Placebo and nocebo effects in itch : from conditioning to psychophysiological effects

Meeuwis, S.H.

Chapter 6

Open- and closed-label placebo and nocebo suggestions about a sham transdermal patch

Submitted as:

Meeuwis, S.H., van Middendorp, H., Lavrijsen, A.P.M., Veldhuijzen, D.S., & Evers. A.W.M. Open- and closed-label placebo and nocebo suggestions about a sham transdermal patch: effects on itch. 2019

ABSTRACT

Objective. Accumulating evidence indicates that placebo effects may also occur when it is known that a placebo is given. It is not yet clear whether these open-label placebo effects are similar to those of concealed (i.e. closed-label) placebo effects for somatic symptoms such as itch or whether nocebo effects can be induced under open-label conditions.

Methods. Healthy volunteers (n=112) were randomized to I) an open-label positive suggestions group, II) a closed-label positive suggestions group, III) an open-label negative suggestions group, or IV) a closed-label negative suggestions group. Participants were told, as cover story, that a transdermal caffeine patch would be applied that positively influences cognitive abilities and, as a side effect, positively or negatively (depending on group allocation) influences itch. Participants in the open-label groups were given a rationale explaining placebo and nocebo effect mechanisms. Itch was induced at baseline and post-suggestions by histamine iontophoresis.

Results. In the positive suggestions groups, significantly lower itch was reported than in the negative suggestions groups for both open- and closed-label contexts (all $p \leq .008$, Cohen's $d \geq 0.47$). Self-rated skin response was rated as less severe following positive versus negative suggestions (all $p \leq .017$, Cohen's $d \geq 0.33$), but no effects on physical skin responses to histamine were found (all $p \geq .23$, Cohen's $d \leq 0.30$).

Conclusion. Itch can be reduced by positive compared to negative suggestions under both open- and closed-label conditions. These findings indicate that open-label suggestions may potentially be a tool to utilize placebo effects for self-reported outcomes in clinical practice, for example by explaining role of expectancy in treatment. It needs to be investigated further under which circumstances an open-label rationale may impact placebo and nocebo effects.

Trial registration. www.trialregister.nl; NTR7174

INTRODUCTION

Placebo effects are beneficial effects that are not attributable to active treatment components such as pharmacological substances [1, 2]. Instead, these effects emerge through expectations about treatment outcomes, that are shaped by information that is provided about a treatment, learning, and environmental and social cues such as a positive patient-clinician interaction [1, 3-5]. Nocebo effects (i.e. adverse treatment outcomes such as side effects, that can be attributed to negative outcome expectations) can be similarly shaped by these pathways [1,6]. Experimental studies have demonstrated that placebo and nocebo effects can be induced in itch [7-9], although some studies show mixed or limited evidence [10-14]. In fact, meta-analyses show that over 30% of symptom improvement in clinical trials for itch and allergic symptoms can be explained by the placebo effect [15,16]. Itch ranks as one of the 50 most common interdisciplinary symptoms which affects an estimated one-fifth of the population, and that it has a debilitating impact on quality of life while existing treatments show limited effects [17-19]. Therefore finding ways to enhance existing treatments for itch becomes increasingly important. Potentially, placebo and nocebo effects may be used to facilitate improvement of existing treatments for itch.

Most studies on placebo and nocebo effects for physical symptoms such as pain or itch have investigated concealed placebo or nocebo induction, in which participants were unaware of receiving a placebo (sham) treatment. Such an approach does not allow for an easy translation towards clinical practice, mostly due to ethical considerations [20]. In the past decade, accumulating evidence shows that placebo effects can also occur when patients are fully informed about receiving placebos. Studies have shown that providing an inert pill in combination with a rationale on how placebo effects can impact medical conditions can reduce symptoms of a variety of medical conditions, amongst which irritable bowel syndrome, low back pain, and symptoms of allergic rhinitis [21-30].

There is limited literature available on whether a non-deceptive (open-label) approach can induce placebo effects for itch specifically, or how these effects relate to concealed (closed-label) placebo effects. Likewise, while we do know that nocebo effects often present as side effects to active treatments (e.g., induced by reading the leaflet of a pharmacological substance) [31-33], not much is known about whether these effects can also be induced using an open-label approach. A single study investigated open-label and closed-label placebo and nocebo effects induced by verbal suggestions about sham cutaneous treatment, and found that both open-label and closed-label suggestions influenced itch after, but not

during, histamine application on the skin [34]. The current study builds on these previous findings and investigates whether positive and negative outcome expectations, induced by a novel suggestive framework (verbal suggestions regarding a transdermal caffeine patch, where positive or negative effects on itch were purported as side effects, provided with either an open-label context or a closed-label context), could influence self-reported itch during an experimental itch induction test using histamine. Secondary outcomes include self-rated and clinical (physical) skin responses to histamine as well as psychological outcomes such as wellbeing. We first examine differences between the combined positive and the combined negative suggestions groups, and next assess effects for open-label and closed-label contexts separately. We expect low itch following positive verbal suggestions compared to high itch following negative verbal suggestions for both open-label and closed-label contexts.

METHODS

The study was approved by the Medical Ethics Committee at the Leiden University Medical Center, The Netherlands (NL64502.058.17) and pre-registered in the Dutch Trial Register on May 6th 2018 (trial ID: NTR7174). The study was conducted in accordance with the Declaration of Helsinki. All participants provided written informed consent. Data for the study were collected between April 2018 and January 2019.

Participants

Healthy male and female volunteers were recruited through advertisements on sites of Leiden University and social media. Participants between 18 and 35 years old that had a good understanding of written and spoken Dutch were included. Exclusion criteria consisted of severe somatic or psychological morbidity (e.g., heart and lung diseases, Diagnostic and Statistical Manual fifth edition (DSM-V) psychiatric disorders); current chronic itch or pain; current use of analgesics, anti-inflammatory drugs, antihistamines, or antibiotics; recent vaccinations; pregnancy; and colour blindness. Participants were asked to refrain from caffeine or nicotine consumption and heavy meals 2h, exercising 12h, and alcohol and drugs 24h prior to participation in the study, which was verified at the start of their appointment.

Study design

A between-subjects, single-blinded, randomized controlled design was applied. Participants were allocated (by block-randomization (n=8/block), online random number generator: www.random.org, Dublin, Ireland) to I) an open-label positive verbal suggestions (VS), II) closed-label positive VS, III) open-label negative VS, or IV) closed-label negative VS group. Allocation was not concealed from the experimenter. Participants were invited to a single laboratory session at the faculty of Social and Behavioural Sciences, Leiden University, The Netherlands. Itch was induced at baseline and post-VS by histamine iontophoresis (see also **Figure 1**).

Materials and Measures

1. Verbal suggestions

The study was advertised as a study that investigated the effects of a transdermal caffeine patch on cognitive abilities and sensitivity to physical stimuli. As part of this cover story, cognitive tasks[3] were conducted before and following suggestions. Following baseline measurements, participants were told that (1) a caffeine-containing patch would be placed on their shoulder, (2) caffeine, like nicotine, can be delivered by this method, and (3) this would influence both cognitive abilities and sensitivity to physical stimuli such as itch. In the positive VS groups, the following suggestion was then given: "*Previous research has shown that itch decreases strongly after applying this patch for most people, i.e. about 95% of people. The caffeine makes your skin less sensitive to physical stimuli. As such, we expect that you will experience less itch, compared to the first test*". In the open-label groups, an additional explanation of the placebo effect was given that stressed the following points: (1) the patch actually did not contain caffeine, (2) the purpose of the study was to test the effects of such positive suggestions, (3) previous research has shown that suggestions can reduce itch, (4) these effects are due to bodily processes, as the brain responds to information about a treatment in the same manner as to the actual treatment, and (5) this may also work when people know that they receive a placebo. For the negative VS groups, positive words were replaced by negative words (i.e. 'more itch' instead of 'less itch', and

[3] Considering that the verbal suggestions were not directly aimed at manipulating the outcomes of the cognitive tasks (but that these were rather included as part of the cover story), the detailed methodology for these tasks, their outcome measures (including related outcomes, e.g., expectations) and their results can be found in the Supplementary Material.

'nocebo' instead of 'placebo'). A 10x10 cm hydrocolloid patch (Medeco B.V., Oud-Beijerland, the Netherlands) was then placed on the non-dominant shoulder.

2. Itch induction: histamine iontophoresis

Itch was induced experimentally by histamine iontophoresis (see Meeuwis, Van Middendorp [13] for detailed methodology). Briefly, itch was induced for 2.5 minutes on the volar side of the forearm. After 2.5 minutes, iontophoresis electrodes were removed, after which a 3-minutes follow-up period commenced. Baseline iontophoresis was conducted on the dominant forearm, and post-VS iontophoresis on the non-dominant forearm.

3. Outcome measures

3.1. Expected itch and expected patch efficacy for skin sensitivity

Prior to each itch induction, participants rated expected itch on a Numeric Rating Scale (NRS) from 0 ('no itch') to 10 ('worst imaginable itch'). In addition, participants rated (post-VS, but prior to iontophoresis) the extent to which they believed the patch would influence skin sensitivity during the itch induction test on a NRS (0 'no effect', 10 'very effective').

3.2. Self-rated itch

Self-rated itch was assessed every 30 seconds during both iontophoresis tests and their follow-up period, using the same NRS as described in **section 3.1.** Participants were asked to rate mean itch experienced during iontophoresis (the primary study outcome) immediately upon removal of the iontophoresis electrodes. Correlations between mean itch assessed following iontophoresis and itch scores assessed every 30 seconds during iontophoresis were calculated to assess reliability of the primary outcome: self-rated mean itch (as assessed directly following the test) was significantly associated with all other itch measurements during iontophoresis for both baseline and post-VS measurements (all $r \geq .35$, all $p < .001$).

3.3. Self-rated and clinical skin response to histamine

As a measure of self-rated skin response, participants were asked to fill in a version of the Sensitive Scale-10 (SS-10) questionnaire [35] that was adjusted for use with histamine iontophoresis (see also [13]). In the current study, Cronbach's alphas for the post-iontophoresis SS-10 were .85 and .86, respectively. Wheal size and flare response to histamine were assessed following both iontophoresis tests by tracing the outer edges on a transparent, 1 cm^2-gridded sheet. Images were uploaded and retraced in ImageJ [36], and wheal and flare areas were calculated (in cm^2). In addition, skin temperature measurements were taken with a handheld infrared digital thermometer pre- and post-iontophoresis. Rise in skin temperature due to iontophoresis (Δ-temperature) was calculated as an outcome measure by subtracting the pre- from post-iontophoresis measurements.

3.4. Wellbeing: the Positive and Negative Affect Schedule

To assess the effects of suggestions on wellbeing, participants filled out the Positive and Negative Affect Schedule (PANAS [37]) at four moments during the laboratory session (see **Figure 1**). In the current study, Cronbach's alpha ranged .88 – .91 for the PANAS positive affect (PA) scale. Considering the scores on negative affect were very low at all measurement points (M_{range} = 11.49–12.32; with variances between 4.09 – 9.60, while the scale ranges from 0–50), group differences for this scale were not analysed. Two additional scales for wellbeing were assessed at the same moments as the PANAS, and are discussed in the **Supplementary Material**.

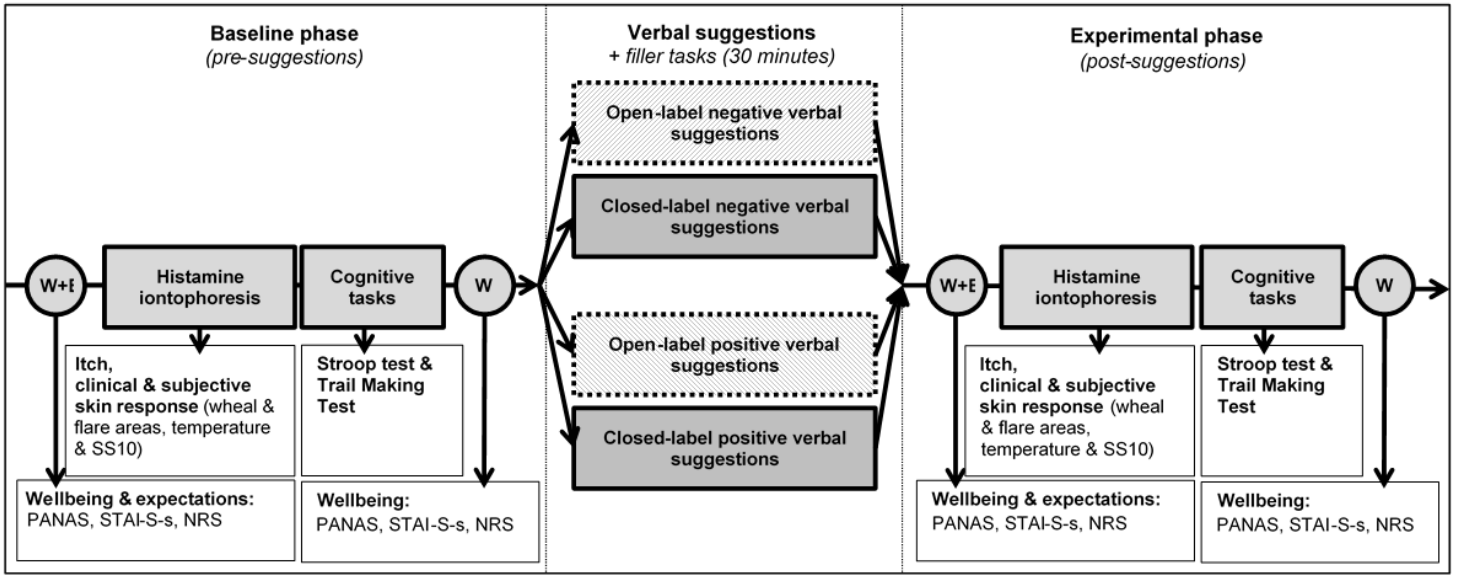


Figure 1. Overview of study design and measurement schedule for the laboratory session. W = wellbeing; E = expectations; PANAS = Positive and Negative Affect Schedule [37]; STAI-S-s = Spielberger State Trait Anxiety Inventory, State anxiety short scale [48]; NRS = Numeric Rating Scale; SS10 = Sensitive Scale-10 [35]; Stroop test, [49]; Trail Making Test [50,51].

Procedure

Prior to participation, volunteers filled out an online screening questionnaire. Eligible volunteers were invited for a single 2-hours laboratory session at the research site of the Social and Behavioural Sciences Department, Leiden University, The Netherlands. Upon arrival, the general procedures were explained and participants provided written informed consent (for the online screening questionnaire, separate online consent was given). Briefly, the in- and exclusion criteria were checked and adherence to lifestyle rules was verified. Next, the baseline phase started and participants filled out questionnaires for wellbeing and expectations. Demographics and personality factors were assessed (the latter were not related to the current study purpose and will be reported elsewhere). Histamine iontophoresis was conducted on the dominant arm, during and following which participants rated itch. Clinical and self-rated skin responses were assessed, followed by cognitive tests, and assessment of wellbeing. Verbal suggestions were given (depending on group allocation) and the inert patch was placed on the participant's shoulder. Participants were asked to perform some neutral filler tasks (i.e., Sudoku's, word & picture search puzzles) while the experimenter left the room, with a twofold purpose: 1) so that carry-over effects in itch could be minimized, an 2) so that the cover story of testing effects of the patch on cognitive tasks could be further reinforced. Thirty minutes after the baseline phase ended, the experimenter returned, and wellbeing and expectations were assessed. Histamine iontophoresis was conducted on the non-dominant forearm, followed by the cognitive tests and wellbeing questionnaires. Finally, participants filled out a closing questionnaire. They were debriefed on the true purpose of the study (in the open-label groups, the study purpose was reconfirmed) by the experimenter. Participants received a compensation of €20,- for the laboratory session.

Statistical analysis

Power analysis was conducted in G*Power [38] to determine the optimal sample size for detecting between-group differences in mean itch, controlled for baseline. The estimated effect size was based on a meta-analysis of open-label placebo [39], which found an average effect of d=0.88 for open-label placebo effect induction in patient samples, compared to a no-treatment control group. As the current study investigated effects in healthy volunteers rather than patients, a more conservative effect size of d=0.78 was used. An a priori power analysis for analysis of covariance (ANCOVA), with α=.05 and β=.80, indicated that, taking into account an additional 5% missing data rate, 28 participants per

group were needed to detect differences between the positive and negative verbal suggestion groups (for separate analysis of open-label and closed-label contexts).
All analyses were conducted in SPSS 23.0 for Windows (IBM SPSS Inc., Chicago, Illinois, US) with an alpha level of α=.05. Normal distribution of the variables, baseline differences, and assumptions were checked prior to data analysis. As was a priori determined, open-label and closed-label groups were first combined to detect differences between the effects of positive verbal suggestions and negative verbal suggestions and to increase power for these analyses. General linear model (GLM) analyses of covariance (ANCOVAs) were conducted for each outcome measure of itch and self-rated and clinical skin response, in which baseline measures were controlled. Within-group baseline-to-post-VS change was explored for each group by paired-sample t-tests (Bonferroni corrected: α/2=.025) to assess impact of each type of verbal suggestions on itch, and self-rated and clinical skin response. Effects of group on wellbeing were explored by mixed between-within repeated measures ANOVA. For itch expectations, GLM ANOVA was used. As an effect size, Cohen's d was calculated from (covariate adjusted) group means and SD's, with the following categories for interpretations: 0.2 small effect, 0.5 medium effect, 0.8 large effect [40]. All analyses were repeated for the separate open-label groups, and the separate closed-label groups. For these secondary analyses, a Bonferroni correction for multiple comparisons was applied (α/2=.025 for ANCOVA and (α/2)/2=.0125 for further within-group t-tests). Data of one participant was excluded from the analyses, as technical issues with the iontophoresis device prevented a baseline measurement of itch. Group means are described as Mean±SD, unless stated otherwise.

RESULTS

Participants

In total, 236 potential participants expressed interest in the study, of whom 79 volunteers refrained from participating for reasons unknown (e.g., no response following invitation), and of whom 43 were excluded (30 for somatic and/or psychological conditions, 7 for medication use, and 6 for having trouble understanding Dutch). Two participants dropped out during the laboratory session, resulting in a final sample of 112 participants (16.1% male) aged between 18 and 31 years old (M_{age}=21.88±2.77). No group differences were found for demographic factors, baseline itch expectation and baseline iontophoresis

outcome parameters for either the combined open- and closed-label groups (see **Table 1**, all $p \geq .16$) or separate groups (see **Supplementary Table E1**; all $p \geq .13$).

Expected itch and expected patch efficacy for skin sensitivity

Expected itch during iontophoresis was significantly lower following suggestions in the combined positive VS groups (M = 4.00±1.87) compared to the combined negative VS groups (M = 5.69±2.16); F(1,109)=19.23, $p<.001$, Cohen's d=0.84. When analyses were repeated for open-label and closed-label contexts separately, group differences in the same direction as for the combined groups were found, with larger effect sizes found for the open-label rather than closed-label context (open-label: F(1,53)=15.00, $p<.001$, Cohen's d=1.04; closed-label: F(1,54)=6.67, $p=.013$, Cohen's d=0.69; see **Figure 2A and B**). Expected patch efficacy for skin sensitivity was somewhat lower in the combined positive VS groups (M=3.43±2.11) compared to the combined negative VS groups (M=4.28±2.55), however, effects were marginal and small; F(1,109)=3.64, $p=.059$, Cohen's d=0.36. When groups were separated for open-label and closed-label context, no differences were found (both $p \geq .13$; see **Figure 2C and D**).

Self-rated mean itch

Self-rated mean itch during iontophoresis was significantly lower in the combined positive VS groups (M=3.29±1.53) compared to the combined negative VS groups (M=4.21±1.96); $F(1,108)=17.14$, $p<.001$, Cohen's d=0.51. Similar group differences were found when analyses were repeated for open-label and closed-label contexts separately, with medium-sized differences for the closed-label context (F(1,53)=9.02, $p=.004$, Cohen's d=0.54), and small-to-medium-sized differences for the open-label context (F(1,52)=7.62, $p=.008$, Cohen's d=0.47; see **Figure 3A and B)**. Within-group analysis of baseline-to-post-VS-change for itch indicated that mean itch reduced significantly following positive VS (both combined and separate groups: all $p \leq .007$), while it did not change in the negative VS groups (all $p \geq .22$) (see **Table 2** for the combined-groups analyses, and **Supplementary Table E2** for the separate-groups analyses).

Table 1. Means ± standard deviations, and analysis of (co)variance (AN(C)OVA) outcomes for the combined open- and closed-label positive verbal suggestions (VS) groups and the combined open- and closed-label negative VS groups.

	Combined open- and closed-label contexts			
			AN(C)OVA	
	Positive VS (*n*=55)	**Negative VS (*n*=56)**	***p*-value**	**Cohen's d**
Demographics				
Sex [male: n (%)]	8 (14.55)	10 (17.86)	.64	
Age	21.89 ± 2.49	21.93 ± 3.02	.94	0.01
Baseline histamine iontophoresis				
Mean itch	3.98 ± 1.43	4.00 ± 1.73	.94	0.01
Self-rated skin response (SS-10) [a]	30.92 ± 13.26	29.79 ± 12.90	.65	0.09
Wheal area [cm^2]	8.92 ± 3.38	9.28 ± 3.87	.61	0.10
Flare area [cm^2]	43.36 ± 15.70	42.17 ± 13.01	.19	0.08
Change in skin temperature [°C] [b, c]	1.39 ± 1.15	1.68 ± 1.00	.16	0.27
Post-VS expectation outcomes for itch				
Expected itch	4.00 ± 1.87	5.69 ± 2.16	< .001	0.84
Expected patch effectiveness for skin sensitivity	3.43 ± 2.11	4.28 ± 2.55	.059	0.36
Post-VS histamine iontophoresis				
Mean itch	3.29 ± 1.53	4.21 ± 1.96	< .001	0.51
Self-rated skin response (SS-10) [a, d]	23.60 ± 11.88	27.56 ± 12.71	< .001	0.39
Wheal area [cm^2]	8.19 ± 3.18	7.92 ± 3.42	.24	0.15
Flare area [cm^2]	41.66 ± 13.33	41.71 ± 13.82	.65	0.05
Change in skin temperature [°C] [c, e]	1.20 ± 1.19	1.20 ± 1.09	.39	0.14

Note (**Table 1**). [a] Misery et al. [35]. [b] n=1 missing due to technical difficulties with the infrared thermometer. [c] calculated as post-iontophoresis temperature – pre-iontophoresis temperature. [d] n=1 missing on the post-VS SS-10. [e] n=2 missing due to technical difficulties with the infrared thermometer.

Clinical and self-rated skin response to histamine

Participants in the combined positive VS groups rated their skin response as less severe compared to the combined negative VS groups, as indicated by small-to-medium-sized significantly lower scores on the SS-10 in the positive VS groups (M=23.60±11.88) compared to the negative VS groups (M=27.56±12.72); $F(1,107)=13.58$, $p<.001$, Cohen's d=0.39. When open-label and closed-label contexts were separated, similar group differences were found, with somewhat larger effects found for the closed-label context (closed-label: $F(1,52)=7.23$, $p=.010$, Cohen's d=0.45; open-label: $F(1,52)=6.09$, $p=.017$, Cohen's d=0.33; see **Supplementary Table E1**). No differences were found for clinical skin response outcomes of wheal and flare area, or skin temperature change between the combined positive and combined negative VS groups (all $p\geq.24$) or between the separate open- and closed-label groups (all $p\geq.23$). An overview of the within-group baseline-to-

post-VS-change for each variable is provided in **Table 2** (combined groups) and **Supplementary Table E2** (separate groups). In short, no significant within-group changes were found for clinical skin response in the combined groups ($p \geq .063$), except for wheal area and skin temperature change in the combined negative VS groups, which decreased significantly from baseline (both $p \leq .001$). When open-label and closed-label contexts were separated, similar decreases were demonstrated in the negative VS groups ($p \leq .009$), except for change in skin temperature in the open-label context ($p = .071$).

Wellbeing: Positive Affect (PA)

No effect of the combined-groups x time interaction on PA was found ($p = .81$), indicating that verbal suggestions did not influence affect during the laboratory session. No main effect of group was found ($p = .51$), but PA changed significantly over time ($p < .001$, see **Supplementary Figure S1**). Post-hoc Bonferroni tests indicated that PA following baseline iontophoresis was significantly higher compared to all other measurements (all $p < .002$), and that other measurement moments did not differ significantly over time (all $p > .99$). Next, analyses were separated for open-label and closed-label contexts. In the open-label context, PA following baseline iontophoresis was higher compared to the two subsequent measurements (all $p < .001$), whereas in the closed-label context, PA following baseline iontophoresis was higher compared to the first and third (post-VS) measurements (all $p \leq .017$; see also **Supplementary Figure S1**). Results for two additional wellbeing scales are discussed in the **Supplementary Material**.

Table 2. Within-group baseline-to-post-verbal suggestions (VS) changes on histamine iontophoresis outcomes for the combined open- and closed-label positive VS groups and the combined negative VS groups.

	Combined open- and closed- label positive VS groups (*n*=55)				**Combined open- and closed-label negative VS groups (*n*=56)**			
	n	Mean change	*t*	*p*	*n*	Mean change	*t*	*p*
Mean itch	55	-0.68	4.97	< .001	56	0.22	-1.24	.22
Self-rated skin response (SS-10) [a]	55	-7.32	6.92	< .001	55	-2.29	2.48	.016
Wheal area [cm^2]	55	-0.73	1.90	.063	56	-1.35	4.64	< .001
Flare area [cm^2]	55	-1.25	1.14	.26	56	-0.46	0.44	.66
Change in skin temperature [°C] [b, c]	54	-0.19	1.18	.25	56	-0.47	3.46	.001

Note (**Table 2**). Mean change was calculated as post-verbal suggestions score – baseline score, with negative values indicating a decrease from baseline, and positive scores indicating an increase from baseline. [a] Misery et al. [35]. [b] n=2 missing due to technical difficulties with the infrared thermometer. [c] calculated as post-iontophoresis temperature – pre-iontophoresis temperature.

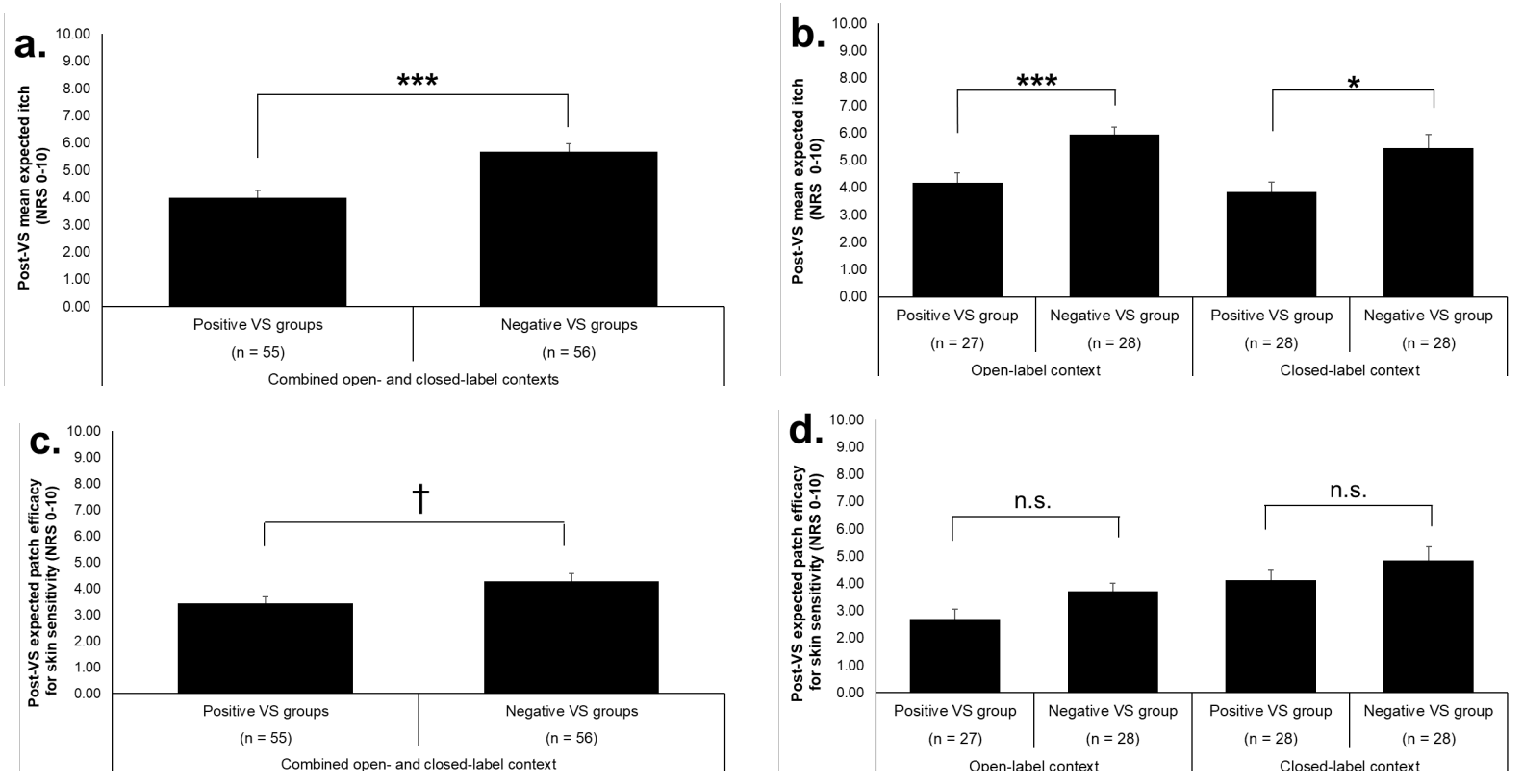
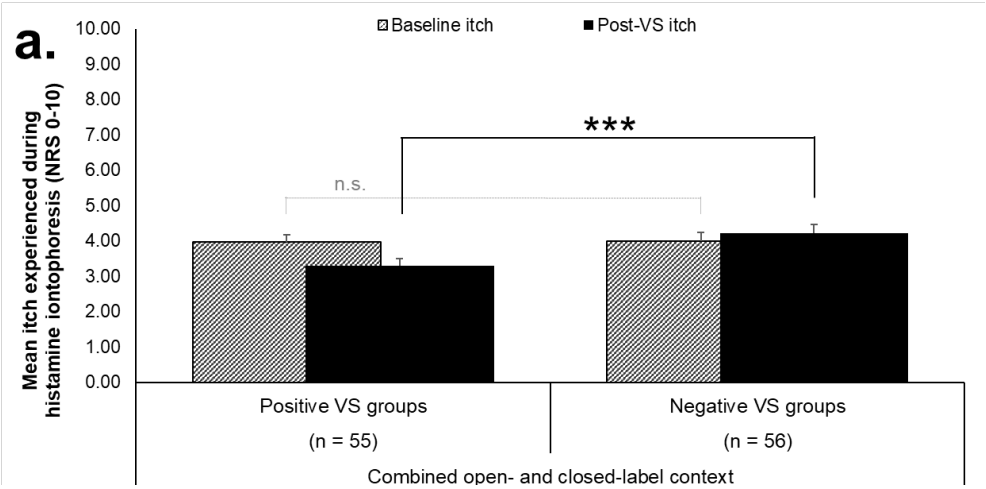


Figure 2. Mean Numeric Rating Scale (NRS) score with the standard error of the mean (SEM) for post-verbal suggestions (VS) itch expectation in [A] the combined open- and closed-label positive and negative VS groups, and [B] the separate open-label and closed-label positive and negative VS groups; and mean NRS with SEM for post-VS expected patch efficacy for skin sensitivity in [C] the combined open- and closed-label positive and negative VS groups, and [D] the separate open-label and closed-label positive and negative VS groups. *** $p<.001$, ** $p<.01$, * $p<.05$, † $p<.10$, n.s. non-significant.



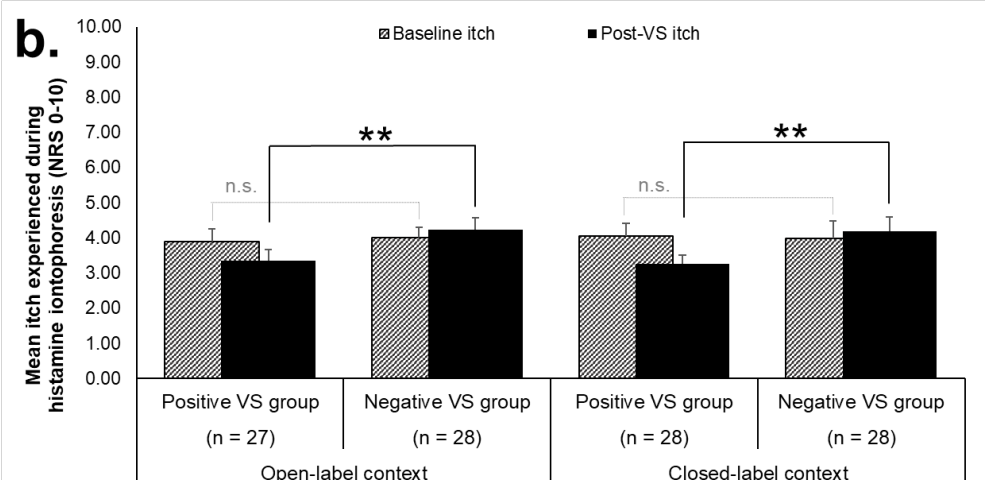


Figure 3. Mean Numeric Rating Scale (NRS) score for itch experienced during histamine iontophoresis for the baseline and post-verbal suggestions (VS) measurements, with the standard error of the mean (SEM) for [A] the combined open- and closed-label positive and negative VS groups, and [B] the separate open-label and closed-label positive and negative VS groups. *** p<.001, ** p<.01, * p<.05, † p<.10, n.s. non-significant.

DISCUSSION

The current study investigated whether positive and negative verbal suggestions regarding a sham transdermal patch for both open-label and closed-label contexts were able to influence self-reported itch during an experimental histamine test. Overall, the study findings illustrate that both open- and closed-label positive suggestions are able to influence expectations for itch and mean itch experienced during an experimental itch induction test compared to negative suggestions. The effects on itch expectations appear larger for the open-label context, whereas for self-rated perceived itch, the effects were larger when suggestions were given for the closed-label context. Secondary analyses indicated that itch decreased significantly following positive suggestions for both open-label and closed-label contexts, but that negative suggestions failed to increase itch. No effects on clinical skin response were found, but participants rated their own skin response as less severe following positive compared to negative suggestions for both open-label and closed-label contexts.

That positive suggestions are able to reduce itch is in line with findings of some, but not all previous studies [8,10-12,14]. The discrepancies in study findings in the literature may be explained by the strength and duration of verbal suggestions. Most of the studies on placebo effects in itch induce positive expectations by using brief suggestions of low or reduced itch [11,12,14]. In line with this, Bartels et al. [10] demonstrated that a combination of learning and suggestions was able to induce placebo effects, but brief suggestions alone could not. On the other hand, Darragh et al. [8] combined verbal suggestions with an information leaflet, which may have contributed to the strength of suggestions. The current study combined positive suggestions about itch with the cover story that a caffeine patch would influence cognitive abilities. That caffeine is able to impact, for example, focus and attention may be commonly accepted knowledge, which may in turn have contributed to the believability of the suggestions for itch.

Negative verbal suggestions did not increase experienced itch, which is not in line with previously conducted research [7,9,41-43]. However, previous studies have induced nocebo-like effects by giving suggestions directly about the itch elicitation methods. Potentially, suggestions regarding a sham treatment method may not elicit equally strong nocebo effects. A previous study, in which suggestions were given about a sham topical treatment, likewise failed to elicit significant increases in itch following negative suggestions [34]. Moreover, most participants were unfamiliar with the itch induction method, which may have resulted in higher itch scores during the baseline test. This may

complicate the estimation of the nocebo response, as the suggestions could have negated a naturally occurring decrease in itch. Future research may consider adding a no-suggestions (natural history) group to control for such effects and to more explicitly evaluate the size of placebo and nocebo effects.

Self-rated skin response was rated as less severe following both open-label and closed-label positive suggestions compared to negative suggestions. Indications that suggestions may be able to influence self-rated skin response have been found in previous research [13] and are further supported here. Clinical – or physical – skin response to histamine on the other hand was generally not influenced by verbal suggestions, which is in line with existing literature [8,13,44]. Wheal area and skin temperature decreased significantly in the negative VS groups. No differences between positive and negative suggestion groups were found, however, making it unlikely that these decreases were related to the manipulation used in the current study. A single previous study showed medium-sized increases in skin temperature following negative suggestions [34], but these findings could not be replicated here. Overall, the findings further support the notion that verbal suggestions may be more likely to impact subjective sensations such as pain or itch, whereas learning (i.e., conditioning) may be needed in addition to instructions in order to induce placebo effects for physical or physiological parameters.

While open-label suggestions appear particularly effective in inducing expectations for itch in the current study, effects on experienced itch were somewhat lower than for the closed-label context, though still medium-sized. The current study is one of the first to investigate similar verbal suggestions for both an open-label and a closed-label context. A previous study showed mixed evidence for the effects of open-label and closed-label suggestions on itch, but effect sizes did indicate that verbal suggestions had lower efficacy for itch in an open-label context as well [34]. Findings of the current study are in line with this. Most open-label studies report higher effect sizes than those reported in the current study though [39]. These studies have often used a rationale in which placebo effects were explained as learned Pavlovian responses [21-30]. The rationale in the current study differs from the one used previously, as only placebo and nocebo effects induced by positive or negative information (suggestions) and conscious expectancy were explained. These differences in rationale may as a consequence impact expectations in a different manner. Moreover, the open-label rationale in the current study was added onto a concealed positive or negative verbal suggestion (i.e. that the patch contained caffeine that would impact perception of itch, whereas in truth the patch contained no caffeine). This differs from previous work:

open-label rationales have either been provided immediately and without prior concealed suggestions [21-30], or have been added as an extended explanation of mechanisms onto a very succinct suggestion about to-be-expected effects [34]. Potentially, such a 'placebo-reveal' (i.e., explaining that you provided deceptive information first) may have resulted in smaller placebo responses in the open-label context compared to the closed-label (concealed) context. It has been shown that conditioned analgesia persists after it is revealed that subjects are in fact receiving a placebo [45]. A similar mechanism (i.e., first a placebo effect induction, which persists after the open-label rationale) may have played a role in the current study. Future research could aim to investigate how variations in the open-label rationale could impact its efficacy, for example by immediately integrating the open-label rationale in the suggestions or by investigating the efficacy of various open-label explanations of the placebo effect. Alternatively, participants may have responded differently to the negative suggestions, when they are given under concealed (closed-label) or open-label conditions. This may explain differences in effect size found under the open-label and closed-label contexts in the current study. There is evidence that information framing can influence the size of nocebo effects, with positive framing reducing the occurrence of (nocebo) side effects compared to negative framing [33]. Hypothetically, explaining how nocebo effects are formed may likewise impact how nocebo effects are formed, though this cannot be concluded exclusively based on data of the current study. Rather, future research may aim to clarify the impact of open-label information on the formation of nocebo effects. If it can be shown that open-label information can impact the formation of nocebo effects, this may be a potential method to prevent nocebo effects occurring in clinical practice. Moreover, an open-label rationale and suggestions may then be used to enhance placebo effects and inhibit nocebo effects simultaneously, for example by providing an explanation on the role of expectancy and context in treatment of medical conditions.

Some limitations need to be taken into account for the current study. The study was conducted single-blinded, with the experimenter giving the suggestions also being the one that conducted the tests. Potentially, this may have (unconsciously) impacted the participants' rating of itch during iontophoresis. Future research might consider using a double-blinded approach, for example, by having iontophoresis performed by a second experimenter who is blinded to allocated conditions. Second, participants received histamine iontophoresis twice within two hours, which may have caused habituation. However, the itch stimuli were relatively short (2.5 minutes) and presented almost one hour

apart. Moreover, by design, baseline iontophoresis took place on the dominant arm, and post-suggestions iontophoresis on the non-dominant arm. There are indications that handedness may affect sensory threshold and pain sensitivity, with the non-dominant arm being more sensitive [46,47]. It is likely that differences between both arms in sensitivity to itch would have negated habituation effects. Regardless, future research may aim to further control for these factors, including handedness. The lack of a no-treatment group in the current study complicates an estimation of the true placebo or nocebo response, as itch may have changed from baseline to post-VS regardless of suggestions. Including a no-treatment group, or counterbalancing the baseline and post-suggestion tests, may be a valuable contribution in future research to more explicitly evaluate placebo and nocebo effect sizes.

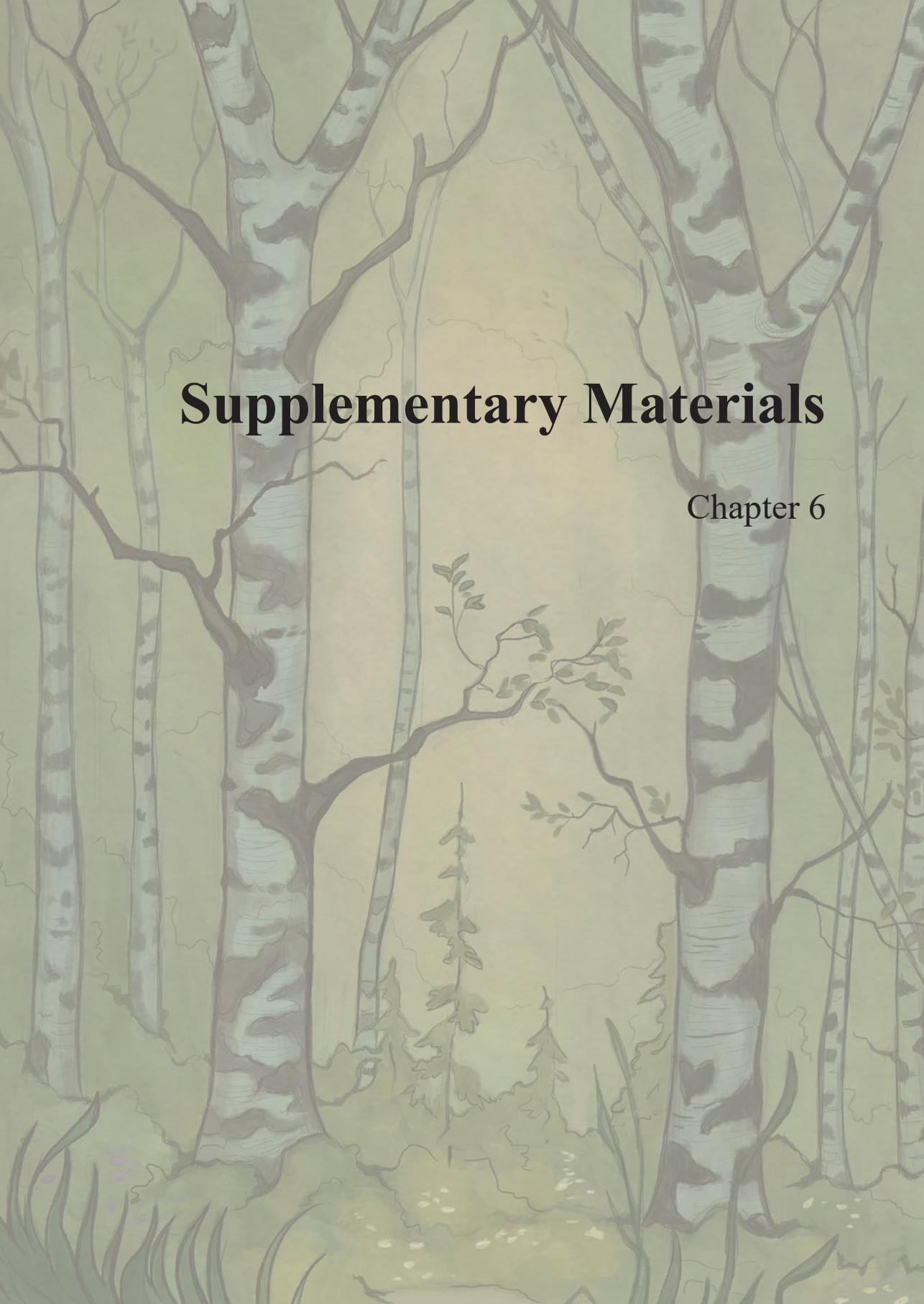
In short, the current study provides evidence that positive verbal suggestions regarding a sham transdermal patch for both open-label and closed-label contexts can influence expectations, itch experienced during, and self-reported skin response following an experimental histamine test. Future research may aim to investigate how variations in open-label rationale may impact the efficacy of positive and negative suggestions for itch. Potentially, open-label rationales may then be used to enhance placebo effects and inhibit nocebo effects in clinical practice, for example by explaining role of expectancy in treatment.

REFERENCES

1. Petrie KJ, Rief W. Psychobiological mechanisms of placebo and nocebo effects: pathways to improve treatments and reduce side effects. Annu Rev Psychol 2019; 70: 599-625.
2. Evers AWM, Colloca L, Blease C, Annoni M, Atlas LY, Benedetti F, Bingel U, Buchel C, Carvalho C, Colagiuri B, Crum AJ, Enck P, Gaab J, Geers AL, Howick J, Jensen KB, Kirsch I, Meissner K, Napadow V, Peerdeman KJ, Raz A, Rief W, Vase L, Wager TD, Wampold BE, Weimer K, Wiech K, Kaptchuk TJ, Klinger R, Kelley JM. Implications of placebo and nocebo effects for clinical practice: expert consensus. Psychother Psychosom 2018; 87: 204-10.
3. Peerdeman KJ, van Laarhoven AIM, Peters ML, Evers AWM. An integrative review of the influence of expectancies on pain. Front Psychol 2016; 7: 1270.
4. Smits RM, Veldhuijzen DS, Wulffraat NM, Evers AWM. The role of placebo effects in immune-related conditions: mechanisms and clinical considerations. Expert Rev Clin Immunol 2018; 14: 761-70.
5. Benedetti F. How the doctor's words affect the patient's brain. Eval Health Prof 2002; 25: 369-86.
6. Bartels DJP, van Laarhoven AIM, van de Kerkhof PCM, Evers AWM. Placebo and nocebo effects on itch: effects, mechanisms, and predictors. Eur J Pain 2016; 20: 8-13.
7. Bartels DJP, van Laarhoven AIM, Stroo M, Hijne K, Peerdeman KJ, Donders ART, van de Kerkhof PCM, Evers AWM. Minimizing nocebo effects by conditioning with verbal suggestion: a randomized clinical trial in healthy humans. PLoS One 2017; 12(9): e0182959.
8. Darragh M, Chang JWH, Booth RJ, Consedine NS. The placebo effect in inflammatory skin reactions: The influence of verbal suggestion on itch and weal size. J Psychosom Res 2015; 78: 489-94.
9. Stumpf A, Zerey V, Heuft G, Stander S, Pfleiderer B, Schneider G. Itch perception and skin reactions as modulated by verbal suggestions: role of participant's and investigator's sex. Acta Derm Venereol 2016; 96: 619-23.
10. Bartels DJP, van Laarhoven AIM, Haverkamp EA, Wilder-Smith OH, Donders ART, van Middendorp H, van de Kerkhof PCM, Evers AWM. Role of conditioning and verbal suggestion in placebo and nocebo effects on itch. PLoS One 2014; 9(3): e91727.
11. Peerdeman KJ, Van Laarhoven AIM, Donders ART, Hopman MTE, Peters ML, Evers AWM. Inducing expectations for health: effects of verbal suggestion and imagery on pain, itch, and fatigue as indicators of physical sensitivity. PLoS One 2015; 10(10):e0139563.
12. van Laarhoven AIM, Vogelaar ML, Wilder-Smith OH, van Riel PLCM, van de Kerkhof PCM, Kraaimaat FW, Evers AWM. Induction of nocebo and placebo effects on itch and pain by verbal suggestions. Pain 2011; 152: 1486-94.
13. Meeuwis SH, van Middendorp H, Veldhuijzen DS, van Laarhoven AIM, De Houwer J, Lavrijsen APM, Evers AWM. Placebo effects of open-label verbal suggestions on itch. Acta Derm Venereol 2018; 98: 268-74.
14. Skvortsova A, Veldhuijzen DS, van Middendorp H, van de Bergh O, Evers AWM. Enhancing placebo effects in somatic symptoms through oxytocin. Psychosom Med 2018; 80: 353-60.
15. Abramowicz M, Kruszewski J, Chcialowski A. Evaluation of the placebo effect in the trials of allergen immunotherapy effectiveness: meta-analysis of randomized and placebo-controlled trials. Postepy Dermatol Alergol 2018; 35: 620-5.
16. van Laarhoven AIM, van der Sman-Mauriks IM, Donders AR, Pronk MC, van de Kerkhof PC, Evers AWM. Placebo effects on itch: a meta-analysis of clinical trials of patients with dermatological conditions. J Invest Dermatol 2015; 135: 1234-43.
17. Pereira MP, Ständer S. Assessment of severity and burden of pruritus. Allergol Int 2017; 66: 3-7.
18. Ständer S, Weisshaar E. Medical treatment of pruritus. Expert Opin Emerg Drugs 2012; 17: 335-45.
19. Seth D, Cheldize K, Brown D, Freeman EF. Global burden of skin disease: inequities and innovations. Curr Dermatol Rep 2017; 6: 204-10.
20. Blease C, Colloca L, Kaptchuk TJ. Are open-label placebos ethical? Informed consent and ethical equivocations. Bioethics 2016; 30: 407-14.
21. Carvalho C, Caetano JM, Cunha L, Rebouta P, Kaptchuk TJ, Kirsch I. Open-label placebo treatment in chronic low back pain: a randomized controlled trial. Pain 2016; 157: 2766-72.
22. El Brihi J, Horne R, Faasse K. Prescribing placebos: an experimental examination of the role of dose, expectancies, and adherence in open-label placebo effects. Ann Behav Med 2019; 53: 16-28.

23. Hoenemeyer TW, Kaptchuk TJ, Mehta TS, Fontaine KR. Open-label placebo treatment for cancer-related fatigue: a randomized-controlled clinical trial. Sci Rep 2018; 8: 2784.
24. Kaptchuk TJ, Friedlander E, Kelley JM, Sanchez MN, Kokkotou E, Singer JP, Kowalczykowski M, Miller FG, Kirsch I, Lembo AJ. Placebos without deception: a randomized controlled trial in irritable bowel syndrome. PLoS One 2010; 5: e15591.
25. Kelley JM, Kaptchuk TJ, Cusin C, Lipkin S, Fava M. Open-label placebo for major depressive disorder: a pilot randomized controlled trial. Psychother Psychosom 2012; 81: 312-4.
26. Leibowitz KA, Hardebeck EJ, Goyer JP, Crum AJ. The role of patient beliefs in open-label placebo effects. Health Psychol 2019.
27. Locher C, Frey Nascimento A, Kirsch I, Kossowsky J, Meyer A, Gaab J. Is the rationale more important than deception? A randomized controlled trial of open-label placebo analgesia. Pain 2017; 158: 2320-8.
28. Sandler AD, Glesne C, Bodfish JW. Conditioned placebo dose reduction: a new treatment in attention-deficit hyperactivity disorder? J Dev Behav Pediatr 2010; 31: 369-75.
29. Schaefer M, Harke R, Denke C. Open-label placebos improve symptoms in allergic rhinitis: a randomized controlled trial. Psychother Psychosom 2016; 85: 373-4.
30. Schaefer M, Sahin T, Berstecher B. Why do open-label placebos work? A randomized controlled trial of an open-label placebo induction with and without extended information about the placebo effect in allergic rhinitis. PLoS One 2018; 13: e0192758.
31. Devlin EJ, Whitford HS, Denson LA, Potter AE. "Just as I expected": a longitudinal cohort study of the impact of response expectancies on side effect experiences during radiotherapy for prostate cancer. J Pain Symptom Manage 2019; 57: 273-81.e4.
32. Howick J, Webster R, Kirby N, Hood K. Rapid overview of systematic reviews of nocebo effects reported by patients taking placebos in clinical trials. Trials 2018; 19:674.
33. Faasse K, Huynh A, Pearson S, Geers AL, Helfer SG, Colagiuri B. The influence of side effect information framing on nocebo effects. Ann Behav Med 2018.
34. Meeuwis SH, van Middendorp H, van Laarhoven AIM, Veldhuijzen DS, Lavrijsen APM, Evers AWM. Effects of open-and closed-label nocebo and placebo suggestions on itch and itch expectations. Front Psychiatry 2019; 10: 436.
35. Misery L, Jean-Decoster C, Mery S, Georgescu V, Sibaud V. A new ten-item questionnaire for assessing sensitive skin: the sensitive scale-10. Acta Derm Venereol 2014; 94: 635-9.
36. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 2012; 9: 671-5.
37. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS Scales. J Pers Soc Psychol 1988; 54.
38. Faul F, Erdfelder E, Lang A-G, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007; 39: 175-91.
39. Charlesworth JEG, Petkovic G, Kelley JM, Hunter M, Onakpoya I, Roberts N, Miller FG, Howick J. Effects of placebos without deception compared with no treatment: a systematic review and meta-analysis. J Evid Based Med 2017; 10: 97-107.
40. Cohen J. Statistical power analysis for the behavioral sciences, 2nd ed. Hillsdale, NJ: Erlbaum; 1988.
41. Bartels DJP, van Laarhoven AIM, van de Kerkhof PCM, Evers AWM. Nocebo effects and scratching behaviour on itch. Acta Derm Venereol 2018; 98: 943-50.
42. van de Sand MF, Menz MM, Sprenger C, Büchel C. Nocebo-induced modulation of cerebral itch processing—an fMRI study. NeuroImage 2018; 166: 209-18.
43. Napadow V, Li A, Loggia ML, Kim J, Mawla I, Desbordes G, Schalock PC, Lerner EA, Tran TN, Ring J, Rosen BR, Kaptchuk TJ, Pfab F. The imagined itch: brain circuitry supporting nocebo-induced itch in atopic dermatitis patients. Allergy 2015; 70: 1485-92.
44. Darragh M, Booth RJ, Koschwanez HE, Sollers J, Broadbent E. Expectation and the placebo effect in inflammatory skin reactions. A randomised-controlled trial. J Psychosom Res 2013; 74: 439-43.
45. Schafer SM, Colloca L, Wager TD. Conditioned placebo analgesia persists when subjects know they are receiving a placebo. J Pain 2015; 16: 412-20.

46. Friedli WG, Fuhr P, Wiget W. Detection threshold for percutaneous electrical stimuli: asymmetry with respect to handedness. J Neurol Neurosurg Psychiatry 1987; 50: 870.
47. Pud D, Golan Y, Pesta R. Hand dominancy—a feature affecting sensitivity to pain. Neurosci Lett 2009; 467: 237-40.
48. Spielberger CD. Manual for the State-Trait Anxiety Inventory. Revised edition ed. Palo Alto, CA: Consulting Psychologists Press; 1983.
49. Hammes J. De Stroop kleur-woord test. Handleiding, tweede gewijzigde druk [The Stroop color-word test. Manual, second revised edition]. Lisse: Swets & Zeitlinger; 1978.
50. Bowie CR, Harvey PD. Administration and interpretation of the Trail Making Test. Nat Protoc 2006; 1: 2277.
51. Reitan RM. The relation of the trail making test to organic brain damage. J Consult Psychol 1955; 19: 393.



Supplementary Materials

Chapter 6

SUPPLEMENTARY METHODS

Materials and Measures

1. Stroop test and Trail Making Test

As part of the cover story of the patch positively influencing cognition, the Stroop test [1] and Trail Making Test [2,3] were assessed at baseline and following suggestions. Stroop interference scores and percentile scores were calculated controlling for age, sex, and education level. As a large inter-individual variability in the execution of the Trail Making Test was noted (e.g., on noticing and dealing with mistakes during the test – some participants did not correct mistakes whereas others did, thus causing differences on the time spent taking the test and the associated outcome measure), these data were not analysed.

2. Expectations for the cognitive tasks

Prior to the cognitive tasks (both baseline and post-verbal suggestions) participants were asked how well they expected to perform during the tasks by rating the following items on a Numeric Rating Scale (NRS) ranging from 0 ("not good at all") to 10 ("very good"): focus, attention, performance, and speed during the tasks. In addition, expected patch efficacy for focus, attention, and speed was rated following the verbal suggestions using the same NRS.

3. Wellbeing: State Anxiety and General Wellbeing Scales

The State Trait Anxiety Inventory – State Anxiety short scales (STAI-S-s [4]), and seven Numeric Rating Scales (NRS) measuring general wellbeing (relaxed, anxious, serene, agreeable, tense, worried, stressed, see also [5]) on a scale from 0 ('not at all') to 10 ('very much so') were assessed at four moments during the laboratory session: I. pre-baseline iontophoresis (baseline 1), II. post-baseline iontophoresis (baseline 2), III. post-verbal suggestions (post-VS 1), and IV. at the end of the session (post-VS 2). Of the NRS items, negative items were recoded and a total score was calculated by summing all items (with higher scores reflecting higher general wellbeing). Cronbach's alpha for state anxiety ranged from .80 to .83 in the current study. For general wellbeing, Cronbach's alpha ranged .86 to .87.

Statistical Analysis

All analyses were conducted in SPSS 25.0 for Windows (IBM SPSS Inc., Chicago, Illinois, US). As was a priori determined, open-label and closed-label groups were first combined to detect differences between the effects of positive verbal suggestions and negative verbal suggestions with optimal power, and second, repeated for the separate open-label context and the separate closed-label context. Normal distribution of the variables, baseline differences, and assumptions were checked prior to data analysis. General linear model (GLM) analyses of covariance (ANCOVAs) were used to assess group differences in expectations regarding the cognitive tasks. For the expected efficacy of the patch, GLM analysis of variance (ANOVA) was used. Effects of suggestions on Stroop scores were analysed using GLM ANCOVA, and effects of group on wellbeing scales were assessed by mixed between-within-subject RMA. The critical alpha used was α=.05 for the combined group analyses. For the separate open-label and closed-label context analyses, a Bonferroni correction for multiple comparisons was applied ($\alpha/2$=.025).

SUPPLEMENTARY RESULTS

Results

1. Expectations and expected patch efficacy for the cognitive tasks

At baseline, expected attention and performance during the tasks were significantly higher in the combined open- and closed-label negative VS groups (attention: M=7.00±1.39; performance: M=6.80±1.07) compared to the combined positive VS groups (attention: M=6.33±1.56; performance: M=6.26±1.44); $F_{att.}(1,109)=5.55$, p=.020, $F_{perf.}(1,109)=5.09$, p=.026; see **Supplementary Table E3**. Marginal differences were found for expected focus (p=.052) and expected speed (p=.073), respectively. When open-label and closed-label contexts were separated, no baseline differences could be found (all $p\geq$.061; see also **Supplementary Table E4**).

When open- and closed-label groups were combined, no group differences were found for any of the post-suggestions expectation measures (all p>.055), with the exception of expected patch influence on speed; F(1,109)=4.11, p=.045, Cohen's d=0.39 (see **Supplementary Table E3**). Participants in the combined negative VS groups expected the patch to be more effective for speed during the cognitive tasks (M=3.70±2.38), compared to

participants in the combined positive VS groups (M=2.86±1.94). No group differences were found in the separate open-label or closed-label context (all p≥.091; see **Supplementary Table E4**). Within-group baseline-to-post-VS-change indicated no significant changes in expectations regarding the cognitive tasks following suggestions after applying the Bonferroni correction for multiple comparisons (all p≥.042), with the exception of expected speed in the combined positive VS groups. Within these groups, a significant decrease in expected speed was noted (M_{change}=-0.48, t(54)=2.54, p=.014 (see **Supplementary Table E5**). When open-label and closed-label contexts were separated, no significant changes from baseline to post-VS were noted (all p≥.13).

2. Effects on Stroop Test

The combined positive and the combined negative VS groups did not differ on Stroop interference scores or percentile scores at baseline, or following verbal suggestions (all p≥.10). Similar findings were demonstrated when analyses were repeated for the separate open-label and closed-label contexts (all p≥.044). Within-group baseline-to-post-VS-change indicated Stroop interference and percentile scores improved in both the combined positive and the combined negative groups (all p<.001). When groups were separated, similar reductions in Stroop interference and percentile scores were found for the closed-label negative VS and the open-label positive VS groups (all p≤.020). In the open-label negative VS group, no significant change in interference score was found after applying the Bonferroni correction for multiple comparisons (p=.039), however, the percentile score did reduce significantly (p=.011). Stroop interference and percentile scores did not change in the closed-label positive VS group (both p≥.037).

3. Subjective wellbeing: State anxiety and General Wellbeing Scales

No combined-group x time interactions, or main effects of the combined groups, were found for STAI total score or general wellbeing scales (p≥.22). Both state anxiety and general wellbeing changed significantly over time (p≤.012, see **Supplementary Figure S2A** (state anxiety) and **Supplementary Figure S3A** (general wellbeing)). For state anxiety, the baseline measurement was significantly lower than the second measurement (baseline post-iontophoresis; p=.006). No significant differences over time for the other measurements were found (all p≥.080). General wellbeing at the final measurement point

was significantly higher compared to the second and third measurements (both $p \leq .015$). When analyses were conducted for the separate open-label and closed-label contexts, no effects were found for state anxiety (see **Supplementary Figure S2B**). For general wellbeing, the final measurement was significantly higher compared to the baseline post-VS measurement only for the open-label context (see **Supplementary Figure S3B**).

Concluding note on the effects of open-label and closed-label suggestions on expectations regarding cognition, outcomes of the cognitive tasks and wellbeing

Overall, verbal suggestions did not induce differences in expectations regarding the cognitive tasks and patch efficacy for such tasks, with the exception of expected speed during the tasks. The participants in the negative suggestions groups expected the patch to be more effective for speed during the cognitive tasks compared to the positive VS groups. It should be noted though that on general, both groups scored low on expected efficacy for speed (mean of 3.70 and 2.86, respectively, on a 0-10 scale). When groups were separated for open-label and closed-label contexts, no differences in expectations were found. Notably, expectations for the cognitive tasks generally did not reduce after suggestions, indicating that the positive suggestions about the tasks may not have impacted participants' expectations. However, expectations were assessed at baseline before any test was conducted. It may be possible that participants found the Stroop test, Trail Making Test, and filler tasks (i.e. Sudoku's and other puzzles) more challenging than previously anticipated, which may have impacted the scores given post-suggestions. In general, some improvement on Stroop test scores was demonstrated for all groups, which could be due to training effects. Regarding wellbeing, significant changes over time could be demonstrated, but positive and negative verbal suggestions did not significantly impact either state anxiety or general wellbeing. This is in line with the findings for positive affect, which are described in the paper of the current study.

Supplementary Table E1. Means, standard deviations, and analysis of (co)variance (AN(C)OVA) outcomes for demographic factors and histamine iontophoresis-related measures for the separate open-label and closed-label positive and negative verbal suggestions (VS) groups.

	Open-label context				Closed-label context			
			AN(C)OVA				AN(C)OVA	
	Positive VS (*n*=27)	**Negative VS (*n*=28)**	***p*-value**	**Cohen's d**	**Positive VS (*n*=28)**	**Negative VS (*n*=28)**	***p*-value**	**Cohen's d**
Demographics								
Sex [male: n (%)]	4 (14.81)	5 (17.86)	.76		4 (14.29)	5 (17.86)	.72	
Age	21.67 ± 2.60	22.29 ± 3.10	.43	0.22	22.11 ± 2.39	21.57 ± 2.95	.46	0.20
Baseline histamine iontophoresis								
Mean itch	3.89 ± 1.35	4.01 ± 1.70	.78	0.08	4.06 ± 1.53	3.99 ± 1.78	.87	0.04
Subjective skin response (SS-10) [a]	30.40 ± 13.38	32.53 ± 13.99	.57	0.16	31.43 ± 13.37	27.05 ± 11.31	.19	0.35
Wheal area [cm^2]	9.15 ± 3.71	8.95 ± 3.52	.84	0.06	8.70 ± 3.08	9.60 ± 4.24	.47	0.24
Flare area [cm^2]	42.25 ± 15.69	42.48 ± 12.54	.95	0.02	44.44 ± 15.92	41.87 ± 13.70	.52	0.17
Change in skin temperature [°C] [b, c]	1.02 ± 1.05	1.44 ± 1.00	.13	0.41	1.76 ± 1.15	1.93 ± 0.95	.56	0.16
Post-VS expectations outcomes for itch								
Expected itch	4.17 ± 1.86	5.93 ± 1.51	< .001	1.04	3.84 ± 1.91	5.44 ± 2.66	.013	0.69
Expected patch effectiveness for skin sensitivity	2.70 ± 2.12	3.71 ± 2.75	.13	0.41	4.13 ± 1.89	4.84 ± 2.24	.20	0.34
Post-VS histamine iontophoresis								
Mean itch	3.34 ± 1.66	4.24 ± 1.76	.008	0.47	3.25 ± 1.42	4.19 ± 2.17	.004	0.54
Subjective skin response (SS-10) [a, d]	23.37 ± 11.73	29.07 ± 12.68	.017	0.33	23.82 ± 12.22	26.00 ± 12.80	.010	0.45
Wheal area [cm^2]	8.15 ± 3.27	7.83 ± 3.02	.75	0.06	8.23 ± 3.15	8.01 ± 3.84	.17	0.25
Flare area [cm^2]	41.25 ± 11.84	42.21 ± 12.28	.90	0.02	42.09 ± 14.88	41.22 ± 15.41	.50	0.09
Change in skin temperature [°C] [c, e]	1.19 ± 1.48	1.07 ± 1.09	.23	0.30	1.21 ± 1.12	1.34 ± 1.10	.92	0.02

Note. [a] Misery and colleagues (6), [b] n=1 missing due to technical difficulties with the infrared thermometer, [c] calculated as post-iontophoresis temperature – pre-iontophoresis temperature. [d] n=1 missing for the Sensitive Scale-10 (SS-10) in the closed-label negative VS group, [e] n=2 missing due to technical difficulties with the infrared thermometer. Cohen's d for ANCOVA was calculated using the covariate-adjusted means (not depicted in the table).

Supplementary Table E2. Within-group baseline-to-post-verbal suggestions (VS) changes on histamine iontophoresis outcomes for the separate open-label and closed-label positive and negative verbal suggestions (VS) groups.

	Open-label context								**Closed-label context**							
	Positive VS group				**Negative VS group**				**Positive VS group**				**Negative VS group**			
	n	Mean change	*t*	*p*	*n*	Mean change	*t*	*p*	*n*	Mean change	*t*	*p*	*n*	Mean change	*t*	*p*
Mean itch	27	-0.55	3.11	.004	28	0.24	-1.02	.32	28	-0.81	3.87	.001	28	0.20	-0.74	.46
Subjective skin response (SS-10) [a]	27	-7.03	5.38	< .001	28	-3.46	2.74	.011	28	-7.60	4.54	< .001	27	-1.09	0.81	.42
Wheal area [cm^2]	27	-1.00	1.73	.095	28	-1.12	2.82	.009	28	-0.47	0.91	.37	28	-1.12	3.69	.001
Flare area [cm^2]	27	-0.16	0.09	.92	28	-0.27	0.19	.86	28	-2.34	1.89	.37	28	-0.65	0.42	.68
Change in skin temperature [°C] [b, c]	27	0.17	-0.67	.51	28	-0.37	1.88	.071	27	-0.55	3.14	.004	27	-0.58	3.03	.005

Note. Mean change was calculated as post-verbal suggestions score – baseline score, with negative values indicating a decrease from baseline, and positive scores indicating an increase from baseline. [a] Misery and colleagues (6), [b] n=2 missing due to technical difficulties with the infrared thermometer. [c] calculated as post-iontophoresis temperature – pre-iontophoresis temperature.

Supplementary Table E3. Means, standard deviations, and analysis of (co)variance (AN(C)OVA) outcomes for expectations and outcomes of the cognitive tests in the combined open- and closed-label positive and negative verbal suggestions (VS) groups.

	Combined open- and closed-label contexts			
			AN(C)OVA	
	Positive VS (*n*=55)	**Negative VS (*n*=56)**	***p*-value**	**Cohen's d**
Baseline expectation outcomes for cognitive tasks				
Expected focus	6.47 ± 1.48	6.97 ± 1.17	.052	0.38
Expected attention	6.33 ± 1.56	7.00 ± 1.39	.020	0.45
Expected performance	6.26 ± 1.44	6.80 ± 1.07	.026	0.43
Expected speed	5.73 ± 1.71	6.25 ± 1.29	.073	0.34
Baseline Stroop test				
Interference score	56.67 ± 7.80	55.38 ± 8.45	.40	0.16
Percentile score	69.91 ± 20.87	65.41 ± 22.26	.28	0.21
Post-VS expectation outcomes for cognitive tasks				
Expected focus	6.33 ± 1.40	6.99 ± 1.41	.10	0.27
Expected attention	6.42 ± 1.60	7.11 ± 1.33	.21	0.20
Expected performance	5.93 ± 1.64	6.57 ± 1.37	.29	0.16
Expected speed	5.24 ± 2.00	6.13 ± 1.43	.055	0.28
Expected patch influence on focus	3.75 ± 2.19	3.81 ± 2.29	.88	0.03
Expected patch influence on attention	3.70 ± 2.28	3.78 ± 2.10	.84	0.04
Expected patch influence on speed	2.86 ± 1.94	3.70 ± 2.38	.045	0.39
Post-VS Stroop test				
Interference score	61.93 ± 9.54	58.86 ± 7.86	.10	0.27
Percentile score	80.87 ± 22.59	75.82 ± 19.00	.43	0.13

Note. Cohen's d for ANCOVA was calculated using the covariate-adjusted means (not depicted in the table).

Supplementary Table E4. Means, standard deviations, and analysis of (co)variance (AN(C)OVA) outcomes for expectations and outcomes of the cognitive tests in the separate open-label and closed-label positive and negative verbal suggestions (VS) groups.

	Open-label context				Closed-label context			
			AN(C)OVA				AN(C)OVA	
	Positive VS (*n*=27)	**Negative VS (*n*=28)**	***p*-value**	**Cohen's d**	**Positive VS (*n*=28)**	**Negative VS (*n*=28)**	***p*-value**	**Cohen's d**
Baseline expectation outcomes for cognitive tasks								
Expected focus	6.26 ± 1.60	6.82 ± 1.33	.17	0.38	6.68 ± 1.34	7.12 ± 0.97	.16	0.38
Expected attention	6.14 ± 1.59	6.75 ± 1.51	.15	0.39	6.53 ± 1.54	7.24 ± 1.25	.061	0.51
Expected performance	6.16 ± 1.45	6.76 ± 1.06	.084	0.47	6.35 ± 1.45	6.83 ± 1.09	.16	0.37
Expected speed	5.71 ± 1.58	6.03 ± 1.41	.44	0.21	5.74 ± 1.86	6.46 ± 1.14	.084	0.47
Baseline Stroop test								
Interference score	57.67 ± 7.32	55.82 ± 7.67	.37	0.25	55.71 ± 8.25	54.93 ± 9.29	.74	0.09
Percentile score	73.48 ± 18.96	67.00 ± 19.97	.22	0.33	66.46 ± 22.36	63.82 ± 24.60	.68	0.11
Post-VS expectation outcomes for cognitive tasks								
Expected focus	6.09 ± 1.53	6.87 ± 1.45	.16	0.35	6.57 ± 1.25	7.11 ± 1.37	.41	0.18
Expected attention	6.34 ± 1.80	7.05 ± 1.30	.34	0.21	6.50 ± 1.40	7.16 ± 1.39	.45	0.17
Expected performance	5.91 ± 1.79	6.68 ± 1.20	.35	0.21	5.94 ± 1.52	6.46 ± 1.55	.61	0.11
Expected speed	5.29 ± 2.22	6.12 ± 1.25	.13	0.32	5.20 ± 1.80	6.14 ± 1.62	.27	0.23
Expected patch influence on focus	3.00 ± 2.22	3.19 ± 2.48	.77	0.08	4.47 ± 1.93	4.44 ± 1.92	.95	0.02
Expected patch influence on attention	2.86 ± 2.11	3.20 ± 2.21	.56	0.16	4.51 ± 2.17	4.36 ± 1.84	.79	0.08
Expected patch influence on speed	2.18 ± 1.79	2.98 ± 2.53	.18	0.37	3.52 ± 1.89	4.42 ± 2.02	.091	0.46
Post-VS Stroop test								
Interference score	64.52 ± 9.24	59.18 ± 8.31	.044	0.48	59.43 ± 9.31	58.54 ± 7.53	.80	0.06
Percentile score	86.07 ± 20.17	76.68 ± 18.86	.21	0.28	75.86 ± 24.00	74.96 ± 19.46	.94	0.02

Note. The critical alpha used following Bonferroni correction for multiple comparisons was $\alpha = (.05/2) = .025$. Cohen's d for ANCOVA was calculated using the covariate-adjusted means (not depicted in the table).

Supplementary Table E5. Within-group baseline-to-post-verbal suggestions (VS) changes for expectations and outcomes of the cognitive tests in the combined open- and closed-label positive and negative verbal suggestions (VS) groups.

	Combined open- and closed- label positive VS groups (*n*=55)				**Combined open- and closed-label negative VS groups (*n*=56)**			
	n	Mean change	*t*	*p*	*n*	Mean change	*t*	*p*
Expectation outcomes for cognitive tasks								
Expected focus	55	-0.14	0.78	.44	56	0.03	-0.15	.89
Expected attention	55	0.09	-0.45	.65	56	0.11	-0.70	.49
Expected performance	55	-0.33	2.09	.042	56	-0.23	1.32	.19
Expected speed	55	-0.48	2.54	.014	56	-0.12	0.68	.50
Stroop test								
Interference score	55	5.26	-4.70	<.001	56	3.48	-3.30	.002
Percentile score	55	10.96	3.95	< .001	56	10.41	-4.01	< .001

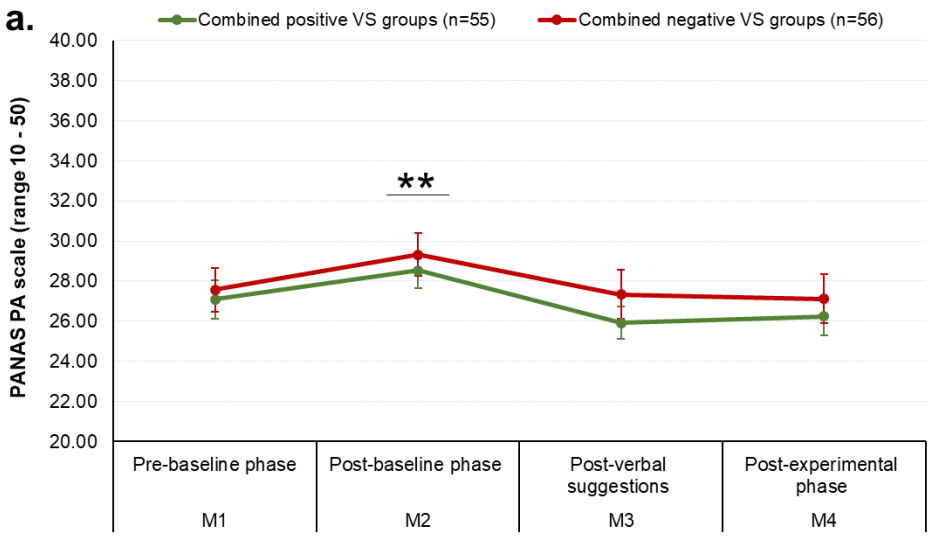
Note. Mean change was calculated as post-verbal suggestions score – baseline score, with negative values indicating a decrease from baseline, and positive scores indicating an increase from baseline.

Supplementary Table E6. Within-group baseline-to-post-verbal suggestions (VS) changes for expectations and outcomes of the cognitive tests in the separate open-label and closed-label positive and negative verbal suggestions (VS) groups.

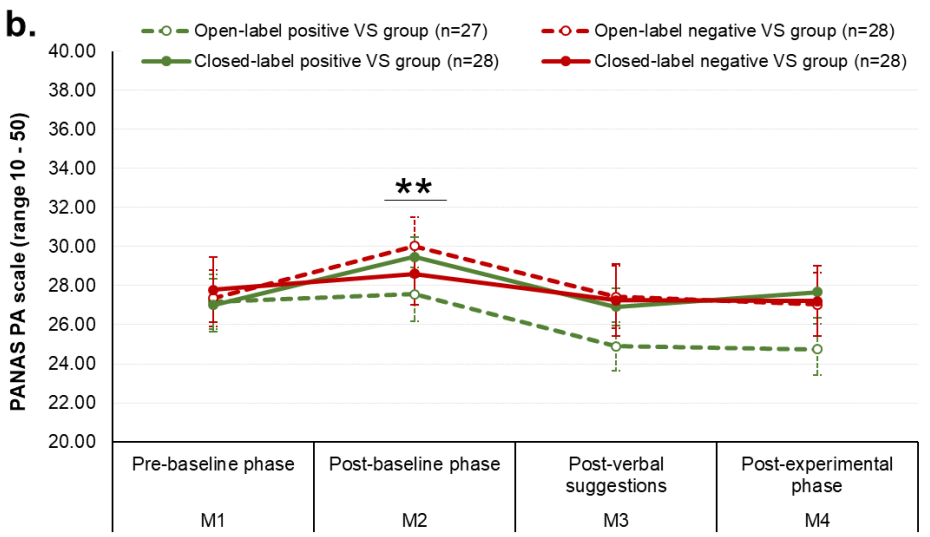
	Open-label context								**Closed-label context**							
	Positive VS group				**Negative VS group**				**Positive VS group**				**Negative VS group**			
	n	Mean change	*t*	*p*	*n*	Mean change	*t*	*p*	*n*	Mean change	*t*	*p*	*n*	Mean change	*t*	*p*
Expectation outcomes for cognitive tasks																
Expected focus	27	-0.17	0.57	.58	28	0.05	-0.20	.85	28	-0.11	0.54	.59	28	-0.003	0.02	.99
Expected attention	27	0.21	-0.73	.47	28	0.30	-1.20	.24	28	-0.03	0.09	.93	28	-0.08	0.42	.68
Expected performance	27	-0.25	1.14	.26	28	-0.09	0.35	.73	28	-0.41	1.76	.089	28	-0.37	1.55	.13
Expected speed	27	-0.43	1.62	.12	28	0.09	-0.31	.76	28	-0.54	1.93	.064	28	-0.33	1.49	.15
Stroop test																
Interference score	27	6.85	-4.82	< .001	28	3.36	-2.17	.039	28	3.71	-2.20	.037	28	3.61	-2.47	.020
Percentile score	27	12.59	-4.02	< .001	28	9.68	-2.75	.011	28	9.39	-2.05	.050	28	11.14	-2.88	.008

Note. Mean change was calculated as post-verbal suggestions score – baseline score, with negative values indicating a decrease from baseline, and positive scores indicating an increase from baseline. The critical alpha used following Bonferroni correction for multiple comparisons was $\alpha = (.05/2)/2 = .0125$.

Supplementary Figure S1. Means + SEMs of the Positive and Negative Affect Schedule (PANAS) subscale 'positive affect' and mixed between-within repeated measures ANOVA (RMA) outcomes for (A) the combined open- and closed label positive VS groups and the combined negative VS groups, and (B) the separate groups.



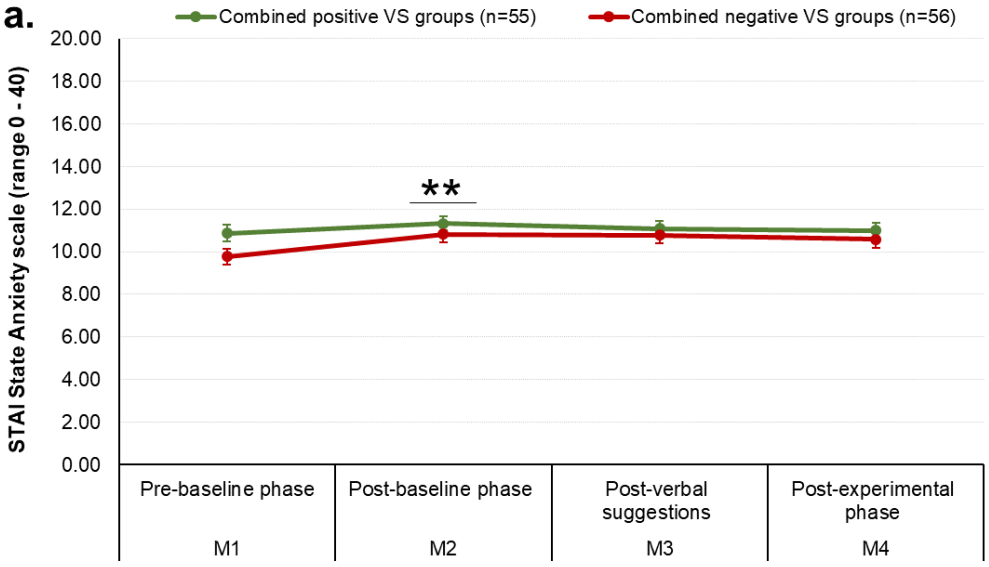
Combined-groups Mixed RMA: group x time n.s.; group n.s.; time $p<.001$.
** Bonferroni post-hoc effect of time: M2 > M1+M3+M4 (all $p<.002$)



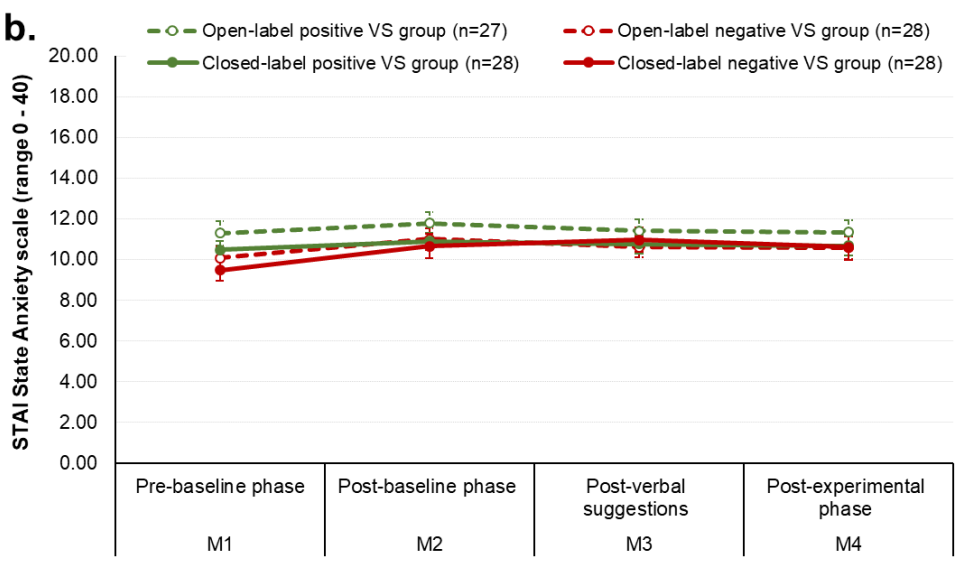
Open-label mixed RMA: group x time n.s.; group n.s.; time $p<.001$.
** Bonferroni post-hoc effect of time: M2 > M3+M4 (all $p<.001$)

Closed-label mixed RMA: group x time n.s.; group n.s.; time $p=.001$.
** Bonferroni post-hoc effect of time: M2 > M1+M3 (all $p<.017$)

Supplementary Figure S2. Means + SEMs of the State and Trait Anxiety Inventory (STAI) subscale 'state anxiety' and mixed between-within repeated measures ANOVA (RMA) outcomes for (A) the combined open- and closed label positive VS groups and the combined negative VS groups, and (B) the separate groups.



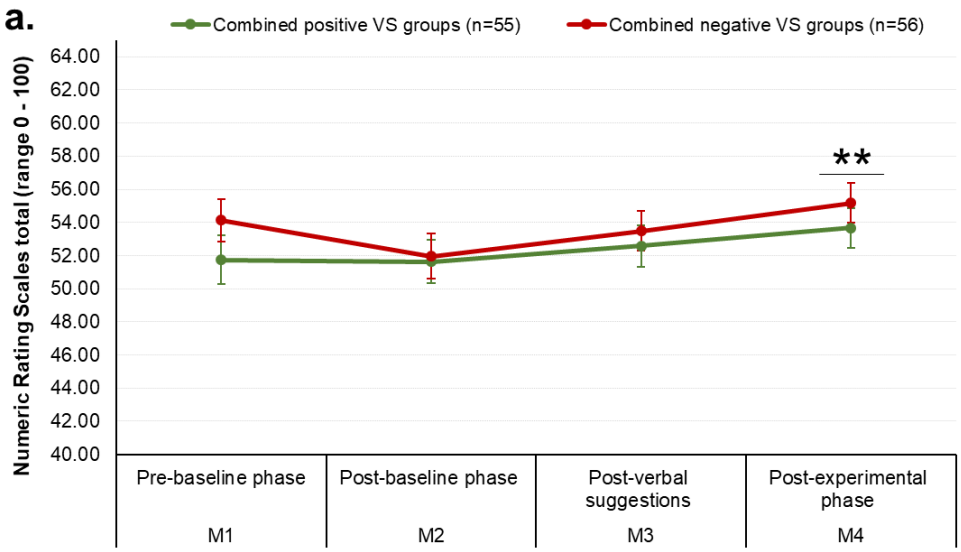
Combined-groups Mixed RMA: group x time n.s.; group n.s.; time *p*=.012
** Bonferroni post-hoc effect of time: M2 > M1 (*p*=.006)



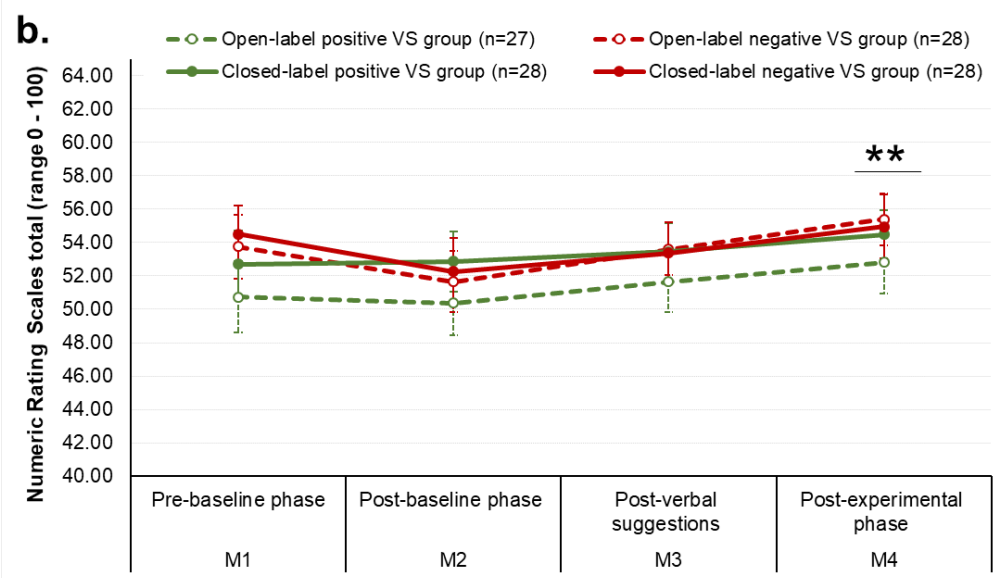
Open-label mixed RMA: n.s.
** Bonferroni post-hoc effect of time: n.s.

Closed-label mixed RMA: n.s.
** Bonferroni post-hoc effect of time: n.s.

Supplementary Figure S3. Means + SEMs of the Numeric Rating Scales (NRS) total score for 'general wellbeing' and mixed between-within repeated measures ANOVA (RMA) outcomes for (A) the combined open- and closed label positive VS groups and the combined negative VS groups, and (B) the separate groups.



Combined-groups Mixed RMA: group x time n.s.; group n.s.; time p=.001
** Bonferroni post-hoc effect of time: M2+M3 < M4 (p<.015)



Open-label mixed RMA: group x time n.s.; group n.s.; time p=.010
** Bonferroni post-hoc effect of time: M2 < M4 (p=.012)

Closed-label mixed RMA: n.s.
** Bonferroni post-hoc effect of time: n.s.

SUPPLEMENTARY REFERENCES

1. Hammes J, *De Stroop kleur-woord test. Handleiding, tweede gewijzigde druk [The Stroop color-word test. Manual, second revised edition]*. 1978, Lisse: Swets & Zeitlinger.
2. Bowie CR, Harvey PD. Administration and interpretation of the Trail Making Test. Nat Protoc 2006; 1: 2277.
3. Reitan RM. The relation of the trail making test to organic brain damage. J Consult Psychol 1955; 19: 393.
4. Spielberger CD, *Manual for the State-Trait Anxiety Inventory*. Revised edition ed. 1983, Palo Alto, CA: Consulting Psychologists Press.
5. Meeuwis SH, van Middendorp H, Pacheco-Lopez G, Ninaber MK, Lavrijsen APM, van der Wee N, Veldhuijzen DS & Evers AWM. Antipruritic placebo effects by conditioning H1-antihistamines. Psychosom Med 2019; 81; 841-50.
6. Misery L, Jean-Decoster C, Mery S, Georgescu V, Sibaud V. A new ten-item questionnaire for assessing sensitive skin: the sensitive scale-10. Acta Derm Venereol 2014; 94: 635-9.