedimonte, et al.	**2021**	19	21	157 (73M/84F)	23.1 (2.1)	Electrical	VS	We analyzed the 5-min condition	n/a	Mean	Within subjects	Not reported	1
26	Geers, Close,	**2019**	36	36	146	19.7	Cold pressor	VS		n/a	Mean	Between	English	3

	et al.				(92F/54M)	(3.2)						subjects	(native)	
27	Nir, Yarnitsky, et al.	**2012**	12	12	48 (48M/0F)	25.8 (3.2)	Hot water bath	VS		n/a	Mean	Between subjects	Not reported	3
28	van den Broeke, Geene, et al.	**2014**	15	15	30 (11M/19F)	23.5 (2.2)	Mechanical stimulation	VS		n/a	First	Within subjects	Dutch (native)	4
29	Vögtle, Barke, et al	**2013**	26	26	80 (0M/80F)	22.5 (4.4)	Pressure	VS		n/a	Mean	Within subjects	German (native)	2
30-31	van Laarhoven, Vogelaar, et al.	**2011**	33pain & 36itch	16pain & 20itch	105 (0M/105F)	21.8 (2.2)	Electrical, Mechanical, Histamine	VS	Three types of stimulations averaged together across pain and across itch	n/a	Mean	Between subjects	Dutch (native)	1
ITCH														
32	Bartels, van Laarhoven, et al.	**2014**	23 + 23	25	95 (22M/73F)	22.7 (3.2)	Electrical	CC+VS & VS	VS condition treated and analyzed as a separate arm	12 (6/6)	Mean	Between subjects	Dutch (native)	4
33	Bartels, van Laarhoven, et al.	**2017**	99	n/a	99 (21M/78F)	20.3 (2.5)	Electrical	CC+VS		16 (10/6)	Mean	Within subjects	Dutch (native)	4
34	Blythe, Peerdeman, et al.	**2021**	19	19	39 (0M/39F)	21.9 (2.4)	Cowhage	CC+VS		4 (2/2)	Mean	Within subjects	English (mixed)	2
35	van de Sand, Menz, et al.	**2018**	30	30	30 (12M/18F)	25.5	Histamine skin scrub	CC+VS		40 (20/20)	Mean	Within subjects	Not reported	5
36	Meeuwis, van Middendorp, et al.	**2019**	24	n/a	92 (16M/76F)	21.8 (2.7)	Histamine iontophoresis	VS		n/a	Mean	Within subjects	Dutch (native)	4
37	Meeuwis, van Middendorp, et al.	**2021**	28	n/a	111 (18M/93F)	21.9 (2.8)	Histamine iontophoresis	VS		n/a	Mean	Within subjects	Dutch (native)	4

Note: the study by van Laarhoven et al., 2011, included both itch and pain manipulations and is listed under pain. When the sample size of a control group is listed as n/a, this suggests that the study used a within-subjects controlled design. In language of assessment, the note 'native' indicates that the local native language of participants was used; when known, the note 'mixed' indicates that the sample was of mixed nationalities and the language of assessments was native for some but not for others. Studies are listed separately for pain and itch and first based on the learning manipulation (VS, verbal suggestions, or CC+VS, combination of classical conditioning and verbal suggestions) and then alphabetically. N, Nocebo; C, Control; M, Male; F, Female.