



Published in final edited form as:

*Pain*. 2014 August ; 155(8): 1426–1434. doi:10.1016/j.pain.2014.04.016.

## The magnitude of nocebo effects in pain: A meta-analysis

Gitte Laue Petersen<sup>a,b,\*</sup>, Nanna Brix Finnerup<sup>b</sup>, Luana Colloca<sup>c</sup>, Martina Amanzio<sup>d</sup>, Donald D. Price<sup>e</sup>, Troels Staehelin Jensen<sup>b</sup>, and Lene Vase<sup>a,b</sup>

<sup>a</sup>Department of Psychology and Behavioural Sciences, School of Business and Social Sciences, Aarhus University, Aarhus, Denmark

<sup>b</sup>Danish Pain Research Center, Aarhus University Hospital, Aarhus, Denmark

<sup>c</sup>National Center for Complementary and Alternative Medicine (NCCAM), National Institute of Mental Health (NIMH), and Clinical Center, Department of Bioethics, National Institutes of Health, Bethesda, MD, USA

<sup>d</sup>Department of Psychology, University of Turin, National Institute of Neuroscience (INN) Turin, Italy

<sup>e</sup>Division of Neuroscience, Department of Oral and Maxillofacial Surgery, University of Florida, Gainesville, FL, USA

### Abstract

The investigation of nocebo effects is evolving and a few literature reviews have emerged, however, so far without quantifying such effects. This meta-analysis investigated nocebo effects in pain. We searched the databases PubMed, EMBASE, Scopus, and the Cochrane Controlled Trial Register with the term “nocebo”. Only studies that investigated nocebo effects as the effects that follow the administration of an inert treatment along with verbal suggestions of symptom worsening and that included a no-treatment control condition were eligible. Ten studies fulfilled the selection criteria. The effect sizes were calculated using Cohen’s *d* and Hedges’ *g*. The overall magnitude of the nocebo effect was moderate to large (lowest  $g = 0.62$  (0.24-1.01) and highest  $g = 1.03$  (0.63-1.43)) and highly variable (range of  $g = -0.43$ -4.05). The magnitudes and range of effect sizes was similar to those of placebo effects ( $d = 0.81$ ) in mechanistic studies. In studies where nocebo effects were induced by a combination of verbal suggestions and conditioning, the effect size was larger (lowest  $g = 0.76$  (0.39-1.14) and highest  $g = 1.17$  (0.52-1.81)) than in studies where nocebo effects were induced by verbal suggestions alone (lowest  $g = 0.64$  (-0.25-1.53) and highest  $g = 0.87$  (0.40-1.34)). These findings are similar to those in the placebo literature. Since the magnitude of the nocebo effect is variable and sometimes large, this meta-analysis demonstrates the importance of minimizing nocebo effects in clinical practice.

\*Corresponding author: Gitte Laue Petersen, Department of Psychology and Behavioural Sciences, School of Business and Social Sciences, Aarhus University, Bartholins Allé 9, DK-8000 Aarhus C, Denmark, Tel: + 45 8716 5791; fax: + 45 8716 4401, gittel@psy.au.dk.

#### Conflicts of interest

The authors declare no conflicts of interest.

## Keywords

nocebo; placebo; pain; magnitude; verbal suggestion; conditioning

---

## 1. Introduction

The nocebo effect is the opposite phenomenon of the placebo effect, which is defined as the effect that follows the administration of a placebo treatment [7]. A placebo treatment is typically given along with verbal suggestions of clinical improvement, thereby making the patient expect symptom relief. The nocebo effect was originally introduced to describe the negative side effects of a placebo treatment [29,41,42]. Negative side effects occur when expectations of symptom relief result in symptom worsening. Today, the nocebo effect is an independent phenomenon. Accordingly, the nocebo effect is defined as the effect that follows the administration of an inert treatment along with verbal suggestions of symptom worsening. The patient expects to feel worse and eventually will [6,7]. In order to differentiate nocebo effects and placebo effects from spontaneous remission and other confounding factors, they are calculated as the symptom difference between a nocebo-treated or placebo-treated group or condition and a no-treatment group or condition [27]. Both nocebo effects and placebo effects are at least partly mediated by the patient's expectations of pain increase or pain relief from a treatment. Expectations may be influenced by prior experiences as in classical conditioning and/or by verbal suggestions given by, for example, health care providers [6,55].

Several meta-analyses on the magnitude of placebo analgesia effects have been conducted [4,35,36,37,54,62,63]. Across meta-analyses, the magnitude of placebo effects has been shown to vary from Cohen's  $d = -0.28$  [35] to  $d = 1.14$  [54] and to depend on how the placebo effect was induced. When placebo effects are induced by verbal suggestions alone, an average effect size of  $d = 0.85$  has been found as opposed to an average effect size of  $d = 1.45$  when verbal suggestions and conditioning are combined [62].

Nocebo effects, especially in pain, have recently received increasing interest, and a few literature reviews have emerged [e.g., 10,18,22]. However, so far no quantification of the effects has been conducted. The nocebo literature is to a large extent based on studies involving healthy volunteers exposed to different types of experimental pain manipulations and therefore the designs are likely to vary across studies. Still, it is of interest to analyze the magnitude and variation of the nocebo effect and to put these findings into perspective in relation to the placebo effect. With this aim in mind, we provide a meta-analysis on nocebo effects in experimental pain studies to answer the following questions:

1. How large are the magnitudes and heterogeneity of nocebo effects?
2. Do the magnitudes of nocebo effects vary according to whether they are induced by verbal suggestions alone or by verbal suggestions combined with conditioning?

## 2. Methods

### 2.1. Sample of studies

Studies were identified by searching the electronic databases PubMed, EMBASE, and Scopus and the Cochrane Controlled Trial Register (the Cochrane Library) using the search term “nocebo”. With regard to EMBASE and Scopus, it was possible to search for articles only and this strategy was therefore chosen. The search terms “nocebo effect” and “nocebo hyperalgesia” were also applied, but they did not generate further studies. No limits, except humans, were applied as no meta-analysis on nocebo effects has been carried out before. The last database search was run on May 31, 2013.

### 2.2 Selection criteria

The study had to be published as a new and full article and therefore abstracts, reviews, and double publications were not considered. As the nocebo effect can be conceptualized as increased negative pain symptom(s) that result from learning procedures (classical conditioning or social observation) and/or verbal suggestions of symptom worsening, the following selection criteria were applied:

1. The purpose of the study should be experimental investigation of nocebo effects in pain. Therefore, adverse effects of (placebo) treatments, for example, following information disclosing potential side effects, were not considered, as the purpose of such verbal information was not to increase pain. Likewise, manipulations and verbal suggestions given to increase pain outside a treatment setting or without administration of an inert treatment were not considered.
2. The study should include a nocebo treatment. The nocebo treatment was conceptualized as administration of an inert agent/intervention along with verbal suggestions for pain increase and/or a learning procedure (either classical conditioning or social observation) which aimed to increase pain levels.
3. The study should include information on no treatment, so the nocebo effect could be calculated as the difference in pain between a nocebo treatment and no treatment. The information on no treatment could come either from a no-treatment group or condition or from the change between minimum and maximum pain levels.
4. Only pain studies (both experimental pain and clinical pain) including numerical rating of pain intensity were included (i.e., both the visual analogue scale (VAS) and the numerical rating scale (NRS)). To allow for consistency across the sample, only pain intensity ratings were obtained and in case of several outcome measures on pain intensity, the one most clinically relevant was chosen as this was considered to be the best test of the existence of nocebo effects.
5. The study should be randomized or counterbalanced and involve blinding procedures.

Studies that investigated nocebo effects but did not fulfill one or more of the selection criteria were excluded from the meta-analysis. Excluded studies are listed in Appendix A with reasons for the exclusion.

### 2.3. Study selection and assessment

Eligibility assessment was performed independently by 2 of the authors (G.L.P. and L.V.). The quality of the studies was assessed by 4 of the authors (G.L.P., L.V., N.B.F., and D.D.P.), who read and assessed the quality of each study independently. Data extraction was performed by 2 of the authors (G.L.P. and L.V.). Any disagreements were resolved by consensus among all authors.

### 2.4. Trial flow

A total of 540 potential articles were identified through the database search, and 272 articles remained for consideration after removal of duplicates. The articles were screened on the basis of the title and abstract. Of these, 61 articles were excluded as they were reviews and 195 articles were discarded due to the first selection criterion as they were not empirical investigations of nocebo effects in pain. Six articles were excluded due to the second selection criterion as they did not administer an inert nocebo treatment/intervention. Ten articles were examined in detail and included in the meta-analysis as they met all the selection criteria [8,9,11,20,21,26,39,44,61,65]. No additional articles were identified by checking the references of the included articles. See Fig. 1 for a flow diagram of the study selection according to the PRISMA statement [51].

### 2.5. General considerations in data collection

The majority of studies investigated the nocebo effect via a nocebo-treated group or condition and a no-treatment group or condition [9,11,39,61] or minor variations of this classical clinical design [8,26,44,65]. One study [8] applied the open/hidden design, but it was comparable to the other studies as inactive saline was administered in full view of the patients along with verbal suggestions for pain increase in the open condition and without the patient's knowledge in the hidden condition. Two studies [20,21] deviated from the classical clinical design in the investigation of nocebo effects as no absolute control condition was included. In these studies, the nocebo effect was calculated as the difference between the minimum and maximum pain levels in the nocebo condition. It is theoretically and practically difficult to avoid any influence on the no-treatment group [33,34] so in that sense it is acceptable to include control groups that minimize the influence.—Also, the studies did administer an inert treatment/intervention with the purpose of increasing participants' pain levels (e.g., by means of a sham electrode) and induced pain via verbal suggestions alone or verbal suggestions combined with conditioning in accordance with the classical clinical design studies. For further considerations concerning data collection, see Appendix B.

### 2.6. Statistics

We conducted the meta-analysis using the Comprehensive Meta-Analysis Program, version 2.2.057 (Comprehensive Meta-Analysis, 2006). As the nocebo studies used different designs

and were conducted in different populations, the analyses were based on a random-effects model in order to take into account the within study error and the between study variance [14]. Means and standard deviations were used for the calculation of effect sizes. If only standard errors were reported, they were converted to standard deviations. The effect size was computed as the mean for the no-treatment group or condition minus the mean for the placebo-treated group or condition, divided by the pooled standard deviation (Cohen's  $d$ ) [16]. The magnitudes were reported as effect sizes weighted according to the number of subjects ( $N$ ) in each study (Hedges'  $g$ ) [29,30]. Positive values of  $d$  and  $g$  indicate higher pain ratings in the placebo condition compared with the no-treatment condition. Negative values of  $d$  and  $g$  indicate lower pain ratings in the placebo condition compared with the no-treatment condition. A value of 0.2 was considered a small effect, a value of 0.5 a medium effect, and a value of 0.8 a large effect [16].

If data were insufficiently, unclearly, or not reported in the articles, the authors were contacted and asked to provide us with the data. We were unable to obtain original data from three studies [8,9,11], in which we based our calculations on reported data or readouts of figures in the article. In one of the studies [9], our readouts were not directly comparable to the pain ratings reported in the article, and we therefore adjusted our read values to the reported values in order to optimize the accuracy of our calculations (for further information, see Appendix B). Some studies [9,39] have several reports of how the magnitudes of placebo effects vary over time and if possible, we preferred data for each subject's pain rating at each measured time point. This allowed original data to be entered into the analysis even when the numbers ( $N$ ) changed over time. In other studies [20,65],  $N$  changed because the placebo effects were induced by means of verbal suggestions alone or in combination with conditioning in groups with different  $N$ . In order to facilitate a more precise comparison across studies the lowest and highest effect sizes of placebo effects were calculated instead of an average. This strategy was chosen in order to avoid manipulation of the original data. If pain was only measured at one time point, this measure was used in the respective calculations of the lowest and highest magnitude of placebo effects.

Based on the  $N$  reported in the studies, some of the placebo effects could be calculated both as "between-subject" and "within-subject" comparisons. In all studies, we entered the placebo effect as a "between-subject" comparison if this was done in the original study. Likewise, we entered the placebo effect as a "within-subject" comparison if this was done in the original study. For the studies applying "within-subject" comparisons,  $N$  was considered the  $N$  for the paired sample and for the studies applying "between-subject" comparisons the exact number of observations for the separate samples was chosen. Data from studies applying "within-subject" comparisons were adjusted using the approach adopted by the Comprehensive Meta-Analysis Program and this included calculations of effect sizes based on pre-post correlations of measures. The change between pre and post measures was based on the Pearson's correlation coefficients ( $r$ ).

In the program, the type of induction of placebo effects was treated as a subgroup moderator analysis with verbal suggestions, verbal suggestion combined with conditioning, and social observation as three subgroups. Significance of heterogeneity in the sample of studies was evaluated by  $Q$ -statistics [16], and  $I^2$  statistics assessed the variance accounted for by

heterogeneity.  $I^2$  values of 25%, 50% and 75% indicate low, moderate, and high degrees of heterogeneity, respectively [32].

### 3. Results

#### 3.1. Characteristics of included studies

Descriptive data for each of the included studies are presented in Table 1. The included studies involved 334 participants divided between 49 patients in 2 studies and 285 healthy volunteers in 8 studies. In studies of patients, they either had mild pain (NRS  $\leq 3$ ) following video-assisted thoracoscopy [8] or were female patients with persistent pain due to irritable bowel syndrome [61]. In studies of healthy volunteers, 1 study was exclusively based on males, 4 studies on females, and 5 studies on a mix of males and females.

Regarding the verbal suggestions for pain increase, all studies except the one by Vögtle et al. [65] gave moderate to strong suggestions for pain increase. None of the studies gave verbal suggestions to the extent that the participant would have a 50/50 probability for pain increase.

#### 3.2. Overall magnitude of nocebo effects

Results for each of the 10 included studies are presented in Table 2. Across the studies, the lowest effect size was  $g = 0.62$  (0.24-1.01) whereas the highest effect size was  $g = 1.03$  (0.63-1.43). The effect sizes varied from no effect to very large effects as indicated by the range of  $g$  values from  $-0.43$  to  $4.05$ . The confidence intervals depicted in Table 2 also show this variability. Thus, the overall magnitude of nocebo effects was moderate to large and highly variable. Also, there did not appear to be any systematic differences in effect sizes as a function of population, pain stimuli, agent, blindness procedure, pain measure, or time of pain measurement but this was not investigated via a statistical test due to the low number of included studies in the meta-analysis. The heterogeneity analyses showed that there was a high degree of heterogeneity between studies for the lowest effect size of  $g$  ( $Q(9) = 34.103$ ,  $P < 0.000$ ,  $I^2 = 73.61\%$ ) as well as for the highest effect size of  $g$  ( $Q(9) = 37.330$ ,  $P < 0.000$ ,  $I^2 = 75.89\%$ ).

#### 3.3. The magnitude of nocebo effects according to the way they were induced

In 6 studies, nocebo effects were induced by verbal suggestions alone, resulting in the lowest effect size of  $g = 0.64$  ( $-0.25$ - $1.53$ ) and the highest effect size of  $g = 0.87$  ( $0.40$ - $1.34$ ). In 5 studies, nocebo effects were induced by verbal suggestions combined with a conditioning procedure resulting in the lowest effect size of  $g = 0.76$  ( $0.39$ - $1.14$ ) and the highest effect size of  $g = 1.17$  ( $0.52$ - $1.81$ ). One study could not be classified as inducing nocebo effects via verbal suggestion alone or verbal suggestion combined with conditioning procedure as it performed a social observation procedure without verbal suggestions [65] and the effect size in this study was  $g = 0.35$  ( $-0.18$ - $0.89$ ). This yielded a lowest effect size of  $g = 0.63$  ( $0.34$ - $0.92$ ) and a highest effect size of  $g = 0.77$  ( $0.46$ - $1.07$ ). None of the included studies applied a conditioning procedure alone (see Table 3). The reason that these numbers add up to 11 studies is that one study [21] investigated the induction of nocebo effects by both verbal suggestions alone and verbal suggestions combined with conditioning.

With regard to the subgroup moderator analysis of how the placebo effects were induced, no statistically significant differences between subgroups were found ( $Q(2) = 1.541$ ,  $P = 0.463$  for the lowest effect size of  $g$  and  $Q(2) = 3.984$ ,  $P = 0.136$  for the highest effect size of  $g$ ).

## 4. Discussion

This meta-analysis shows that the magnitude of the placebo effect is moderate to large and highly variable. Also, the placebo effect is higher when induced by conditioning combined with verbal suggestions as compared to verbal suggestions alone.

### 4.1. The magnitude of placebo effects in pain

This meta-analysis demonstrates a moderate to large magnitude of placebo effects in relation to pain as indicated by the lowest effect size of  $g = 0.62$  (0.24-1.01) and the highest effect size of  $g = 1.03$  (0.63-1.43). A large variation from no effect to very large effects as indicated by a range of  $g$  from  $-0.43$  to  $4.05$  was found. Also, large confidence intervals across studies were observed. The heterogeneity analyses showed that the included studies were highly heterogeneous suggesting that the overall effect sizes may not be a precise estimate. Therefore, the generally large dispersion of effect sizes across studies should be emphasized instead of an exclusive focus on the overall summary outcome. This is highlighted in the presentation of the lowest and highest magnitudes of effect sizes. The high degree of heterogeneity across studies is not unexpected taking the differences in applied experimental designs and placebo manipulations into account.

The large variability in the magnitude of placebo effects does not seem to be explained by differences between healthy subjects and patients. This is not surprising as the patients in the two studies only experienced mild pain or pain that did not require pain medication. Nor did the magnitude appear to vary as a function of pain stimuli, agent, pain measure, blindness procedure, or time of pain measurement. Due to the low number of included studies, no statistical test was performed on the potentially moderating effects of these factors. However, based on a thorough look at the data, there were no indications that the magnitude varied systematically according to these factors and hence, they do not seem to predict the size of the placebo effect across the included pain studies. Nonetheless, future studies of these aspects including formal tests are warranted.

### 4.2. The magnitude of placebo effects as a function of how the effects were induced

The magnitude of placebo effects varied according to how the effects were induced. Our results showed that verbal suggestions of pain increase alone lead to a moderate to large placebo effect (lowest  $g = 0.64$  ( $-0.25$ - $1.53$ ) and highest  $g = 0.87$  ( $0.40$ - $1.34$ )) and a larger placebo effect when verbal suggestions were combined with conditioning (lowest  $g = 0.76$  ( $0.39$ - $1.14$ ) and highest  $g = 1.17$  ( $0.52$ - $1.81$ )). Even though the magnitudes differed according to how the effects were induced, the difference was not statistically significant. Still, it is important to note that verbal suggestions without conditioning appear to be able to produce moderate to large placebo effects, so the presence of a placebo effect may not depend on a preconditioning procedure. However, prior experience may enhance the magnitude of the placebo effect. One of the included studies [21] investigated whether the

number of learning trials influenced the persistence of nocebo effects. It was found that several trials of conditioning induced a more robust and persistent nocebo effect ( $g = 0.79$  (0.46-1.13)) than one trial of conditioning ( $g = 0.51$  (0.20-0.82)). Thus, prior negative experience with treatments may have an effect on how subsequent treatments are perceived as recently illustrated by Kessner and colleagues [x]. Only 1 study [65] applied social observation and no verbal suggestions for pain increase, resulting in a small to medium effect size. It is, however, preliminary to evaluate the influence of social observation on the magnitude of nocebo effects based on a single study.

### 4.3. Nocebo effects and placebo effects

As the applied conceptualization of the nocebo effect is analogous to the general conceptualization of the placebo effect and as similar selection criteria were used in one of the meta-analyses on the magnitude of placebo effects [63], it is possible to discuss the magnitude of nocebo effects and placebo effects. In the meta-analysis by Vase and colleagues including 24 studies published between 2002 and 2007, the average weighted effect size was  $d = 0.81$  (range = 0.12-2.51). This is comparable to the magnitudes of nocebo effects (i.e., lowest  $d = 0.65$  (0.24-1.05) and highest  $d = 1.07$  (0.65-1.48)). Thus, the overall magnitudes of nocebo and placebo effects appear to be roughly similar.

In a previous meta-analyses on placebo effects [62], placebo effects induced via verbal suggestions had an average effect size of  $d = 0.85$ , whereas in studies where verbal suggestions and conditioning were combined, the average effect size was  $d = 1.45$ . These findings are also comparable to the present findings on nocebo effects (verbal suggestions alone ( $d = 0.90$ ) and verbal suggestions combined with conditioning  $d = 1.22$ ). The comparable magnitudes for nocebo and placebo effects could support the hypothesis that similar mechanisms are involved in the opposite effects. For example, the healthy subject or patient perceives the agent as more powerful when the intervention involves both verbal suggestions and conditioning as opposed to verbal suggestions alone. However, further experimental studies are warranted to clarify the specific mechanisms involved in the effects. Overall, the findings from meta-analyses on placebo and nocebo effects suggest that roughly speaking it may be equally easy or difficult to obtain nocebo effects and placebo effects.

### 4.4. Nocebo effects in clinical research and practice

Compared with placebo research, little attention has been directed to clinical studies of nocebo effects and their implications for clinical practice [1]. The nocebo effects analyzed in this meta-analysis represent the effects which can be attributed to a nocebo treatment (i.e., an inert treatment along with verbal suggestions of pain increase) compared with a no-treatment group or condition. It is important to differentiate these effects from the *apparent nocebo effects* observed as adverse side effects in the placebo group in a randomized, controlled clinical trial [3,22] or in clinical practice. For example, Amanzio and colleagues [2] have investigated apparent nocebo effects in the placebo groups of randomized, controlled trials with anti-migraine drugs. In this analysis, all patients received inactive placebo treatments, but there was a high rate of adverse side effects and, interestingly, the side effects matched the side effects in the active treatment groups. This finding suggests



that simple verbal suggestions of potential side effects during the informed consent process may in itself lead to the experience of aversive side effects.

Information about potential side effects or verbal suggestions of pain increase may induce apparent nocebo effects, not only in relation to administration of inert treatments, but also in relation to an active pain treatment. Bingel et al. [12] exposed healthy volunteers to experimental pain stimuli and gave them an active pain-reducing medication and told them that the medication would worsen their pain when the infusion of medication ceased. Interestingly, the participants experienced a pain increase to the extent that the effect of the active treatment was abolished. Thus, the influence of verbal suggestions for pain worsening may be so potent that they outmatch the effect of active pain-reducing medication.

#### 4.5. Conclusions

The findings from the present meta-analysis and the related studies on apparent nocebo effects may have important implications for clinical practice. The moderate to large but highly variable magnitude of nocebo effects produced by verbal suggestions alone demonstrates the importance of how information should be framed in order to minimize the harm of nocebo effects. Overall, the results of this meta-analysis and the findings from the previous meta-analyses of placebo effects suggest that roughly speaking it may be equally easy or difficult to obtain nocebo and placebo effects. Similar to placebo effects, nocebo effects have been shown to be especially large when verbal suggestions (of increased pain) are combined with conditioning. Therefore, it is likely that the efficacy of future pain treatments may be enhanced if *both* positive and negative experiences with treatments are addressed in pain patients.

#### Acknowledgments

We wish to thank Professor Michael Væth for helpful discussions of statistical issues and all the authors of the included studies who have taken the time to provide us with their raw data.

This work is part of the Europain Collaboration and funded by the Innovative Medicines Initiative Joint Undertaking (IMI JU) Grant No 115007, resources which are composed of financial contribution from the European Union's Seventh Framework Program (FP7/2007-2013) and EFPIA companies in kind contribution ([www.imi.europa.eu](http://www.imi.europa.eu)). It was also supported by the MINDLab UNIK initiative at Aarhus University, which is funded by the Danish Ministry of Science, Technology and Innovation.

#### Appendix A

The listed studies gave rise to special considerations in relation to the selection criteria. For transparency, the reasons for exclusion are specified below.

Exclusion due to selection criterion 1: the purpose of the study (not in relation to pain)	
Chooi et al., 2010 [15]	The influence of negative words on pain following caesarean section
De la Cruz et al., 2010 [23]	Nocebo effects in relation to fatigue in cancer
Drici et al., 1995 [24]	Nocebo effects in relation to personality

Keitel et al., 2013 [40]	Nocebo effects in relation to Parkinson's disease
Klosterhalfen et al., 2009 [43]	Nocebo effects in relation to rotation procedure
Levine et al., 2006 [45]	Nocebo effects in relation to gastric symptoms and nausea
Link et al., 2006 [47]	Nocebo effects in relation to cognitive performance
Pollo et al., 2012 [53]	Nocebo effects in relation to motor training
Stovner et al., 2008 [57]	Not purpose to investigate nocebo effects (nocebo-related interpretation of mobile phone headache)
Swider & Babel, 2013 [58]	Purpose to investigate placebo effects in relation to pain but finding of nocebo effect; no inert agent administered (criterion 2)
Vernia et al., 2010 [64]	Nocebo effects in relation to lactose intolerance
<b>Exclusion due to selection criterion 1: the purpose of the study (side effects of drug or placebo)</b>	
Beedie et al., 2007 [5]	Nocebo effects as side effects in relation to sport
Flaten et al., 1999 [28]	Nocebo effects as side effects of drugs
Liccardi et al., 2004 [46]	Nocebo effects as side effects of drugs
Lombardi et al., 2008 [48]	Nocebo effects as side effects of drugs
Manchikanti et al., 2005 [50]	Nocebo effects as side effects of drugs
Scott et al., 2008 [56]	Nocebo effects as side effects of a placebo intervention
<b>Exclusion due to selection criterion 2: nocebo treatment</b>	
Bjørkedal & Flaten, 2012 [13]	No inert agent administered (conditioned pain modulation)
Jensen et al., 2012 [38]	No inert agent administered (face cues)
Lorenz et al., 2005 [49]	No inert agent administered
Nicolodi & Torrini, 2009 [52]	No inert agent administered (nocebo in relation to headache; colored cues)
Van Laarhoven et al., 2012 [59]	No inert agent administered
Varelmann et al., 2010 [60]	No inert agent administered

## Appendix B

### Specific considerations in relation to study design

The study by Kong and colleagues [44] used a baseline and test measurement on both a control and test area, but only data for the test area for baseline and test measurements were entered into the systematic. The study by Elsenbruch and colleagues [25] used a baseline and test measurement on separate days for both the nocebo-treated group and the no-treatment group, but only data for the test measurements for both groups were entered into the systematic review. The study by Vögtle et al. [65] used a measurement of pain with and without ointment for both the nocebo-treated group and the no-treatment group. Only data for pain with ointment for both groups were entered into the meta-analysis. These adjustments facilitated more precise comparisons with the other studies.

### **Specific considerations in relation to pain stimuli and measures**

In one study, the same participant was exposed to more than 1 painful stimulus [61], which led to the report of several nocebo effects. In this study, each irritable bowel syndrome patient was exposed to 2 different pain stimuli (evoked rectal distention and heat pain i.e. immersion of foot into hot water). Only the nocebo effect on rectal distention was entered into the meta-analysis, as this represents the most clinically relevant type of pain for this type of participants. In one of the studies by Colloca and colleagues [21], groups 1-4 were exposed to non-painful stimuli (low and high tactile stimuli) and not included in the meta-analysis as only painful stimuli were accepted. In one of the studies by Benedetti and colleagues [11], only tolerance measures were reported and these measures were entered into the meta-analysis.

### **Specific considerations in relation to how the nocebo effects were induced**

Across the studies, nocebo effects were induced by verbal suggestions of pain increase and sometimes in combination with a classical conditioning procedure. In one study [26], the verbal suggestions did not relate to the agent, but to the sensitization of pain over time. In another study [65], the verbal suggestion was related to sensitization of the skin, not in patients with pain, but with sexual dysfunction. Both studies were accepted for inclusion because the conceptualization of nocebo and the design were in agreement with previous experimental nocebo studies. In two studies [21,65], nocebo effects were induced both by verbal suggestions alone in one group and by verbal suggestions combined with conditioning or via social observation in another group. In the calculation of the overall magnitude of the nocebo effect, both groups were entered into the analysis, whereas the groups were analyzed separately in the calculation of the lowest and highest effect size as a function of how the nocebo effect was induced. In two studies by Benedetti and colleagues, group 5 [11] and groups 3 and 4 [8] were not included in the meta-analysis because they involved conditioning with pharmacological agents (ketorolac, 2003 [11] and proglumide [8] interfering with the development of the nocebo effect.

### **Specific considerations in relation to pain measures over time**

In some studies, the nocebo effect was measured over time in a certain interval (e.g., every 5 min for 30 min). In Johansen and colleagues [39], pain was measured over time at 5, 10, 15, and 20 minutes, but after 15 minutes only 6 out of 20 subjects rated their pain. Thus, the pain measures at 5 (pain reports of 17 subjects) and 10 minutes (pain reports of 13 subjects) were only entered into the meta-analysis. Also, some of the participants dropped out over time: 3 participants dropped out after 5 minutes in both the no-treatment group and the nocebo-treated group and 7 participants in the no-treatment group and 6 participants in the nocebo-treated group dropped out after 10 minutes. Thus, *N* is different for the two time measurements.

## Specific considerations in relation to calculation of nocebo effects

The study by Benedetti and colleagues from 2006 [9] only reported means and standard deviations for the last pain ratings, but we were interested in the means and standard deviations for each pain rating in order to calculate the lowest, highest, and average effect sizes of the nocebo effect. As we did not have access to raw data, we read the means and standard deviations for each pain rating of Figs. 2A and 3A in the article. However, when we compared the last pain rating reported in the article with our readouts, the standard deviations were different. Therefore, we adjusted all our standard deviations with 0.24 in order to match our calculations with the accurate data.

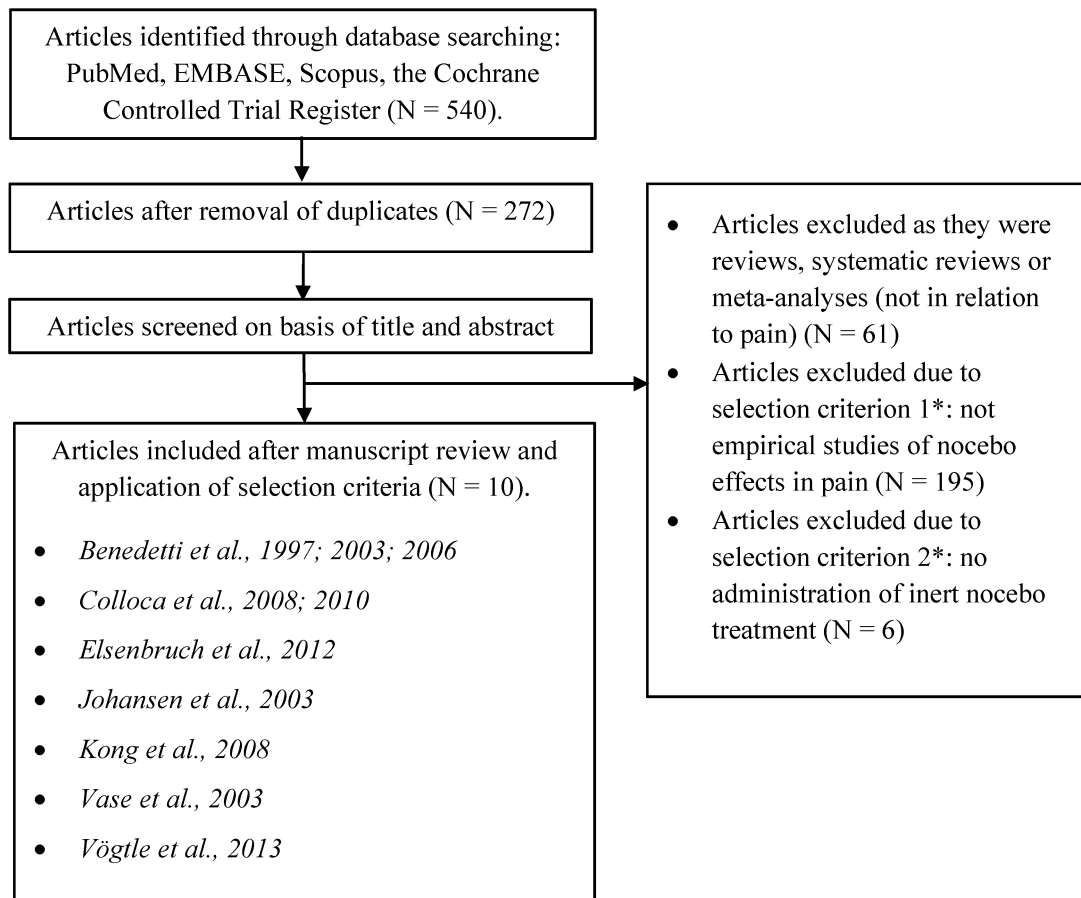
## References

- [1]. Amanzio M. Do we need a new procedure for the assessment of adverse events in anti-migraine clinical trials. *Recent Pat CNS Drug Disc.* 2011; 6:41–7.
- [2]. Amanzio M, Corazzini LL, Vase L, Benedetti F. A systematic review of adverse effects in placebo groups of anti-migraine clinical trials. *Pain.* 2009; 146:261–9. [PubMed: 19781854]
- [3]. Barsky AJ, Saintfort R, Rogers MP, Borus JF. Nonspecific medication side effects and the nocebo phenomenon. *JAMA.* 2002; 287:622–7. [PubMed: 11829702]
- [4]. Beecher HK. The powerful placebo. *J Am Med Assoc.* 1955; 159(17):1602–6. [PubMed: 13271123]
- [5]. Beedie CJ, Coleman DA, Foad AJ. Positive and negative placebo effects resulting from the deceptive administration of an ergogenic aid. *Int J Sport Nutr Exerc Metab.* 2007; 17(3):259–69. [PubMed: 17693687]
- [6]. Benedetti F. Mechanisms of placebo and placebo-related effects across diseases and treatments. *Annu Rev Pharmacol Toxicol.* 2008; 48:33–60. [PubMed: 17666008]
- [7]. Benedetti, F. Understanding the mechanisms in health and disease. Oxford University Press; Oxford: 2009. Placebo effects.
- [8]. Benedetti F, Amanzio M, Casadio O, Oliaro A, Maggi G. Blockade of nocebo hyperalgesia by the cholecystokinin antagonist proglumide. *Pain.* 1997; 71:135–40. [PubMed: 9211474]
- [9]. Benedetti F, Amanzio M, Vighetti S, Asteggiano G. The biochemical and neuroendocrine basis of the hyperalgesic nocebo effect. *J Neurosci.* 2006; 26:12014–22. [PubMed: 17108175]
- [10]. Benedetti F, Lanotte M, Lopiano L, Colloca L. When words are painful: unraveling the mechanisms of the nocebo effect. *Neuroscience.* 2007; 147:260–71. [PubMed: 17379417]
- [11]. 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; 23(10):4315–23. [PubMed: 12764120]
- [12]. Bingel U, Wanigasekera V, Wiech K, Mhuircheartaigh RN, Lee MC, Ploner M, Tracey I. The effect of treatment expectation on drug efficacy: imaging the analgesic benefit of the opioid remifentanyl. *Sci Trans Med.* 2011; 3:70ra14.
- [13]. Bjørkedal E, Flaten MA. Expectations of increased and decreased pain explain the effect of conditioned pain modulation in females. *J Pain Res.* 2012; 5:289–300. [PubMed: 23049277]
- [14]. Borenstein, M.; Hedges, LV.; Higgins, JPT.; Rothstein, HR. Introduction to meta-analysis. John Wiley & Sons Ltd.; West Sussex, UK: 2009.
- [15]. Chooi CSL, Nerlekar R, Raju A, Cyna AM. Placebo vs. nocebo questioning for pain evaluation following caesarean section. *Anaesth Intens Care.* 2010; 38(4):762–3.
- [16]. Cochran WG. The combination of estimates from different experiments. *Biometrics.* 1954; 10:101–29.
- [17]. Cohen, J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed.. Lawrence Erlbaum Associates; New Jersey, USA: 1988.
- [18]. Colloca L, Benedetti F. Nocebo hyperalgesia: how anxiety is turned into pain. *Curr Opin Anaesthesiol.* 2007; 5:435–9. [PubMed: 17873596]

- [19]. Colloca L, Lopiano L, Lanotte M, Benedetti F. Overt versus covert treatment for pain, anxiety, and Parkinson's disease. *Lancet Neurol.* 2004; 3:679–84. [PubMed: 15488461]
- [20]. Colloca L, Sigauco M, Benedetti F. The role of learning in nocebo and placebo effects. *Pain.* 2008; 136:211–8. [PubMed: 18372113]
- [21]. Colloca L, Petrovic P, Wager TD, Ingvar M, Benedetti F. How the number of learning trials affects placebo and nocebo responses. *Pain.* 2010; 151:430–9. [PubMed: 20817355]
- [22]. Colloca L, Miller FG. The nocebo effect and its relevance for clinical practice. *Psychosom Med.* 2011; 73:598–603. [PubMed: 21862825]
- [23]. De La Cruz M, Hui D, Parsons HA, Bruera E. Placebo and nocebo effects in randomized double-blind clinical trials of agents for the therapy for fatigue in patients with advanced cancer. *Cancer.* 2010; 116(3):766–74. [PubMed: 19918921]
- [24]. Drici M-D, Raybaud F, De Lunardo C, Iacono P, Gustovic P. Influence of the behaviour pattern on the nocebo response of healthy volunteers. *Brit J Clin Pharmacol.* 1995; 39(2):204–6.
- [25]. Dworkin RH, Turk DC, Katz NP, Rowbotham MC, Peirce-Sandner S, Cerny I, Clingman CS, Eloff BC, Farrar JT, Kamp C, McDermott MP, Rappaport BA, Sanhai WR. Evidence-based clinical trial design for chronic pain pharmacotherapy: a blueprint for ACTION. *Pain.* 2011; 152:107–15.
- [26]. Elsenbruch S, Schmid J, Bäsler M, Ceso E, Schedlowski M, Benson S. How positive and negative expectations shape the experience of visceral pain: an experimental pilot study in healthy women. *Neurogastroenterol Motil.* 2012 doi: 10.1111/j.1365-2982.2012.01950.x.
- [27]. Fields HL, Levine JD. Pain-mechanics and management. *West J Med.* 1984; 141(3):347–57. [PubMed: 6209862]
- [28]. Flaten MA, Simonsen T, Olsen H. Drug-related information generates placebo and nocebo responses that modify the drug response. *Psychosom Med.* 1999; 61(2):250–5. [PubMed: 10204979]
- [29]. Hahn RA. The nocebo phenomenon: concept, evidence, and implications for public health. *Prev Med.* 1997; 26:607–11. [PubMed: 9327466]
- [30]. Hedges LV. Distribution theory for Glass' estimator of effect size and related estimators. *J Educ Stat.* 1981; 6:107–28.
- [31]. Hedges LV. Estimation of effect size from a series of independent experiments. *Psychol Bull.* 1982; 92:490–9.
- [32]. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med.* 2002; 21:1539–58. [PubMed: 12111919]
- [33]. Hróbjartsson A. The uncontrollable placebo effect. *Eur J Clin Pharmacol.* 1996; 50:345–8. [PubMed: 8839653]
- [34]. Hróbjartsson, A. Placeboeffekten i lægevidenskab og klinik. In: Andersen, OK.; Claësson, MH.; Hróbjartsson, A.; Sørensen, AN., editors. *Placebo. Historie, biologi, og effekt.* Akademisk Forlag A/S; 1997. p. 11-67.
- [35]. Hróbjartsson A, Gotzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *N Engl J Med.* 2001; 344(21):1594–1602. [PubMed: 11372012]
- [36]. Hróbjartsson A, Gotzsche PC. Placebo interventions for all clinical conditions. *Cochrane Database Syst Rev.* 2004; 3:CD003974. [PubMed: 15266510]
- [37]. Hróbjartsson A, Gotzsche PC. Placebo interventions for all clinical conditions. *Cochrane Database Syst Rev.* 2010; 1(1):CD003974. [PubMed: 20091554]
- [38]. Jensen KB, Kaptchuk TJ, Kirsch I, Raicek J, Lindstrom KM, Berna C, Gollub RL, Ingvar M, Kong J. Nonconscious activation of placebo and nocebo pain responses. *Proc Natl Acad Sci.* 2012; 109(39):15959–64. [PubMed: 23019380]
- [39]. Johansen O, Brox J, Flaten MA. Placebo and nocebo responses, cortisol, and circulating beta-endorphin. *Psychosom Med.* 2003; 65:786–90. [PubMed: 14508021]
- [40]. Keitel A, Wojtecki L, Hirschmann J, Hartmann CJ, Ferrea S, Südmeyer M, Schnitzler A. Motor and cognitive placebo-/nocebo responses in Parkinson's disease patients with deep brain stimulation. *Behav Brain Res.* 2013; 250:199–205. [PubMed: 23651878]
- [41]. Kennedy WP. The nocebo reaction. *Med World.* 1961; 95:203–5. [PubMed: 13752532]

- [42]. Kissel, P.; Barrucand, D. Placebos et effet placebo en médecine. Masson; Paris, France: 1964.
- [43]. Klosterhalfen S, Kellerman S, Brain S, Kowalski A, Schrauth M, Zipfel S, Enck P. Gender and the nocebo response following conditioning and expectancy. *J Psychosom Res.* 2009; 66(4):323–8. [PubMed: 19302890]
- [44]. Kong J, Gollub RL, Polich G, Kirsch I, LaViolette P, Vangel M, Rosen B, Kaptchuk TJ. A functional magnetic resonance imaging study on the neural mechanisms of hyperalgesic nocebo effect. *J Neurosci.* 2008; 28:13354–62. [PubMed: 19052227]
- [45]. Levine ME, Stern RM, Koch KL. The effects of manipulating expectations through placebo and nocebo administration on gastric tachyarrhythmia and motion-induced nausea. *Psychosom Med.* 2006; 68(3):478–86. [PubMed: 16738082]
- [46]. Liccardi G, Senna G, Russo M, Bonadonna P, Crivellaro M, Dama A, D'Amato M, Canonica GW, Passalacqua G. Evaluation of the nocebo effect during oral challenge in patients with adverse drug reactions. *J Invest Allergy and Clin Immun.* 2004; 14(2):104–7.
- [47]. Link J, Haggard R, Kelly K, Forrer D. Placebo/nocebo symptom reporting in a sham herbal supplement trial. *Eval Health Prof.* 2006; 29(4):394–406. [PubMed: 17102062]
- [48]. Lombardi C, Gargioni S, Canonica GW, Passalacqua G. The nocebo effect during oral challenge in subjects with adverse drug reactions. *Eur Ann Allergy Clin Immunol.* 2008; 40(4):138–41. [PubMed: 19227649]
- [49]. Lorenz J, Hauck M, Paur RC, Nakamura Y, Zimmermann R, Bromm B, Engel AK. Cortical correlates of false expectations during pain intensity judgments – a possible manifestation of placebo/nocebo cognitions. *Brain, Behav and Immun.* 2005; 19:283–95.
- [50]. Manchikanti L, Pampati V, Damron K. The role of placebo and nocebo effects of perioperative administration of sedatives and opioids in interventional pain management. *Pain Phys.* 2005; 8(4):349–55.
- [51]. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA group. Preferred reporting items for systematic reviews and meta-analysis: the PRISMA statement. *J Clin Epidemiol.* 2009; 62(10): 1006–12. [PubMed: 19631508]
- [52]. Nicolodi M, Torrini A. From nocebo effect to the hypothesis of nocebo and psychometric tests as entry criteria in headache trials. *Cephalalgia.* 2009; 29(SUPPL. 1):39.
- [53]. Pollo A, Carlino E, Vase L, Benedetti F. Preventing motor training through nocebo suggestions. *Eur J Appl Physiol.* 2012; 112(11):3893–3903. [PubMed: 22411454]
- [54]. Price DD, Riley JL 3rd, Vase L. Reliable differences in placebo effects between clinical analgesic trials and studies of placebo analgesia mechanisms. *Pain.* 2003; 104:715–6.
- [55]. Price DD, Finniss DG, Benedetti F. A comprehensive review of the placebo effect: recent advances and current thought. *Annu Rev Psychol.* 2008; 59:565–90. [PubMed: 17550344]
- [56]. Scott DJ, Stohler CS, Egnatuk CM, Wang H, Koeppel RA, Zubieta J-A. Placebo and nocebo effects are defined by opposite opioid and dopaminergic responses. *Arch Gen Psychiatry.* 2008; 65:220–31. [PubMed: 18250260]
- [57]. Stovner LJ, Oftedal G, Straume A, Johnsson A. Nocebo as headache trigger: Evidence from a sham-controlled provocation study with RF fields. *Acta Neurol Scand.* 2008; 117:67–71.
- [58]. Swider K, Babel P. The effect of the sex of a model on nocebo hyperalgesia induced by social observational learning. *Pain.* 2013 doi:pii: S0304-3959.
- [59]. Van Laarhoven AIM, Vogelaar MK, Wilder-Smith OH, van Riel PICM, van der Kerkhof PCM, Kraaimaat FW, Evers AWM. Induction of nocebo and placebo effects on itch and pain by verbal suggestions. *Pain.* 2011; 152:1486–94. [PubMed: 21353388]
- [60]. Varelmann D, Pancaro C, Cappiello EC, Camann WR. Nocebo-induced hyperalgesia during local anesthetic injection. *Anesth Analg.* 2010; 110(3):868–70. [PubMed: 20042440]
- [61]. Vase L, Robinson ME, Verne GN, Price DD. The contribution of suggestion, desire, and expectation to placebo effects in irritable bowel syndrome patients. An empirical investigation. *Pain.* 2003; 105:17–25. [PubMed: 14499416]
- [62]. Vase L, Riley JL III, Price DD. A comparison of placebo effects in clinical analgesic trials versus studies of placebo analgesia. *Pain.* 2002; 99:443–52. [PubMed: 12406519]

- [63]. Vase L, Petersen GL, Riley JL 3rd, Price DD. Factors contributing to large analgesic effects in placebo mechanism studies conducted between 2002 and 2007. *Pain*. 2009; 145:36–44. [PubMed: 19559529]
- [64]. Vernia P, Di Camillo M, Foglietta T, Avallone VE, De Carolis A. Diagnosis of lactose intolerance and the “nocebo” effect: the role of negative expectations. *Digest Liver Dis*. 2010; 42(9):616–9.
- [65]. Vögtle E, Barke A, Kröner-Herwig B. Nocebo hyperalgesia induced by social observational learning. *Pain*. 2013 doi: 10.1016/j.pain.2013.04.041.

**Fig. 1.**

Trial flow of the selection process of studies including number of events and reasons for exclusion.

\*See Appendix A for special considerations in relation to excluded articles.



**Table 1**

Characteristics of included studies, including title, number of participants, population, pain, design, agent, and framing of suggestions.

Study	N <sup>*1</sup>	Population	Pain	Design	Agent	Exact framing of suggestions
<b>Benedetti et al. (1997)</b>	36	Thoracic patients	Post-operative pain	Double-blind Randomized Between subjects <sup>#2</sup>	Open saline injection	“Told that it [the injection] produced a pain increase within 30 min”.
<b>Benedetti et al. (2003)</b>	26	Healthy volunteers	Tourniquet ischemic pain	Double-blind Randomized Between subjects	Saline injection	“Told that it was a drug that increased pain”.
<b>Benedetti et al. (2006)</b>	25	Healthy volunteers	Tourniquet ischemic pain	Double-blind Randomized Between subjects	Inert talc pill 5 min. before pain stimulation	“It was a powerful vasoconstrictor further increasing the tourniquet-induced ischemia [...] told that, because of the quick vasoconstriction, this would induce a faster and larger increase of pain intensity, so that a quite strong hyperalgesic effect should be expected [...] told that they could give up at any time”.
<b>Colloca et al. (2008)</b>	29	Healthy women	Electrical shock	Single-blind Randomized Within subjects	Sham electrode	“Told that a green light would anticipate a stimulus that was made painful by the stimulation of the middle finger [...] that the green light anticipated the activation of the electrode that in turn induced a hyperalgesic effect”.
<b>Colloca et al. (2010)</b>	46	Healthy volunteers	Electrical shock	Single-blind Pseudo-randomized Within subjects	Sham electrode	“Informed that the activation of electrodes attached to the ankle would [...] increase their [...] painful perception when a red light was displayed”. Control: “Told that a yellow light would indicate the deactivation of the ankle electrodes and thus that no treatment would be given”.
<b>Eisenbruch et al. (2012)</b>	32	Healthy women	Rectal distension	Single-blind Randomized Between subjects	Inert substance (sodium chloride)	“Informed that it [the injection] contained sodium chloride, but were also told that the aim of the study was to verify our previous observation that pain ratings increased significantly in the majority of subjects over time, i.e., that sensitization occurred, over the course of the 2-day experiment”.
<b>Johansen et al. (2003)</b>	34 <sup>#3</sup>	Healthy men	Tourniquet ischemic pain	Double-blind Randomized Between subjects	Saline injection	“Told that it was a substance that would increase the pain”.
<b>Kong et al. (2008)</b>	13	Healthy volunteers	Heat pain	Single-blind Pseudo-randomized Within subjects	Sham acupuncture	“After treatment we told subjects they would be receiving the same [increased] stimuli series administered before treatment”.
<b>Vase et al. (2003)</b>	13	Women with IBS	Rectal distension	Double-blind Randomized Within subjects	Sterile surgical lubricant/saline jelly	“The agent you have just been giving is known to significantly increase pain in some patients”.

Study	N <sup>*1</sup>	Population	Pain	Design	Agent	Exact framing of suggestions
Viggle et al. (2013)	80	Healthy volunteers	Pressure pain	Single-blind Randomized between subjects	Ointment	Verbal suggestion: "The ointment was designed to intensify the sensitivity of the skin in patients with sexual dysfunction". Social observation: "The influence of a nonverbal instruction on the experimental procedure was being tested".

\*1 N = only for the subjects included in the calculations. N may thus be different from the total N in some studies.

\*2 "Between subjects" and "within subjects" designs as reported in the study and included in the calculations.

\*3 Dropout of participants over time, After 5 min, 17 participants in each group rated their pain and after 10 min, 14 and 13 participants rated their pain.

**Table 2**

Magnitudes of placebo effects.

Study	Pain measure	Time of pain measurement	Cohen's <i>d</i> lowest (CI 95% <sup>*2</sup> )	Cohen's <i>d</i> highest (CI 95% <sup>*2</sup> )	Hedges' <i>g</i> lowest (CI 95% <sup>*2</sup> )	Hedges' <i>g</i> highest (CI 95% <sup>*2</sup> )
Benedetti et al. (1997)	NRS	Before the injection and after 30 min	2.10 (1.29-2.91)	2.10 (1.29-2.91)	2.05 (1.26-2.85)	2.05 (1.26-2.85)
Benedetti et al. (2003)	Tolerance <sup>*1</sup>	From last squeeze to unbearable pain	1.12 (0.29-1.95)	1.12 (0.29-1.95)	1.09 (0.28-1.89)	1.09 (0.28-1.89)
Benedetti et al. (2006)	NRS	Every minute for 10 min	1.40 (0.53-2.28)	4.19 (2.79-5.59)	1.35 (0.51-2.20)	4.05 (2.70-5.41)
Colloca et al. (2008)	NRS	After each stimulation	0.62 (-0.35-1.60)	1.13 (0.61-1.64)	0.59 (-0.33-1.50)	1.06 (0.57-1.55)
Colloca et al. (2010)	VAS	After each trial	0.53 (0.20-0.85)	0.82 (0.48-1.17)	0.51 (0.20-0.82)	0.79 (0.46-1.13)
Eisenbruch et al. (2012)	VAS	After each stimulation	0.11 (-0.59-0.80)	0.23 (-0.47-0.92)	0.11 (-0.57-0.78)	0.22 (-0.46-0.90)
Johansen et al. (2003)	VAS	5 and 10 min after pain had reached 7 on a VAS	-0.45 (-1.13-0.24)	0.59 (-0.18-1.36)	-0.43 (-1.1-0.23)	0.58 (-0.17-1.32)
Kong et al. (2008)	VAS	After each stimulation	1.03 (0.45-1.61)	1.03 (0.45-1.61)	0.96 (0.42-1.50)	0.96 (0.42-1.50)
Vase et al. (2003)	M-VAS	After each stimulation	-0.04 (-0.83-0.76)	0.48 (-0.26-1.23)	-0.03 (-0.78-0.71)	0.45 (-0.25-1.15)
Vigtle et al. (2013)	NRS	After each stimulation	0.36 (-0.19-0.90)	0.60 (0.05-1.14)	0.35 (-0.18-0.89)	0.59 (0.05-1.13)
<b>Total <i>d</i> and <i>g</i></b>			<b>0.65 (0.24-1.05)</b>	<b>1.07 (0.65-1.48)</b>	<b>0.62 (0.24-1.01)</b>	<b>1.03 (0.63-1.43)</b>

<sup>\*1</sup> Pain was measured via tolerance and thus the sign of the effect sizes was changed.

<sup>\*2</sup> CI = Confidence intervals.

<sup>\*3</sup> Effect sizes were weighted according to the number of subjects (*N*) in each study (Hedges' *g*) [30,31].

**Table 3**

Induction of placebo effects by verbal suggestions alone, verbal suggestions and conditioning or social observation.

Study	Pain measure	Time of pain measurement	Cohen's <i>d</i> lowest (CI 95%)*2	Cohen's <i>d</i> highest (CI 95%)*2	Hedges' <i>g</i> lowest (CI 95%)*2	Hedges' <i>g</i> highest (CI 95%)*2
<i>Verbal suggestions alone</i>						
Benedetti et al. (1997)	NRS	Before the injection and after 30 min	2.10 (1.29-2.91)	2.10 (1.29-2.91)	2.05 (1.26-2.85)	2.05 (1.26-2.85)
Benedetti et al. (2003)	Tolerance	*1 From last squeeze to unbearable pain	1.12 (0.29-1.95)	1.12 (0.29-1.95)	1.09 (0.28-1.89)	1.09 (0.28-1.89)
Colloca et al. (2008) <sup>*4</sup>	NRS	After each stimulation	0.62 (-0.35-1.60)	0.62 (-0.35-1.60)	0.59 (-0.33-1.50)	0.59 (-0.33-1.51)
Johansen et al. (2003)	VAS	5 and 10 min after pain had reached 7 on a VAS	-0.45 (-1.13-0.24)	0.59 (-0.18-1.36)	-0.43 (-1.1-0.23)	0.58 (-0.17-1.32)
Vase et al. (2003)	M-VAS	After each stimulation	-0.04 (-0.83-0.76)	0.48 (-0.26-1.23)	-0.03 (-0.78-0.71)	0.45 (-0.25-1.15)
Vögtle et al. (2013) <sup>*5</sup>	NRS	After each stimulation	0.60 (0.05-1.14)	0.60 (0.05-1.14)	0.59 (0.05-1.13)	0.59 (0.05-1.13)
<b>Average <i>d</i> and <i>g</i></b>						
			<b>0.66 (-0.26-1.59)</b>	<b>0.90 (0.42-1.38)</b>	<b>0.64 (-0.25-1.53)</b>	<b>0.87 (0.40-1.34)</b>
<i>Verbal suggestions and conditioning</i>						
Benedetti et al. (2006)	NRS	Every minute for 10 min	1.40 (0.53-2.28)	4.19 (2.79-5.59)	1.35 (0.51-2.20)	4.05 (2.70-5.41)
Colloca et al. (2008) <sup>*4</sup>	NRS	After each stimulation	1.13 (0.61-1.64)	1.13 (0.61-1.64)	1.06 (0.57-1.55)	1.06 (0.57-1.55)
Colloca et al. (2010)	VAS	After each trial	0.53 (0.20-0.85)	0.82 (0.48-1.17)	0.51 (0.20-0.82)	0.79 (0.46-1.13)
Elsenbruch et al. (2012)	VAS	After each stimulation	0.11 (-0.59-0.80)	0.23 (-0.47-0.92)	0.11 (-0.57-0.78)	0.22 (-0.46-0.90)
Kong et al. (2008)	VAS	After each stimulation	1.03 (0.45-1.61)	1.03 (0.45-1.61)	0.96 (0.42-1.50)	0.96 (0.42-1.50)
<b>Average <i>d</i> and <i>g</i></b>						
			<b>0.80 (0.41-1.20)</b>	<b>1.22 (0.55-1.90)</b>	<b>0.76 (0.39-1.14)</b>	<b>1.17 (0.52-1.81)</b>
<i>Social observation</i>						
Vögtle et al. (2013) <sup>*5</sup>	NRS	After each stimulation	0.36 (-0.19-0.90)	0.36 (-0.19-0.90)	0.35 (-0.18-0.89)	0.35 (-0.18-0.89)
<b>TOTAL <i>d</i> and <i>g</i></b>						
			<b>0.65 (0.35-0.95)</b>	<b>0.79 (0.47-1.10)</b>	<b>0.63 (0.34-0.92)</b>	<b>0.77 (0.46-1.07)</b>

\*<sup>1</sup> Pain was measured via tolerance and thus the sign of the effect sizes was changed.

\*<sup>2</sup> CI = confidence intervals.

\*<sup>3</sup> Effect sizes were weighted according to the number of subjects in each study (Hedges'  $g$ ) [30,31].

\*<sup>4</sup> Total  $d$  and  $g$  vary from the overall magnitude of placebo effects (Table 2) as both verbal suggestions alone and verbal suggestions and conditioning are measured in Colloca et al. [21] and therefore the study is represented in both categories.

\*<sup>5</sup> In Vögtle et al. [65], the lowest effect size is induced by means of social observation and thus the effect size is only represented once (i.e., in the category of social observation).