Universität Basel
Fakultät für Psychologie

Exploring the ethical use of placebo effects in affective states: current evidence and next steps

Inauguraldissertation zur Erlangung der Würde einer Doktorin der Philosophie vorgelegt der Fakultät für Psychologie der Universität Basel von

Dilan Sezer

aus Duggingen BL

Basel, 2023

Universität Basel

Fakultät für Psychologie

Genehmigt von der Fakultät für Psychologie auf Antrag von

Prof. Dr. Jens Gaab

Prof. Dr. Wilhelm Ruppen

Datum des Doktoratsexamen: 04.07.2023

Prof. Dr. Jens Gaab

Dekan der Fakultät für Psychologie

Universität Basel
Fakultät für Psychologie

Erklärung zur wissenschaftlichen Lauterkeit

Ich erkläre hiermit, dass die vorliegende Arbeit ohne die Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel selbstständig verfasst habe. Zu Hilfe genommene Quellen sind als solche gekennzeichnet. Des Weiteren versichere ich, sämtliche Textpassagen, die unter Zuhilfenahme KI-gestützter Programme verfasst wurden, entsprechend unter «auxiliary means» gekennzeichnet sowie mit einem Hinweis auf das verwendete KI-gestützte Programm versehen zu haben. Die veröffentlichten oder zur Veröffentlichung in Zeitschriften eingereichten Manuskripte wurden in Zusammenarbeit mit den Koautoren erstellt und von keinem der Beteiligten an anderer Stelle publiziert, zur Publikation eingereicht, oder einer anderen Prüfungsbehörde als Qualifikationsarbeit vorgelegt. Es handelt sich dabei um folgende Manuskripte:

1. **Study I:** Sezer, D., Locher, C., & Gaab, J. (2022). Deceptive and open-label placebo effects in experimentally induced guilt: a randomized controlled trial in healthy subjects. *Scientific Reports, 12*(1), 21219. https://doi.org/10.1038/s41598-022-25446-1
2. **Study II:** Buergler, S., Sezer, D., Bagge, N., Kirsch, I., Locher, C., Carvalho, C., & Gaab, J. (2023). Imaginary pills and open-label placebos can reduce test anxiety by means of placebo mechanisms. *Scientific Reports, 13*(1), 2624. https://doi.org/10.1038/s41598-023-29624-7
3. **Study III:** Buergler, S.*, Sezer, D.*, Gaab, J., Locher, C. (in Review). The role of population, expectation, modality, and comparator on open-label placebo effects: A network meta-analysis.
   **shared first authorship*

Spezifizierung des eigenen Forschungsbeitrags zu den Manuskripten:

1. Eigener Beitrag nach CRediT[1]:

☐ Conceptualization	☒ Data curation	☒ Formal Analysis
☐ Funding acquisition	☐ Investigation	☒ Methodology
☐ Project administration	☐ Resources	☒ Software
☐ Supervision	☐ Validation	☒ Visualization
☒ Writing – original draft		
☒ Writing – review & editing		

Das Manuskript wurde bisher für keine anderen Qualifikationsarbeiten eingereicht.

2. Eigener Beitrag nach CRediT[1]:

☒ Conceptualization	☒ Data curation	☒ Formal Analysis
☐ Funding acquisition	☒ Investigation	☒ Methodology
☐ Project administration	☐ Resources	☒ Software
☐ Supervision	☐ Validation	☒ Visualization
☒ Writing – original draft		
☒ Writing – review & editing		

Das Manuskript wurde bisher für keine anderen Qualifikationsarbeiten eingereicht.

3. Eigener Beitrag nach CRediT[1]:

☒ Conceptualization	☒ Data curation	☒ Formal Analysis
☐ Funding acquisition	☒ Investigation	☒ Methodology
☒ Project administration	☐ Resources	☒ Software
☒ Supervision	☒ Validation	☒ Visualization
☒ Writing – original draft		
☒ Writing – review & editing		

Das Manuskript wurde bisher für keine anderen Qualifikationsarbeiten eingereicht.

[1] https://casrai.org/credit/

Universität Basel

Fakultät für Psychologie

Open-Science Aspekte der Manuskripte:

1. Preregistration: ☐ ja ☒ nein
   Open-Access-Publikation: ☒ ja ☐ nein
   Open-Access-Data/Analyse: ☐ ja ☒ nein

2. Preregistration: ☒ ja ☐ nein
   Open-Access-Publikation: ☒ ja ☐ nein
   Open-Access-Data/Analyse: ☐ ja ☒ nein

3. Preregistration: ☒ ja ☐ nein
   Open-Access-Publikation: ☒ ja ☐ nein
   Open-Access-Data/Analyse: ☐ ja ☒ nein
   Ort/URL der Daten und Analysen: Meta-Analyse, Daten in einzelnen Studien

Ort, Datum Basel, 27. April 2023

Signatur

Vorname Nachname Dilan Sezer

Danksagung

Eine Dissertation lässt sich nicht ohne ein tragendes Umfeld schreiben. Daher möchte ich an dieser Stelle gerne wichtigen Wegbegleiter:innen danken. Ohne die unermessliche Unterstützung, die ich durch sie alle erfahren durfte, würde ich heute nicht am selben Punkt in meinem Leben stehen.

Mein erster Dank gilt Jens Gaab. Dank seinem Vertrauen habe ich erst die Möglichkeit erhalten, mich auf diesen Weg zum Doktortitel zu begeben und durch seine inspirierende, persönliche und stets zielgerichtete Begleitung habe ich die Orientierung auf diesem Weg nie verloren. Auch über die fachliche Begleitung hinaus hat mich die Zusammenarbeit geprägt. So kenne ich nun nicht nur wunderbare und einfachzuzubereitende Gerichte, sondern habe Mut mich einzusetzen und Bestehendes stets weiter zu denken. Weiter gilt mein Dank Wilhelm Ruppen, der mir in gleichem Ausmass stets Vertrauen und Zuversicht in mich und den Prozess geschenkt hat und ohne seine wertvolle Unterstützung diese Doktorarbeit nicht möglich geworden wäre. Durch ihn sowie durch Tobias Schneider konnte ich in den letzten 4 Jahren wertvolle Einblicke in die Behandlung von Schmerzen an der Schnittstelle zwischen Medizin und Psychologie gewinnen. Diese horizonterweiternde und stets sehr warme interdisziplinäre Zusammenarbeit hat mich sehr bereichert.

Auch mein Team hat mich durch diese intensive und in vielerlei Hinsicht sehr lehrreiche Zeit begleitet. Ob an der Fakultät für Psychologie oder am Universitätsspital, überall hatte ich das Glück, neben dem beruflichen Austausch auch Raum für persönliche Höhen und Tiefen zu haben. Dieser Kontext hat es mir immer ermöglicht, an meinem Arbeitsplatz nicht nur professionell, sondern auch authentisch zu sein. Ich bin sehr dankbar für diese Freiheit und die Freundschaften, die daraus entstanden sind.

Auch meinem Umfeld ausserhalb der Fakultät und des Unispitals möchte ich herzlich für die wertvolle Begleitung und Unterstützung in den letzten Jahren danken. Meinen Eltern und meiner Schwester danke ich für mein Fundament, das sie mit Liebe, Sicherheit, Neugierde und kritischem Denken angereichert haben und welches mir Vertrauen in mich mit auf den Weg gegeben hat. Zu wissen, dass Sie immer für mich da

sind, ist ein riesiges Geschenk. Neben meiner Familie gilt mein Dank auch meinen lieben Freundinnen. Das Lachen mit euch zaubert immer wieder Leichtigkeit in mein Leben. Ein spezieller und unmessbarer Dank gilt meinem Partner Philipp. Seitdem wir uns kennen inspiriert er mich jeden Tag aufs Neue und schenkt mir mit seiner Liebe und Unterstützung einen Boden, auf dem ich mich entfalten kann. Seine humorvolle, initiative und mitreissende Art durchs Leben zu gehen trägt mich durch jeden Tag.

Zu formulieren, wie unendlich dankbar ich für die tiefste Freundschaft, die ich je erleben durfte, und ja, man kann sagen, das gemeinsame durchs Leben gehen, bin, übersteigt diese Danksagung. Ohne Sarah wäre diese Erfahrung nicht dieselbe gewesen. Ohne sie hätte ich sie nicht machen wollen. Ihre mitreissende Begeisterungsfähigkeit, ihre Liebe für das Schöne im Leben und ihre durchdringende Herzlichkeit hat mich in jedem Bereich meines Lebens berührt und inspiriert, und mir ermöglicht neue Seiten an mir kennenzulernen. Danke für alles!

Table of contents

Abstract

Placebo effects have been recognized as essential psychobiological factors that significantly influence mental and physical well-being. Affective states, in particular, have demonstrated substantial placebo responses in clinical trials of antidepressants and a wide range of placebo effects in experimental studies. Nevertheless, utilizing these placebo effects without deceiving patients has only recently been considered, with initial studies showing promising results. However, the current body of research on such open-label placebos (OLPs) in affective states is limited, and conflicting results highlight the need for further research to strengthen and broaden our knowledge in this field. This thesis strives to contribute to addressing this research gap in multiple ways.

Firstly, it expands the existing pool of experimental affect paradigms for conducting basic placebo research by demonstrating its feasibility in a randomized controlled trial involving healthy volunteers (**Study I**). Besides successfully testing a new paradigm, this study reveals that OLPs can be as efficacious as deceptive placebos in reducing experimentally induced guilt. Secondly, a large randomized controlled trial replicates previous findings on the efficacy of OLPs in the treatment of preclinical test anxiety (**Study II**). It also provides evidence that placebo effects can be harnessed when participants imagine taking a pill, implying that placebo effects can be harnessed without the ingestion of physical pills. Thirdly, this thesis takes a meta-perspective by synthesizing the results of various studies in a network meta-analysis (**Study III**). The results of both preclinical and clinical studies provide a robust database demonstrating the efficacy of OLPs in regulating psychological complaints. These analyses further highlight that the treatment rationale provided with OLP interventions is indispensable for their efficacy.

The promising results of the three studies contribute collectively to the growing evidence base supporting the use of OLPs as a viable and ethical approach to harnessing placebo effects in the management of affect-related conditions. Moreover, they provide not only a new experimental paradigm for conducting experimental research but also valuable insights into which components of the OLP treatment regime are most critical. Crucially, they highlight that placebo effects extend beyond the effects of the mere intake of inactive remedies but instead emphasize the importance of both treatment explanations and the power of imagination. As such, these insights can inform future placebo research, leading to more efficacious and ethical interventions for individuals experiencing psychological distress.

Introduction

Traditionally, placebo effects were viewed as a nuisance that needed to be controlled for or minimized in clinical trials. However, they are now acknowledged as crucial psychobiological factors that manage a wide range of clinical and nonclinical symptoms and that should be maximized in treating mental and physical health complaints (Evers et al., 2018). However, due to their deceptive nature, the use of placebo treatments in clinical practice is problematic, and it is important to explore ethical treatment options that harness the promising effects of covertly administered placebos in affective disorders. The present work aims to strengthen the body of evidence regarding the potential of openly administered placebos (so-called open-label placebos; OLPs) in the context of affective states by investigating their efficacy, contributing to a better understanding of treatment mechanisms, and expanding methodological approaches to the study of OLPs.

To better understand why it is worthwhile to investigate the potential of OLPs to regulate affective states, it is useful to review some mechanistic explanations for deceptive placebos. One of which emphasizes the importance of the interplay between associative learning and appraisals in response to treatment context. According to Ashar et al. (2017) appraisals shape how individuals perceive their future health and the meaning of their symptoms, including expectations, self-evaluations, and beliefs about others. Thus, the treatment context and the treatment provided could promote a positive evaluation of the development of one's symptoms. As a result, automatic positive feedback loops can be activated, acting as self-fulfilling prophecies. For instance, expecting less pain can lead to less experienced pain, reinforcing the initial assumption of expecting less pain (Ashar et al., 2017).

Following this line of thought further, given the positive effects of optimistic expectations about future health and thus enhanced positive emotions, emotion and stress regulation are thought to be involved in the creation of the placebo effect (Colloca & Benedetti, 2007; Flaten et al., 2013; Flaten et al., 2011; Geers et al., 2021). This notion is supported by research indicating a close neurological connection between the appraisal and emotion regulation networks (Ashar et al., 2017). The involvement of emotion regulation in placebo effects can be exemplified by placebo analgesia: There, it has been shown that reduced brain activity in areas involved in pain processing following a placebo intervention co-occurs with decreased activity in areas that represent negative emotions (Wager & Atlas, 2015). This association is an intriguing finding further supporting the importance of emotion regulation when it comes to placebo effects and ultimately to

managing physical and mental health in general (DeSteno et al., 2013; Sheppes et al., 2015). Based on the neurological involvement of emotion regulation in placebo analgesia, Flaten et al. (2011) hypothesized that placebo-induced treatment expectations and thereby reduced negative emotions could be important mediators of the placebo effect. Placebos may automatically activate these processes (Ashar et al., 2017) by providing a plausible explanation for why a treatment should work. Such explanations can elicit a reassuring and hopeful perspective on the future evolvement of the experienced symptoms. Thus, one part of placebo responses could be explained by evaluating the helpfulness of a situation in managing one's symptoms and generating an emotionally colored expectation response to it (Kirsch, 2018).

Based on this emotional pathway of placebo effects it is unsurprising that affective states have not only been shown to mediate but are also prone to substantial placebo effects. Meta-analytical findings suggest this holds particularly for depression and anxiety, where substantial placebo response rates have been observed in first-line treatments like antidepressants (Bandelow et al., 2015; Khan & Brown, 2015; Kirsch, 2019; Locher, Koechlin, et al., 2017; Munkholm et al., 2019). Notably, in depression, only about 15 out of 100 patients treated with antidepressants show an antidepressant effect that exceeds the one of those treated with placebo (Stone et al., 2022). Similarly, only a difference in clinical response of 10% equaling approximately 2 points on the 17-item Hamilton depression rating scale (range 0–52 points) between placebo and antidepressant is reported (Khan et al., 2012; Khan et al., 2000; Munkholm et al., 2019). These negligible differences in clinical response across verum an placebo group are not only insignificant from a clinical perspective (Kirsch, 2019; Munkholm et al., 2019) but are also sided by significantly lower dropout rates due to fewer adverse side effects (Locher, Koechlin, et al., 2017). The considerable placebo effect in antidepressant trials along with various side effects, questions the use of antidepressants in the treatment of both anxiety and major depression and shed light on the clinical potential of placebo effects in affective and anxiety disorders.

Therefore, it is unsurprising that it has been voiced to utilize placebos in clinical contexts and to regulate daily affective experiences with high prevalence and burden through the use of placebo interventions (Geers et al., 2021). Up until the year of 2010, doing so would have entailed deception, as it has long been believed that deceiving patients is a necessary component of successful placebo interventions (Miller & Colloca, 2009). However, it is unethical to deceive patients, as deception is a violation of autonomy

and the right to make informed decisions about one's own healthcare (Annoni, 2018). In addition, and of equally significant importance, it has been pointed out that clinicians need to adequately inform about the mechanisms at play when using treatments that vastly work through placebo mechanisms, as is the case in antidepressants (Gaab et al., 2016).

Clinicians have thus two ethical options for treating patients based on current research: they can inform patients that a significant portion of the effects of antidepressants are due to non-specific effects rather than the active ingredient, or they can use the power of placebo effects in an open and collaborative manner. With regard to the latter, OLPs hold the promise of an exciting line of research. They involve openly informing patients that they receive a placebo without any active ingredients, yet may still experience improvements due to placebo mechanisms (e.g., positive treatment expectations, learning effects, and a patient-provider interaction). The OLP approach aims to harness the power of placebo effects while being transparent with patients about the nature of the treatment. Studies on OLPs have shown promising results in various conditions, including, chronic pain, irritable bowel syndrome, and ADHD (Charlesworth et al., 2017; von Wernsdorff et al., 2021).

With regard to affective disorders, however, a pilot study with a diagnosed sample of major depression found that the OLP group did not significantly differ from the no-treatment control group (Kelley et al., 2012). This finding can possibly be explained by a lack of power due to a small sample size of only 20 participants. A second study investigated OLPs as an add-on to treatment as usual in 38 depressed patients (Nitzan et al., 2020). There, symptoms of depression only decreased significantly in a subgroup of non-geriatric patients with an early onset of depression compared to the treatment as usual control group alone. Further evidence on the efficacy of OLP in depression stems from the study by Schienle and Jurinec (2022), which randomized 60 patients diagnosed with major depressive disorder to a 4-week cognitive-behavioral therapy (CBT) program with or without daily OLP treatment. The study concluded that, while the OLP add-on reduced symptoms of depression to a greater extent, the changes were not clinically meaningful in comparison to CBT alone, and the high dropout rate within the CBT + OLP group (27% vs. 7%) raised questions about the acceptance of OLPs in depression treatment. In summary, the current evidence base for OLP in depressed patients is weak, which is – in light of the well-documented placebo effects in antidepressant trials – surprising and raises the need for further investigations into OLP effects in depression.

Consequently, several studies have already aimed at zooming in on the complex phenomenology of OLP effects in depression and have focused on preclinical populations or individual symptoms. In contrast to the available clinical OLP findings in depression, they found OLPs to have significant effects on emotions, including sadness (Haas et al., 2020; Hahn et al., 2022), general emotional well-being (El Brihi et al., 2018), and subjective and objective emotional distress (Guevarra et al., 2020). Along these lines, OLP interventions have also been shown to be efficacious for preclinical test anxiety in students (Schaefer et al., 2019) and in state anxiety and acute stress in individuals who believed in the effectiveness of placebos (Schaefer et al., 2021).

Despite these promising findings in experimental and preclinical trials, there are also studies that failed to replicate the positive findings of earlier studies testing the efficacy of OLPs in affective states: For instance, Bräscher et al. (2022) in emotional well-being and Friehs et al. (2022) in experimentally induced sadness. In the case of Bräscher et al. (2022), slight methodological differences (i.e., cross-over design vs. parallel study, prospective vs. retrospective symptom assessment, and video vs. personal interaction) are being discussed by the authors as potential explanations contributing to differences in effects across studies. The aforementioned examples emphasize the crucial role of replication (Shrout & Rodgers, 2018) – especially in a young research field such as OLPs – in determining which trial design features promote or hinder the detection of potential effects.

Not only replication of studies is crucial in determining the robustness of findings, but also the statistical aggregation of effects across different studies is of great importance. The network meta-analytic approach offers a unique way to enhance the analytical power by incorporating both direct and indirect evidence from a network of studies to estimate the effects (Salanti, 2012). This is especially crucial in fields where the number of studies and therefore direct comparisons is limited, such as the one of OLPs. For instance, when a specific intervention is only compared to a no-treatment control group in a single study, the network meta-analytic approach enables a comparison of this intervention with other control groups used in other studies, leveraging indirect evidence. Furthermore, it is worth noting that the currently available meta-analyses have either grouped all clinical (Charlesworth et al., 2017; von Wernsdorff et al., 2021) or all nonclinical (Spille et al., 2023) conditions together. As such, there is a lack of evidence on the specific effects of OLPs in distinct populations, such as pain and psychological conditions. Furthermore, comparing the effects across clinical and nonclinical samples is

impaired, as different inclusion criteria are employed across meta-analyses. Therefore, updating the existing analyses with the continually expanding study base on OLPs and performing subgroup analyses tailored to specific populations is essential.

Not only with regard to the robustness of results but also with regard to the underlying mechanisms of OLP effects there is a lot yet unknown. To close this research gap, researchers have tried to model key affects by means of various experimental models, thereby enabling to gain further insights into mediating processes and crucial intervention components. Various laboratory studies have indicated that it is not only possible to induce affective states but also that deceptive placebos can mitigate the negative emotional impact of these different stimuli. Examples include watching disturbing images (Schienle et al., 2016; Schienle et al., 2014) or sorrowful movie clips (Glombiewski et al., 2019), listening to mood-suggestive music and engaging in autobiographic recall of upsetting memories (Rebstock et al., 2020), anticipating a painful stimulus (Meyer et al., 2019), or public speaking tasks (Abrams et al., 2001). Despite this range of experimental paradigms, still, not all affects can be mimicked, raising the need for more experimental models.

For example, depression is a complex and multifaceted condition that encompasses a wide range of symptoms with strong emotional components, such as low self-esteem, feelings of hopelessness, self-blame, worthlessness, and guilt (American Psychiatric Association, 2013). To gain a deeper understanding of the key processes of both deceptive and OLP effects in affective states, it is essential to establish additional experimental models. These models would allow employing sophisticated research methods such as neuroimaging techniques, to broaden our understanding of these phenomena. In addition to the experimental affect induction, pre-clinical studies offer a good means to investigate placebo effects while reducing the burden of study participation or the challenge of delayed obtainment of treatment for patients (Benedetti, 2021). Furthermore, in comparison to experimental models, a subclinical population allows the investigation of the impact over more extended periods and thus increases the external validity (Vase et al., 2005). However, although both proposed alternative means to clinical trials to conduct placebo research in affective states hold practical and ethical advantages, naturally, they also diminish the generalizability of results to clinical populations. Despite the potential drawbacks, the benefits of gaining a deeper understanding of the mechanisms involved and the systematic investigation of critical intervention features may still outweigh the disadvantages.

In summary, compelling evidence supports the theoretical potential of OLPs to improve the treatment of affective states by reducing side effects, upholding efficacy, and maintaining ethical standards. On an empirical level, the current evidence base for the efficacy of OLP in affect regulation is mixed, both in clinical and experimental studies. As the research field of OLPs in affective states is still in its infancy, with the first study published in 2012 (Kelley et al., 2012), there is a need for more research. This research should aim at investigating which affects and within which populations OLPs can effectively alleviate symptoms as compared to control conditions. Experimental and preclinical studies might, in particular, help shed light on the potential of OLP in depressed and anxious affect regulation. To gain a deeper understanding of the robustness of the effects of individual studies, it is furthermore essential to replicate them and aggregate individual studies by means of conducting network meta-analyses. Such analyses not only increase analytical power but can elucidate mechanistic and methodological features, such as differences across control groups or intervention kinds, thereby going beyond mere efficacy evaluation.

Research Goal

This thesis aims to contribute to the growing body of research investigating the potential of transparently administered placebos (i.e., OLPs) to harness placebo effects in affective states. Thereby contributing to understanding how treatments for affective and anxiety disorders can be optimized by employing state-of-the-art informed consent and thereby upholding patient autonomy. In addition, this thesis addresses key OLP research challenges by first investigating a new model to induce a widely common affective state (guilt) in healthy individuals to study the acute effects of both deceptive and open-label placebos (Study I). Furthermore, it aims at replicating findings that suggest the possibility that OLPs can be used as a treatment for preclinical test anxiety while meanwhile investigating whether a physical pill is necessary to induce OLP effects (Study II). Finally, it paves novel statistical avenues by network meta-analytically aggregating evidence across different fields of OLP research with a particular focus on the effects of OLP within psychological conditions (Study III).

Methods and Results

Three different studies were conducted to gain a deeper understanding of the potential of OLPs to affect regulation. From a methodological point of view, each study addressed this question on a different level of generalizability, from basic research over a preclinical trial to a network meta-analytic approach. The following sections briefly summarize the employed methods and key findings and provide additional results that are relevant to this thesis research questions.

Study I was conducted in a laboratory setting and aimed at expanding the range of experimental paradigms to conduct basic placebo research in affective disorders. In doing so, this study served as a proof-of-concept study for an experimental guilt induction paradigm used for the first time within the context of placebo research and based on an autobiographic writing task. In addition, this study aimed to examine the potential of deceptive and open-label placebo effects to reduce experimentally induced guilt in healthy students. Therefore, following the experimental induction of guilt, participants were randomized to receive a deceptive placebo (DP; n = 35), an OLP (n = 35), or no treatment (NT; n = 39). The primary outcome was subjective guilt responses assessed in the area under the curve (AUC) across several time points within the duration of the experiment. Secondary outcomes were shame, pride, and affect. We hypothesized that DP and OLP would reduce guilt compared to NT. Results showed that guilt responses were higher in the NT group than in the placebo groups, estimate = 2.03, 95% CI = 0.24–3.82, d = 0.53, whereas AUC guilt did not differ significantly between the placebo groups, estimate = −0.38, 95% CI = −2.52–1.76, d = −0.09. This effect was not observed for any of the secondary outcomes (see the original publication in Appendix A for more details on the study rationale, methods, and results).

In **Study II**, we aimed to open up the lens by moving on from the short-term laboratory setting to a preclinical sample of test-anxious individuals to replicate previous findings (Schaefer et al., 2019) on the efficacy of OLP across the time span of three weeks. In addition, this study set out to investigate one of many unanswered questions regarding the crucial components of OLP efficacy by evaluating placebo effects without using a physical pill, i.e., by imagining taking a pill. To pursue this research goal, we applied in healthy students with self-reported test anxiety either an OLP intervention (n = 59) or a newly developed imaginary pill intervention (IP; n = 55) that was based on knowledge derived from placebo and imagination research to compare them to a control group (CG;

n = 59). Both intervention groups were instructed to take two pills daily for three weeks. The primary outcome was test anxiety, and the secondary outcomes were sleep quality, general well-being, and test performance. Groups' test anxiety differed already after one week of intervention, $p_{adj} < .001$, and this effect was maintained through the end of week three, $F(2,169) = 11.50$, $p < .001$. Test anxiety was lower in the intervention groups compared to the CG, $t(169) = -4.44$, $p < .001$, $d = -0.71$, and the interventions did not differ significantly, i.e., both were similarly efficacious, $t(169) = 0.61$, $p = .540$, d = 0.11. Regarding the secondary outcomes, a statistically significant interaction was found between group and time in general well-being, $F(5, 422.71) = 3.58$, $p = .004$, but not on sleep quality, $F(5, 443.85) = 0.90$, $p = .485$ (see the original publication in Appendix B for more details on the study rationale, methods, and results).

Finally, **Study III** further zoomed out and took a meta-perspective on the critical determinants for OLP efficacy across different studies and populations. This serves to understand better why, under what circumstances (e.g., through which routes of administration or against which control groups), and in which populations (e.g., healthy vs. clinical and pain vs. psychological) OLPs work. Within this thesis, a special focus was placed on a subgroup of studies that investigated the effects of OLPs on psychological complaints. To tackle these research goals, we conducted the first network meta-analyses (NMA) in the field of OLPs. Our analyses revealed that OLPs can be beneficial compared to NT in nonclinical (12 trials; 1'015 participants) and preclinical and clinical populations combined (25 trials; 2'006 participants), with a trend for larger effects in clinical conditions. Included preclinical/clinical studies in the psychological roam investigated cancer-related fatigue (*n* = 3), depression (*n* = 2), well-being (*n* = 2), insomnia (*n* = 1), test anxiety (*n* = 1), and relaxation (*n* = 1). The nonclinical studies investigating OLP effects on psychological states focused on acute stress (*n* = 1) and sadness (*n* = 2). While in the whole network, the kind of OLP administration route had no substantial impact on OLP effects, positive treatment expectations delivered either as verbal suggestions or through conditioning were found to be essential. Further, our analyses showed that OLP effects within clinical samples can vary depending on the comparator used. Within the psychological subsamples, the results of the whole network were generally replicated. For instance, within the clinical sample, OLP pills were significantly better than NT (SMD = 0.44) and waitlist (WL; SMD = 0.47) but only when provided with a plausible treatment rationale (see Table 1). However, within the nonclinical sample, OLP pills did not outperform NT (SMD = -0.02), whereas OLPs administered as nasal spray were

significantly better than NT (SMD = 0.62), which was the only comparator in the network (see Table 2). Reducing the variance of clinical conditions within the subgroup also lead to a decrease in heterogeneity in both subgroup networks, preclinical/clinical: I^2 = 26.5% (all) to I^2 = 0% (psychological only); nonclinical: I^2 = 66% (all) to I^2 = 0% (psychological only; see the original publication and its supplement in Appendix C for more details on the study rationale, methods, and results).

Table 1. Head-to-head comparisons of network meta-analysis on preclinical/clinical psychological studies only

Pair-wise meta-analysis			
NT	**-0.44** [-0.68; -0.21]	.	.
-0.44 [-0.68; -0.21]	**OLP pills**	0.14 [-0.45; 0.72]	**0.47** [0.22; 0.72]
-0.30 [-0.93; 0.33]	0.14 [-0.45; 0.72]	**OLP-**	.
0.03 [-0.32; 0.37]	**0.47** [0.22; 0.72]	0.33 [-0.31; 0.96]	**WL**
Network meta-analysis			

Note. Displayed are SMDs and 95% confidence intervals. Significant comparisons are marked in bold. Green cells represent combined evidence of indirect and direct comparisons, and orange cells are direct comparisons only, NT = no treatment, OLP = open-label placebo, OLP- = open-label placebos without rationale, WL = waitlist.

Table 2. Head-to-head comparisons of network meta-analysis on nonclinical psychological studies only

Pair-wise meta-analysis			
DP	**0.86** [0.40; 1.32]	**0.39** [0.03; 0.76]	.
0.97 [0.56; 1.37]	**NT**	**-0.62** [-0.94; -0.31]	0.02 [-0.53; 0.56]
0.34 [-0.01; 0.70]	**-0.62** [-0.94; -0.31]	**OLP nasal**	.
0.98 [0.31; 1.66]	0.02 [-0.53; 0.56]	**0.64** [0.01; 1.26]	**OLP pills**
Network meta-analysis			

Note. Displayed are SMDs and 95% confidence intervals. Significant comparisons are marked in bold. Green cells represent combined evidence of indirect and direct comparisons, and orange cells are direct comparisons only, DP = deceptive placebo, NT = no treatment, OLP = open-label placebo.

General Discussion

The overarching aim of this thesis was to contribute to the growing literature that explores the potential of OLP to be an ethical and efficacious treatment approach for affective disorders. The investigation of this question through three different studies each on a different level of generalizability has revealed promising effects within the experimental and preclinical setting. In addition, several important contributions to a better understanding of key OLP components can be derived from the studies, and two new methodological research approaches were tested that advance the field of OLP studies in affective disorders.

In **Study I**, a novel experimental method was effectively tested, utilizing an autobiographical writing task to evoke guilt. The findings of this study suggest that guilt can be intentionally induced in healthy students through experimental means and that placebos are efficacious in mitigating immediate guilt responses (d = 0.53), irrespective of whether they are administered openly or deceptively. Furthermore, we observed narrative-specific effects with significant changes in guilt but not shame, pride, or affect. These results indicate not only that guilt is amenable to placebos but also that placebos can be administered in an ethical and potentially emotion-specific manner. The **second study's** findings complement the notion that OLPs can significantly reduce affective states. This could be shown by replicating previously found effects of OLP in a large sample (N = 173) of students that suffered from preclinical test anxiety. Intriguingly, beyond the mere replication, this is the first study to show that the sheer imagination of the pill-taking act in combination with a plausible treatment rationale is sufficient to induce medium to large effects (combined OLP and IP effect size in comparison to CG: d = 0.71). This finding emphasizes the significant influence of the psychological component of the OLP effect on test anxiety. Finally, **Study III** consolidated the results of the first two studies by means of the network meta-analytical aggregation of previously published studies that investigated the efficacy of ethical placebo administration including a separate analysis for studies with psychological complaints as outcomes. Results indicate a solid evidence base for preclinical and clinical studies that use OLP pills to regulate affective states ($N_{network}$ = 10). In addition, nasally administered OLPs appear promising in reducing experimentally induced affective distress ($N_{network}$ = 3). The network as a whole, including various conditions, suggests that OLPs can be an efficacious intervention for nonclinical ($N_{network}$ = 12) and preclinical/clinical populations ($N_{network}$ = 25), with preclinical/clinical samples benefiting to a greater extent from the treatment altogether. Notably, a salient

and clear finding was that embedding the pill administration in a narrative (i.e., rationale) is indispensable and, thus, an essential component of the success of OLPs, and that simply delivering a pill is not sufficient. Finally, there is evidence within the clinical network that the efficacy of OLP pills does not exceed those of standard treatments as they elicited statistically greater effects compared to no treatment but not as compared to treatment as usual.

With regard to the empirical question of whether OLPs are efficacious for managing affective states, these three studies contribute collectively to the growing literature supporting the use of OLP as a viable, efficacious, and ethical approach to harnessing placebo effects in affective states. Specifically, the two randomized-controlled studies demonstrate the efficacy of OLPs in two distinct emotions. In the context of test anxiety (Study II), the effects of the intervention were not only robust but also medium to large in size and consistent across subscales, such as worry, emotionality, interference, and lack of confidence. Importantly, these benefits were sustained over the three-week intervention period, with a notable improvement observed as early as the first week of the intervention. Intriguingly, the observed effects (d = 0.71) are comparable in size to existing psychological interventions for test anxiety (g = -0.76; Huntley et al., 2019). In addition, drawing on the result from the experimental guilt study (Study I) OLP interventions appear to be equally efficacious as DPs. These findings align with studies reporting that DPs and OLPs delivered with a treatment rationale have similar effects in healthy subjects enduring experimentally induced pain (Locher, Frey Nascimento, et al., 2017). This finding highlights the potential of OLPs to harness placebo effects in these conditions ethically. However, the results from some studies provide conflicting evidence (see discussion in Study I for more details) suggesting that the comparative efficacy of DPs and OLPs may depend on the target condition. In conclusion, the findings, along with the evidence from the NMAs of the subgroups with psychological conditions (Study III), strongly support the efficacy of OLPs in reducing psychological distress, particularly when accompanied by a plausible treatment explanation.

With their promising results, the three studies align with an increasing number of individual investigations in which OLPs have significant effects on emotions, including anxiety (Schaefer et al., 2019; only in those that believed strongly in placebos (Schaefer et al., 2021)), sadness (Haas et al., 2020; Hahn et al., 2022), general emotional well-being (El Brihi et al., 2018; Kleine-Borgmann et al., 2021), and emotional distress (Guevarra et al., 2020). In addition, emerging evidence suggests that the effects of OLPs are not limited

to subjective outcomes but can also be observed at the neurological level. For instance, a study by Guevarra and colleagues (2020) demonstrated that subjectively perceived reductions in emotional arousal in response to OLPs correlate with the reduction of an objective neural marker of emotional distress. Further evidence for OLP effects on objective markers of psychological distress stems from a study by (Schaefer et al., 2021) that found significantly reduced cortisol responses in healthy individuals that suffered from acute stress and that believed in the power of the ingested placebo pill. As stated by Guevarra et al. (2020), these findings of OLP efficacy on objective markers provide preliminary evidence that OLPs may not be solely driven by response bias but may reflect "genuine psychobiological effects". These studies refute a key criticism of the effects of placebo studies (Hróbjartsson et al., 2011) and open up new avenues for practical applications of OLPs, paving the way for further exploration of their use in clinical practice.

This exploration of OLP effects can extend beyond the management of emotional well-being alone. Acute episodes of emotional distress can also impact various physical (e.g., cardio vascular disease: DeSteno et al., 2013) and psychiatric conditions (Sheppes et al., 2015). Effective management of emotional distress through placebo interventions can therefore have far-reaching implications for both mental and physical health. One practical application of OLPs could be their use as co-interventions with existing therapies, for example, psychotherapy. By incorporating OLPs into existing treatment plans, clinicians can potentially enhance the therapeutic effects of traditional interventions and improve patient outcomes. Moreover, OLPs have the potential to influence the accompanying emotional challenges of various somatic conditions, especially those involving pain, as emotional distress can exacerbate existing pain and increase the likelihood of chronic pain (Lumley et al., 2011). Therefore, if the evidence base for both subjective and objective outcomes regarding affect management increases further, leading to more robust effects, OLPs hold the potential for improved management of both emotional and physical symptoms, particularly for conditions characterized or sided by negative emotions.

Despite this promising outlook, it remains to be seen whether the findings of experimental and preclinical studies (e.g., the ones testing objective markers as mentioned above and Study I and II of this thesis) can be applied to clinical populations, such as those with major depression. The available literature on OLP efficacy in depression presents conflicting evidence, with only one of the three small clinical studies (N = 20 – 60; Kelley et al., 2012; Nitzan et al., 2020; Schienle & Jurinec, 2022) reporting

a reasonable effect size (e.g., d = 0.54 in Kelley et al., 2012) and all of them finding no statistically significant differences in comparison to control groups. In light of the reasonably sized and well-researched placebo response rates in antidepressant trials, these findings are surprising. Moreover, this raises questions about whether these rates are largely attributed to regression to the mean and natural history effects, potentially resulting in an overestimation of the placebo and drug effect's magnitude. However, it is important to note, as mentioned above, that all three studies had small sample sizes, suggesting that they may have been underpowered. Aggregated results from the NMAs (Study III) provide evidence that as sample sizes increase, either within individual trials or meta-analytic aggregation, more significant benefits of OLP on psychological complaints may be observed. Thus, increasing statistical power in clinical studies could help transfer the effects observed in experimental and preclinical studies to clinical populations.

The findings of this thesis go beyond the evidence for the efficacy of OLPs in affective states. From a mechanistic point of view, this thesis also sheds light on important key components of OLP interventions. Firstly, Study I and III highlight the importance of the treatment explanation with regard to different aspects. Regarding the content of the rationale, the finding of Study I that the symptom-specific rationale might have led to a symptom-specific effect points to an exciting line of research that needs to be systematically addressed in future studies. Suppose future trials confirm the findings of symptom specificity, the differential effects of placebos across disorders, populations, and settings (Benedetti, 2021) could be seen as specific to the rationales employed (questioning whether placebo effects actually are unspecific effects). To date, only three studies have investigated the extent to which different types of rationales can lead to varying effects in OLPs. Yet, two of these studies found no significant OLP effects of either rationale condition compared to controls. This negative finding makes it difficult to infer the influence of the different kinds of employed rationales. These rationales included hope vs. expectations (Kube et al., 2020) and personal-emotional vs. scientific-matter-of-fact style (Friehs et al., 2022). Only one study has tested the effects of different target symptoms that are mentioned within the rationale. This study showed that pain and mood could be positively affected by OLP treatment, regardless of whether the treatment was intended to treat pain or improve mood, suggesting that rationale-specific effects might not be observable in osteoarthritis (Olliges et al., 2022). Thus, while there is inconclusive and very limited evidence on the importance of the content and the information style of the rationale, there is compelling evidence from the NMAs of Study III that explaining why

the given treatment is intended to help manage symptoms is critical. Several studies have shown that OLPs provided without explaining the treatment and its expected effects are generally ineffective (e.g., Barnes et al., 2023; Locher, Frey Nascimento, et al., 2017), except for one study (Schaefer et al., 2018). Secondly, the findings of Study II add to the notion that the intrapersonal aspect of the OLP intervention is more critical than its physical constituents, by showing the efficacy of an imaginary pill intake procedure in reducing experienced test anxiety over a period of three weeks. With this finding, this study aligns with an increasing number of reports that show that placebo effects can also be harnessed without physical treatment components (e.g., by changing mindsets, using psychological placebos, or boosting treatment expectancies; Crum & Langer, 2007; Gaab et al., 2019; Kong et al., 2018). Thus, while it appears that physical treatment components can be substituted by psychological interventions and mental rituals, the provision of a treatment rationale seems to be indispensable for OLP efficacy.

Adding to the findings regarding the efficacy and important intervention constituents, this thesis also explored new methodological approaches in OLP research. First, we found that the employed guilt paradigm (Study I) exerted its intended effects by inducing guilt as a consequence of writing (“guilt induction”) and thinking (“guilt boost”) about an interpersonally unfair behavior toward another person. These results align with other studies testing this approach (Schär et al., 2022) and open new possibilities for conducting experimental placebo research on affective states. For example, experimental paradigms facilitate the systematic manipulation of the treatment setting and application, which can aid our understanding of the mechanisms involved in how OLPs influence affective states without burdening patients. The network meta-analytical approach employed in Study III is another important methodological contribution. This approach allowed us to assess the importance of the comparator group selection through indirect evidence across multiple studies. Specifically, we found that some OLP interventions resulted in larger effects when compared to NT or WL groups as opposed to treatment as usual groups. Additionally, we were able to compare different types of OLP modalities and constituents using this methodology (i.e., identifying the deciding factor of the rationale, see above). As the number of studies per node within the network grows, future research could investigate other treatment aspects, such as whether OLP is used as a stand-alone or an add-on treatment or the effect of different dosages. Further implications for future studies arising from Study III are discussed in the next section.

Future Studies

The three conducted studies have aided our understanding of OLP effects in general, but in particular concerning their efficacy in affective states. However, beyond answering several crucial questions, they also raise many new and thereby point towards interesting future research avenues. In particular, the adoption of a meta-perspective through the conductance of the NMAs (Study III) identified several research gaps that are of particular importance to the investigation of the efficacy of OLPs in affective disorders: First, larger studies should be conducted, as sample sizes are often relatively small (range clinical and preclinical psychological studies: 20 - 154). Second, the population should be more representative (i.e., a high percentage of females and a mean age < 40 often characterize the samples). Third, it would be crucial to conduct future studies by more independent research teams with less allegiance to OLP research and in countries other than Germany and the USA (i.e., all the experimental psychological paradigms included in Study III have been conducted in Germany). Fourth, in future studies, the control group used should be chosen deliberately (e.g., no treatment or treatment as usual) because depending on the type of control group – as Study III shows – different sizes of effects can result. Fifth, further experimental studies should be designed more according to the needs of clinical populations: For instance, the modalities OLP nasal (e.g., sprays) were only studied in nonclinical populations in sadness and not in any of the preclinical and clinical psychological conditions. This possibly indicates that this route of administration is not suitable for clinical conditions. Sixth, to minimize the impact of response bias and social desirability bias, future research should consider expanding the range and type of affect measures beyond self-report measures. This could involve incorporating physiological and behavioral measures in addition to self-report measures (Mauss & Robinson, 2009).

Finally, as there is network meta-analytic evidence that OLPs do not outperform currently employed treatment as usual, non-inferiority trials could aid in understanding whether this is also the case for OLPs compared to current first-line therapies for affective disorders. By comparing the effects of OLP with those of for instance antidepressants or psychotherapy within one study, meaningful information on the comparative efficacy of these treatments can be obtained. By that, avoiding some of the ethical and practical challenges associated with traditional superiority trials (i.e., in noninferiority trials both groups receive an effective treatment). However, it is essential to note that non-inferiority trials also have their unique features that require careful consideration from an ethical standpoint (Garattini, 2007). Nevertheless, a direct comparison of OLPs with first-line

treatments is essential in establishing whether OLPs hold any practical implications. Therefore, it is crucial to recognize that merely demonstrating superiority to no treatment may not be sufficient to do so. To be considered an efficacious treatment option, OLPs must demonstrate non-inferiority, or equivalence, to existing treatments to be regarded as a safe and efficacious alternative to conventional treatments.

Limitations

Naturally, each of the studies conducted as part of this thesis holds its limitations that have been discussed in the respective manuscripts. However, two crucial aspects warrant critical investigation as they apply to each of the three studies: First, as already touched upon within the main body of the general discussion, the generalizability of the presented results beyond the studied populations is limited due to the selective inclusion criteria of the randomized controlled trials (e.g., healthy students in Study I and II) that lead to predominantly female and on average very young samples (i.e., < 25 years) that can be assumed to have good strategies to deal with aversive emotions. Upon inspection of the samples included within the subgroup of psychological studies in Study III, this criticism also applies to the studies included in the network. Second, both randomized controlled trials (i.e., Study I and II), as well as all studies that tested a psychological outcome and were thus included in the subgroup of the network (Study III), relied on self-reported subjective outcomes, which could favor the emergence of various biases (e.g., response bias or social desirability bias).

Conclusion

Due to substantial side effects and limited response rates beyond placebo, first-line pharmacological treatments for anxiety and mood disorders have been subject of debate, and the need for alternative treatment approaches has been voiced. Therefore, this thesis focused on exploring the potential of OLP interventions in regulating affective states, while simultaneously investigating some underlying mechanisms and testing novel methodological approaches. To address the complexity of these goals, a multi-method approach was chosen, involving three studies that address this question of efficacy at different levels of generalizability toward clinical populations.

Collectively the three studies support the idea of OLP interventions as a promising and ethical alternative to traditional treatments for managing affective states. Moreover, the research presented in this thesis suggests that placebo effects go beyond the mere

ingestion of inactive remedies (Benedetti et al., 2003). Instead, the findings emphasize the significance of treatment explanations and the power of imagination in eliciting placebo effects. To determine whether effects from experimental and preclinical studies can be transferred to multi-symptom diseases such as major depression, larger future studies in clinical populations are necessary. Additionally, an increasing number of experimental paradigms, such as the one that was successfully tested within this thesis, will allow researchers to answer questions about the importance of treatment explanation content and style.

The growing body of present and future research contributes to the investigation of ethically justifiable usage of placebo effects for the benefit of patients. If the evidence base for both subjective and objective outcomes regarding affect management further increases, leading to more robust effects, OLPs hold the potential to improve the management of both emotional and physical symptoms, particularly for conditions characterized or sided by negative emotions.

References

Abrams, K., Kushner, M., Medina, K. L., & Voight, A. (2001). The pharmacologic and expectancy effects of alcohol on social anxiety in individuals with social phobia. *Drug and Alcohol Dependence*, *64*(2), 219-231. https://doi.org/10.1016/S0376-8716(01)00125-9

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Association.

Annoni, M. (2018). The ethics of placebo effects in clinical practice and research. *International Review of neurobiology*, *139*, 463-484. https://doi.org/10.1016/bs.irn.2018.07.031

Ashar, Y. K., Chang, L. J., & Wager, T. D. (2017). Brain mechanisms of the placebo effect: An affective appraisal account. *Annual Review of Clinical Psychology*, *13*, 73-98. https://doi.org/10.1146/annurev-clinpsy-021815-093015

Bandelow, B., Reitt, M., Röver, C., Michaelis, S., Görlich, Y., & Wedekind, D. (2015). Efficacy of treatments for anxiety disorders: A meta-analysis. *International clinical psychopharmacology*, *30*(4), 183-192. https://doi.org/10.1097/YIC.0000000000000078

Barnes, K., Babbage, E., Barker, J., Jain, N., & Faasse, K. (2023). The role of positive information provision in open-label placebo effects. *Applied Psychology: Health and Well-Being*. https://doi.org/10.1111/aphw.12444

Benedetti, F. (2021). *Placebo effects: Understanding the mechanisms in health and disease* (Third edition. ed.). Oxford University Press.

Benedetti, F., Maggi, G., Lopiano, L., Lanotte, M., Rainero, I., Vighetti, S., & Pollo, A. (2003). Open versus hidden medical treatments: The patient's knowledge about a therapy affects the therapy outcome. *Prevention & treatment*, *6*(1), 1a. https://doi.org/10.1037/1522-3736.6.1.61a

Bräscher, A.-K., Ferti, I.-E., & Witthöft, M. (2022). Open-label placebo effects on psychological and physical well-being: A conceptual replication study. *Clinical Psychology in Europe*, *4*(4), 1-16. https://doi.org/10.32872/cpe.7679

Charlesworth, J. E. G., Petkovic, G., Kelley, J. M., Hunter, M., Onakpoya, I., Roberts, N., Miller, F. G., & Howick, J. (2017). Effects of placebos without deception compared with no treatment: A systematic review and meta-analysis. *Journal of Evidence-Based Medicine*, *10*(2), 97-107. https://doi.org/10.1111/jebm.12251

Colloca, L., & Benedetti, F. (2007). Nocebo hyperalgesia: How anxiety is turned into pain. *Current Opinion in Anesthesiology*, *20*(5), 435-439. https://doi.org/10.1097/ACO.0b013e3282b972fb

Crum, A. J., & Langer, E. J. (2007). Mind-set matters: Exercise and the placebo effect. *Psychological science*, *18*(2), 165-171.

DeSteno, D., Gross, J. J., & Kubzansky, L. (2013). Affective science and health: The importance of emotion and emotion regulation. *Health Psychology*, *32*(5), 474. https://doi.org/10.1037/a0030259

El Brihi, J., Horne, R., & Faasse, K. (2018). Prescribing placebos: An experimental examination of the role of dose, expectancies, and adherence in open-label placebo effects. *Annals of Behavioral Medicine*, *53*(1), 16-28. https://doi.org/10.1093/abm/kay011

Evers, A. W., Colloca, L., Blease, C., Annoni, M., Atlas, L. Y., Benedetti, F., Bingel, U., Büchel, C., Carvalho, C., & Colagiuri, B. (2018). Implications of placebo and nocebo effects for clinical practice: Expert consensus. *Psychotherapy and psychosomatics*, *87*(4), 204-210. https://doi.org/10.1159/000490354

Flaten, M. A., Aslaksen, P. M., & Lyby, P. S. (2013). Positive and negative emotions and placebo analgesia. In *Placebo and Pain* (pp. 73-81). Elsevier. https://doi.org/10.1016/B978-0-12-397928-5.00008-8

Flaten, M. A., Aslaksen, P. M., Lyby, P. S., & Bjørkedal, E. (2011). The relation of emotions to placebo responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1572), 1818-1827. https://doi.org/10.1098/rstb.2010.0407

Friehs, T., Rief, W., Glombiewski, J. A., Haas, J., & Kube, T. (2022). Deceptive and non-deceptive placebos to reduce sadness: A five-armed experimental study. *Journal of Affective Disorders Reports*, *9*, 100349. https://doi.org/10.1016/j.jadr.2022.100349

Gaab, J., Blease, C., Locher, C., & Gerger, H. (2016). Go open: A plea for transparency in psychotherapy. *Psychology of Consciousness: Theory, Research, and Practice*, *3*(2), 175. https://doi.org/10.1037/cns0000063

Gaab, J., Kossowsky, J., Ehlert, U., & Locher, C. (2019). Effects and components of placebos with a psychological treatment rationale–three randomized-controlled studies. *Scientific reports*, *9*(1), 1-8. https://doi.org/10.1038/s41598-018-37945-1

Garattini, S. (2007). Non-inferiority trials are unethical because they disregard patients' interests. *The Lancet*, *370*(9602), 1875-1877. https://doi.org/10.1016/S0140-6736(07)61604-3

Geers, A. L., Faasse, K., Guevarra, D. A., Clemens, K. S., Helfer, S. G., & Colagiuri, B. (2021). Affect and emotions in placebo and nocebo effects: What do we know so far? *Social and Personality Psychology Compass*, *15*(1), e12575. https://doi.org/10.1111/spc3.12575

Glombiewski, A. J., Rheker, J., Wittkowski, J., Rebstock, L., & Rief, W. (2019). Placebo mechanisms in depression: An experimental investigation of the impact of expectations on sadness in female participants. *Journal of affective disorders*, *256*, 658-667. https://doi.org/10.1016/j.jad.2019.06.070

Guevarra, D. A., Moser, J. S., Wager, T. D., & Kross, E. (2020). Placebos without deception reduce self-report and neural measures of emotional distress. *Nature communications*, *11*(1), 1-8. https://doi.org/10.1038/s41467-020-17654-y

Haas, J. W., Rief, W., Glombiewski, J. A., Winkler, A., & Doering, B. K. (2020). Expectation-induced placebo effect on acute sadness in women with major depression: An experimental investigation. *Journal of affective disorders*, *274*, 920-928. https://doi.org/10.1016/j.jad.2020.05.056

Hahn, A., Göhler, A. C., Hermann, C., & Winkler, A. (2022). Even when you know it is a placebo, you experience less sadness: First evidence from an experimental open-label placebo investigation. *Journal of affective disorders*. https://doi.org/10.1016/j.jad.2022.02.043

Hróbjartsson, A., Kaptchuk, T. J., & Miller, F. G. (2011). Placebo effect studies are susceptible to response bias and to other types of biases. *Journal of Clinical Epidemiology*, *64*(11), 1223-1229. https://doi.org/10.1016/j.jclinepi.2011.01.008

Huntley, C. D., Young, B., Temple, J., Longworth, M., Smith, C. T., Jha, V., & Fisher, P. L. (2019). The efficacy of interventions for test-anxious university students: A meta-analysis of randomized controlled trials. *Journal of anxiety disorders*, *63*, 36-50. https://doi.org/10.1016/j.janxdis.2019.01.007

Kelley, J. M., Kaptchuk, T. J., Cusin, C., Lipkin, S., & Fava, M. (2012). Open-label placebo for major depressive disorder: A pilot randomized controlled trial. *Psychotherapy and psychosomatics*, *81*(5), 312-314. https://doi.org/10.1159/000337053

Khan, A., & Brown, W. A. (2015). Antidepressants versus placebo in major depression: An overview. *World Psychiatry*, *14*(3), 294-300. https://doi.org/10.1002/wps.20241

Khan, A., Faucett, J., Lichtenberg, P., Kirsch, I., & Brown, W. A. (2012). A systematic review of comparative efficacy of treatments and controls for depression. https://doi.org/10.1371/journal.pone.0041778

Khan, A., Warner, H. A., & Brown, W. A. (2000). Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials: An analysis of the Food and Drug Administration database. *Archives of General Psychiatry*, *57*(4), 311-317. https://doi.org/10.1001/archpsyc.57.4.311

Kirsch, I. (2018). Response expectancy and the placebo effect. *International Review of neurobiology*, *138*, 81-93. https://doi.org/10.1016/bs.irn.2018.01.003

Kirsch, I. (2019). Placebo effect in the treatment of depression and anxiety. *Frontiers in Psychiatry*, 407. https://doi.org/10.3389/fpsyt.2019.00407

Kleine-Borgmann, J., Schmidt, K., Billinger, M., Forkmann, K., Wiech, K., & Bingel, U. (2021). Effects of open-label placebos on test performance and psychological well-being in healthy medical students: A randomized controlled trial. *Scientific reports*, *11*(1), 1-11. https://doi.org/10.1038/s41598-021-81502-2

Kong, J., Wang, Z., Leiser, J., Minicucci, D., Edwards, R., Kirsch, I., Wasan, A. D., Lang, C., Gerber, J., & Yu, S. (2018). Enhancing treatment of osteoarthritis knee pain by boosting expectancy: A functional neuroimaging study. *NeuroImage: Clinical*, *18*, 325-334. https://doi.org/10.1016/j.nicl.2018.01.021

Kube, T., Rief, W., Vivell, M.-B., Schäfer, N. L., Vermillion, T., Körfer, K., & Glombiewski, J. A. (2020). Deceptive and nondeceptive placebos to reduce pain: An experimental study in healthy individuals. *The Clinical Journal of Pain*, *36*(2), 68-79. https://doi.org/10.1097/AJP.0000000000000781

Locher, C., Frey Nascimento, A., Kirsch, I., Kossowsky, J., Meyer, A., & Gaab, J. (2017). Is the rationale more important than deception? A randomized controlled trial of open-label placebo analgesia. *Pain*, *158*(12), 2320-2328. https://doi.org/10.1097/j.pain.0000000000001012

Locher, C., Koechlin, H., Zion, S. R., Werner, C., Pine, D. S., Kirsch, I., Kessler, R. C., & Kossowsky, J. (2017). Efficacy and safety of selective serotonin reuptake inhibitors, serotonin-norepinephrine reuptake inhibitors, and placebo for common psychiatric disorders among children and adolescents: A systematic review and meta-analysis. *JAMA Psychiatry*, *74*(10), 1011-1020. https://doi.org/10.1001/jamapsychiatry.2017.2432

Lumley, M. A., Cohen, J. L., Borszcz, G. S., Cano, A., Radcliffe, A. M., Porter, L. S., Schubiner, H., & Keefe, F. J. (2011). Pain and emotion: A biopsychosocial review of recent research. *Journal of Clinical Psychology*, *67*(9), 942-968. https://doi.org/10.1002/jclp.20816

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion*, *23*(2), 209-237. https://doi.org/10.1080/02699930802204677

Meyer, B., Yuen, K. S., Saase, V., & Kalisch, R. (2019). The functional role of large-scale brain network coordination in placebo-induced anxiolysis. *Cerebral cortex*, *29*(8), 3201-3210. https://doi.org/10.1093/cercor/bhy188

Miller, F. G., & Colloca, L. (2009). The legitimacy of placebo treatments in clinical practice: Evidence and ethics. *The American Journal of Bioethics*, *9*(12), 39-47. https://doi.org/10.1080/15265160903316263

Munkholm, K., Paludan-Müller, A. S., & Boesen, K. (2019). Considering the methodological limitations in the evidence base of antidepressants for depression: A reanalysis of a network meta-analysis. *BMJ open*, *9*(6), e024886. https://doi.org/10.1136/bmjopen-2018-024886

Nitzan, U., Carmeli, G., Chalamish, Y., Braw, Y., Kirsch, I., Shefet, D., Krieger, I., Mendlovic, S., Bloch, Y., & Lichtenberg, P. (2020). Open-Label placebo for the

treatment of unipolar depression: Results from a randomized controlled trial. *Journal of affective disorders*, *276*, 707-710. https://doi.org/10.1016/j.jad.2020.07.077

Olliges, E., Stroppe, S., Haile, A., Reiß, F., Malhis, M., Funke, S. A., & Meissner, K. (2022). Open-label placebo administration decreases pain in elderly patients with symptomatic knee osteoarthritis–A randomized controlled trial. *Frontiers in Psychiatry*, *13*.

Rebstock, L., Schäfer, L. N., Kube, T., Ehmke, V., & Rief, W. (2020). Placebo prevents rumination: An experimental study. *Journal of affective disorders*, *274*, 1152-1160. https://doi.org/10.1016/j.jad.2020.06.010

Salanti, G. (2012). Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: Many names, many benefits, many concerns for the next generation evidence synthesis tool. *Research Synthesis Methods*, *3*(2), 80-97. https://doi.org/10.1002/jrsm.1037

Schaefer, M., Denke, C., Harke, R., Olk, N., Erkovan, M., & Enge, S. (2019). Open-label placebos reduce test anxiety and improve self-management skills: A randomized-controlled trial. *Scientific reports*, *9*(1), 1-6. https://doi.org/10.1038/s41598-019-49466-6

Schaefer, M., Hellmann-Regen, J., & Enge, S. (2021). Effects of open-label placebos on state anxiety and glucocorticoid stress responses. *Brain Sciences*, *11*(4), 508. https://doi.org/10.3390/brainsci11040508

Schaefer, M., Sahin, T., & Berstecher, B. (2018). 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*, *13*(3), 1-14. https://doi.org/10.1371/journal.pone.0192758

Schär, S., Vehlen, A., Ebneter, J., Schicktanz, N., de Quervain, D. J., Wittmann, L., Götzmann, L., grosse Holtforth, M., Protic, S., & Wettstein, A. (2022). Guilt is effectively induced by a written auto-biographical essay but not reduced by experimental pain. *Frontiers in Behavioral Neuroscience*, *16*. https://doi.org/10.3389/fnbeh.2022.891831

Schienle, A., Gremsl, A., Übel, S., & Körner, C. (2016). Testing the effects of a disgust placebo with eye tracking. *International Journal of Psychophysiology*, *101*, 69-75. https://doi.org/10.1016/j.ijpsycho.2016.01.001

Schienle, A., & Jurinec, N. (2022). Open-label placebos as adjunctive therapy for patients with depression. *Contemporary Clinical Trials Communications*, *28*, 100948. https://doi.org/10.1016/j.conctc.2022.100948

Schienle, A., Übel, S., Schöngaßner, F., Ille, R., & Scharmüller, W. (2014). Disgust regulation via placebo: An fMRI study. *Social Cognitive and Affective Neuroscience*, *9*(7), 985-990. https://doi.org/10.1093/scan/nst072

Sheppes, G., Suri, G., & Gross, J. J. (2015). Emotion regulation and psychopathology. *Annual Review of Clinical Psychology*, *11*, 379-405. https://doi.org/10.1146/annurev-clinpsy-032814-112739

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual review of psychology*, *69*, 487-510. https://doi.org/10.1146/annurev-psych-122216-011845

Spille, L., Fendel, J. C., Seuling, P. D., Göritz, A. S., & Schmidt, S. (2023). Open-label placebos—a systematic review and meta-analysis of experimental studies with non-clinical samples. *Scientific reports*, *13*(1), 3640. https://doi.org/10.1038/s41598-023-30362-z

Stone, M. B., Yaseen, Z. S., Miller, B. J., Richardville, K., Kalaria, S. N., & Kirsch, I. (2022). Response to acute monotherapy for major depressive disorder in randomized,

placebo controlled trials submitted to the US Food and Drug Administration: Individual participant data analysis. *BMJ*, *378*. https://doi.org/10.1136/bmj-2021-067606

Vase, L., Robinson, M. E., Verne, G. N., & Price, D. D. (2005). Increased placebo analgesia over time in irritable bowel syndrome (IBS) patients is associated with desire and expectation but not endogenous opioid mechanisms. *Pain*, *115*(3), 338-347. https://doi.org/10.1016/j.pain.2005.03.014

von Wernsdorff, M., Loef, M., Tuschen-Caffier, B., & Schmidt, S. (2021). Effects of open-label placebos in clinical trials: A systematic review and meta-analysis. *Scientific reports*, *11*(1), 3855. https://doi.org/10.1038/s41598-021-83148-6

Wager, T. D., & Atlas, L. Y. (2015). The neuroscience of placebo effects: Connecting context, learning and health. *Nature Reviews Neuroscience*, *16*(7), 403-418. https://doi.org/10.1038/nrn3976

Auxiliary means

ChatGPT version March 23, OpenAI: https://chat.openai.com

Help with the linguistic improvement of the entire work. Especially in terms of clarity of phrases and sentence structure.

DeepL Translate, DeepL SE: https://www.deepl.com/translator

Translation of sentences and text passages independently formulated in German throughout the entire work.

Appendices

Appendix A: Study I

Sezer, D., Locher, C., & Gaab, J. (2022). Deceptive and open-label placebo effects in experimentally induced guilt: a randomized controlled trial in healthy subjects. *Scientific Reports, 12*(1), 21219. https://doi.org/10.1038/s41598-022-25446-1

scientific reports

Check for updates

OPEN

Deceptive and open-label placebo effects in experimentally induced guilt: a randomized controlled trial in healthy subjects

Dilan Sezer[1]✉, Cosima Locher[2,3] & Jens Gaab[1]

Placebos are known to yield significant effects in many conditions. We examined deceptive and open-label placebo effects on guilt, which is important for self-regulation and a symptom of mental disorders. Following an experimental induction of guilt, healthy subjects were randomized to deceptive placebo (DP; *n* = 35), open-label placebo (OLP; *n* = 35), or no treatment (NT; *n* = 39). The primary outcome was guilt responses assessed in area under the curve (AUC). Secondary outcomes were shame, guilt, and affect. We hypothesized that DP and OLP would reduce guilt compared to NT. Guilt responses were higher in the NT group than in the placebo groups (estimate = 2.03, 95% CI = 0.24–3.82, *d* = 0.53), whereas AUC guilt did not differ significantly between the placebo groups (estimate = −0.38, 95% CI = −2.52–1.76, *d* = −0.09). Placebos are efficacious in reducing acute guilt responses, regardless of the placebo administration (i.e., open vs. deceptive). Furthermore, we observed narrative-specific effects with significant changes of guilt but not shame, pride, or affect. These results indicate not only that guilt is amenable to placebos but also that placebos can be administered in an ethical and potentially emotion-specific manner.

Placebos have been found to have clinically significant effects on subjective and objective outcomes in a variety of conditions[1,2]. This especially holds true for acute and chronic pain, where the administration of a placebo has led to analgesia in healthy and clinical populations[3–5], as well as for depressive disorders, for which placebo responses have been found to be so substantial that differences between a placebo and antidepressant medication are a subject of constant debate[6,7].

Placebo effects have also been demonstrated in a number of nonclinical psychological domains, such as in reducing social pain[8]; facilitating social trust and approach behavior[9]; increasing happiness and reducing stress and depression[10,11]; increasing short- and midterm subjective well-being[12]; reducing unpleasantness, sadness and rumination[13–16]; diminishing disgust[17]; and increasing the subjective pleasantness of wine[18]. However, in contrast to the plethora of established experimental pain paradigms, such as the Cold Pressure Test e.g.[19–21], experimentally induced heat pain[22,23], or intracutaneous electrical stimulation[24,25], comparable experimental paradigms are scarce in placebo research on psychological and behavioral outcomes. For example, experimentally inducing sadness by watching a sad movie[15,26], reading self-deprecating statements[27], listening to sad music[28,29], or inducing anxiety by looking at fearful pictures[30,31] are rare examples of experimental paradigms in nonpain placebo research. Given that comparable experimental paradigms would enable important insights into the inner workings of clinically relevant phenomena it is of vital importance for placebo research to extend the range of experimental nonpain paradigms.

One area in current placebo research where experimental paradigms would be of great importance is research into the ethical application of placebo interventions. This field of research has recently gained continuous attention and has provided initial evidence that placebos can also work when they are fully disclosed and administered transparently[32]. Such open-label placebos (OLPs) have been found to have significant effects, for example, in pain conditions (e.g.,[33–35]) and for test anxiety[36], with mixed results for depression[37,38]. In a pilot study with a diagnosed sample of major depression[37], the OLP group did not significantly differ compared to the no treatment control group, which can possibly be explained by the lack of power due to a small sample size of only 20 participants. The

[1]Division of Clinical Psychology and Psychotherapy, Faculty of Psychology, University of Basel, Missionsstrasse 62, 4055 Basel, Switzerland. [2]Department of Consultation-Liaison Psychiatry and Psychosomatic Medicine, University Hospital Zurich, Zurich, Switzerland. [3]Faculty of Health, University of Plymouth, Plymouth, UK. ✉email: dilan.sezer@unibas.ch

second study investigated OLPs as an add-on to treatment as usual in 38 depressed patients[38]. There, symptoms of depression only decreased significantly in a subgroup of non-geriatric patients with an early onset of depression compared to the treatment-as-usual control group alone. In the light of the well-documented placebo effects in antidepressant trials, these findings are surprising and raise the need for further investigations into OLP effects in depression. Experimental studies might in particular help shed light on the underlying OLP mechanisms.

Depression is unquestionably a multifaceted disease. Nevertheless, the experimental induction of single symptoms of depression in healthy and clinical populations may be a promising approach for better understanding the efficacy of OLPs in the symptom picture of depression[15,16,29,39]. In this context, self-conscious emotions like guilt and shame are of interest[40]. Although they may at first sight seem very similar, the emotion shame focuses on the perceived shortcomings of the self, while guilt focuses on the negative consequences of specific actions[41]. In their adaptive form, these emotions are conceptualized as important moral emotions[42]. As such, guilt in particular can function as a relationship enhancer[43,44] and can motivate reparative actions like apologies and confessions[45]. However, in their maladaptive forms, guilt and shame have also been linked to perfectionism[46], which has long been conceptualized as a pathology-causing personality trait[47]. Feeling guilty in everyday life has been associated with heightened aversive arousal states, social distress (e.g., rejection, and loneliness), fewer pleasant and relaxed states[48], and, in the absence of opportunities for compensation, with self-punishment[49]. In addition, guilt can be found at the core of many psychological disorders, such as major depressive disorder[50,51] and of posttraumatic stress disorder[52,53]. Given the relevance and high prevalence of guilt in the general[54] and psychiatric population, examining the possible effects of placebos on guilt is of interest.

In the present study, we set out to test the efficacy of placebos in reducing experimentally induced feelings of guilt in a randomized controlled trial with healthy subjects. To pursue this research question, we employed an autobiographic writing task to evoke acute feelings of guilt[55,56]. To test the potential of an ethically feasible placebo intervention for guilt, we used both a deceptive placebo (DP) and an OLP. Interestingly, direct comparisons of OLPs with DPs have led to inconclusive evidence. Whereas some studies have reported comparable symptom reduction with both OLPs and DPs[21,22,57–59], other studies have found OLPs to be inferior to DPs[15,16]. Despite conflicting evidence, we expected no difference between the efficacies of the DP and that of the OLP in reducing the experience of experimentally induced guilt. Finally, we hypothesized that both the DP and the OLP would lead, when provided with plausible and symptom-specific treatment explanations, to a symptom-specific reduction of the emotional response to experimentally induced guilt as compared to no treatment (NT).

Materials and methods

Study design. Between August 2019 and March 2020, we conducted a randomized controlled parallel-group trial at the Division of Clinical Psychology and Psychotherapy (Faculty of Psychology, University of Basel, Switzerland). Written delayed informed consent was obtained from each subject before participation in the study. The Ethics Committee of the Faculty of Psychology at the University of Basel, Switzerland, approved the design and the informed consent of the study. The study was carried out in accordance with the protocol and principles enunciated in the current version of the Declaration of Helsinki. It was registered retrospectively as a clinical trial on the German Clinical Trials Register (DRKS00029098; 25/05/2022) and follows the reporting guidelines of the Consolidated Standards of Reporting Trials (CONSORT).

Study population. In total, 112 subjects were recruited through the online recruitment system of the Faculty of Psychology (BAPS-Sona, http://baps.sona-systems.com) and through advertisements in lectures at the University of Basel. On the flow of subjects through the study and assessments, see Fig. 1. Interested subjects registered online for the study. Subjects received study credits for their participation. To participate, they had to be healthy by self-report, aged between 18 and 40 years, and be sufficiently proficient in German. Exclusion criteria were self-reported acute or chronic somatic diseases or psychiatric disorders, being in psychological or psychiatric treatment, and taking psychotropic drugs.

Study procedure, guilt induction, and guilt boost. Upon arrival, subjects received a description of the study and were informed that they would not receive all information on the nature of the treatment before the start of the study due to the studies research design, but that this missing information would be fully disclosed after the termination of the study. After providing delayed informed consent, inclusion and exclusion criteria were checked, subjects' demographics were registered, and baseline measures of guilt proneness, state guilt, shame, pride, and emotional valence (for a description of all assessments, see section "Measures and questionnaires"; T0) were assessed. Meanwhile, investigators opened a sequentially numbered sealed envelope to determine the treatment assignment of the subject and kept the group allocation to themselves. Then the subjects in all the groups were invited to write on paper about an experience in which the subject had behaved unfairly toward an intimate person, infringed important rules of conduct, or hurt or even harmed a trusted person through their behavior. We specified that subjects should choose a situation that still emotionally burdened them (for a detailed description of the guilt-induction instructions, see the supplementary material). Similar autobiographic approaches have previously been shown to be efficacious in eliciting guilt in healthy subjects[55,56,60–62]. The guilt induction had a duration of 10 min, and subjects kept their writing to themselves. Afterward, state guilt, shame, pride, and emotional valence were assessed again (T1). Subjects then received either a DP or an OLP (for descriptions, see below), whereas the NT subjects were invited to read travel magazines such as Geo Roadtrips and Terra Mater.

After the DP, the OLP, or NT, all subjects of each group were instructed to answer a question regarding their expected guilt reduction in response to the DP, the OLP or NT before reading a neutral travel magazine for 5 min. Subsequently, state guilt, shame, pride, and emotional valence were assessed again (T2). However,

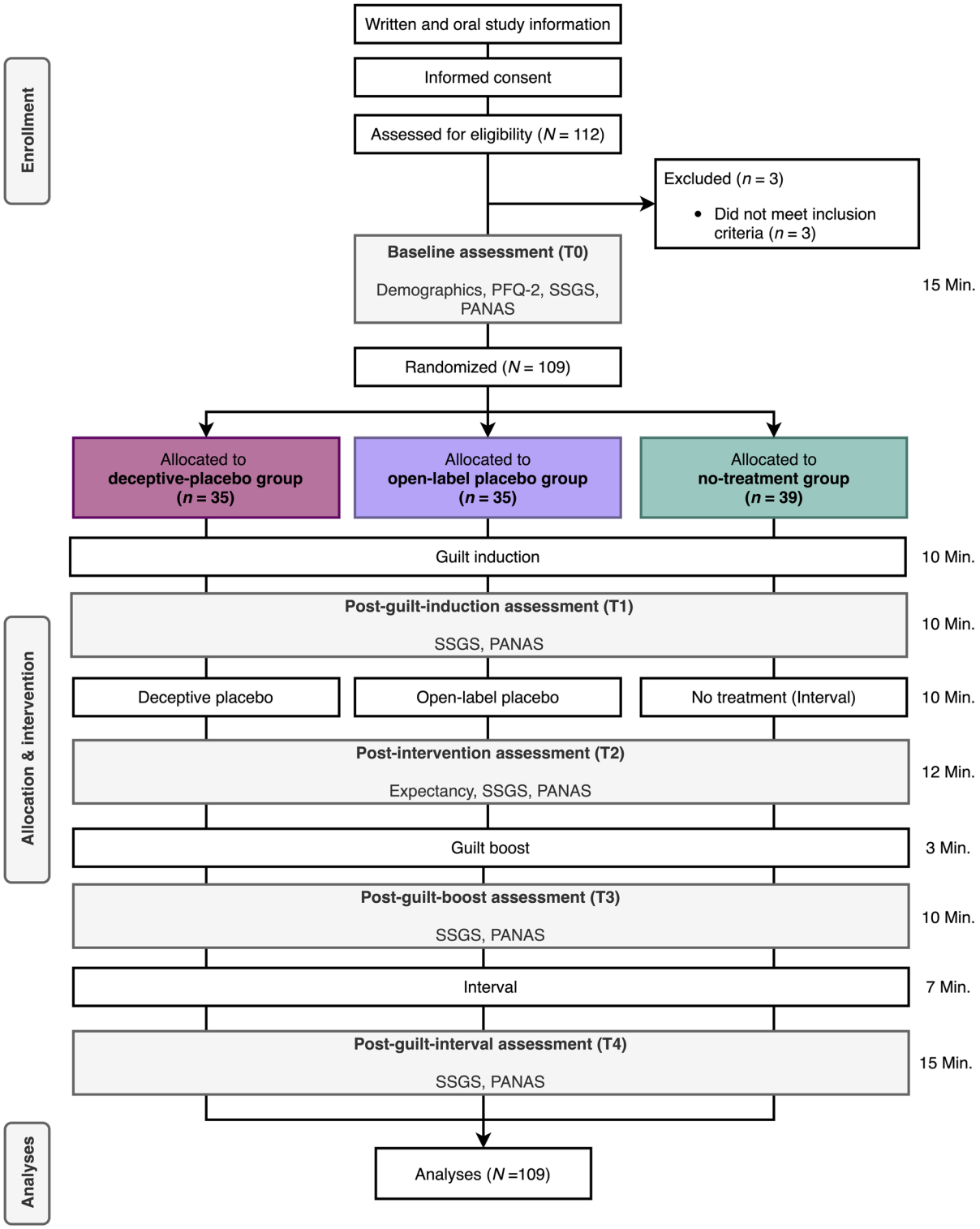


Figure 1. Study design and flow of subjects. Note: DP, deceptive placebo; OLP, open-label placebo; NT, no treatment; PFQ-2, Personal Feelings Questionnaire-2; SSGS, State Shame and Guilt Scale; PANAS, Positive and Negative Affect Schedule; CMQ, Context Model Questionnaire.

we did not expect to observe any treatment effects immediately after treatment because inductions of negative affects in healthy subjects are known to be of short duration[63]. To observe possible treatment effects, we therefore implemented a guilt boost: subjects were instructed to think back to the event they had written down during the guilt induction for 1 min with closed eyes (see the supplementary material for details on the guilt boost). Following the guilt boost, state guilt, shame, pride, and emotional valence were quantified again (T3). The final assessment of state guilt, shame, pride, and emotional valence followed after an interval of about 7 min (T4). Finally, in order to terminate the study with a positive feeling, all subjects were asked to write down three things they were thankful for.

Upon termination of the study in March 2020, all study subjects were debriefed about the aims of the experiment and the deception in the DP group and were provided with the opportunity to withdraw their data.

Treatments. Subjects in the DP group received a blue medium-sized placebo pill (P-dragee, blau, Lichtenstein manufactured by Zentiva Pharma GmbH). A study team member told them that the pill contained a phytopharmacon that supposedly reduces the feeling of guilt through its calming and comforting properties and that this effect would occur within 3–5 min (see the supplementary material for a translation of the German script). Subjects in the OLP group received the same pill but were provided with the rationale used by Kaptchuk et al.[33]: they were told that placebos are efficacious, that they work through expectation and previous conditioning, and that an open attitude toward the treatment could be helpful but was not necessary for its effect. The instructions were identical in terms of structure and format in both placebo groups, but they differed in content. Furthermore, in order to foster the expectation of relief, both the deceptive and open-label rationales included information on the expected efficacy of the given treatment (see supplementary material for the scripted instructions).

Randomization and blinding. The random allocation sequence was created by an independent research assistant prior to the study start using www.randomizer.org. To implement the random allocation sequence (allocation ratio: 1/3:1/3:1/3), investigators opened a sealed envelope containing the group allocation of a subject after the baseline assessment (T0). Due to the nature of the interventions, only subjects in the DP condition were blind to their treatment allocation.

Measures and questionnaires. To measure the primary and secondary outcomes the State Shame and Guilt Scale (SSGS[64]) and the German version of the Positive and Negative Affect Schedule (PANAS[65]) were applied. The SSGS consists of three subscales measuring state shame, guilt, and pride with five items each that are rated on a 5-point Likert scale. For the purpose of this study, we translated the SSGS from English into German. The PANAS consists of two subscales measuring positive and negative affect with 10 items each that are rated on a 5-point Likert scale. The SSGS subscale "guilt" served as the primary outcome of this study, whereas SSGS "shame" and "pride" and the PANAS "positive" and "negative" subscales served as secondary outcomes. All the subscales of the SSGS and the PANAS were applied in all assessments (i.e., T0–T4).

Throughout the experiment additional variables and potential predictors of primary and secondary outcomes were assessed. At the baseline assessment (T0), demographic variables (e.g., age, sex) and a measurement of guilt proneness (German version of the Personal Feelings Questionnaire, PFQ-2[66,67]) were applied. Finally, the expectation of relief was measured once in all groups at T2 right after administration of the placebo, by asking subjects the following question: "On a scale of 1–10, how much do you expect your guilt to be reduced? (1 = not at all, 10 = completely)". Higher numbers indicated a greater expectation. See Fig. 1 for an overview of all the assessments and their respective time points.

Statistical analyses. All analyses were carried out using RStudio for Mac. To examine the validity of the experimental guilt induction and the guilt boost, two-way mixed analyses of variance (ANOVAs) were computed for the time points T0–T2 (guilt induction) and T2–T4 (guilt boost). However, whenever the assumptions for a two-way mixed ANOVA were not met, a robust two-way mixed ANOVA with 20% trimmed means using WRS2 package[68] was calculated with the independent between-subject factor "group" and the within-subject factor "time." Separate analyses were carried out for each subscale of the SSGS and the PANAS.

To detect differences between the groups, area-under-the-curve (AUC) parameters were calculated for the SSGS and PANAS subscales between T0 and T2 (guilt induction validation check) and between T2 and T4 (treatment effects); the AUC of the SSGS guilt subscale from T2–T4 was defined as the primary outcome. Using the AUC to assess group differences across different time points offers the unique possibility of simplifying the statistical analysis without the losing of the information contained in multiple measurements while also increasing the power[69]. Following the trapezoid formula, the AUC was calculated with respect to increase (AUCi), which refers to changes over time[69]. AUCi values were calculated for the different time intervals between measurements (see Fig. 1) and were compared between conditions with a one-factor between-subject ANOVA. If the normality assumption for the ANOVA was not met, a Kruskal–Wallis test was used. If there were significant extreme outliers, as assessed by above quartile 3 + 3 times the interquartile range or below quartile 1 − 3 times the interquartile range, a robust ANOVA using the WRS2 package was applied. To test our hypotheses, the following two a priori contrasts were calculated: DP & OLP vs. NT (C1); DP vs. OLP (C2). Contrasts are reported as mean differences (estimates) and confidence intervals (CI). Despite nonnormal AUCi scores in each of the two subscales of the PANAS, all a priori contrast analyses were performed on the untrimmed data.

To investigate the influence of different variables (e.g., guilt and shame proneness, and expectation of relief), Pearson correlations with AUCi sizes for each outcome were calculated. Differences across groups regarding the scores of predictors were assessed using a one-factor between-subject ANOVA or, if appropriate, a Kruskal–Wallis

	DP	OLP	NT	F/X^2
n (% female)	35 (80.00%)	35 (74.29%)	39 (64.10%)	$X^2(2, 109) = 0.29$, $p = .864$
Age in years, mean (*SD*)	22.89 (3.62)	24.03 (5.56)	21.67 (3.13)	$F(2, 106) = 2.93$, $p = .058$
SSGS				
Guilt, mean (*SD*)	2.05 (0.80)	2.06 (1.00)	1.92 (0.70)	$F(2, 106) = 0.32$, $p = .729$
Shame, mean (*SD*)	1.53 (0.55)	1.67 (0.79)	1.34 (0.44)	$F(2, 106) = 2.67$, $p = .074$
Pride, mean (*SD*)	3.45 (0.45)	3.47 (0.55)	3.62 (0.58)	$F(2, 106) = 1.05$, $p = .353$
PANAS				
Positive, mean (*SD*)	3.05 (0.58)	3.27 (0.60)	3.25 (0.61)	$F(2, 106) = 1.51$, $p = .226$
Negative, mean (*SD*)	1.30 (0.36)	1.44 (0.42)	1.52 (0.42)	$F(2, 106) = 2.64$, $p = .076$
PFQ-2				
Guilt, mean (*SD*)	21.71 (3.16)	21.14 (4.03)	21.28 (3.02)	$F(2, 106) = 0.27$, $p = .766$
Shame, mean (*SD*)	32.66 (3.32)	33.31 (2.97)	32.18 (3.49)	$F(2, 106) = 1.11$, $p = .334$

Table 1. Baseline between-group comparisons on demographic and outcome measures. *SD*, standard deviation; DP, deceptive placebo; OLP, open-label placebo; NT, no treatment; SSGS, State Shame and Guilt Scale; PANAS, Positive and Negative Affect Schedule; PFQ-2, Personal Feelings Questionnaire 2.

test. For pairwise comparisons of secondary outcomes (e.g., expectation of relief), a pairwise Wilcoxon test with a BH adjustment[70] was used.

An alpha level of 0.05 was used for all tests. There was no missing data. Unless indicated, all results shown are means + /− standard deviations (*SD*). Using the statistical software G*Power, we conducted a conservative power calculation on the basis of an *F* test for a multivariate analysis of variance (MANOVA) with a within-and-between-factor interaction for three groups. This analysis showed that we would need a sample size of $N = 110$ for a power of 0.9 to detect a medium to large effect size of $f = 0.3$ (based on observed effect sizes in previous clinical[32] and experimental OLP studies[22]) with a one-sided alpha level of 0.05.

Results

Sample characteristics and general overview of data. In total, 112 subjects signed up for the study. Three subjects had to be excluded because they did not meet the inclusion criteria (see Fig. 1). Thus, 109 subjects were included in the analysis. Baseline characteristics did not differ significantly across the groups (see Table 1). Figure 2 displays the temporal course of the SSGS subscales. A complete overview of mean values per group for each outcome at each assessment time point can be found in Table S1 in the online supplementary material.

Validation check of guilt induction and guilt boost in primary and secondary outcomes. To examine the validity of the experimental guilt induction, two-way mixed ANOVAs were calculated for the subjective ratings of guilt, shame, pride, and positive affect, and negative affect from T0 to T2. The assumptions for a standard two-way mixed ANOVA were only met for the analyses of pride and positive affect. For all outcomes, there was a highly significant effect of time from T0 to T2 (all $ps < 0.001$; see Table S2), which indicates that the guilt induction led to significant responses in all the assessed affective states, with most pronounced changes from T0 to T1 in guilt (see Table S2).

Regarding the guilt boost, two-way mixed ANOVAs were calculated for subjective ratings of guilt, shame, pride, and positive affect, and negative affect from T2 to T4. The assumptions for standard two-way mixed ANOVA were not met for all analyses. For all outcomes, there was a highly significant effect of time from T2 to T4 (all $ps < 0.001$; see Table S3), which indicates that the guilt boost successfully changed all the assessed affective states, with most pronounced changes from T2 to T3 in guilt and pride (see Table S3).

To assess possible group differences in their responses to the guilt induction (T0–T1), AUCi sizes were compared for the time points of T0 and T2 across groups using a one-factor ANOVA for guilt and a Kruskal–Wallis test for all the other outcomes. As expected, the mean size of the AUCi between T0 and T2 did not differ significantly across the groups (all $ps > 0.121$; see Tables S4 and S5), which indicates that the groups had comparable responses to the initial guilt induction.

Group differences in primary and secondary outcomes. For possible differences in emotional responses following the guilt boost between subjects receiving a DP, an OLP, or NT, the AUCi from T2 to T4 was compared across groups with a one-factor ANOVA. These analyses showed significantly different AUCi sizes for guilt ($F(2, 106) = 3.38$, $p = 0.038$) but not for shame, pride, positive affect, or negative affect (all $ps > 0.191$; see Table S4). A priori orthogonal contrasts of guilt showed significantly smaller AUCi guilt scores for the two treatment groups taken together in comparison to the NT scores (DP & OLP vs. NT: estimate = 2.03, 95% CI = 0.24–3.82, $d = 0.53$), which indicates a smaller increase in guilt following the guilt boost. No significant difference in AUCi sizes between the two treatment groups was found (DP vs. OLP: estimate = −0.38, 95% CI = −2.52–1.76, $d = -0.09$). Table 2 shows mean AUCi values from T2 to T4 for each group and subscale and the differences in the means of each calculated contrast.

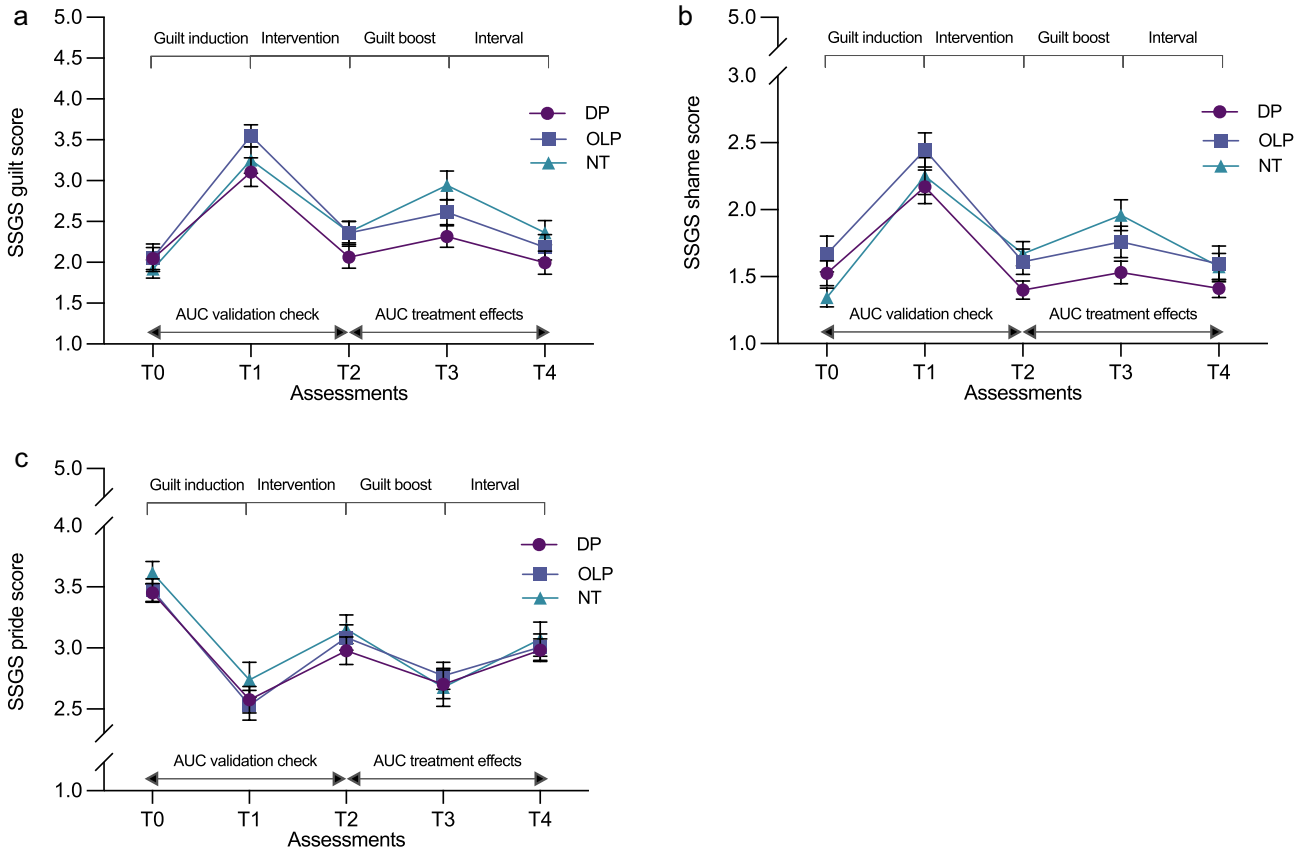


Figure 2. Temporal course of the SSGS guilt (**a**), shame (**b**) and pride (**c**) scale scores across experimental groups. Note: Displayed are means per group: error bars represent the standard error of the mean. DP, deceptive placebo; OLP, open-label placebo; NT, no treatment; SSGS, State Shame and Guilt Scale; AUC, area under the curve.

	DP ($n=35$)	OLP ($n=35$)	NT ($n=39$)	DP & OLP vs. NT	DP vs. OLP
SSGS	Mean (*SD*)			Mean difference (CI)	
Guilt	1.02 (3.16)	0.64 (4.04)	2.85 (4.48)	2.03 (0.24–3.82)*, $d=0.53$	− 0.38 (− 2.52–1.76), $d=-0.09$
Shame	0.69 (3.06)	0.68 (3.52)	1.10 (3.23)	0.41 (− 1.08–1.89), $d=0.13$	− 0.01 (− 1.79–1.76), $d=0.0$
Pride	− 1.35 (2.65)	− 1.85 (3.31)	− 2.67 (3.37)	− 1.07 (− 2.49–0.35), $d=-0.35$	− 0.50 (− 2.20–1.20), $d=-0.16$
PANAS					
Positive	− 0.83 (3.00)	− 1.26 (3.02)	− 1.59 (3.22)	− 0.54 (− 1.94–0.86), $d=-0.18$	0.42 (− 1.25–2.09), $d=0.14$
Negative	0.70 (2.02)	0.22 (2.35)	0.75 (2.62)	0.29 (− 0.78–1.36), $d=0.12$	0.49 (− 0.79–1.76), $d=0.21$

Table 2. Area-under-the-curve SSGS and PANAS scores and between-group contrasts for T2–T4. *SD*, standard deviation; DP, deceptive placebo; OLP, open-label placebo; NT, no treatment; SSGS, State Shame and Guilt Scale; AUCi, area under the curve with respect to increase; CI, confidence interval, $^*p<0.05$.

Associations of additional variables of interest with outcomes. The mean expectation of relief, guilt proneness, and shame proneness, including their correlation with the AUCi values of the SSGS and PANAS subscales from T2 to T4 are shown in Table S6 for all groups.

Omnibus tests showed that the groups differed in their expectation of guilt relief following the treatments (Kruskal–Wallis test $p=0.021$): the OLP group ($M=4.49$, $SD=2.11$) displayed significantly higher expectations of guilt relief than the DP group ($M=3.23$, $SD=1.72$; post hoc Wilcoxon test p adj. $=0.031$). The expectation of guilt relief in the NT group ($M=4.23$, $SD=2.10$) did not significantly differ from that in the OLP group (Wilcoxon test, p adj. $=0.544$) but differed significantly from that in the DP group (Wilcoxon test, p adj. $=0.045$). However, despite significant group differences in the expectation of relief, there was no significant correlation with any primary or secondary outcomes (see Table S6). The groups did not differ with regard to guilt and shame proneness (guilt: Kruskal–Wallis test, $p=0.671$; shame: Kruskal–Wallis test, $p=0.241$).

Discussion

Given the high prevalence of guilt as a self-conscious emotion that is associated with a variety of unpleasant psychological states in everyday life, its relevance in depression and other psychological disorders, and the substantial magnitude of placebo effects in pharmacological and psychotherapeutic treatments of depressive disorders, we set out to assess the effects of deceptive and open-label placebos on experimentally induced guilt responses in healthy subjects in comparison to a no-treatment condition.

First, our experimental guilt induction and a subsequent guilt boost elicited robust emotional responses of guilt as well as—although to a lower degree—of shame, pride, and positive affect, and negative affect. Second, and importantly, the administration of the placebo—either deceptive or open—significantly reduced the guilt responses to the guilt boost in comparison to no treatment with a medium effect size of $d = 0.53$. Interestingly, this effect was not observed for any other outcome, which suggests the possibility that the symptom-specific placebo rationales led to symptom-specific placebo effects.

In the following, the observed effects will be discussed from an empirical, and a methodological perspective. Empirically, our findings show that deceptive and open placebos were equally efficacious in reducing the self-conscious emotion of guilt. These findings are in line with a growing number of reports that have found OLPs to have significant effects on emotions, including anxiety[36,71], depression[37,38], sadness[16,28], general emotional well-being[72,73], and emotional distress[74]. Furthermore, our results are also in line with studies reporting that DPs and OLPs have equal effects in healthy subjects[21,22,57], which highlights the potential of OLPs as a means of ethically harnessing placebo effects in these conditions. But there is also contradicting evidence: for example, studies have found that DPs lead to greater heat-pain tolerance than OLPs did in healthy subjects[59] or that the placebo effect disappears when it is openly administered to treat motion-induced nausea[75]. With regard to nonanalgesic paradigms, only one placebo study has compared OLPs to DPs for experimentally induced sadness in depressed subjects[16], and it found greater placebo effects from DPs. However, while the DPs decreased sadness from before to after the induction of sadness, OLPs were also efficacious at preventing an increase in sadness while there was an increase in the NT group. In summary, the evidence on the comparative efficacy of DPs and OLPs is promising even if it is, to some extent, mixed and seems to depend on the target condition. Further studies are needed to fully understand the similarities and differences of the efficacy and mechanisms of DPs and OLPs across different fields of application and populations. Despite the inconclusive evidence, even if OLPs are found to be less efficacious than DPs in some cases, the effects of OLPs are, in contrast to those of DPs, ethically acceptable and thus suitable to use in practice[76]. Regarding the underlying mechanisms of deceptive and open-label placebos, there is some evidence that optimism is not of the same importance in OLPs as it is in DPs[12], which suggests that the mechanisms operating in DPs and OLPs are not entirely the same. This finding is complemented by the results of the present study, which found no association between the expectation of guilt relief—a well-studied mechanism of deceptive placebos[77]—and the response to the guilt induction. However, since the pattern of the expectation of relief across the groups, differed from what we expected[78] (i.e., the DP group displayed significantly lower expectations of relief as compared to the two other groups), it is questionable, whether the scale we employed was capable of reliably measuring expectations of guilt reduction. Another possible explanation for this finding could be that the rationale used in the DP group (i.e., that it is a phytopharmaceutical) might not have been entirely convincing, leaving subjects of that group with fewer expectations towards guilt reduction. Thus, more research using validated scales is needed in order to establish the importance of expectations of relief in OLP effects.

From a methodological point of view, we found that the employed guilt paradigm exerted its intended effects by inducing guilt as a consequence of writing ("guilt induction") and thinking ("guilt boost") about an interpersonally unfair behavior toward another person. The tasks did not only impact guilt but also all the other assessed affective states. Yet as indicated by the amount of change between the baseline and the measurement after guilt induction (T0–T1), the effects were most pronounced for guilt. These promising results are in line with other studies testing this approach[56] and open new possibilities for conducting experimental placebo research on affective states. For example, the nature of the experimental design, in which the intervention is delivered prior to the guilt induction of interest (i.e. the guilt boost), offers the unique possibility of testing the short-term preventive effect of a placebo intervention. Furthermore, in the context of the ethical application of placebo interventions, experimental paradigms facilitate the systematic manipulation of the treatment setting and application, which can aid our understanding of the mechanisms involved in how OLPs influence affective states. In this regard, the finding that the symptom-specific rationale might have led to a symptom-specific effect points to an interesting line of research which needs to be systematically addressed in future studies. If future randomized controlled trials testing differential effects of symptom-specific rationales were to support the observation of this study, the various and different effects of placebos across disorders, populations, and settings could be seen as specific to the rationales employed.

This study corroborates important findings on the efficacy of OLPs on affective states. In addition, we successfully tested a guilt-inducing paradigm, which will enable further research on placebo effects on psychological parameters. However, several aspects of the study require critical examination. First, within the study design, only a single medication intake was simulated and assessed for its immediate effects, so we cannot draw any conclusions regarding the durability of the effects we found. Second, the measurements of the outcomes were subjective rather than objective, which raises the question of report and social-desirability bias. Nevertheless, self-report measures are standard outcomes in trials of affective outcomes, and research indicates that placebo treatments are most efficacious for such subjective complaints[79]. Third, since the absence of a significant difference is not the same as equivalence[80], future studies should use noninferiority comparisons of DP and OLP treatments to answer the question of the equivalence of both treatments. Fourth, in the current study the observation of a symptom-specific placebo response following a symptom-specific rationale might be biased, as this was not systematically tested in a randomized fashion. Last, guilt in healthy individuals and guilt in patients might not

be comparable. In our study, guilt was experimentally induced in healthy subjects, who can be assumed to have good strategies for dealing with negative emotions. Furthermore, a meta-analytical review on the association of different forms of guilt and depressive symptoms found that maladaptive guilt correlates substantially with depressive symptoms[81] but that contextually legitimate or adaptive guilt does not ($r = 0.06$). There is thus a need to replicate the findings of our study in clinical populations.

Guilt can be a burdensome emotion, in both healthy and clinical populations. The present study investigated whether a deceptive and an open-label placebo could reduce experimentally induced guilt in healthy subjects. The results show that placebos are efficacious in reducing acute experimentally induced guilt responses in comparison to no treatment, regardless of the placebo administration (i.e., open vs. deceptive). This indicates that placebos can have demonstrable effects on guilt and that these effects can be employed while respecting important ethical principles.

Data availability

The protocol and datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Received: 6 May 2022; Accepted: 30 November 2022
Published online: 08 December 2022

References

1. Evers, A. W. *et al.* Implications of placebo and nocebo effects for clinical practice: Expert consensus. *Psychother. Psychosom.* **87**, 204–210. https://doi.org/10.1159/000490354 (2018).
2. Evers, A. W. *et al.* What should clinicians tell patients about placebo and nocebo effects? Practical considerations based on expert consensus. *Psychother. Psychosom.* **90**, 49–56. https://doi.org/10.1159/000510738 (2021).
3. Forsberg, J. T., Martinussen, M. & Flaten, M. A. The placebo analgesic effect in healthy individuals and patients: A meta-analysis. *Psychosom. Med.* **79**, 388–394. https://doi.org/10.1097/PSY.0000000000000432 (2017).
4. Cragg, J. J. *et al.* Meta-analysis of placebo responses in central neuropathic pain: Impact of subject, study, and pain characteristics. *Pain* **157**, 530–540. https://doi.org/10.1097/j.pain.0000000000000431 (2016).
5. Vase, L., Riley, J. L. & Price, D. D. A comparison of placebo effects in clinical analgesic trials versus studies of placebo analgesia. *Pain* **99**, 443–452. https://doi.org/10.1016/S0304-3959(02)00205-1 (2002).
6. Cipriani, A. *et al.* Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: A systematic review and network meta-analysis. *Focus* **16**, 420–429. https://doi.org/10.1176/appi.focus.16407 (2018).
7. Munkholm, K., Paludan-Müller, A. S. & Boesen, K. Considering the methodological limitations in the evidence base of antidepressants for depression: A reanalysis of a network meta-analysis. *BMJ Open* **9**, e024886. https://doi.org/10.1136/bmjopen-2018-024886 (2019).
8. Koban, L., Jepma, M., Geuter, S. & Wager, T. D. What's in a word? How instructions, suggestions, and social information change pain and emotion. *Neurosci. Biobehav. Rev.* **81**, 29–42. https://doi.org/10.1016/j.neubiorev.2017.02.014 (2017).
9. Yan, X., Yong, X., Huang, W. & Ma, Y. Placebo treatment facilitates social trust and approach behavior. *Proc. Natl. Acad. Sci. USA* **115**, 5732–5737. https://doi.org/10.1073/pnas.1800779115 (2018).
10. Woodworth, R. J., O'Brien-Malone, A., Diamond, M. R. & Schüz, B. Web-based positive psychology interventions: A reexamination of effectiveness. *J. Clin. Psychol.* **73**, 218–232. https://doi.org/10.1002/jclp.22328 (2017).
11. Darragh, M. *et al.* A take-home placebo treatment can reduce stress, anxiety and symptoms of depression in a non-patient population. *Aust. N. Z. J. Psychiatry* **50**, 858–865. https://doi.org/10.1177/0004867415621390 (2016).
12. Gaab, J., Kossowsky, J., Ehlert, U. & Locher, C. Effects and components of placebos with a psychological treatment rationale–three randomized-controlled studies. *Sci. Rep.* **9**, 1–8. https://doi.org/10.1038/s41598-018-37945-1 (2019).
13. Petrovic, P. *et al.* Placebo in emotional processing—induced expectations of anxiety relief activate a generalized modulatory network. *Neuron* **46**, 957–969. https://doi.org/10.1016/j.neuron.2005.05.023 (2005).
14. Rebstock, L., Schäfer, L. N., Kube, T., Ehmke, V. & Rief, W. Placebo prevents rumination: An experimental study. *J. Affect. Disord.* **274**, 1152–1160. https://doi.org/10.1016/j.jad.2020.06.010 (2020).
15. Glombiewski, A. J., Rheker, J., Wittkowski, J., Rebstock, L. & Rief, W. Placebo mechanisms in depression: An experimental investigation of the impact of expectations on sadness in female participants. *J. Affect. Disord.* **256**, 658–667. https://doi.org/10.1016/j.jad.2019.06.070 (2019).
16. Haas, J. W., Rief, W., Glombiewski, J. A., Winkler, A. & Doering, B. K. Expectation-induced placebo effect on acute sadness in women with major depression: An experimental investigation. *J. Affect. Disord.* **274**, 920–928. https://doi.org/10.1016/j.jad.2020.05.056 (2020).
17. Schienle, A., Übel, S., Schöngaßner, F., Ille, R. & Scharmüller, W. Disgust regulation via placebo: An fMRI study. *Soc. Cogn. Affect. Neurosci.* **9**, 985–990. https://doi.org/10.1093/scan/nst072 (2014).
18. Werner, C. P. *et al.* Price information influences the subjective experience of wine: A framed field experiment. *Food Qual. Prefer.* **92**, 104223. https://doi.org/10.1016/j.foodqual.2021.104223 (2021).
19. Mitchell, L. Cold pressor test. *Encycl. Behav. Med.* https://doi.org/10.1007/978-1-4419-1005-9_239 (2013).
20. Geers, A. L., Wellman, J. A., Fowler, S. L., Helfer, S. G. & France, C. R. Dispositional optimism predicts placebo analgesia. *J. Pain* **11**, 1165–1171. https://doi.org/10.1016/j.jpain.2010.02.014 (2010).
21. Disley, N., Kola-Palmer, S. & Retzler, C. A comparison of open-label and deceptive placebo analgesia in a healthy sample. *J. Psychosom. Res.* **140**, 110298. https://doi.org/10.1016/j.jpsychores.2020.110298 (2021).
22. Locher, C. *et al.* Is the rationale more important than deception? A randomized controlled trial of open-label placebo analgesia. *Pain* **158**, 2320–2328. https://doi.org/10.1097/j.pain.0000000000001012 (2017).
23. Krummenacher, P. *et al.* Expectancy-induced placebo analgesia in children and the role of magical thinking. *J. Pain* **15**, 1282–1293. https://doi.org/10.1016/j.jpain.2014.09.005 (2014).
24. Schneider, T., Luethi, J., Mauermann, E., Bandschapp, O. & Ruppen, W. Pain response to open label placebo in induced acute pain in healthy adult males. *Anesthesiology* **132**, 571–580. https://doi.org/10.1097/ALN.0000000000003076 (2020).
25. Koppert, W. *et al.* A new model of electrically evoked pain and hyperalgesia in human skin: The effects of intravenous alfentanil, S (+)-ketamine, and lidocaine. *J. Am. Soc. Anesthesiol.* **95**, 395–402. https://doi.org/10.1097/00000542-200108000-00022 (2001).
26. Gross, J. J. & Levenson, R. W. Eliciting emotions using films. *Cogn. Emot.* **9**, 87–108. https://doi.org/10.1080/02699939508408966 (1995).
27. Velten, E. A laboratory task for induction of mood states. *Behav. Res. Ther.* **6**, 473–482. https://doi.org/10.1016/0005-7967(68)90028-4 (1968).

28. Hahn, A., Göhler, A. C., Hermann, C. & Winkler, A. Even when you know it is a placebo, you experience less sadness: First evidence from an experimental open-label placebo investigation. *J. Affect. Disord.* https://doi.org/10.1016/j.jad.2022.02.043 (2022).
29. Göhler, A. C., Haas, J. W., Sperl, M. F., Hermann, C. & Winkler, A. Placebo nasal spray protects female participants from experimentally induced sadness and concomitant changes in autonomic arousal. *J. Affect. Disord.* **295**, 131–138. https://doi.org/10.1016/j.jad.2021.07.037 (2021).
30. Geeraerts, B. *et al.* Influence of experimentally induced anxiety on gastric sensorimotor function in humans. *Gastroenterology* **129**, 1437–1444. https://doi.org/10.1053/j.gastro.2005.08.020 (2005).
31. Morris, J. S. *et al.* A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature* **383**, 812–815. https://doi.org/10.1038/383812a0 (1996).
32. von Wernsdorff, M., Loef, M., Tuschen-Caffier, B. & Schmidt, S. Effects of open-label placebos in clinical trials: A systematic review and meta-analysis. *Sci. Rep.* **11**, 1–14. https://doi.org/10.1038/s41598-021-83148-6 (2021).
33. Kaptchuk, T. J. *et al.* Placebos without deception: A randomized controlled trial in irritable bowel syndrome. *PLoS ONE* **5**, 1–7. https://doi.org/10.1371/journal.pone.0015591 (2010).
34. Carvalho, C. *et al.* Open-label placebo treatment in chronic low back pain: A randomized controlled trial. *Pain* **157**, 2766. https://doi.org/10.1097/j.pain.0000000000000700 (2016).
35. Kleine-Borgmann, J., Schmidt, K., Hellmann, A. & Bingel, U. Effects of open-label placebo on pain, functional disability, and spine mobility in patients with chronic back pain: A randomized controlled trial. *Pain* **160**, 2891–2897. https://doi.org/10.1097/j.pain.0000000000001683 (2019).
36. Schaefer, M. *et al.* Open-label placebos reduce test anxiety and improve self-management skills: A randomized-controlled trial. *Sci. Rep.* **9**, 1–6. https://doi.org/10.1038/s41598-019-49466-6 (2019).
37. Kelley, J. M., Kaptchuk, T. J., Cusin, C., Lipkin, S. & Fava, M. Open-label placebo for major depressive disorder: A pilot randomized controlled trial. *Psychother. Psychosom.* https://doi.org/10.1159/000337053 (2012).
38. Nitzan, U. *et al.* Open-Label placebo for the treatment of unipolar depression: Results from a randomized controlled trial. *J. Affect. Disord.* **276**, 707–710. https://doi.org/10.1016/j.jad.2020.07.077 (2020).
39. Friehs, T., Rief, W., Glombiewski, J. A., Haas, J. & Kube, T. Deceptive and non-deceptive placebos to reduce sadness: A five-armed experimental study. *J. Affect. Disord.* **9**, 100349. https://doi.org/10.1016/j.jadr.2022.100349 (2022).
40. Tangney, J. P. E. & Tracy, J. L. Self-conscious emotions. (2012).
41. Tangney, J. P. & Dearing, R. L. *Shame and guilt*. (Guilford Press, 2003).
42. Tangney, J. P. Moral affect: The good, the bad, and the ugly. *J. Pers. Soc. Psychol.* **61**, 598. https://doi.org/10.1037/0022-3514.61.4.598 (1991).
43. Baumeister, R. F., Stillwell, A. M. & Heatherton, T. F. Guilt: An interpersonal approach. *Psychol. Bull.* **115**, 243. https://doi.org/10.1037/0033-2909.115.2.243 (1994).
44. Cryder, C. E., Springer, S. & Morewedge, C. K. Guilty feelings, targeted actions. *Pers. Soc. Psychol. Bull.* **38**, 607–618. https://doi.org/10.1177/0146167211435796 (2012).
45. Graton, A. & Ric, F. How guilt leads to reparation? Exploring the processes underlying the effects of guilt. *Motiv. Emot.* **41**, 343–352. https://doi.org/10.1007/s11031-017-9612-z (2017).
46. Fedewa, B. A., Burns, L. R. & Gomez, A. A. Positive and negative perfectionism and the shame/guilt distinction: Adaptive and maladaptive characteristics. *Pers. Individ. Differ.* **38**, 1609–1619. https://doi.org/10.1016/j.paid.2004.09.026 (2005).
47. Hewitt, P. L. & Flett, G. L. Perfectionism in the self and social contexts: Conceptualization, assessment, and association with psychopathology. *J. Pers. Soc. Psychol.* **60**, 456. https://doi.org/10.1037/0022-3514.60.3.456 (1991).
48. Baumeister, R. F., Reis, H. T. & Delespaul, P. A. Subjective and experiential correlates of guilt in daily life. *Pers. Soc. Psychol. Bull.* **21**, 1256–1268. https://doi.org/10.1177/014616729521120 (1995).
49. Nelissen, R. & Zeelenberg, M. When guilt evokes self-punishment: Evidence for the existence of a Dobby effect. *Emotion* **9**, 118. https://doi.org/10.1037/a0014540 (2009).
50. Ghatavi, K., Nicolson, R., MacDonald, C., Osher, S. & Levitt, A. Defining guilt in depression: a comparison of subjects with major depression, chronic medical illness and healthy controls. *J. Affect. Disord.* **68**, 307–315. https://doi.org/10.1016/S0165-0327(01)00335-4 (2002).
51. Singh, S. M. & Sharma, A. The prevalence and correlates of guilt in depression: A study from North India. *Asian J. Psychiatr.* **6**, 622–623. https://doi.org/10.1016/j.ajp.2013.05.010 (2013).
52. Bovin, M. J. *et al.* Does guilt mediate the association between tonic immobility and posttraumatic stress disorder symptoms in female trauma survivors?. *J. Trauma. Stress* **27**, 721–724. https://doi.org/10.1002/jts.21963 (2014).
53. Beck, J. G. *et al.* Exploring negative emotion in women experiencing intimate partner violence: Shame, guilt, and PTSD. *Behav. Ther.* **42**, 740–750. https://doi.org/10.1016/j.beth.2011.04.001 (2011).
54. Luck, T. & Luck-Sikorski, C. Feelings of guilt in the general adult population: Prevalence, intensity and association with depression. *Psychol. Health Med.* **26**, 1143–1153. https://doi.org/10.1080/13548506.2020.1859558 (2021).
55. Mills, C. & D'Mello, S. On the validity of the autobiographical emotional memory task for emotion induction. *PLoS ONE* **9**, e95837. https://doi.org/10.1371/journal.pone.0095837 (2014).
56. Schär, S. *et al.* Guilt is effectively induced by a written auto-biographical essay but not reduced by experimental pain. *Front. Behav. Neurosci.* https://doi.org/10.3389/fnbeh.2022.891831 (2022).
57. Mundt, J. M., Roditi, D. & Robinson, M. E. A comparison of deceptive and non-deceptive placebo analgesia: Efficacy and ethical consequences. *Ann. Behav. Med.* **51**, 307–315. https://doi.org/10.1007/s12160-016-9854-0 (2017).
58. Lembo, A. *et al.* Open-label placebo vs double-blind placebo for irritable bowel syndrome: A randomized clinical trial. *Pain* **162**, 2428–2435. https://doi.org/10.1097/j.pain.0000000000002234 (2021).
59. Kube, T. *et al.* Deceptive and nondeceptive placebos to reduce pain: An experimental study in healthy individuals. *Clin. J. Pain* **36**, 68–79. https://doi.org/10.1097/AJP.0000000000000781 (2020).
60. Bastian, B., Jetten, J. & Fasoli, F. Cleansing the soul by hurting the flesh: The guilt-reducing effect of pain. *Psychol. Sci.* **22**, 334. https://doi.org/10.1177/0956797610397058 (2011).
61. Inbar, Y., Pizarro, D. A., Gilovich, T. & Ariely, D. Moral masochism: On the connection between guilt and self-punishment. *Emotion* **13**, 14 (2013).
62. De Hooge, I. E., Nelissen, R., Breugelmans, S. M. & Zeelenberg, M. What is moral about guilt? Acting "prosocially" at the disadvantage of others. *J. Pers. Soc. Psychol.* **100**, 462. https://doi.org/10.1037/a0021459 (2011).
63. Frost, R. O. & Green, M. L. Velten mood induction procedure effects: Duration and postexperimental removal. *Pers. Soc. Psychol. Bull.* **8**, 341–347. https://doi.org/10.1177/0146167282082024 (1982).
64. Marschall, D., Sanftner, J. & Tangney, J. P. The state shame and guilt scale. *George Mason University*, (1994).
65. Krohne, H. W., Egloff, B., Kohlmann, C.-W. & Tausch, A. Untersuchungen mit einer deutschen version der" positive and negative affect schedule"(PANAS). *Diagnostica* **42**, 139–156 (1996).
66. Harder, D. H. & Zalma, A. Two promising shame and guilt scales: A construct validity comparison. *J. Pers. Assess.* **55**, 729–745. https://doi.org/10.1080/00223891.1990.9674108 (1990).
67. Rüsch, N. *et al.* Measuring shame and guilt by self-report questionnaires: A validation study. *Psychiat. Res.* **150**, 313–325. https://doi.org/10.1016/j.psychres.2006.04.018 (2007).

68. Mair, P. & Wilcox, R. Robust statistical methods in R using the WRS2 package. *Behav. Res. Methods* **52**, 464–488. https://doi.org/10.3758/s13428-019-01246-w (2020).
69. Pruessner, J. C., Kirschbaum, C., Meinlschmid, G. & Hellhammer, D. H. Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology* **28**, 916–931. https://doi.org/10.1016/S0306-4530(02)00108-7 (2003).
70. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x (1995).
71. Schaefer, M., Hellmann-Regen, J. & Enge, S. Effects of open-label placebos on state anxiety and glucocorticoid stress responses. *Brain Sci.* **11**, 508. https://doi.org/10.3390/brainsci11040508 (2021).
72. Kleine-Borgmann, J. *et al.* Effects of open-label placebos on test performance and psychological well-being in healthy medical students: A randomized controlled trial. *Sci. Rep.* **11**, 1–11. https://doi.org/10.1038/s41598-021-81502-2 (2021).
73. 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.* **53**, 16–28. https://doi.org/10.1093/abm/kay011 (2019).
74. Guevarra, D. A., Moser, J. S., Wager, T. D. & Kross, E. Placebos without deception reduce self-report and neural measures of emotional distress. *Nat. Commun.* **11**, 1–8. https://doi.org/10.1038/s41467-020-17654-y (2020).
75. Barnes, K., Yu, A., Josupeit, J. & Colagiuri, B. Deceptive but not open label placebos attenuate motion-induced nausea. *J. Psychosom. Res.* **125**, 109808. https://doi.org/10.1016/j.jpsychores.2019.109808 (2019).
76. Blease, C., Colloca, L. & Kaptchuk, T. J. Are open-label placebos ethical? informed consent ethical equivocations. *Bioethics* **30**, 407–414. https://doi.org/10.1111/bioe.12245 (2016).
77. Kirsch, I. Response expectancy and the placebo effect. *Int. Rev. Neurobiol.* **138**, 81–93. https://doi.org/10.1016/bs.irn.2018.01.003 (2018).
78. Haas, J. W., Rief, W. & Doering, B. K. Open-label placebo treatment: Outcome expectations and general acceptance in the lay population. *Int. J. Behav. Med.* **28**, 444–454. https://doi.org/10.1007/s12529-020-09933-1 (2021).
79. Benedetti, F., Frisaldi, E. & Shaibani, A. Thirty years of neuroscientific investigation of placebo and nocebo: The interesting, the good, and the bad. *Annu. Rev. Pharmacol.* **62**, 323–340. https://doi.org/10.1146/annurev-pharmtox-052120-104536 (2022).
80. Mauri, L. & D'Agostino, R. B. Sr. Challenges in the design and interpretation of noninferiority trials. *N. Engl. J. Med.* **377**, 1357–1367. https://doi.org/10.1056/NEJMra1510063 (2017).
81. Kim, S., Thibodeau, R. & Jorgensen, R. S. Shame, guilt, and depressive symptoms: A meta-analytic review. *Psychol. Bull.* **137**, 68. https://doi.org/10.1037/a0021466 (2011).

Acknowledgements

The authors thank Emil Nissen, Mareike Rytz, Saruga Kunaratnam, Tim Kilb, Maya Lüthi, Samuel Jenny, Rebecca Bendel, and Julia Gottstein for conducting the studies, for their input on designing of the experiment and on processing and analyzing the data, and for their help with the references. We also thank Kyrill Schwegler for his valuable input and help with establishing the guilt-induction paradigm.

Author contributions

J.G. had the idea for the study. J.G. and C.L. designed the study. J.G. contributed to the data acquisition. D.S. carried out the analysis. D.S. wrote the manuscript, which was revised by all the authors.

Funding

CL received funding from the Swiss National Science Foundation (SNSF): PZ00P1_201972.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-25446-1.

Correspondence and requests for materials should be addressed to D.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary material: Deceptive and open-label placebo effects in experimentally-induced guilt: A randomized controlled trial in healthy subjects

Dilan Sezer[1*], Cosima Locher, and Jens Gaab[1]

[1]Division of Clinical Psychology and Psychotherapy, Faculty of Psychology, University of Basel, Switzerland

* Corresponding author

E-Mail: dilan.sezer@unibas.ch (DS)

Content

1. Instructions Guilt Paradigm (all groups)

“This task is about writing down a personal memory where you think you misbehaved or hurt someone close to you and therefore still feel **guilty** today, i.e. when you think about it, you still feel bad. It is important that you not only think about it when you write it down, but that you also think back intensively to this situation, to this moment.

You have the next 10 minutes for this task. You can write about your memory here on this paper, it should help you to put yourself in the place of what happened and also to relive the feelings of that time. However, this note is only for you and we will not read your note and you can take it with you after the end of the study if you wish. I will bring it to your attention when the time is up.”

2. Intervention Scripts

Deceptive Placebo (translated from German)

"Now that the feelings about the experience you described are present again, we would like to try to reduce these feelings. For this purpose, you will receive a pill from me. It contains a phytopharmacological active ingredient." (Study personnel shows the package and removes the pill from the blister). "So, it is purely herbal and does not require a prescription. The active ingredient has been empirically proven to have a positive emotional effect on depressive moods and we expect that this active ingredient will also noticeably reduce the activated feelings of guilt again. In addition to phytopharmacological substances such as butterbur, essential oils and St. John's wort, the pill also contains povidone and macrogol6000; these are valerian derivatives that have a calming and relaxing effects. Via the bloodstream, they reach serotonin receptors in the allocortex, more precisely in the amygdala, where they develop their relaxing and calming effect within 3-5 minutes and then work for at least the next 45 minutes. In the clinical field, this pill is mainly used for depression, anxiety and sleep disorders. We use it because we want to take advantage of its positive effect on feelings of guilt. We do not expect any side effects. Here is the pill." (Study participant takes the pill)

Open-label Placebo (translated from German)

"Now that the feelings about the experience you described are present again, we would like to try to reduce these feelings. After this instruction you will receive a pill from me. The dragée is a placebo and therefore does not contain any medicinal active ingredient." (Study personnel shows the package and removes the pill from the blister). "The pills consists only of sugar (lactose, sucrose, glucose) and stabilisers. However, we know that placebos are very effective. Openly-administered placebos have been used for various clinical disorders and problems and have shown very good efficacy. This has been shown in studies on pain and depression, among others. In the case of depression, feelings of guilt in particular are a major component. Placebos work through expectation and learning processes. The body reacts automatically and symptoms are reduced. An open, positive attitude towards placebos can be helpful, but this is not necessary for a positive effect. We expect the activated guilt feelings to be noticeably reduced again and use it because

we want to use its positive emotional effect on the guilt feelings. In doing so, we do not expect any side effects. Here is the pill" (Study participant takes the pill.)

No Treatment (translated from German)

"You were randomly assigned to the control group, which does not receive any intervention and thus enables the comparison to other study groups, which, unlike you, received an experimental treatment. Such control groups are necessary in research to find out whether a treatment really has a specific effect or whether a possible effect can be attributed to the natural course of symptoms or to measurement errors. Hence, your data can be used as a reference for the other study groups."

3. Instructions Guilt Boost (all groups)

“At the beginning of this experiment you wrote a text. We now ask you to remember again exactly the event that really happened and in which you behaved badly and unfairly towards a person close to you. Please try again to put yourself emotionally back into this unpleasant situation. To be able to concentrate better, please close your eyes for 1 minute. After one minute you can open your eyes again and continue.”

4. Results not displayed in the manuscript

Table S1

Overview of mean values per group for all assessed outcomes and time points.

		T0	T1	T2	T3	T4
SSGS		group			Mean (SD)	
Guilt						
	DP (35)	2.05 (0.80)	3.10 (1.05)	2.06 (0.79)	2.31 (0.79)	1.99 (0.84)
	OLP (35)	2.06 (1.00)	3.55 (0.79)	2.36 (0.84)	2.61 (0.91)	2.18 (0.93)
	NT (39)	1.92 (0.70)	3.25 (1.02)	2.37 (0.83)	2.94 (1.09)	2.36 (0.94)
Shame						
	DP (35)	1.53 (0.55)	2.17 (0.74)	1.4 (0.41)	1.53 (0.50)	1.41 (0.41)
	OLP (35)	1.67 (0.79)	2.45 (0.75)	1.61 (0.57)	1.76 (0.69)	1.59 (0.79)
	NT (39)	1.34 (0.44)	2.25 (0.87)	1.67 (0.57)	1.9 (0.73)	1.57 (0.62)
Pride						
	DP (35)	3.45 (0.45)	2.58 (0.65)	2.98 (0.66)	2.70 (0.69)	2.98 (0.55)
	OLP (35)	3.47 (0.55)	2.53 (0.72)	3.09 (0.61)	2.77 (0.65)	3.01 (0.64)
	NT (39)	3.62 (0.58)	2.74 (0.90)	3.15 (0.74)	2.68 (0.98)	3.07 (0.88)
PANAS						
Positive						
	DP (35)	3.05 (0.58)	2.43 (0.58)	2.71 (0.61)	2.61 (0.65)	2.62 (0.53)
	OLP (35)	3.27 (0.60)	2.58 (0.63)	2.92 (0.61)	2.75 (0.71)	2.8 (0.83)
	NT (39)	3.25 (0.61)	2.66 (0.74)	2.95 (0.72)	2.71 (0.81)	2.8 (0.72)
Negative						
	DP (35)	1.3 (0.36)	1.94 (0.72)	1.34 (0.34)	1.47 (0.39)	1.36 (0.39)
	OLP (35)	1.44 (0.42)	2.23 (0.69)	1.52 (0.42)	1.63 (0.39)	1.41 (0.39)
	NT (39)	1.52 (0.42)	2.19 (0.76)	1.78 (0.56)	1.96 (0.71)	1.73 (0.60)

Note. SD, standard deviation; DP, deceptive placebo; OLP, open-label placebo; NT, no treatment; SSGS, State Shame and Guilt Scale; PANAS, Positive and Negative Affect Schedule

Table S2

Time effects of (robust) two-way mixed ANOVAs for SSGS and PANAS scores across T0 – T2 and mean changes from T0 – T1.

	time	**Change T0 – T1** (Mean (SD))			
SSGS		**Overall**	**DP (N = 35)**	**OLP (N = 35)**	**NT (N = 39)**
Guilt	$Q(2, 49) = 92.18$, ***p* < .001** (robust)	1.29 (0.98)	1.06 (0.88)	1.49 (0.96)	1.33 (1.06)
Shame	$Q2, 52) = 85.81$, ***p* < .001** (robust)	0.78 (0.72)	0.65 (0.64)	0.78 (0.63)	0.91 (0.86)
Pride	$F(2, 212) = 108..45$, ***p* < .001**	-0.89 (0.69)	-0.87 (0.59)	-0.94 (0.77)	-0.88 (0.73)
PANAS					
Positive	$F(2, 212) = 80.121$, ***p* < .001**	-0.63 (0.56)	-0.62 (0.57)	-0.69 (0.53)	-0.59 (0.58)
Negative	$Q(2, 51) = 16.29$, ***p* < .001** (robust)	0.69 (0.60)	0.64 (0.57)	0.79 (0.49)	0.67 (0.71)

Note. SSGS, State Shame and Guilt Scale; PANAS, Positive and Negative Affect Schedule; Q values indicate robust analysis.

Table S3

Time effects of robust two-way mixed ANOVAs for SSGS and PANAS scores across T2 – T4 and mean changes from T2 – T3.

	time	**Change T2 – T3** (Mean (SD))			
SSGS		**Overall**	**DP (N = 35)**	**OLP (N = 35)**	**NT (N = 39)**
Guilt	Q(2, 48) = 19.12, p < .001 (robust)	0.36 (0.61)	0.25 (0.54)	0.25 (0.57)	0.57 (0.66)
Shame	Q(2, 49) = 16.62, p < .001 (robust)	0.19 (0.46)	0.13 (0.42)	0.15 (0.44)	0.29 (0.51)
Pride	Q(2, 51) = 22.87, p < .001 (robust)	-0.36 (0.45)	-0.27 (0.36)	-0.31 (0.44)	-0.48 (0.51)
PANAS					
Positive	Q(2, 48.1) = 11.36, p <.001 (robust)	-0.17 (0.38)	-0.10 (0.32)	-0.17 (0.34)	-0.24 (0.46)
Negative	Q(2, 51.1) = 16.29, p < .001 (robust)	0.14 (0.33)	0.13 (0.25)	0.12 (0.31)	0.18 (0.39)

Note. SSGS, State Shame and Guilt Scale; PANAS, Positive and Negative Affect Schedule; Q values indicate robust analysis.

Table S4

Comparison of AUCi sizes using Kruskal Wallis test or one-way ANOVA for T0 – T2 and T2 – T4.

	T0 – T2	**T2 – T4**
SSGS		
Guilt	$F(2, 106) = 2.16$, $p = .121$	$F(2, 106) = 3.38$, ***p* = .038**
Shame	Kruskal Wallis test, $p = .173$	$F(2, 106) = 0.19$, $p = .826$
Pride	Kruskal Wallis test, $p = .923$	$F(2, 106) = 1.681$, $p = .191$
PANAS		
Positive	Kruskal Wallis test, $p = .499$	Kruskal Wallis test, $p = .581$
Negative	Kruskal Wallis, $p = .34$	$Q(2, 40.09) = 0.43$, $p = .652$ (robust)

Note. SSGS, State Shame and Guilt Scale; PANAS, Positive and Negative Affect Schedule; AUCi, Area und the Curve increase.

Table S5

Area under the curve SSGS and PANAs scores and between-group contrasts for T0 – T2.

	DP (N = 35)	**OLP (N = 35)**	**NT (N = 39)**	**DP & OLP vs. NT**	**DP vs. OLP**
SSGS	Mean (SD)			Mean difference (CI)	
Guilt	10.66 (11.19)	16.43 (12.89)	15.59 (13.56)	2.05 (-3.67 – 7.77), d = 0.16	5.77 (-1.07 – 12.62), d = 0.46
Shame	5.83 (8.25)	7.49 (9.10)	10.72 (10.28)	4.06 (-0.15 – 8.27), d = 0.44	1.66 (-3.38 – 6.69), d = 0.18
Pride	-11.11 (8.11)	-11.37 (9.86)	-11.08 (8.99)	0.17 (-3.92 – 4.25), d = 0.02	-0.26 (-5.14 – 4.63), d = -0.03
PANAS					
Positive	-7.91 (7.88)	-8.70 (6.54)	-7.42 (6.75)	0.88 (-2.32 – 4.09), d = 0.13	-0.79 (-4.62 – 3.05), d = -0.11
Negative	6.60 (6.14)	8.30 (6.22)	8.01 (8.89)	0.56 (-2.73 – 3.86), d = 0.08	1.7 (-2.24 – 5.64), d = -0.23

Note. SD, standard deviation; DP, deceptive placebo; OLP, open-label placebo; NT, no treatment; SSGS, State Shame and Guilt Scale; PANAS, Positive and Negative Affect Schedule; AUCi, area under the curve with respect to increase; CI, confidence interval, * p adj. < 0.05

Associations of additional variables with outcomes

There was no significant difference among baseline guilt and shame proneness across groups (PFQ guilt: Kruskal Wallis test, p = .671; PFQ shame: Kruskal Wallis test, p = .241).

Table S6

Overview of mean values per group for all assessed predictors and their correlation with AUCi sizes from T2 – T4.

	DP (N = 35)	OLP (N = 35)	NT (N = 39)	*guilt*	*shame*	*pride*	*positive*	*negative*
Expectation of Relief (T2)								
Expectation of Symptom intensity after treatment (NRS)	3.23 (1.72)	4.49 (2.11)	4.23 (2.10)	0.12	0.1	0.04	0.05	0.02
PFQ (T0)								
Guilt	21.71 (3.16)	21.14 (4.03)	21.28 (3.02)	0.07	-0.03	-0.04	0.09	0.12
Shame	32.66 (3.32)	33.31 (2.97)	32.18 (3.49)	0.15	-0.12	0.06	0.09	**-0.21***

Note. SD, standard deviation; DP, deceptive placebo; OLP, open-label placebo; NT, no treatment; CMQ, Context Model Questionnaire; PFQ-2, Personal Feelings Questionnaire 2; NRS (1-10), Numeric Rating Scale; * $p < 0.05$

Appendix B: Study II

Buergler, S., Sezer, D., Bagge, N., Kirsch, I., Locher, C., Carvalho, C., & Gaab, J. (2023). Imaginary pills and open-label placebos can reduce test anxiety by means of placebo mechanisms. *Scientific Reports, 13*(1), 2624. https://doi.org/10.1038/s41598-023-29624-7

scientific reports

Check for updates

OPEN

Imaginary pills and open-label placebos can reduce test anxiety by means of placebo mechanisms

Sarah Buergler[1]✉, Dilan Sezer[1], Niels Bagge[2], Irving Kirsch[3], Cosima Locher[4,5], Claudia Carvalho[6] & Jens Gaab[1]

Placebos have been shown to be beneficial for various conditions even if administered with full transparency. Hence, so-called open-label placebos (OLPs) offer a new way to harness placebo effects ethically. To take this concept one step further, this study aimed at evaluating placebo effects without the use of a physical placebo, i.e., by imagining taking a pill. Healthy students (*N* = 173) with self-reported test anxiety were either randomized to an imaginary pill (IP; n = 55), an OLP (n = 59) or a control group (CG; n = 59). Both intervention groups were instructed to take two pills daily for three weeks. Primary outcome was test anxiety, secondary outcomes were sleep quality, general well-being and test performance. Groups test anxiety differed at study-endpoint, $F(2,169) = 11.50$, $p < .001$. Test anxiety was lower in the intervention groups compared to the CG, $t(169) = -4.44$, $p < .001$, $d = -0.71$. The interventions did not differ significantly, i.e., both were similarly efficacious, $t(169) = 0.61$, $p = .540$, $d = 0.11$. The interaction between group and time in explaining test anxiety was significant, $F(5,407.93) = 6.13$, $p < .001$. OLPs and IPs reduced test anxiety in healthy participants compared to the CG. This finding opens the door for a novel and ethical method to harness placebo effects.

Placebo effects are clinically highly relevant and the need to harness these effects has been voiced[1]. In this regard, open-label placebos (OLPs) administered with full disclosure and transparency can be deemed both ethical and feasible as they avoid the use of deception[2]. Interestingly, meta-analyses show medium sized to large clinically relevant effects of OLPs in patients with various clinical conditions compared to control groups[3,4]. Thus, if placebos also work without deception, it implies that it is not necessarily the pill serving as a symbol for a real medication that triggers these effects. The investigation of underlying mechanisms by eliminating the physical treatment constituent (i.e., the pill itself) can reveal the power of the purely psychological component of a placebo. For this reason, we aimed to evaluate placebo effects without the use of a placebo by having participants imagine taking a pill rather than actually taking one.

The concept of an imaginary pill (IP) was first introduced by De Shazer in 1984 in the context of clinical hypnosis[5]. More recently, Niels Bagge, a Danish clinician, independently introduced the same idea without hypnosis[6]. Although seemingly farfetched, recent data supports its plausibility: For instance, pharmacological placebos can be effective even when only possessed, but not applied[7]. Also, psychotherapeutic, non-pharmacological placebos have been shown to be effective[8] and the idea of triggering placebo effects without a placebo pill is discussed in sports performance[9], healthcare[10] and in research on the moderating role of mind-sets[11]. Additionally, a study by Peerdeman et al.[12] indicated that mental imagery of reduced pain can induce placebo-like expectancy effects on pain. Thus, placebos can also be purely psychological in nature and still produce beneficial effects. With regard to the underlying mechanisms of such psychological placebos, it yet needs to be investigated, whether their efficacy is purely mediated by the meaning that is attributed to these rituals or the expectations of improvement that are being formed as a consequence[13,14]. Despite the elimination of the physical stimulus, it is plausible that an IP relies in principle on the same underlying mechanisms as an OLP. Besides expectation, conditioning could for instance play a role, as even imagining something can activate corresponding brain areas and associated learning mechanisms (e.g.[15]). In addition, placebo mechanisms have also been discussed in relation to the theory of embodied cognition, which states that our experiences are not only consciously stored

[1]Division of Clinical Psychology and Psychotherapy, Faculty of Psychology, University of Basel, Basel, Switzerland. [2]Institute for Emotion-Focused Therapy, Roskilde, Denmark. [3]Program in Placebo Studies, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, USA. [4]Department of Consultation-Liaison Psychiatry and Psychosomatic Medicine, University Hospital Zurich, University of Zurich, Zurich, Switzerland. [5]Faculty of Health, University of Plymouth, Plymouth, UK. [6]Department of Clinical and Health Psychology, Instituto Superior de Psicologia Aplicada (ISPA), Lisbon, Portugal. ✉email: sarah.buergler@unibas.ch

as memories, but also directly imprinted in our bodies without any cognitive process being involved[16]. Thus, a placebo effect could result from the unconscious internal act of imagining a specific change in body state[16], which underlines the potential of IPs to harness placebo effects. In conclusion, this study provides a next step towards the quantification of the combined effect of various plausible psychological mechanisms within placebo research, omitting the physical treatment component.

In light of the high prevalence of test anxiety, affecting for example 53% of German freshman medical students[17], and its negative impact on educational performance[18,19], this condition is suitable to test the effects of an IP and OLP intervention. Evidence suggests that OLPs can effectively reduce test anxiety in healthy college students[20] and can have a positive impact on subjective well-being, whereas no improvement of exam performance by the intervention was found[21]. Furthermore, placebo effects in psychopharmacological treatments of anxiety disorders in general[22,23], social anxiety[24,25], generalized anxiety disorders[26] and panic disorders[27,28] are moderate to large.

In the present study, we set out to test the efficacy of an IP and OLP intervention in reducing test anxiety in a randomized controlled trial with healthy participants. To pursue this research question, we applied a previously used OLP intervention (e.g. in[29,30]) and further developed an IP intervention that was based on knowledge derived from placebo and imagination research to compare them to a control group (CG). We hypothesized that students receiving the OLP and IP intervention would show greater decreases in test anxiety from baseline to study endpoint (shortly before the exam) compared to students in the CG. We further expected students in the intervention groups to show higher general well-being, higher sleep quality and higher test performance than students in the CG.

Results

Sample characteristics and study flow. As shown in Fig. 1, of the 283 interested participants, 33 did not provide an e-mail contact and six did not give informed consent. The remaining 244 participants completed the online screening, of which 15 did not fulfill at least one inclusion criteria and 18 were excluded due to other reasons. Hence, 211 participants were randomized, of whom 178 received the intervention and completed the baseline assessment (T1; see Fig. 1 for reasons of exclusion). Five participants were excluded from the analyses as there was missing data (mostly due to nonattendance at exams because of COVID-19). Hence, an *N* of 173 was used for the final analyses (IP = 55, OLP = 59, CG = 59).

Table 1 depicts participants demographic and baseline characteristics. Participants' age ranged from 18 to 47 years, with a mean of 22.70 (± 4.18) years. The majority were female (85.55%) and undergraduate psychology students (87.86%). The three groups did not significantly differ in any of the demographic characteristics or primary and secondary outcomes at baseline. All outcomes were within the normal range of scores in the anxiety, well-being, and sleep questionnaires, indicating that our sample was healthy displaying an average test anxiety score.

Primary and secondary outcomes at study endpoint (T4). Figure 2 shows mean improvement from baseline (T1) to endpoint (T4) per group on the primary outcome. Table 2 depicts all primary and secondary outcomes for all groups and assessments.

The overall analyses of covariance (ANCOVA) showed that groups significantly differed in test anxiety at study endpoint T4, $F(2, 169) = 11.50$, $p < .001$. Planned contrasts indicated that the mean changes in test anxiety were significantly greater in the intervention groups (OLP/IP) compared to the CG at study endpoint T4, $t(169) = -4.44$, $p < .001$, $d = -0.71$. However, changes in test anxiety did not differ between the two intervention groups, $t(169) = 0.61$, $p = .540$, $d = 0.11$. These results held true for all subscales of the test anxiety questionnaire (see supplementary Table S1).

Regarding secondary outcomes, the groups differed significantly in terms of general well-being at study endpoint T4, $F(2, 169) = 9.37$, $p < .001$ with the same pattern across all subscales. Changes in general well-being were significantly greater in the intervention groups (OLP/IP) compared to the CG at study endpoint T4, $t(169) = -3.98$, $p < .001$, $d = -0.64$, but did not differ between the two intervention groups, $t(169) = 0.38$, $p = .707$, $d = 0.07$.

No significant between-group effect was found for total sleep quality, $F(2, 169) = 0.902$, $p = .408$, or in any of its component subscales. Nevertheless, although the overall between-group effect on the subjective sleep quality at study endpoint failed to reach statistical significance, $F(2, 169) = 2.73$, $p = .068$, contrasts indicated that both intervention groups showed better subjective sleep quality compared to the CG, $t(169) = -2.40$, $p = .017$, $d = -0.39$, with no significant difference between the intervention groups, $t(169) = 0.06$, $p = .952$, $d = 0.01$.

With respect to the test performance, 120 participants had a continuous test score (IP = 41, OLP = 35, CG = 44). Mean grade was 4.82 (± 0.83) ranging from 2.5 to 6.0 (IP = 4.94 ± 0.83, OLP = 4.92 ± 0.76, CG = 4.62 ± 0.87). Figure 3 depicts the participants grades per group. The overall ANOVA showed no significant group effect on test score, $F(2, 117) = 1.98$, $p = .143$. The contrasts, however, indicated that the intervention groups (OLP/IP) had higher test scores compared to the CG, $t(117) = 1.98$, $p = .050$, $d = 0.38$, whereas the intervention groups did not differ, $t(117) = -0.12$, $p = .908$, $d = -0.03$. Binary test scores (pass/fail) revealed that 155 (89.60%) of all participants passed the exam (IP = 87.27%, OLP = 96.61%, CG = 84.75%).

Primary and secondary outcomes over time (T1–T4). Figure 4 shows the course of test anxiety outcomes over time. There was a statistically significant interaction between group and time (T1–T4) for test anxiety, $F(5, 407.93) = 6.13$, $p < .001$. Bonferroni adjusted post-hoc p-values showed that the simple main effect of group was significant after one week (T2; $p_{adj} < .001$), two weeks (T3; $p_{adj} < .001$) and three weeks (T4; $p_{adj} < .001$) after

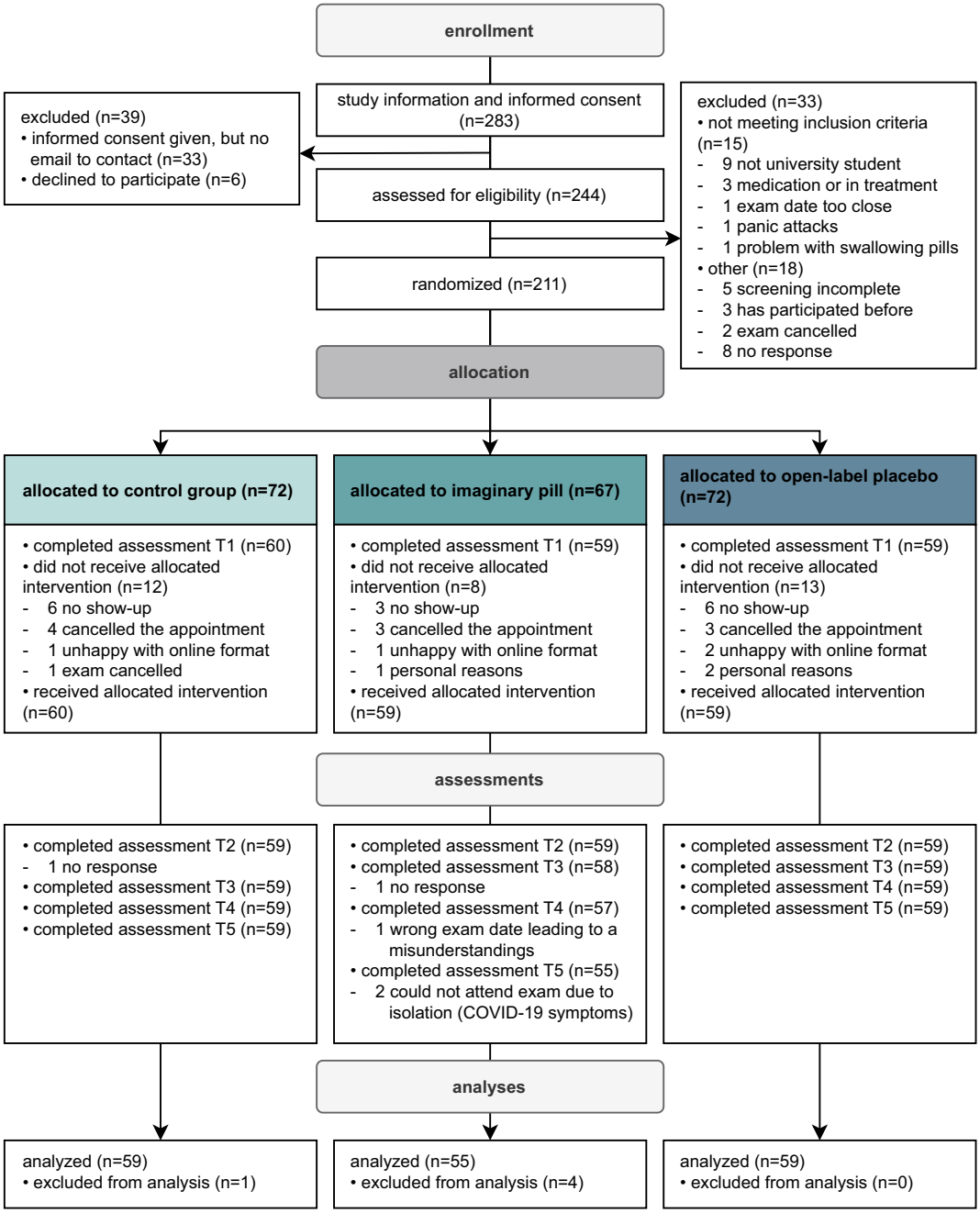


Figure 1. CONSORT diagram. Flow of the study, including reasons for exclusions.

	IP	OLP	CG
N (% female)	55 (82%)	59 (90%)	59 (85%)
Age in years, M (SD)	23.20 (4.30)	22.00 (3.48)	22.95 (4.67)
Psychology students, N (%)	49 (89%)	52 (88%)	51 (86%)
Test anxiety (PAF), M (SD)	45.53 (6.81)	48.34 (6.94)	47.36 (6.37)
Sleep quality (PSQI), M (SD)	5.69 (2.83)	6.02 (2.92)	5.85 (3.05)
General well-being (ASS-SYM), M (SD)	45.40 (20.90)	49.03 (22.52)	48.86 (20.76)

Table 1. Demographics and baseline scores of primary and secondary outcomes per group. *ASS-SYM* Änderungssensitive Symptomliste (general well-being), *CG* control group, *IP* imaginary pill, *M* mean, *OLP* open-label placebo, *PAF* Prüfungsangstfragebogen (test anxiety questionnaire), *PSQI* pittsburgh sleep quality index, *SD* standard deviation.

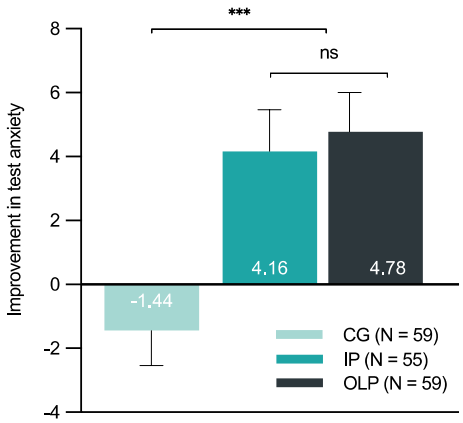


Figure 2. Mean improvement in test anxiety (PAF: test anxiety questionnaire) from baseline (T1) to endpoint (T4) per group. Results indicate a significant improvement for the OLP and IP group compared to the CG. *Note. CG* control group, *IP* imaginary pill, *ns* = not significant, *OLP* open-label placebo, ***$p < .001$. Error bars represent standard error of the mean.

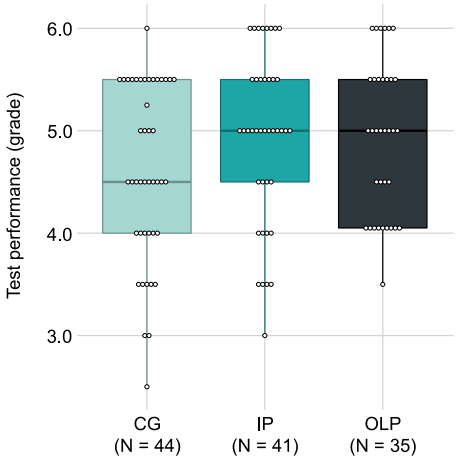


Figure 3. Boxplot showing continuous grades of the participants per group. Every dot represents a participants' grade with higher grades being better (ranging from 1.0 to 6.0). *Note.* Median is represented by the bold line within the box and upper/lower quartiles mark the end of the box. *CG* control group, *IP* imaginary pill, *OLP* open-label placebo.

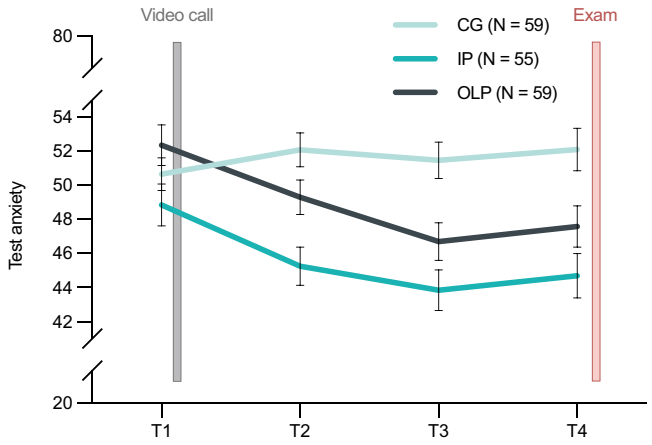


Figure 4. Course of test anxiety over time. Mean test anxiety per group from baseline (T1) through midpoints (T2, T3) to study endpoint (T4). *Note.* Error bars represent standard error of the mean. *CG* control group, *IP* imaginary pill, *OLP* open-label placebo.

		T1	T2	T3	T4	T5
	Group (*N*)	M (SD)				
PAF	IP (55)	48.85 (9.20)	45.25 (8.26)	43.84 (8.86)	44.69 (9.72)	39.85 (10.18)
	OLP (59)	52.36 (9.18)	49.31 (7.81)	46.70 (8.49)	47.58 (9.39)	41.64 (10.60)
	CG (59)	50.66 (7.37)	52.08 (7.70)	51.47 (8.23)	52.10 (9.56)	45.78 (9.97)
ASS-SYM	IP (55)	45.40 (20.90)	42.55 (22.57)	37.29 (21.9)3	39.25 (23.04)	
	OLP (59)	49.03 (22.52)	46.56 (19.48)	44.58 (19.96)	42.85 (21.99)	
	CG (59)	48.86 (20.76)	52.85 (24.36)	51.81 (22.02)	55.80 (28.58)	
PSQI	IP (55)	5.69 (2.83)	5.69 (2.48)	5.49 (2.46)	5.54 (2.71)	
	OLP (59)	6.02 (2.92)	5.49 (2.52)	5.88 (3.08)	5.86 (2.82)	
	CG (59)	5.85 (3.05)	6.36 (2.94)	6.36 (2.90)	6.22 (3.00)	

Table 2. Mean values for primary and secondary outcomes per group at all assessed timepoints. *ASS-SYM* Änderungssensitive Symptomliste (general well-being), *CG* control group, *IP* imaginary pill, *M* mean, *OLP* open-label placebo, *PAF* Prüfungsangstfragebogen (test anxiety questionnaire), *PSQI* pittsburgh sleep quality index, *SD* standard deviation.

randomization, but not at baseline (T1; $p_{adj} = .098$). Furthermore, there was also a statistically significant effect of time on test anxiety scores for the IP ($p_{adj} < .001$) and OLP ($p_{adj} < .001$) group, but not for the CG ($p_{adj} = .318$).

Regarding the secondary outcomes, a statistically significant interaction was found between group and time (T1–T4) in general well-being, $F(5, 422.71) = 3.58$, $p = .004$. Considering the Bonferroni adjusted p-values, the simple main effect of group was significant at T3 ($p_{adj} = .004$) and T4 ($p_{adj} = .004$), but not at T1 ($p_{adj} = .598$) and T2 ($p_{adj} = .061$). Also, the effect of time was significant with an increase of general well-being in the IP ($p_{adj} < .01$), but not in the OLP ($p_{adj} = .191$) group, whereas general well-being in the CG showed a trend to decrease ($p_{adj} = .071$). There was no significant interaction between group and time on sleep quality scores, $F(5, 443.85) = 0.90$, $p = .485$.

Rating of test anxiety at follow up, opinion on treatment idea, side-effects and adherence. Regarding the retrospective evaluation of the test situation (T5), the overall ANCOVA showed a significant overall effect of group, $F(2, 169) = 5.89$, $p = .003$. Planned contrasts indicated that mean retrospective test anxiety scores were rated significantly lower in the intervention groups (OLP/IP) compared to the CG at T5, $t(169), = -3.29$, $p = .001$, $d = -0.53$. However, retrospective test anxiety scores did not differ between the two intervention groups, $t(169) = 0.10$, $p = .918$, $d = 0.02$.

Table 3 provides an overview of the evaluation of the idea (positive, negative, neutral) towards the two interventions in the context of the open questions. The two independent raters had concordant judgments for 91.2% of the answers. A third rater was included for the remaining 8.8%.

No negative side-effects were reported, other than in the IP group, in which three subjects mentioned additional effects immediately after pill intake (i.e., dry mouth, goose bumps, warmth radiating from the abdomen). These effects were suggested during the pill intake in the study contact and were part of the IP response to demonstrate the effect of the pill (see supplementary material).

Regarding adherence, one participant (1.7%) in the OLP group and five participants (9.1%) in the IP group reported less than 80% adherence (i.e., forgot 9 or more pills).

Influence of study contact duration, treatment provider and moderation of treatment expectancy on primary outcome. Study contact duration was significantly associated with changes in test anxiety from T1 to T4, $F(1, 168) = 5.84$, $p = .017$. However, when including treatment group as an additional factor in the model, contact duration was no longer significant, $F(1, 166) = 0.01$, $p = .942$ and group remained significant, $F(2, 166) = 8.00$, $p < .001$. Treatment provider was not associated with changes in test anxiety, $F(1, 171) = 0.80$, $p = .373$.

Mean expectancy of relief across the 20 items of the test anxiety questionnaire was significantly different across the three groups, $F(2, 170) = 14.86$, $p < .001$. Participants receiving an intervention (IP/OLP) expected less symptoms compared to the CG, $t(169) = -5.76$, $p < .001$, $d = -0.92$, whereas scores of the two intervention groups were comparable, $t(169) = 1.47$, $p = .144$, $d = 0.28$. Mean expectancy of relief measures significantly correlated with endpoint test anxiety (T4; $r = 0.56$, $p < .001$). When including expectancy of relief as an additional covariate into

	OLP (*N* = 59) *N* (%)	IP (*N* = 55) *N* (%)
Positive	39 (66.1%)	38 (69.1%)
Negative	8 (13.6%)	10 (18.2%)
Neutral	12 (20.3%)	7 (12.7%)

Table 3. Ratings of the open-ended questions. Judgement of the idea regarding the respective interventions. *IP* imaginary pill, *OLP* open-label placebo.

the overall model, expectancy of relief was significantly associated with test anxiety, $F(1, 168) = 21.14$, $p < .001$, but treatment group remained significant, $F(2, 168) = 12.87$, $p < .001$.

Discussion

The present randomized controlled trial tested the effects of an IP against an OLP intervention and a CG on test anxiety in healthy students. We found that both IP and OLP significantly reduced test anxiety compared to the CG with a moderate-to-large effect size (d = − 0.71). These findings were comparable across all subscales of the test anxiety questionnaire (i.e., worry, emotionality, interference and lack of confidence). Interestingly, the beneficial effect was apparent over the course of the three weeks, starting after only one week of intervention. While study contact duration and treatment provider did not appear to be critical for changes in test anxiety, the observed effects were associated with treatment expectancy as this measure positively correlated with changes in test anxiety (r = 0.56). The retrospective assessment of the exam situation (follow-up T5) supports the superiority of the two interventions over the CG, as it indicated less retrospectively perceived anxiety during the exam situation. Consistent with the effects on our primary outcome, general well-being was significantly augmented in both intervention groups compared to the CG with a moderate to large effect (d = − 0.64). Overall sleep quality, however, was not affected by the intervention, i.e., all three groups showed comparable sleep quality during the three weeks. Test performances (i.e., continuous grades) were significantly better in the intervention groups compared to the CG with a small effect (d = 0.38). Overall, OLP and IP showed comparable results on all assessed outcomes. These findings question the necessity of the pill to produce positive treatment effects.

The effect sizes of the two interventions in the present study are slightly higher compared to a previous OLP trial testing openly prescribed placebos in test anxiety against no treatment with a between group effect size of d = 0.54[20], whereby test anxiety scores in both studied populations indicate average, non-clinical test anxiety[31]. The remarkable and rapid decreases in test anxiety in the intervention groups of the present study are noteworthy. The observed effect is comparable to the moderate-to-large effect (g = − 0.76) of a meta-analysis on various psychological interventions for test-anxious university students (i.e., psychological, study skill training, and/or combined intervention packages) against control conditions[32].

Extended or different placebo paradigms such as IPs aid to understand the mechanisms of OLP by systematically manipulating the treatment setting and application. As OLP and IP groups showed comparable results in all outcomes, the necessity of a physical placebo to produce positive treatment effects is called into question. Psychological components, for their part, may be sufficient on their own to exploit placebo effects which is supported by studies showing that triggering placebo effects without a physical treatment component is possible[8,11,33]. Research on placebo-like expectancy effects in pain analgesia is consistent with this: Peerdeman et al. (2017) showed less experienced pain in participants receiving instructions to vividly imagine a warm and impermeable glove preventing pain from cold before a cold pressor test, compared to a control imagery group instructed to imagine their hand without any reference to pain or cold water. This effect was mediated by expected pain[12]. Along these lines, expectancy of relief was also significantly associated with test anxiety in our study. However, the treatment group remained significant even after expectancy was included in our linear model, implying that not only expectancy but also other factors must account for the group-specific improvement in test anxiety. The effects can, for example, be discussed in the context of the embodied cognition theory, which states that a placebo effect can result unconsciously from embodied experiences by an internal act of imagining a particular state change in the body[16]. Similarly, conditioning effects may have played a role in our study, as even imagining something can activate corresponding brain areas (e.g.[15]). The Western cultural understanding of a pill underpins this line of reasoning as a pill in itself has a therapeutic meaning—learning from an early age to associate the pill and its effects, whereas no physical pill is required to trigger positive processes. Notably, mental imagery relies on similar neural processes to those of actual perception[34,35]. The ability to generate internal representations that retain the essential features of a perceptual experience suggests that mental imagery may have similar effects to actual experiences[12]. Consistent with the response expectancy theory[14] the findings of this study extend previous research on the mechanisms of placebo effects by showing that placebo effects on test anxiety can be induced not only by a physical cue, but also by imagining a pill and its effects. Overall, it can be suggested that OLP and IP may rely on the same underlying mechanisms (e.g., expectations, conditioning, embodied cognition), whereas these mechanisms can be triggered even in the absence of a physical pill.

Due to the COVID-19 pandemic, the study contact took place by means of a virtual clinical encounter. The present study is not the first to provide the OLP treatment remotely: Kube and colleagues[36], however, failed to replicate previous findings of OLP effects on allergic rhinitis[37,38]. They concluded that remote OLP provision is feasible, yet their effectiveness might be lower, as a physical encounter between patient and provider might be a prerequisite for OLPs to be effective[36]. However, our findings demonstrate that providing OLP and IP remotely is not only feasible but can also yield significant effects. A potential reason for the better effects in this online intervention compared to Kube et al. might be the younger sample (22.7 vs. 31.1 years) consisting only of students who may be more accustomed to online interactions. Whether the effects would be different with physical contact remains unclear and should be tested in a follow-up study.

This is the first study to conceptually extend ethically feasible placebo treatments by testing an IP intervention for test anxiety, taking OLP research a step further. It moreover corroborates important findings on OLP efficacy in a remote setting on a large sample. A manual including a five-step procedure was developed by our team to implement the IP intervention (see supplementary material). Manualized instructions used in the study further allowed for the control of many incidental factors to make accurate inferences about the interventions tested. Weekly assessments of primary and secondary outcomes moreover enabled observation of placebo effects over time. Also, there were less than 3% participants with missing data and reported nonadherence was low, especially in the OLP group.

Nevertheless, several aspects of this study need to be considered: Due to sample restriction and recruitment locations and routes, a largely female, young, academic sample resulted, limiting generalizability of the findings. Also, most outcome measures were self-reported and rather subjective than objective, raising questions of report and social desirability bias. Disappointment effects may have further played a role in the CG as they were not offered future treatment. In fact, 52.5% reported to be disappointed due to being allocated to the CG. However, given that test anxiety can be assumed to increase as the exam approaches[39], but the CG showed stable scores over time, it seems that despite disappointment this group also benefited from taking part in the study. In addition, adherence was self-reported, so we had no option to verify the reported values, eliminating the influence of social desirability bias. Further, because of planning reasons, a short time gap between study contact and start of the intervention occurred in some participants. However, this gap was kept to a minimum. Also, the conduct of the present study coincided with the start of the COVID-19 pandemic, which necessitated meeting with participants virtually. Due to the remote setting, the participants in the OLP group received the envelope with the placebo pills in advance: although not knowing about the contents of the envelope and being instructed not to open the envelope until the study contact occurred, we had no way of controlling this behavior. Nevertheless, positive effects of interventions could be observed and implementation remotely was feasible. Considering this, we assessed changes in test anxieties due to the pandemic-related circumstances which were almost evenly distributed across participants—with some reporting unchanged (33.5%), higher (34.1%) or less (32.4%) anxiety. Comparisons of within changes of participants should, however, control for these complicated circumstances. Future investigations should test OLP and IP with physical contact and no pandemic-related restrictions.

The present study is the first to conceptually expand on previous OLP studies by eliminating the physical pill as a treatment component and testing an IP intervention. Results indicate a moderate-to-large effect of both interventions on test anxiety and general well-being in a large cohort of 173 healthy students. These findings demonstrate that placebo effects can be harnessed without the use of a physical pill. The IP intervention could thus serve as a stand-alone or adjunct treatment to maximize and boost placebo effects in clinical practice, as indicated by the ethical principle of "beneficence"[1,40,41]. As an ethical, cost-effective, easily applicable and fully patient-centered method, the IP intervention has potential and should be tested in other settings, conditions and populations.

Methods

Experimental design. Between March 2020 and July 2021, we conducted an online randomized controlled, parallel group trial at the Division of Clinical Psychology and Psychotherapy at the Faculty of Psychology, University of Basel, Switzerland, in order to test the effects of an IP and OLP intervention compared to a CG in healthy students with test anxiety. Written informed consent was obtained from each participant before participation in the study. The Ethics Committee of the Faculty of Psychology, University of Basel approved the design and informed consent of the study. This study was carried out in accordance with the protocol and principles enunciated in the current version of the Declaration of Helsinki. The study was registered at ClinicalTrials.gov: NCT04250571 (31/01/2020).

Study population. Participants were recruited via web- and print-based advertisement (title: "Efficacy study of two treatment methods for test anxiety") and registered online for the study. Potential enrollees had to be students of the University of Basel aged between 18 and 65 years. To meet inclusion criteria, participants had to have an exam at the end of the semester, have self-reported test anxiety, being healthy by self-report (i.e., no known current or chronic primary pain disorders or psychiatric disorders) and be sufficiently proficient in German. Exclusion criteria were use of medications (psychoactive or narcotic), being in psychological or psychiatric treatment, taking psychotropic drugs, being a master student in Psychology (due to prior knowledge about placebo mechanisms), allergy to one of the ingredients of the placebo pills (see supplementary material), and problems swallowing pills. All participants were reimbursed either financially or with credit points.

Study procedure. The study procedure is depicted in Fig. 5. Interested participants were directed to an online survey page providing information about the nature and purpose of the study. Upon providing online informed consent, participants were checked online for inclusion and exclusion criteria. Eligible participants were randomly assigned to one of three study group. Baseline assessments of primary and secondary outcomes were completed online two or less days before the study contact (T1). The study contact, in which participants received one of three interventions (four to three weeks before the exam), was held online via the standard video call software of the University of Basel, zoom (https://zoom.us/), the use of which was approved by the ethics committee. Expectancy of relief was assessed immediately after the study contact. Study contacts were distributed over the time period of four to three weeks before the exam for resource management reasons (number of treatment provider). However, treatment started exactly three weeks before the exam as indicated by the receipt of a reminder e-mail in both intervention groups, i.e., the treatment duration was the same. Again, two weeks (T2), one week (T3) and two or less days before the exam (T4) all three groups were asked to complete online assessments of primary and secondary outcomes, as well as to answer one question regarding their intervention adherence. After the exam, there was a final online assessment (T5) to evaluate retrospective experiences of the exam situation, to assess side-effects during the treatment period and to answer open questions respective to the group (e.g., possible feelings of disappointment to be assigned to the CG, see supplementary material). Finally, all participants were asked about their examination grade (approximately two months after the exam).

Study arms. In total, there were three study arms, i.e., CG, OLP, and IP.

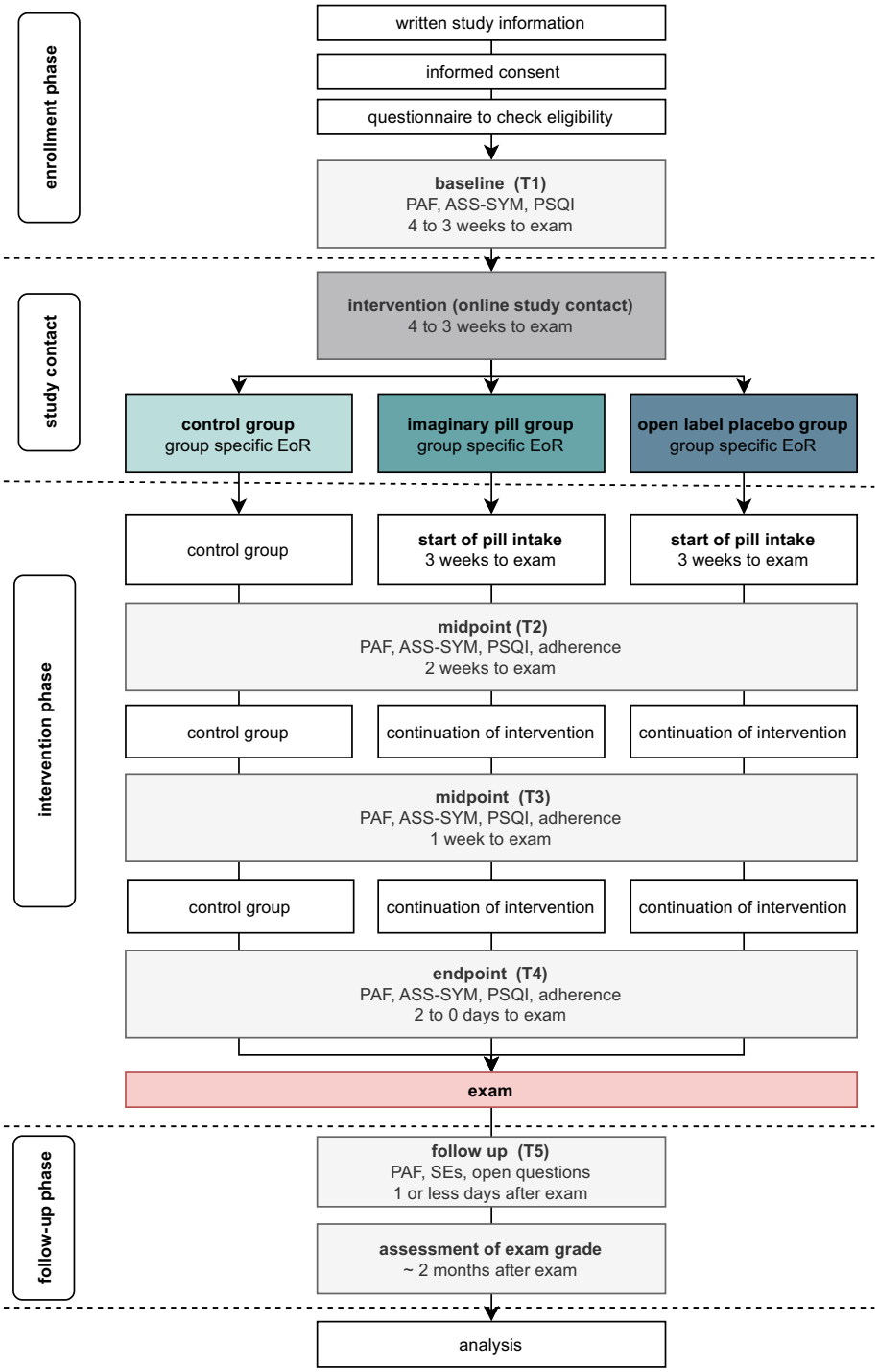


Figure 5. Procedure of the study. *Note: ASS-SYM* Änderungssensitive Symptomliste (general well-being), *EoR* expectancy of test anxiety relief, *PAF* Prüfungsangstfragebogen (test anxiety questionnaire), *PSQI* pittsburgh sleep quality index, *SEs* side-effects, *SDD* sociodemographic data.

Participants allocated to the IP group did not take a physical pill, but imagined taking a pill along with verbal suggestions from the treatment provider during the study contact. Hence, the idea of IPs has resemblance to the clinical application of hypnosis[42]. Participants in this group received a procedure in accordance with the technique by de Shazer[5] and a structure proposed by Niels Bagge[6]. Detailed formulation and translation to German was performed by the local study team (SB, DS, CL, JG). The instruction consisted of a procedure including five steps: (1) identifying the persons' problem and the desired state, (2) building trust in the treatment, (3) constructing a personally meaningful pill, (4) taking the IP, (5) suggestions for self-administration in real life and building adherence (see supplementary material). Importantly, step 2 consisted of teaching participants about findings of (open-label) placebo and imagination research. At the end of the intervention, participants in the IP

group had to describe their individual elaborated pill (size, shape, pill kind, color, packaging) and its effects in an interactive document. They sent the completed document back to the treatment provider and were able to print it out for their own use. Participants were asked not to take any physical aids, such as a candy, to facilitate their imagination, ensuring that the groups remained distinguishable in their specific ingredients. Participants were instructed to take two IPs a day for three weeks until the exam takes place and received daily e-mail reminders during that period to remember their IP intake.

In the OLP group participants obtained the information that they were receiving inert blue pills (i.e. "P-Dragees" blau Lichtenstein manufactured by Zentiva Pharma GmbH) and were given a treatment rationale in accordance with previous OLP studies (e.g.,[29,30]; see supplementary material), that encompassed four discussion points. In order to keep the OLP rationale similar to the one of the IP, a brief introduction was added at the beginning of the intervention, elaborating on what comprises the persons' problem (how do symptoms express themselves; how does the person wish to feel). Hence, the rationale was structured as follows: (1) identifying the OLP-sensitive problem, (2) deceptive as well as OLPs are efficacious, (3) one mechanism of placebo is conditioning, (4) an open attitude towards the treatment can be helpful but is not necessary for its effect, (5) taking the pill faithfully is important. Participants were instructed to take two placebo pills a day for three weeks until the exam takes place. The placebo pills were sent in an envelope to participants via postal mail prior to the online study contact or if participants did not wish to disclose their postal addresses, they were given the option of a personal handover by a member of the study team. Participants did not know about the content of the envelope and were instructed to not open the envelope until the study contact takes place. Daily e-mails were also sent to this group as a reminder to take the placebo pills.

In order to control for factors not considered characteristic for the intervention, the CG was fashioned according to the intervention groups (i.e., characteristic components were the pill intake and intervention-specific rationales[43]). Participants were (1) reminded of the importance of this group, (2) asked about the nature of their exam, (3) about their problem (i.e., test anxiety) and the wished-for state, (4) and about learning strategies. The design of this group attempted to keep interaction time comparable and to account for the structural equivalence between the CG and intervention groups, e.g., by allowing the CG to talk about the problem (i.e., test anxiety) to enable a "fair" comparison of groups[44]. Despite the interventional nature of this study arm, no advice or problem-solving task was given (see supplementary material).

Study contacts on zoom were carried out by five female treatment providers. Although not all treatment providers had the same number of study contact appointments, the proportion of participants per group were evenly distributed among them. Average duration of interventions was 31 minutes (IP = 44 min, OLP = 29 min, CG = 20 min).

Randomization and blinding. A random allocation sequence was created by SB using the built-in random number generator in Microsoft Excel for Mac, version 16.53. Participants were enrolled in the pre-generated list in order of their study registration and assigned by master students to interventions accordingly. All participants were informed about their assigned group at the study contact via zoom. Due to the study design, the providers were unblinded to the treatment they were administering. However, the encounter was kept constant in all groups through a standardized protocol. Also, except for the study contact on zoom, all communication was via e-mail contact (e.g., sending links for online assessments), using the same e-mail templates for all three groups to ensure the same type of interaction.

Outcome measures. The primary outcome was test anxiety measured by means of the "Prüfungs-Angst Fragebogen" (PAF; English: "test anxiety questionnaire"[31]). The questionnaire consists of 20 items with four subscales (worry, emotionality, interference, lack of confidence) with scores ranging from 20 to 80 points. Each item is rated on a 4-point Likert scale (*1—almost never* to *4—almost always*). Secondary outcomes were sleep quality and general well-being. Sleep quality was assessed by means of the "Pittsburgh Sleep Quality Index" (PSQI[45,46]). The PSQI is an 18 item self-rating questionnaire forming 7 subscales. To fit our time frame, we adjusted the time interval to the last week (7 days). To assess general well-being the ASS-SYM symptom list was used (Änderungssensitive Symptomliste[47]). This list is composed of 48 items and 6 subscales. Lower scores indicate less symptoms (i.e., higher general well-being). All measures were assessed four to three weeks prior to the exam (T1; baseline assessment), two and one week prior the exam (T2–T3; midpoint assessments) and two or less days prior the exam (T4; endpoint assessment).

Test performance of each participant was collected as another secondary outcome (approximately two months after the exam). Students received as a test performance either a continuous grade, ranging from a minimum of 1 (very poor) to a maximum of 6 (very good) in the Swiss grading system, or a binary test score (pass or fail), where a grade greater than or equal to 4 (sufficient) is considered a pass. Other outcomes of interests included sociodemographic data (SDD) assessed at T1. Immediately after the intervention, expectancy of test anxiety relief[48] according to the received intervention was assessed using an ad-hoc constructed questionnaire with each item of the primary outcome on a numeric rating scale from 1 to 4 (e.g., based on the intervention you have received, how strong would you expect the following symptoms to be present before your next exam on a scale from 1—*almost never* to 4—*almost always*) as e.g. used in[49]. Furthermore, within the intervention groups, adherence was assessed weekly with a single item asking for how often someone forgot the actual or imagined pill intake in the last week. In total, each participant assigned to one of the two intervention groups had to take 42 pills (i.e., 2 pills × 21 days). A sum score was computed to determine overall adherence. Adherence was defined as > 80% (i.e., 9 or more missed pills). Additional variables were collected on the same day or at most one day after the exam (T5; follow up assessment) including retrospective experience of the exam using the test anxiety questionnaire (i.e., the wording of the introduction was changed as follows: please read through each statement

and choose from the four answers 1—*almost never* to 4—*almost always* the one that indicates best how you were feeling during the exam), side-effects (i.e., (1) did you experience side effects, (2) if yes, give a description, (3) how severe were they from 0—*none* to 100—*very severe*, (4) when was the onset, (5) how long did the they last, (6) was there a connection with participation in our study?) and open questions respective to group allocation for example about the idea of intervention (i.e., what do you think about the idea of taking placebo/imaginary pills?; see supplementary material for all open-ended questions). All outcome variables were assessed by means of online surveys using Limesurvey (limesurvey.org).

Statistical analyses. Statistical analyses were carried out using the open-source software environment RStudio. For all analyses, significance level was set at $\alpha = 5\%$. Using a conservative power analysis on the basis of an *F*-Test and an ANCOVA for three groups, we calculated that a total sample size of $N = 206$ for a power of 0.9 and a total sample size of $N = 158$ for a power of 0.8 would be necessary to detect a medium effect size of $f = 0.25$ (i.e., $d = 0.5$) with an alpha-level of 0.05, using the statistical software G*Power. On this basis we decided on a total sample size of a minimum of 165 participants. Considering dropouts (e.g., due to increased nonattendance because of the COVID-19 pandemic), we planned to include and randomize slightly more than 55 per group ($N \sim 60$). Cohen's *d* was used to assess the size of effects.

Initially planned multiple imputation was not conducted, as there were less than 3% participants with missing data and the missingness appeared to be completely random (e.g., due to nonattendance at exams because of COVID-19). The five participants with missing data were thus not considered for analyses (see Fig. 1 for reasons for exclusion).

To assess differences in changes from baseline (T1) to endpoint (T4; primary analyses) and follow-up (T5) in test anxiety across the three groups, two separate omnibus tests (ANCOVA) using treatment group as the independent factor and baseline (T1) as covariate to control for baseline differences[50] were computed to test for overall effects. Orthogonal contrasts were computed to evaluate intergroup differences in the change from baseline (T1) to study endpoint (T4). The contrasts were: CG < IP + OLP and IP < OLP. To evaluate changes over time, we conducted a two-way mixed analysis of variance (ANOVA) using group as between-subject factor and time (T1-T4) as within-subject factor. Bonferroni adjustments accounted for multiple testing within post-hoc tests.

To analyze test performance across groups, we followed a two-step approach as there were continuous (grades) as well as binary (pass/fail) test scores. First, we performed an analysis only with participants having a continuous test score (1–6) using an ANOVA to test for overall effects and above-mentioned contrasts. Second, all continuous variables were transformed into a binary test score (pass ≥ 4; fail < 4) and reported as percentages of passing.

In order to investigate differences in treatment expectancy of relief across groups, an overall ANOVA was performed using the expectancy scores as outcome and group as between-subject factor. A priori contrasts were then used to explore differences across groups. Furthermore, we calculated correlations in order to investigate possible relationships between treatment expectancy of relief and test anxiety[51] and computed a linear model with test anxiety from T1 to T4 as dependent factor and expectancy of relief as independent factor to investigate their impact on the effects.

To investigate the influence of study contact duration and treatment provider on test anxiety we used a linear model with the corresponding variable as independent factor and changes in test anxiety from T1 to T4 as dependent variable. To analyze the open-ended questions about attitudes toward the idea about the two interventions, two independent raters rated each statement as "positive," "negative," or "neutral". When ratings differed, a consensus was reached by a third rater.

Data availability

Access to data from this study may be obtained by contacting the corresponding author.

Received: 2 November 2022; Accepted: 7 February 2023
Published online: 14 February 2023

References

1. Enck, P., Bingel, U., Schedlowski, M. & Rief, W. The placebo response in medicine: Minimize, maximize or personalize?. *Nat. Rev. Drug Discov.* **12**, 191–204 (2013).
2. Blease, C., Colloca, L. & Kaptchuk, T. J. Are open-label placebos ethical? Informed consent and ethical equivocations. *Bioethics* **30**, 407–414 (2016).
3. Charlesworth, J. E. G. *et al.* Effects of placebos without deception compared with no treatment: A systematic review and meta-analysis. *J. Evid. Based Med.* **10**, 97–107 (2017).
4. von Wernsdorff, M., Loef, M., Tuschen-Caffier, B. & Schmidt, S. Effects of open-label placebos in clinical trials: A systematic review and meta-analysis. *Sci. Rep.* **11**, 3855 (2021).
5. de Shazer, S. The imaginary pill technique. *J. Strateg. Syst. Ther.* **3**, 30–34 (1984).
6. Bagge, N. *The Imaginary Placebo Pill: Isolating Psychological Components of Placebo Pills for Treatment in Clinical Practice* [Conference presentation abstract]. Second official conference of the Society for Interdisciplinary Placebo Studies (SIPS). (Leiden, NL, 2019).
7. Wai-Lan Yeung, V., Geers, A. L. & Colloca, L. Merely possessing a placebo analgesic improves analgesia similar to using the placebo analgesic. *Ann. Behav. Med.* **54**, 637–652 (2020).
8. Gaab, J., Kossowsky, J., Ehlert, U. & Locher, C. Effects and components of placebos with a psychological treatment rationale—Three randomized-controlled studies. *Sci. Rep.* **9**, 1421 (2019).
9. Davis, A. J., Hettinga, F. & Beedie, C. You don't need to administer a placebo to elicit a placebo effect: Social factors trigger neurobiological pathways to enhance sports performance. *Eur. J. Sport Sci.* **20**, 302–312 (2020).
10. Chen, P.-H.A. *et al.* Socially transmitted placebo effects. *Nat. Hum. Behav.* **3**, 1295–1305 (2019).
11. Crum, A. J. & Langer, E. J. Mindset matters: Exercise and the placebo effect. *Psychol. Sci.* **18**, 165–171 (2007).

12. Peerdeman, K. J., van Laarhoven, A. I. M., Bartels, D. J. P., Peters, M. L. & Evers, A. W. M. Placebo-like analgesia via response imagery. *Eur. J. Pain* **21**, 1366–1377 (2017).
13. Moerman, D. E. & Jonas, W. B. Deconstructing the placebo effect and finding the meaning response. *Ann. Intern. Med.* **136**, 471–476 (2002).
14. Kirsch, I. Response expectancy theory and application: A decennial review. *Appl. Prev. Psychol.* **6**, 69–79 (1997).
15. Ogino, Y. *et al.* Inner experience of pain: Imagination of pain while viewing images showing painful events forms subjective pain representation in human brain. *Cereb. Cortex* **17**, 1139–1146 (2007).
16. Thompson, J. J., Ritenbaugh, C. & Nichter, M. Reconsidering the placebo response from a broad anthropological perspective. *Cult. Med. Psychiatry* **33**, 112–152 (2009).
17. Bischofsberger, L. *Prävalenz und Ausprägungen von Prüfungsangst bei Studienanfänger/-Innen der Medizin*. Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) (2022).
18. Ergene, T. Effective interventions on test anxiety reduction: A meta-analysis. *Sch. Psychol. Int.* **24**, 313–328 (2003).
19. von der Embse, N., Jester, D., Roy, D. & Post, J. Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review. *J. Affect. Disord.* **227**, 483–493 (2018).
20. Schaefer, M. *et al.* Open-label placebos reduce test anxiety and improve self-management skills: A randomized-controlled trial. *Sci. Rep.* **9**, 13317 (2019).
21. Kleine-Borgmann, J. *et al.* Effects of open-label placebos on test performance and psychological well-being in healthy medical students: A randomized controlled trial. *Sci. Rep.* **11**, 2130 (2021).
22. Sugarman, M. A., Loree, A. M., Baltes, B. B., Grekin, E. R. & Kirsch, I. The efficacy of paroxetine and placebo in treating anxiety and depression: A meta-analysis of change on the Hamilton rating scales. *PLoS ONE* **9**, e106337 (2014).
23. Sugarman, M. A., Kirsch, I. & Huppert, J. D. Obsessive-compulsive disorder has a reduced placebo (and antidepressant) response compared to other anxiety disorders: A meta-analysis. *J. Affect. Disord.* **218**, 217–226 (2017).
24. Faria, V. *et al.* Do you believe it? verbal suggestions influence the clinical and neural effects of escitalopram in social anxiety disorder: A randomized trial. *EBioMedicine* **24**, 179–188 (2017).
25. Oosterbaan, D. B., van Balkom, A. J. L. M., Spinhoven, P. & van Dyck, R. The placebo response in social phobia. *J. Psychopharmacol.* **15**, 199–203 (2001).
26. Cuijpers, P. *et al.* Psychological treatment of generalized anxiety disorder: A meta-analysis. *Clin. Psychol. Rev.* **34**, 130–140 (2014).
27. Ahmadzad-Asl, M. *et al.* Systematic review and meta-analysis of the placebo effect in panic disorder: Implications for research and clinical practice. *Aust. N. Z. J. Psychiatry.* **56**, 1130–1141 (2022).
28. Kim, S., Wollburg, E. & Roth, W. T. Opposing breathing therapies for panic disorder: A randomized controlled trial of lowering vs raising end-tidal PCO2. *J. Clin. Psychiat.* **73**, 13479 (2012).
29. Kaptchuk, T. J. *et al.* Placebos without deception: A randomized controlled trial in irritable bowel syndrome. *PLoS ONE* **5**, e15591 (2010).
30. Carvalho, C. *et al.* Open-label placebo treatment in chronic low back pain: A randomized controlled trial. *Pain* **157**, 2766–2772 (2016).
31. Hodapp, V., Rohrmann, S. & Ringeisen, T. *PAF–Prüfungsangstfragebogen [the brief German test anxiety inventory]* (Hogrefe, 2011).
32. Huntley, C. D. *et al.* The efficacy of interventions for test-anxious university students: A meta-analysis of randomized controlled trials. *J. Anxiety Disord.* **63**, 36–50 (2019).
33. Kong, J. *et al.* Enhancing treatment of osteoarthritis knee pain by boosting expectancy: A functional neuroimaging study. *NeuroImage Clin.* **18**, 325–334 (2018).
34. Kosslyn, S. M., Ganis, G. & Thompson, W. L. Neural foundations of imagery. *Nat. Rev. Neurosci.* **2**, 635–642 (2001).
35. Mcnorgan, C. A meta-analytic review of multisensory imagery identifies the neural correlates of modality-specific and modality-general imagery. *Front. Hum. Neurosci.* **6**, 285 (2012).
36. Kube, T., Hofmann, V. E., Glombiewski, J. A. & Kirsch, I. Providing open-label placebos remotely—A randomized controlled trial in allergic rhinitis. *PLoS ONE* **16**, e0248367 (2021).
37. Schaefer, M., Harke, R. & Denke, C. Open-label placebos improve symptoms in allergic rhinitis: A randomized controlled trial. *Psychother. Psychosom.* **85**, 373–374 (2016).
38. 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* **13**, e0192758 (2018).
39. Lotz, C. & Sparfeldt, J. R. Does test anxiety increase as the exam draws near?—Students' state test anxiety recorded over the course of one semester. *Pers. Individ. Differ.* **104**, 397–400 (2017).
40. Evers, A. W. *et al.* Implications of placebo and nocebo effects for clinical practice: Expert consensus. *Psychother. Psychosom.* **87**, 204–210 (2018).
41. Evers, A. W. *et al.* What should clinicians tell patients about placebo and nocebo effects? Practical considerations based on expert consensus. *Psychother. Psychosom.* **90**, 49–56 (2021).
42. Lynn, S. J., Laurence, J.-R. & Kirsch, I. Hypnosis, suggestion, and suggestibility: An integrative model. *Am. J. Clin. Hypn.* **57**, 314–329 (2015).
43. Grünbaum, A. The placebo concept. *Behav. Res. Ther.* **19**, 157–167 (1981).
44. Baskin, T. W., Tierney, S. C., Minami, T. & Wampold, B. E. Establishing specificity in psychotherapy: A meta-analysis of structural equivalence of placebo controls. *J. Consult. Clin. Psychol.* **71**, 973 (2003).
45. Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R. & Kupfer, D. J. The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Res.* **28**, 193–213 (1989).
46. Riemann, D. & Backhaus, J. *Behandlung von Schlafstörungen. Ein psychologisches Gruppenprogramm (Materialien für die psychosoziale Praxis). Beltz Psychologie Verlags Union*. (Weinheim, 1996).
47. Krampen, G. *Änderungssensitive Symptomliste zu Entspannungserleben, Wohlbefinden, Beschwerden-und Problembelastungen: ASS-SYM*. (Hogrefe, 2006).
48. Papakostas, G. I. & Fava, M. Does the probability of receiving placebo influence clinical trial outcome? A meta-regression of double-blind, randomized clinical trials in MDD. *Eur. Neuropsychopharmacol.* **19**, 34–40 (2009).
49. Locher, C. *et al.* Is the rationale more important than deception? A randomized controlled trial of open-label placebo analgesia. *Pain* **158**, 2320–2328 (2017).
50. Vickers, A. J. & Altman, D. G. Analysing controlled trials with baseline and follow up measurements. *BMJ* **323**, 1123–1124 (2001).
51. Pearson, J., Naselaris, T., Holmes, E. A. & Kosslyn, S. M. Mental imagery: Functional mechanisms and clinical applications. *Trends Cogn. Sci.* **19**, 590–602 (2015).

Acknowledgements

The authors would like to thank Berfin Bakis for her assistance with the data monitoring, assistance in conducting the study and support in writing the manuscript. We also thank all Masterstudents: Annina Dettwiler, Alina Burkart, Yannick Schönenberger, Marlon Enzmann, Alex Busch, and interns Clara Bruhin, Jana Hochreuter, for aiding with conducting the study and analyzing the data.

Author contributions

Conceptualization: S.B., D.S., N.B., I.K., C.L., C.C., J.G.; Methodology: S.B., D.S., N.B., I.K., C.L., C.C., J.G.; Formal Analyses: S.B., D.S., I.K., C.L.; Investigation: S.B., D.S.; Writing—original draft: S.B., D.S.; Writing—review & editing: S.B., D.S., N.B., I.K., C.L., C.C., J.G.; Visualization: S.B.; Supervision: N.B., I.K., C.L., C.C., J.G.; Project Administration: S.B.

Funding

This work was not supported by any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. CL received funding from Swiss National Science Foundation (SNSF): PZ00P1_201972.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-29624-7.

Correspondence and requests for materials should be addressed to S.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Materials for

Imaginary pills and open-label placebos can reduce test anxiety by means of placebo mechanisms

Sarah Buergler, Dilan Sezer, Niels Bagge, Irving Kirsch, Cosima Locher, Claudia Carvalho, & Jens Gaab

This file includes:

Supplementary Text

Ingredients of placebo pills

Lactose-Monohydrat, Magnesiumstearat, mikrokristalline Cellulose, hochdisperses Siliciumdioxid, weisser Ton, Macrogolglycerolhydroxystearat, Arabisches Gummi, Montanglycolwachs, Povidon, Talkum, Titandioxid, Patentblau-V-Aluminiumsalz, Calciumcarbonat, Sucrose, Glukosesirup, Maisstärke, Macrogol 6000

Rationales

Imaginary Pill (IP) rationale (translated from German in English)	Open-label Placebo (OLP) rationale (translated from German in English)	Control Group (CG) rationale (translated from German in English)
1. Identifying the IP-sensitive problem and the desired state: Before I explain the concept of the open administration of placebos and how we use these placebo effects with the imaginary pill intervention, I would like to know more about your test anxiety and your preparation stress. Could you use a previous exam situation to describe what the symptoms feel like? I ask you now to think back to that bad exam/situation. Can you now describe to me, based on this previous exam situation, what the symptoms of your exam anxiety felt like? What are the sensations in your body? What are your thoughts and emotions when you experience this anxiety? Now, thinking of the upcoming exam, if you had to specify how strong these symptoms are from 0-10 (not at all – very strong) in this moment, what would you say? And now can you describe how you would like to feel in your exam phase/during the exam? Can you describe to me a specific situation	**1. Identifying OLP-sensitive problem** Before I explain the concept of the open administration of placebos and how we use these placebo effects with the imaginary pill intervention, I would like to know more about your test anxiety and your preparation stress. Could you use a previous exam situation to describe what the symptoms feel like? I ask you now to think back to that bad exam/situation. Can you now describe to me, based on this previous exam situation, what the symptoms of your exam anxiety felt like? What are the sensations in your body? What are your thoughts and emotions when you experience this anxiety? Now, thinking of the upcoming exam, if you had to specify how strong these symptoms are from 0-10 (not at all – very strong) in this moment, what would you say? And now can you describe how you would like to feel in your exam phase/during the exam? Can you describe to me a specific situation where you feel this? What are your feelings in	**1. Explaining importance of group** As you already know from the study information, we randomly assign all study participants to one of the three study groups. You have been assigned to the control group, which means you will not receive a treatment. This group and your participation is very important for our study. Only through the control group can we see how symptoms naturally behave when you do not receive a treatment. So, we are also asking you to fill out all the online surveys accordingly. You will still receive weekly surveys.

where you feel this? What are your feelings in this situation? What are your thoughts? What physical sensations do you have in this situation? Try to put yourself in this positive state. [Check to see if the person really knows how they want to feel. Get as precise as possible].	this situation? What are your thoughts? What physical sensations do you have in this situation? Ok, what you say is all very understandable and I hope we can help you with our intervention to reach the positive state you just described. Is it okay if I now explain to you the concept of open placebos?	
2. Building trust in the treatment Is it ok if I now explain the concept of open-label placebos and how you can use this placebo effect with the imaginary pill for your goal – the positive experience, we've just talked about? OK, we know from clinical research that placebos have significant effects on pain, depression and anxiety and that these effects can even be demonstrated in changes in brain activity and the release of neurotransmitters. As mentioned earlier, scientists previously assumed that placebo pills can only help if they are given covertly, i.e. with deception. Now, however, more recent studies suggest that this is not the case. This means that placebos can work even if the patient knows that it is a placebo. We are incorporating this approach in our study. Many double-blind randomized studies show that the placebo effect is very effective for many complaints. This means that placebos can relieve pain, cramps and gastrointestinal complaints, among other problems, and also have a very positive effect on mood. Especially for chronic back pain and irritable	**2. Deceptive and OLPs are effective** We know from clinical research that placebos have significant effects on pain, depression and anxiety and that these effects can even be demonstrated in changes in brain activity and the release of neurotransmitters. As mentioned earlier, scientists previously assumed that placebo pills can only help if they are given covertly, i.e. with deception. Now, however, more recent studies suggest that this is not the case. This means that placebos can work even if the patient knows that it is a placebo. We are incorporating this approach in our study. Many double-blind randomized studies show that the placebo effect is very effective for many complaints. This means that placebos can relieve pain, cramps and gastrointestinal complaints, among other problems, and also have a very positive effect on mood. Especially for chronic back pain and irritable bowel syndrome, the open-label placebo treatment has been shown to be very effective, even in patients where nothing else has worked. Here at the division, a study has already been carried out in which a placebo	**2. Nature of exam** In your case, we would be interested in how the exam stress and anxiety manifests itself and what your general learning strategies are. I will possibly make some notes on this. Before we get to your exam anxiety itself, I'd like to ask you questions about the nature of the exam: - What format does the exam take? Is it written or oral? - Are you generally more afraid of written/oral (repeat what was said) exams, compared to exams that have a different format? - Is it a repetition exam? - *If not yet clear:* Does the test anxiety also have to do with the subject in which the exam takes place? In which subject is the exam? What does this subject involve? - What makes the exam so difficult or scary for you? - How often are you afraid of an exam to this extent or are you stressed because of the exam (in every learning phase or

bowel syndrome, the open-label placebo treatment has been shown to be very effective, even in patients where nothing else has worked. Here at the division, a study has already been carried out in which a placebo cream was used for the treatment of heat-induced pain. And there too we found large placebo effects. A positive placebo effect has also been shown for test anxiety. This has been shown recently by a study from Germany, where they tested open placebos also in students. We are now considering the possibility that if placebos work, even though we know that they are placebos, then we could simply omit the sugar pill and imagine the pill and still have all the placebo effects. A reaction to placebos is not only triggered by the placebo pill itself, but also by the imaginative meaning that is both consciously and automatically attributed to the placebo pill. Imagination research shows, for example, that the idea of something activates the same areas of the brain as when one actually sees or experiences something. A study has also shown that the idea of exercising in a gym can already lead to muscle growth. Accordingly, it is possible to imagine taking this pill and achieve a similar effect as if you were taking a real pill. And this is exactly what I would like to discuss and practice with you. Okay for you?	cream was used for the treatment of heat-induced pain. And there too we found large placebo effects. A positive placebo effect has also been shown for test anxiety. This has been shown recently by a study from Germany, where they tested open placebos also in students.	especially now)? (Possibly why especially now?)

3. Constructing a personally meaningful pill The first step is to find an imaginary pill for you. Recall the positive state, that you described earlier and the experience of relief you would like to feel. Suppose there was a pill that could bring you to that state, what effects would that pill have, how would it help you reach that state? Imagine there was a pill that could have all these positive effects. What would this pill look like (regarding color, shape and size)? And is the pill packaged also? [Wait and trust, that the person will come up with a pill. If a picture of such a pill is cannot be formed, then offer pill characteristics to choose from, for example: "The pill could be round, oval, (...)" etc.]	**3. One mechanism of placebos: Conditioning** Next, I would like to explain in more detail why placebos can alleviate symptoms. A very important explanation is that the body automatically reacts to the intake of medication. From an early age we learn that pills and effects are related, it results in a learning effect, so to speak. Accordingly, swallowing the pill alone can lead to symptom relief. The physiological reaction of our body to placebos is comparable to this. We know that when placebos work, they release neurotransmitters such as endorphins and dopamine, automatically activating specific areas of the brain. These neurotransmitters, in turn, can relieve symptoms or have a positive effect on mood.	**3. Talking about the problem (test anxiety) and the wished-for state** - Can you tell me specifically about a bad exam (it can also be a lecture or something similar) that you have had in the past and where you were very afraid? [Ask person to actually name an exam, the more specific the better]. I ask you now to think back to that bad exam/situation [wait until person remembers]. Now, using that previous exam situation, can you describe to me what the symptoms of your exam anxiety felt like? What were the sensations in your body based on your experience? What are thoughts and emotions that went through your mind? - Could you use a previous exam situation to describe what the symptoms feel like? I ask you now to think back to that bad exam/situation. Can you now describe to me, based on this previous exam situation, what the symptoms of your exam anxiety felt like? What are the sensations in your body? What are your thoughts and emotions when you experience this anxiety? Now, thinking of the upcoming exam, if you had to specify how strong these symptoms are from 0-10 (not at all – very strong) in this moment, what would you say? - And now can you describe how you would like to feel in your exam phase/during the exam?

4. Taking the IP Now imagine the pill described in detail as if it were a real pill. You can ascribe so much reality to the pill that taking it is experienced as if you were swallowing a real pill. It may take some practice. The effect may be stronger and the procedure easier for you if you have done it several times. Now I would suggest that you take your imaginary pill, to try this. You can close your eyes, if you want to. Just think of it as a regular pill. Imagine the pill and how it is packaged. Imagine how you take the pill out of the packaging and how you hold it in your hand. Bring it to your mouth. Swallow the pill slowly. Now it is in your body and starts to work. Maybe you can already feel the effects of the pill. Try to feel what the pill does to you. Maybe the pill has also other effects, such as making your mouth dry. You might get warm or a little dizzy. You have now had your first experience of such an imaginary pill taking. Try to remember this state so that you can recall it on your own. Now, if you had to indicate again after taking your imaginary pill how strong at the moment your symptoms are from 0-10 when you think about the upcoming exam, what would you say?	**4. An open attitude towards the treatment can be helpful but is not necessary** It's also absolutely okay if you have doubts that placebos work. As mentioned before, placebos can work automatically, which means they can work even if you have doubts.	**4. Learning strategies** - Now I'm still wondering what your general learning strategies are: Do you work in study groups or more alone or both? - Do you study with summaries, mind maps, study plans or flashcards? Thank you very much for your answers to the many questions, it is very informative.

5. Suggestions for self-administering in real life and building adherence For the effect of this intervention it is now important that you take such an imaginary pill twice a day from the time when you receive a reminder per e-mail: once in the morning and once in the evening, in order to reach the desired state (up to the exam). Before taking the pill, take a little time to recall the image of the pill you have just described. I also ask you to fill out the announced surveys once a week until the exam. You will receive an e-mail with the link at the right time, so that you remember to do it. Then one more thing: In order for us to really be able to identify what the effects of an imaginary pill are, we ask you not to take sweets like Sugus or Tiktak to make it easier for you to imagine. As said before, it is best to simply take your time and take the imaginary pill twice a day for three weeks. We will also send you daily reminders that you are reminded to take the imaginary pill. I am aware that the concept of the imaginary pill may sound strange to you at first. But we would like to find out whether you can reach the desired state if you imagine taking a pill every day to relieve your test anxiety. We have developed this procedure here at the university in collaboration with experts from all over the world and we really believe in the effectiveness of this treatment. Also, because there are already several cases from the clinic where the imaginary pill treatment has shown very good effectiveness. Therefore, I would	**5. Taking the pill faithfully is important** Therefore, it is important that you take the placebos regularly and according to the prescription. This means for you that you have to take the placebo pills faithfully in order to feel an effect. It is important for you to know that for some people the effects occur earlier and for others later. When you take the pills, we recommend that you also be aware of what the pills are supposed to help you against, i.e. to achieve the positive state you described earlier. I am aware that this may sound unfamiliar to you at first. However, we want to find out what happens to your symptoms when you take placebo pills every day. Therefore, I would like to encourage you to give the open placebo treatment a chance and see what happens. Now you may open the parcel and take out the box with the placebo pills in it. Please take two pills every day for the next three weeks (until the exam) from the time you receive a reminder by e-mail. It is best to take the pills at the same time in the morning and in the evening. There is also a small envelope in the package, which you can open right away. On the envelope you will find information on how to take the pills. As we said, we will send you daily reminders to remember to take your pills. If you take two pills a day for three weeks, that's a total of 42 pills. There are 50 pills in the package, which means there are 8 pills too many. You don't have to send them back to us (you can take them at a later date,	

like to encourage you to give the imaginary pill a chance and see what happens.	for example). During the three weeks please take always two pills per day and not more or less. I also ask you to fill out the announced surveys once a week until the exam. You will receive an e-mail with the link at the right time to make sure you remember to do this.	

Open-ended questions

Open-ended questions in open-label placebo group

1. What do you think about the idea of taking placebo pills? (open-ended question)
2. Do you find the placebo pill has generally helped you to be less anxious/stressed before the exam? Yes/No
3. For which symptoms did the placebo pill help to which extent 0% (the pill did not help at all) - 100% (the pill helped 100%)
   - Concerning excitement (emotional and physical tension)
   - Concern (thoughts about failure, self-doubt)
   - Regarding distraction (distraction from the task by irrelevant thoughts)
   - Regarding confidence (self-worth)
4. Did you assume that the placebo pills would work or were you skeptical? (open-ended question)
5. What do you think was in the placebo pills ? (open-ended question)
6. What did you learn by participating in this treatment study ? (open-ended question)
7. Do you have any other comments ? (open-ended question)

Open-ended questions in imaginary pill group

1. In general, how open are you to taking a pharmacological pill for your test anxiety ? 0% (not at all open) to 100% (very open) (slider).
2. Do you find the imaginary pill helped you to be less anxious/stressed before the exam ? Yes/No
3. For which symptoms did the imaginary pill help to what extent 0% (the pill did not help at all) - 100% (the pill helped 100%)
   - Regarding excitement (emotional and physical tension)
   - Regarding concerns (thoughts about failure, self-doubt)
   - Regarding distraction (distraction from the task by irrelevant thoughts)
   - Regarding confidence (self-worth)
4. How difficult was it for you to imagine the imaginary pill? 1 (very easy) - 7 (very difficult)
5. How well could you imagine the following aspects of the imaginary pill (0 - not at all well to 100 almost identical to a real pill):
   - Seeing the pill (visualization)
   - Tasting the pill
   - Feeling the pill
   - Effects of the pill
6. Did you find it easier to visualize and take the imaginary pill during the study?
   - Yes it was easier
   - It did not change
   - No it became more difficult
8. What do you think about the idea of taking an imaginary pill ? (open-ended question)
9. Did you assume that the imaginary pill would work or were you skeptical ? (open-ended question)
10. Did you learn anything from participating in this treatment study? If yes, what? (open-ended question)
11. Do you have any other comment? (open-ended question)

Open-ended questions in control group

1. Were you disappointed that you were in the control group? Yes/No
2. Is there anything else you would like to comment on? (open-ended question)

Supplementary Table

Table S1

Mean values for subscales of the test anxiety questionnaire for all assessed timepoints.

		T1	T2	T3	T4	T5
	Group (n)	*M (SD)*				
worry	IP (55)	13.27 (3.36)	12.36 (3.65)	12.18 (3.58)	12.36 (3.99)	10.42 (10.18)
	OLP (59)	14.42 (2.96)	13.85 (3.00)	13.00 (3.44)	13.66 (3.47)	10.86 (10.60)
	CG (59)	13.95 (3.50)	14.05 (2.75)	13.89 (3.48)	14.14 (3.65)	11.49 (9.97)
emotionality	IP (55)	11.00 (3.25)	9.69 (2.81)	9.24 (2.84)	9.80 (3.13)	9.58 (3.47)
	OLP (59)	12.08 (3.52)	10.97 (2.78)	10.07 (2.83)	10.24 (3.15)	10.14 (3.33)
	CG (59)	11.25 (2.89)	11.83 (3.39)	11.75 (2.88)	12.10 (3.84)	11.93 (3.99)
interference	IP (55)	10.42 (3.26)	10.29 (3.11)	9.36 (3.25)	9.38 (3.31)	6.85 (2.38)
	OLP (59)	11.34 (3.14)	10.66 (2.82)	9.95 (3.13)	10.20 (3.14)	7.20 (2.72)
	CG (59)	11.31 (3.14)	11.98 (3.17)	11.59 (3.40)	11.69 (3.27)	8.34 (2.91)
lack of confidence	IP (55)	14.16 (2.94)	12.91 (2.71)	13.05 (2.65)	13.15 (3.05)	13.00 (3.49)
	OLP (59)	14.51 (2.47)	13.83 (2.64)	13.68 (2.68)	13.47 (2.93)	13.44 (3.53)
	CG (59)	14.15 (2.32)	14.22 (2.09)	14.24 (2.74)	14.17 (2.83)	14.02 (2.96)

Note. ASS-SYM, Änderungssensitive Symptomliste (general well-being); CG, control group; IP, imaginary pill; M, mean; OLP, open-label placebo; PAF, Prüfungsangstfragebogen (test anxiety questionnaire); PSQI, pittsburgh sleep quality index; SD, standard deviation.

Appendix C: Study III

Buergler, S.*, Sezer, D.*, Gaab, J., & Locher, C. (in Review). The role of population, expectation, modality, and comparator on open-label placebo effects: A network meta-analysis.

** shared first authorship*

The role of population, expectation, modality and comparator on open-label placebo effects: A network meta-analysis

Sarah Buergler[1*°], Dilan Sezer[1*], Jens Gaab[1], & Cosima Locher[2, 3]

[1] Division of Clinical Psychology and Psychotherapy, Faculty of Psychology, University of Basel, Switzerland

[2] Department of Consultation-Liaison Psychiatry and Psychosomatic Medicine, University Hospital Zurich, University of Zurich, Switzerland

[3] Faculty of Health, University of Plymouth, Plymouth, UK

*Shared first authorship

°Corresponding author: Sarah Buergler, MSc, Division of Clinical Psychology and Psychotherapy, Faculty of Psychology, University of Basel, Switzerland. Emails: sarah.buergler@unibas.ch

Keywords: **open-label placebo, network meta-analysis**

Short title: **Open label-placebos – network meta-analysis**

Abstract

Three meta-analyses demonstrate the clinical potential of open-label placebos (OLPs). However, there is a need to synthesize the existing evidence through more complex analyses that allow to answer questions beyond mere efficacy. This serves to better understand why and under what circumstances OLPs work (e.g., in what populations or through which routes of administration). To answer these questions, we conducted the first network meta-analysis in the field of OLPs. Our analyses revealed that OLPs can be beneficial in comparison to NT in nonclinical (12 trials; 1'015 participants) and clinical populations (25 trials; 2'006 participants). The kind of modality had no substantial impact on OLP effects. However, positive treatment expectations were found to be important in order for OLPs to work. Further, OLP effects can vary depending on the comparator used. Thus, the population, modality, expectation and comparator should be considered when designing and interpreting OLP studies.

Introduction

Placebos have been found to have clinically significant effects in a variety of clinical conditions[1,2], but their use in clinical practice is denied as it violates ethical obligations. In this regard, open-label placebos (OLP) administered under full disclosure and transparency can be considered both ethical and feasible[3]. Several studies show medium sized to large clinically relevant effects of OLPs[4–6] that can be comparable in magnitude to deceptively administered placebos (DP)[7–12]. However, given that this field of research is still in its infancy with the first controlled study published in 2008[13], there are still many questions that need to be addressed.

In OLP, no meta-analysis has so far explored the differential effects across distinct populations. Whereas in clinical conditions OLPs were significantly more efficacious compared to NT with moderate to large effects (SMD = 0.88[4] and 0.72[5]), in nonclinical experimental conditions a moderate effect was found in OLPs for self-reported outcomes (SMD = 0.43) and no significant effect was observed for objective outcomes (SMD = -0.02[6]). Thus, it appears that clinical populations may benefit more from OLP treatments than nonclinical populations, a finding known in deceptive placebos where placebo analgesia tends to be higher in patients compared to healthy subjects[14,15].

OLP effects may not only vary across populations but also across treatment modalities. For example, more invasive placebo procedures, such as injections and sham procedures, have been shown to increase expectations towards a treatment's efficacy – and in turn enhance placebo effects[16–18]. However, while placebo effects in itch seem not to differ between oral and injective placebo administration[19], in osteoarthritis intra-articular and topical placebo were more efficacious than orally administered placebo[20]. It is argued that more invasive administrations of placebos have stronger effects than less invasive administration (oral or nasal) in the case of pain, whereas in nonpain conditions such as itch, this might not be the case[21]. Nonetheless, placebo experts strongly agree that clinicians should not prescribe more invasive treatments merely to obtain stronger placebo effects, due to practical and ethical restrictions, higher costs, and higher risk of undesirable side effects[1]. This is especially true for OLPs as it is unclear to date whether the findings on deceptive placebos that more invasive treatments are more beneficial can be applied to the field of OLPs.

Not only the route of administration, but also associative learning (i.e., conditioning) and verbal suggestions that accompany a treatment play a key role in the expected and actual placebo effects[22]. In the majority of OLP studies the administration of the placebo is accompanied with a rationale consisting of four discussion points in order to induce positive treatment expectations (see e.g.,[23]). So far, the impact of positive expectation on OLP effects has been explored in studies comparing OLPs with expectation induction (i.e., through verbal suggestions or conditioning) to OLPs without such expectation-inducing procedures (hereafter, OLP-[8,24–27]). Some authors have concluded that the treatment rationale is crucial when it comes to the efficacy of OLP (e.g.,[8]), however, systematic investigations are limited to the experimental context[6].

Effect sizes may also depend on different control conditions used in trials. For example, it has been found that waitlist (WL) control groups lead to larger effects than no treatment (NT) controls[28]. This result could be due to the fact that subjects who are assigned to a WL group are not actively looking for improvement opportunities during the waiting phase, as might be the case with the NT group. Blease and colleagues (2019)[29] compared this phenomenon with the induction of nocebo effects in the context of OLPs, especially in the case when the experimenter mentions the potential advantages of the OLP intervention before the assignment to the WL. Further, the use of treatment as usual (TAU) controls can be considered problematic as the "treatment as usual" is typically not monitored or sufficiently reported, which may lead to structural inequivalence across studies which apply TAU[30,31]. Thus, it is warranted to take a closer look at the different comparators that are used across OLP studies.

As illustrated above, currently existing meta-analyses on OLP effects did not address several important aspects: (1) In each of the three analyses only one type of population (clinical or nonclinical) was considered. The comparability of effect sizes across different individual meta-analyses, however, might be limited, as these studies used different definitions for eligible OLP interventions and for conditions that qualify for the nonclinical and clinical population. This especially holds true for the question, whether subclinical conditions (e.g., menopausal hot-flushes, self-reported test-anxiety or general well-being) are to be considered nonclinical or clinical. Therefore, there is a need for a clear definition of these samples and for meta-analytic analyses that apply the same inclusion criteria in both areas. (2) These meta-analyses, as well as other meta-analyses of different placebo administration modes, cannot comprehensively examine different routes of administration, in part because interpretation of results from multiple meta-analyses is compromised by indirect comparison via subgroup analyses[21]. Hence, it remains unclear whether different placebo administrations result in different effects to justify the choice of one route of administration over another. (3) Furthermore, no review study has to date systematically, and on the basis of a relatively large database, examined whether the effects of OLPs with positive expectations either through a rationale or other expectation-inducing measures (e.g., conditioning) differ from those without expectation induction. (4) Finally, the current OLP meta-analyses lumped all control groups into one arm and thus did not differentiate between the different control conditions. However, it is of great importance to investigate OLP efficacy in comparison to different kinds of control groups, thereby ensuring that OLP effect sizes are neither over- nor underestimated.

To examine these open questions, a network meta-analysis (NMA) is the method of choice. NMAs allow the comparison of multiple treatment and comparator groups. Further, an NMA produces more accurate effect sizes than a traditional meta-analysis by including both direct and indirect evidence. To the best of our knowledge, this is the first NMA on OLP treatments. On the basis of the above discussed challenges and open questions in OLP research, we derived the following research questions (RQ) that can be answered in a network meta-analytic framework: Is the magnitude of the OLP effects different across (RQ1) clinical vs. nonclinical populations, (RQ2) OLP treatment modalities, (RQ3) treatment expectation, and (RQ4) comparator groups.

Methods

Search Strategy

A systematic review and NMA was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Study population analysis (PRISMA) statement[32,33] (eAppendix 5 in the supplement). The search strategies were conducted in Medline, Embase, and PsycINFO via Ovid, the Cumulative Index of Nursing and Allied Health Literature (CINAHL), clinicaltrials.gov, Open-Trials, and Cochrane Register of Controlled Trials were developed in close collaboration with an information specialist. The four databases and the three registries were searched using text word synonyms and database-specific subject headings for open-label placebos in February 2nd, 2021 (eAppendix 1 in the supplement) and updated on June 8th, 2022 (eAppendix 2 in the supplement). No language restrictions were applied. For Medline and Embase, randomized controlled trial (RCT) filters were applied, and conference abstracts and conference reviews were excluded from Embase. References were exported to Endnote X9 and deduplicated using the Bramer method. Furthermore, additional trials were identified from an existing systematic review on OLPs[6] and a newsletter on placebo studies (https://jips.online/). If data was not available, the corresponding authors of the respective publication were contacted via email. Several reviewers, in pairs of two, independently screened the references based on their titles and abstracts using https://covidence.org. Selected references were retrieved in full-text and independently assessed for eligibility by two reviewers. Any disagreements over eligibility were resolved by consensus or, if necessary, by consultation of a third reviewer. This study was registered with Prospero (CRD42020161696).

Study Selection

We included RCTs comparing OLPs compared to a control group in clinical (e.g., chronic low-back pain, depression, irritable bowel syndrome, allergic rhinitis), subclinical (e.g., menopausal hot flushes or test anxiety), as well as nonclinical (e.g., experimental induced pain or allergic reactions) populations. There were no age restrictions. Our definition of OLP was as follows: (1) The placebo must have been given openly, i.e., the receiver was 100% aware of getting the placebo when applied. Studies needed to state explicitly that the placebo was delivered with the full awareness of the receiver, i.e., solely using the term "open-label" as description of the study was not enough, as this term was used inconsistently sometimes referring to treatment provider being unblind. Also, balanced placebo design studies (with e.g., a 50% chance of receiving a placebo) were excluded. (2) The placebo had to consist of a “pharmacological” property, i.e., was defined as everything that can be swallowed (e.g., pills, capsules, sirups, etc.), applied on the skin or other body parts (such as a cream or eye drops) or injected. Studies testing devices (e.g., deep brain stimulation) as well as placebo exercises, and diets were excluded. Also, studies testing procedures such as placebo massage or acupuncture without including an additional treatment arm fulfilling our placebo definition were not eligible. (3) At least minimal positive expectation needed to be induced alongside the placebo administration (e.g., either through a rationale (i.e., positive suggestions) or conditioning). (4) The placebo needed to be applied with the intention of a positive effect (i.e., therapeutic or well-being enhancing, no nocebo effects). Based on these criteria,

none of the open-label drug trials using OLP as a comparator, which we aimed to also include in these analyses, met our definition.

Crossover studies were only included if we were able to extract the results of the first period of the trial (i.e., before the first cross) separately. This is because data from crossover studies should not be treated as if data stems from parallel-trials[34]. If this data was not reported, authors were contacted. In case of no response, these studies were excluded from the analysis. In order to be included, studies needed to report a baseline and a post measure or alternatively report change scores from baseline to post. Studies reporting only post values or where we were not able to retrieve means and standard deviations (SDs) were excluded. For studies published more than once (i.e., secondary analysis), we included only the entry with the most relevant data to our analysis.

Data Extraction

All relevant data were extracted independently in pairs of two using a standardized excel template. Disagreements were clarified through consensus and by consultation with a third reviewer, if required. Means and standard deviations (SDs) were extracted and in case SDs were not reported, we calculated them from standard errors (SE), confidence intervals (CIs), or interquartile ranges (IQR) and medians were converted to means[34]. If the sample size used for the analysis was not reported, we used the sample size of the baseline data (i.e., participants randomized). If it was not possible to impute appropriate measures for the calculation of effect sizes or if data was missing, we contacted the authors to obtain them. If authors did not provide the respective information, studies were excluded from further analyses.

Primary Outcomes

We applied a hierarchy for the choice of outcomes: (1) As a first choice, we extracted the primary outcome as defined by the study authors. In the presence of two or more primary outcomes, we checked trial registries for additional information and/or contacted authors. If no information could be obtained, the outcome for the present analysis was selected based on (2) the most frequently reported outcome across our data pool (i.e., pain was preferred over medication use) in order to reduce heterogeneity, and if this was not applicable (3) the most informative outcome (e.g., a symptom-related scale preferred over a general quality of life assessment). In the absence of a baseline assessment for the primary outcome, another outcome was chosen according to these rules, avoiding the exclusion of this study (see eTable 1 in the supplement for the rationale of choice for the outcomes). If more than one baseline measure was collected, we chose the timepoint closest to the start of the intervention[35]. If more than one post measurement was reported, we extracted the first assessment after intervention end (i.e., measured at the time point closest to the end of treatment), if no other explanation for the most clinically relevant time point was given in the publication (i.e., a definition of the primary endpoint measurement). In studies including a WL control group, participants additionally received the OLP treatment after study completion. Outcomes for these individuals were not included in the analyses because they lacked a control group for comparison and in order to avoid enrolling participants multiple times.

Sample building

We allocated each study to either the nonclinical or the clinical study pool. Nonclinical studies were defined as studies that: experimentally induced states (i.e., experimentally induced pain, itch, sadness), whereas clinical studies investigated the effects of OLPs in naturally occurring states (e.g., clinical: irritable bowel syndrome, chronic lower back pain; subclinical: test anxiety, well-being, relaxation). One study[25] experimentally induced pain in an IBS patient sample. This study was rubricated as clinical.

Node building

In order to be able to test the effects of different OLP modalities in comparison to different control groups, each group in a study was clustered together with similar other study groups. Our strategy to create the nodes was data-based and with the aim to restrain from a high number of nodes. This lumping approach has the methodological advantage to increase power and to allow for more accurate estimates of the effect sizes[36,37]. The following rules were applied: (1) Nodes were built according to the OLP administration route, i.e., nasal (vapor, spray), dermal (cream, patch), or injection. In the case of oral application, we differentiated between pills (capsules, tablets) and suspensions (drops, solutions). (2) Groups testing different treatment rationales or intervention components alongside with the placebo administration (i.e.,[12,26,38–40]) or different amounts of placebos per day (i.e.,[41]) were merged. However, study groups testing the effect of OLP without the application of any expectation induction (herein referred to as OLP-) were separately entered into the analyses. (3) If there were different comparator groups that fell within one category (e.g., several DP groups), we merged them into one node (i.e.,[25,27,38,40]). (4) To assess the differences of expectation induction (e.g., through verbal suggestion or conditioning paradigms), these nodes were defined separately (e.g., OLP vs. cOLP). (5) In all cases where participants could receive the intervention upon study conclusion, we used the node WL control group. When data of study groups were merged, we used different formulas[34].

Risk of Bias

We assessed the risk of bias of the included studies using the Cochrane risk of bias tool 2[42]. Each study was assessed by two reviewers, with conflicts resolved by consensus. To account for the special nature of included studies in this NMA (i.e., all of them not being blind), we employed some special rules: (1) If we received a “high” risk of bias rating in domain 4 only due to signaling question 4.5 (“Is it likely that assessment of the outcome was influenced by knowledge of intervention received?”), we overwrote the suggestion of the algorithm for this domain to “some concerns”. The rationale for this decision is based on the fact that a single “high” judgment in one of the four domains leads to an overall high risk of bias. Thus, all of our included studies would have received a high overall risk rating and consequently we would have lost all variance in our assessments. (2) When answering signaling questions 2.1 (“Were participants aware of their assigned intervention during the trial?”) and 4.3.question (“Were outcome assessors aware of the intervention received by study participants?”) for the comparison OLP (i.e., being aware of receiving the intervention) and DP group (i.e., being not aware), we judged as if both groups were unblinded as suggested by the authors of the risk of bias tool 2[43]. (3) Because the risk of bias tool

2 requires an assessment of the level of study group comparisons within a study, multiple assessments were performed per study. However, all multiple assessments within a study were identical and thus reported in a single column (see eTable 1 in the supplement).

Statistical analysis

In order to answer our research questions (RQ) we proceeded as follows: (RQ1) Two different networks were conducted separately, one for the clinical and one for the nonclinical population. These networks were then compared qualitatively. (RQ2) OLP treatment modalities were compared directly using head to head comparisons, excluding OLP-. (RQ3) The effect of treatment expectations was assessed using head to head comparisons with OLP- to all other OLP modality groups. (RQ4) To assess the effects of different comparator groups (i.e., NT, TAU, WL) we compared all OLP modalities that were significantly better than NT with the other comparator groups.

Effect sizes of the interventions applying the standardized mean difference (SMD) were calculated, with their magnitude interpreted as small, moderate or large, with 0.20, 0.50, and 0.80 SD units[44]. We decided to employ random-effects models rather than fixed-effects models because the included studies were expected to be heterogeneous. Network meta-analytic methods were applied within a frequentist framework using the package "netmeta" in R[45,46]. Results are presented as SMDs with corresponding 95% confidence intervals.

NMA relies on the assumption of transitivity to estimate indirect treatment effects. This assumption implies that any study participant that meets all inclusion criteria in each network is likely, in principle, to be randomized to any of the interventions in the corresponding network. We addressed the assumption of transitivity[47], by first conducting two separate networks (i.e., nonclinical and clinical) in order for the distribution of potential modifiers (e.g. population) to be more balanced across comparisons and by second checking whether the direct and indirect treatment effects are in statistical agreement (via an assessment for inconsistency). We conducted a statistical evaluation of consistency, i.e., the agreement between direct and indirect evidence, using local (separating direct from indirect evidence[48]) as well as global (design- by-treatment interaction test[49]) approaches.

The various effects of the groups were ranked using *P* scores. *P* scores are values between 0 and 1 and have an interpretation analogous to the surface under the cumulative ranking curve values[50] and measure the extent of certainty that a treatment is better than another treatment, averaged over all competing treatments. The *P* scores result in a ranking of all treatments that essentially follows the ranking of the point estimates but takes precision into account[50].

For all treatment comparisons in a NMA, we assumed a common between-study heterogeneity. Different statistics were used to quantify heterogeneity: the (within design) Q statistic[45], the between-study variance $\tau 2$, and the heterogeneity statistic I^2 [50]. The I^2 value can be interpreted as follows: 0 to 40% might not be important; 30 to 60% may represent moderate heterogeneity; 50 to 90% may represent substantial heterogeneity; 75 to 100% represents considerable heterogeneity[51].

The certainty of evidence for the network estimates of the efficacy outcomes was evaluated by using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) ratings[47], which were conducted in CINeMA (Confidence in Network Meta-Analysis[50]). In GRADE, the quality of a body of evidence is defined as the study limitations, imprecision, inconsistency, indirectness and reporting bias[47]. To assess across-study bias (reporting bias), a comparison-adjusted funnel plot and the Egger test for funnel plot asymmetry were computed[48]. In case of asymmetry the trim and fill method was used to adjust for small-study effects with NT as reference[52,53]. Due to too few comparisons we were not able to use the tool for assessing risk of bias due to missing evidence in a synthesis (ROB-MEN[54]) as initially planned.

We conducted sensitivity analyses excluding studies in which the risk of bias was high. We decided to choose this criterion, as all studies had at least a moderate risk due to the fact that blinding was not given and that most outcomes were patient-reported. We also conducted sensitivity analyses to investigate if results differed within the clinical network when the subclinical studies were excluded. Furthermore, owing to the great variance of included conditions within each of the two networks and due to considerable heterogeneity in the nonclinical network, we performed subgroup analysis for two broad areas – pain and psychological conditions.

Results

A total of 12’991 records were retrieved by bibliographic database and registries searching. After removing duplicates, 6’811 remained and their title and abstracts were screened together with 21 additionally identified records. Subsequently, 731 full-texts were screened. Thirty-seven RCTs (comprising 3’021 participants) conducted between 2010 and 2022 comparing 12 interventions and 3 control groups met all of the eligibility criteria and were included into our analyses. A flow chart detailing the process of study identification and selection is shown in eFigure 1 in the supplement. All studies were reported in English and included an adult sample with a mean (SD) age of 36 (15.3) years (range 19-70 years). All selected outcomes were of continuous nature (see eTable 1 in the supplement for details on selected outcomes). The individual characteristics of the 37 studies included in the analysis are given in eTable 1 in the supplement.

Nonclinical sample

Twelve studies yielded sufficient data to be included in the analysis of the nonclinical sample (comprising 1’015 participants). The sample sizes of individual studies ranged from 21 to 151. The mean (SD) age of this sample was 23.6 (2.1) years (range: 20–28 years), and 67.7% of the sample population were female. Four studies examined experimentally induced pain, three itch, two sadness, one acute stress, one nausea, and one tested muscle strength. All studies were single-center studies except one[38]. Eight trials recruited participants from Europe (Germany, Netherlands, Switzerland and UK), three from North America (USA) and one from Australia. The studies had different routes of placebo administration such as nasal (4 studies), dermal (6 studies) and oral (2 studies). Ten studies used a NT control, nine included a DP and two an OLP- condition. One study used conditioning in order to evoke positive treatment expectations, all others used verbal suggestions.

Figure 1A shows the network of eligible comparisons and figure 2A shows the forest plot of the NMA including all treatments and control groups using NT as a reference. In this network, only nasal OLPs were significantly better than NT (SMD = 0.43, [0.02–0.84]). Dermally applied conditioned and unconditioned OLPs as well as OLP pills were not significantly better compared to NT (SMDs ranging from 0.10, [-0.60–0.80] to 0.47, [-0.33–1.28]). OLP- was worse than NT (SMD = -0.60, [-1.15– -0.05]). (RQ2) The investigation of head to head comparisons (see eTable 2 in the supplement) of different OLP modalities revealed no significant differences with SMDs ranging from 0.04, [-0.83–0.92] to 0.38, [-0.68–1.43]. (RQ3) OLPs without the induction of treatment expectation were statistically worse compared to all other OLP modalities (SMDs ranging from -0.86, [-1.41– -0.31] to -1.07, [-2.02– -0.12]) except for the comparison with OLP pills (SMD = -0.69, [-1.57– 0.19]). (RQ4) Within this network there was only one comparator (i.e., NT). Therefore, differential effects depending on the comparison groups used could not be investigated.

Clinical sample

The analysis of the clinical sample included 25 studies with 2'006 participants and sample sizes of individual studies ranging from 19 to 211. The mean (SD) age of this sample was 43.7 (14.9) years (range: 19–70 years), and 70.7% were female. The different populations used in the 25 included studies were the following: chronic low back pain (4 studies), allergic rhinitis (3 studies), cancer-related fatigue (3 studies), irritable bowel syndrome (2 studies), knee osteoarthritis (2 studies), major depressive disorder (2 studies), acute pain (following spine surgery; 1 study), acute pain (spinal cord injury and polytrauma; 1 study), chronic low back pain + experimental pain (1 study), menopausal hot flushes (1 study), primary insomnia (1 study), relaxation test (1 study), test anxiety (1 study), well-being (1 study), and well-being + cognitive enhancement (1 study). The mean duration of the treatment phase was three weeks (range: 1 day to 12 weeks). No study was multicentered. Thirteen trials recruited patients from Europe (Germany, Austria, Denmark, Portugal), eight from North America (USA), three from Asia (Japan and Israel), and one from Australia. Various routes of placebo administration were used such as nasal (4 studies) and dermal (5 studies) applications as well as injections (2 studies). Furthermore, oral applications included pills (21 studies) and suspensions (2 studies). Nine studies used a NT control condition, five TAU and eight a WL. Furthermore, two studies included a DP, two an OLP- group, one a psychological intervention and one a treatment program (exercise and education intervention) as a comparator group. Overall, three studies used a conditioning paradigm to induce positive treatment expectation.

Figure 1B depicts the clinical network with eligible comparisons and figure 2A shows the Forest plot of the NMA including all treatments and control groups using NT as reference. In the clinical network, conditioned and unconditioned OLP pills outperformed NT (0.89, [0.01–1.76] to 0.46, [0.28–0.65], respectively). Injected OLPs and conditioned and unconditioned OLP suspensions were not statistically better than NT (SMDs ranging from 0.23, [-0.54–1.01] to 0.70, [-0.14–1.54]). (RQ2) The investigation of head to head comparisons (see eTable 2 in the supplement) of different OLP modalities revealed no significant differences with SMDs ranging from -0.08, [-1.18–1.01] to 0.65, [-0.50–1.81]. (RQ3) OLPs without the induction of treatment expectation were not statistically different from any other OLP modality (SMDs ranging from -0.26, [-1.03–0.51] to -0.92, [-1.87–0.04]) except for the comparison with OLP pills, here OLP- was significantly worse (SMD = -0.49, [-0.92– -0.07]). (RQ4) The investigation of the effects of treatment comparators showed that OLP pills was in addition to NT also significantly better than WL (SMD = 0.43, [0.22 – 0.64]) but not TAU (SMD = 0.16, [-0.48–0.80]). In addition, cOLP pills was significantly better than TAU (SMD = 0.58, [0.02–1.15]), and marginally not significant compared with WL, yet the effect was high (SMD = 0.86, [-0.02–1.74]).

Results of sensitivity analyses, adverse events and certainty of evidence assessment can be found in the supplement (see eAppendix 6).

Discussion

This systematic review and NMA of RCTs with 3'021 individuals assessed the efficacy of various OLP interventions in comparison to different types of control groups both in a nonclinical and clinical sample. The aim was to examine whether the size of the OLP effect is different across (RQ1) nonclinical vs. clinical populations, (RQ2) treatment modalities, (RQ3) treatment expectation, and (RQ4) comparator groups. Across both networks, a wide range of conditions was studied with pain and diverse psychological conditions being the most frequent.

Within the nonclinical sample the NMA revealed a significant effect of OLP administered as a spray or vapor (i.e, OLP nasal) compared with NT (SMD = 0.43). All other OLP interventions showed small to medium but insignificant SMDs compared with NT. Similar results were found for the clinical sample, where only OLP pills outperformed NT (SMD conditioned = 0.89; unconditioned= 0.46), with again all other modalities showing insignificant but small to medium effects. Even though only some OLP modalities were significantly better than NT, the comparison of the different employed OLP modalities in both networks showed no significant differences. However, OLPs without the induction of treatment expectation were statistically worse compared to the majority of OLP modalities within the nonclinical network (SMD ranging from -0.86 to -1.07) and compared to OLP pills in the clinical network (SMD = -0.49). Finally, the comparison of treatment comparator groups in the clinical network showed that OLP pills were better than WL (SMD = 0.43) but not better than TAU (SMD = 0.16).

In the following, the observed effects will be discussed with regard to the four distinct research questions. In order to investigate differential effect sizes across the nonclinical and clinical sample (RQ1), we compared the findings of both networks qualitatively. We found that the effect sizes for the comparison of OLP pills to NT yielded smaller and nonsignificant effects within the nonclinical sample (i.e., SMD nonclinical = 0.10; clinical = 0.46). This trend was also exemplified by the comparison of DP against NT (SMD nonclinical = 0.50; clinical = 0.76). Similar observations have previously been reported for both somatic and psychological conditions: For example, studies investigating placebo analgesia have found an average effect size of 1.24 in healthy individuals and an effect size of 1.49 in patients[14]. This finding not only suggests that DPs employed in OLP studies tend to yield smaller effects as compared to studies investigating DPs only, but also sheds light on the difference between the effect sizes of placebo effects across nonclinical and clinical samples. In this regard, our two networks may support the notion that placebo effects tend to be of greater magnitude in clinical as opposed to nonclinical populations. This trend was supported by our sensitivity analysis, where effect sizes were slightly bigger when excluding subclinical studies. A potential explanation could be the more pronounced desire of relief in patients as opposed to healthy individuals[55]. In summary, this finding suggests that clinical and subclinical populations might benefit from OLP treatments to a greater degree than healthy individuals and that experimental studies on healthy individuals may underestimate the magnitude of the OLP effect in patients. However, this comparison is only qualitative in nature and therefore could be further explored as part of a single study.

In terms of OLP modalities (RQ2), none of the direct comparisons were statistically significant, indicating that there might not be a difference in the effect across OLP intervention modalities in either sample. This finding stands in contrast to the roam of DP, where it is known that more invasive routes of administration can yield bigger effects compared to less invasive procedures[19,21]. This discrepancy suggests that findings from DP research might not be valid for the field of OLP. However, SMDs varying up to 0.50 across modalities suggest (especially within the clinical sample) that the current analyses might be underpowered in order to observe statistically significant differences. However, there is also reason to assume that potential differences in OLP modalities may be obscured given that the present analyses investigated the efficacy of OLP treatments across a variety of different somatic and psychological conditions. Supporting this line of reasoning, Peerdeman et al. (2017)[56] found that expectations towards the efficacy of different routes of administration differed for pain and itch, e.g., injected medications were expected to be most effective for relieving pain and topical medications for alleviating itching. These results might reflect the impact of knowledge and prior experience on treatment expectations. Regardless, placebo experts advise against prescribing more invasive treatments to yield stronger effects, as this entails practical and ethical limitation[1]. Especially the yet small database for OLPs calls for a cautious consideration regarding the use of more invasive procedures.

Regarding our research question on the impact of expectation (RQ3), evidence from both networks suggests that OLP interventions delivered without the evocation of at least minimal treatment expectations are less efficacious as compared to OLP interventions with the induction of treatment expectation. This finding was especially pronounced within the nonclinical network, where OLP- was less efficacious as compared to all other groups within the network, even to NT (SMD = -0.60). However, the efficacy of OLP- within the nonclinical network was solely evaluated by two trials investigating dermal placebo applications[8,26]. Nevertheless, it appears that expectancy building is an important component of OLP interventions and that simply prescribing an inert treatment is not sufficient. Hence, OLP treatments might be cost-efficient but not as time-efficient as over the counter medicine. Possible explanations for this observation could be that participants do not feel taken seriously when they are simply told that they are receiving a placebo treatment, or that they are disappointed because they may not know about the power of placebo effects. The effects in the clinical setting (where no differences between OLP- and NT were observed) might potentially be buffered by at least performing a ritual of e.g. taking pills over a period of time. In a single administration, which was often the case in the nonclinical studies, no such ritual could be established. Therefore, the rationale seems to be an essential and potentially indispensable component for the efficacy of OLP[57]. In this sense, OLP conditions that do not include any expectation building component could at best serve as control groups, controlling for the component of the pill. The pill itself might therefore not be necessary to produce positive treatment effects in OLP studies. This finding is supported by a recently published RCT on OLPs and imaginary pills (Buergler et al. 2023).

With respect to the potential impact of different comparators (RQ4), our systematic search showed that due to the experimental setting all nonclinical studies used a NT control group. Differences across control groups could thus only be investigated within the clinical sample. There, we identified three different comparison groups, namely NT, WL and TAU. Comparison of effect sizes across different

comparator groups showed that OLP pills was in addition to NT also significantly better than WL (SMD = 0.43) but not TAU (SMD = 0.16). In other words, this finding could imply that OLP pills are better than "nothing", but not better than "something". Thus, the efficacy of both of these interventions seems to depend on the kind of control group used, a finding in line with psychotherapy research[28]. However, whereas there WL was notably inferior to NT, in the present study both comparison groups yielded comparable effects. Conditioned OLP pills, on the other hand, were significantly better than NT as well as TAU (SMD = 0.58) and tended to be better than WL (SMD = 0.86; *n.s.*). However, these findings are based on two studies only, indicating that the obtained conclusions are not entirely conclusive and should be further explored. Nevertheless, these findings suggest that comparator groups within OLP studies should be chosen carefully as the effects might differ according to the chosen comparator.

Overall, the present analyses confirm the results of previous meta-analyses investigating the efficacy of OLP in clinical populations, which found moderate to high effect sizes[4,5]. In contrast, the results of the nonclinical sample contradict in part the findings by Spille et al. (2022)[6], which found a medium sized effect for subjective outcomes for OLPs in comparison to NT. This difference in findings might be explained by different inclusion criteria and thus another body of studies that contributed to the results (e.g., the inclusion of subclinical studies in their analysis) and might further be fostered by their differentiation between objective and subjective outcomes. Remarkably, the herein found effect sizes for OLP pills in the clinical sample were smaller as compared to previous investigations, which included only OLP pills in comparison to different control conditions (SMD = 0.46 vs. 0.88[4] and 0.72[5]). This trend towards smaller effects across the timespan suggests that in an early state of research, "positive" studies are more likely to be published (reporting bias – which was also present within the clinical sample of this NMA) and with time insignificant results are more likely to be published (time lag bias).

This study has several strengths. First, the direct comparison of different placebo intervention modalities and comparators is of great importance to inform the young research field of OLPs about the comparative efficacy in order to better design future studies. The network meta-analytic approach uniquely allows investigating the effects of different modalities and comparators. Second, this methodology allows combining direct and indirect evidence to get the most precise estimate of the intervention differences. Third, we were able to include 13 more studies than the newest existing meta-analysis on OLP in clinical conditions[5], which strengthens the body of evidence. Fifth, the clear definition of OLPs is a strength of this analysis as well as the several sensitivity analyses that were conducted, which showed comparable results that further supported the trends of the overall analyses. Finally, the application of the same inclusion criteria for the nonclinical and the clinical sample allows to more reliably compare effect sizes across both populations. However, this study has several limitations that should be taken into account when interpreting the results. First, although the network meta-analytic approach allowed to include 12 studies within the nonclinical network and 25 within the clinical, which represents a considerably broad range of studies as compared to previous analysis, the relatively small number of studies in each node and the resulting small power might have led to a lack of significance (large confidence intervals). Second, a major limitation of our NMAs is associated with the fact that most interventions have been tested in less than 100 participants. It is therefore possible that the effect of some of these interventions is owing to a so-called small-study effect: smaller trials show different, often

larger, treatment effects than bigger ones[58,59]. Third, substantial heterogeneity was found in our NMAs. The variety of the studied conditions, the format of the interventions (e.g., duration), and the reported outcomes differed widely, which may have contributed to the statistical heterogeneity and certainly to the clinical heterogeneity. However, we tried to reduce heterogeneity by applying a very precise and strict definition of OLPs, by conducting two separate networks and by choosing the most frequent outcome, in case of the presence of several outcomes. Furthermore, sensitivity analyses suggest that the results remain unchanged when looking at more homogeneous subgroups within the network as for example pain. Fourth, although NMAs have the advantage of making use of all available data, the indirect evidence does not directly stem from randomized comparisons[60]. Fifth, according to the GRADE framework, the within-study bias of many comparisons was assessed as "some concerns", which can be attributed in part by methodological difficulties which arise through the nature of OLPs (i.e., participants being unblinded) and the nature of most outcomes being self-reported. Sixth, funnel plots and accompanying Egger's tests indicated a risk for reporting bias for the clinical network because of the lack of small studies comparing NT versus OLP pills with negative effects. Seventh, we excluded cross-over studies due to analytical concerns regarding comparability with parallel-trials[34], which reduced the body of evidence to parallel trials with accompanying loss in power. Eighth, the clinical sample also includes undiagnosed subclinical conditions, which limits the comparability to other studies, which only included studies with diagnosed samples (e.g.,[5]). However, this was accounted for by conducting a sensitivity analysis which yielded comparable results. Tenth, the relatively early stage of OLP research did not allow to investigate the efficacy of OLPs within distinct conditions. Therefore, the present NMA examined the interventions on meta-level lumping studies with different conditions, which might impair the requirement for NMAs of included populations being in theory jointly randomizable. Finally, the results on the comparable efficacy of OLP modalities might be explained by population specific choices of treatment modalities, obscuring potential differences within each domain.

With this NMA, we were able to identify several research gaps: First, larger studies should be conducted, as sample sizes are often relatively small (range: 19 - 211). Second, the population should be more representative: Currently, the majority of the study population is female (70%) and especially in the nonclinical sample very young (mean age: 23.6 years). This complicates, among other things, the transfer of nonclinical findings to the clinical population, which was on average older (mean age: 43.7 years). Third, adverse events should be reported more structured and consistently. Because of not or inconsistently reported adverse events, we were not able not analyze them in the present study. Fourth, it would be crucial to conduct future studies by more independent research teams with less allegiance to OLP research. Fifth, in future studies, the control group used should be chosen deliberately, because depending on the type of control group – as our study shows – different sizes of effects result. Also, in further meta-analysis, control groups should not be lumped together, as this can obscure possible treatment effects. Sixth, in a further (network) meta-analyses, it would be informative to distinguish between active and non-active OLP, which was not considered in these analyses. Seventh, further experimental studies should be designed more according to the needs of clinical populations: For instance, the OLP modalities OLP nasal (e.g., spays) and OLP dermal (e.g., creams) were only studied in nonclinical populations and not in clinical, possibly indicating that this route of administration is not

suitable for clinical conditions. Eighth, future (network) meta-analyses should take into account that potential differences between OLP modalities may be masked, as their effects may differ depending on the type of disease. Finally, in order to reduce within study bias, future research should include objective outcomes and behavioral markers.

To conclude, OLPs can be beneficial compared to control conditions in nonclinical and clinical conditions. However, the magnitude of effects appears to be smaller compared to previous meta-analyses and further depend on several aspects that we have considered in our NMAs. (1) We identified a trend for greater effect sizes within the clinical network. Hence, research in nonclinical samples may underestimate the magnitude of OLP effects in patients. (2) There were no differences in the effect across OLP modalities in either sample. This finding calls for a cautious consideration regarding the use of more invasive OLP procedures. (3) Inducing positive treatment expectation is of great importance for the efficacy of OLPs. Simply prescribing an OLP seems not to be enough and might even hold the risk of being worse than receiving nothing. (4) Finally, we found that OLP effects can vary depending on the comparator used. In other words, some interventions facilitate relief when compared to “nothing” but their effect appears to vanish when compared to other treatments. With this NMA, we hope to expand the knowledge in the emerging research field of OLPs and inform future studies aimed at exploring ethical ways to use placebo effects for the good of patients.

Additional statements

Acknowledgments: The authors would like to thank all masters students: Laura Tincknell, Sarah Marie Hanig, Jana Hochreuter, Jennifer Lopez and interns: Sebastian Penz, Lea Masanti, Cyrill Hübscher as well as our research assistants: Yannick Schulz for aiding with data screening, extraction, eTable1 and references and Berfin Bakis with her help double-checking the results. Furthermore, the authors want to thank Hannah Ewald and Christian Appenzeller for their help with developing the search strategy and conducting the search and Guido Schwarzer for his help regarding analyzing the data in R.

Contributors: Authors contributed to the work as follows: Conceptualization: SB, DS, CL, JG; Methodology: SB, DS, CL; Formal Analyses: SB, DS, CL; Writing—original draft: SB, DS; Writing—review & editing: SB, DS, CL, JG; Visualization: SB, DS; Supervision: CL, JG; Project Administration: SB, DS

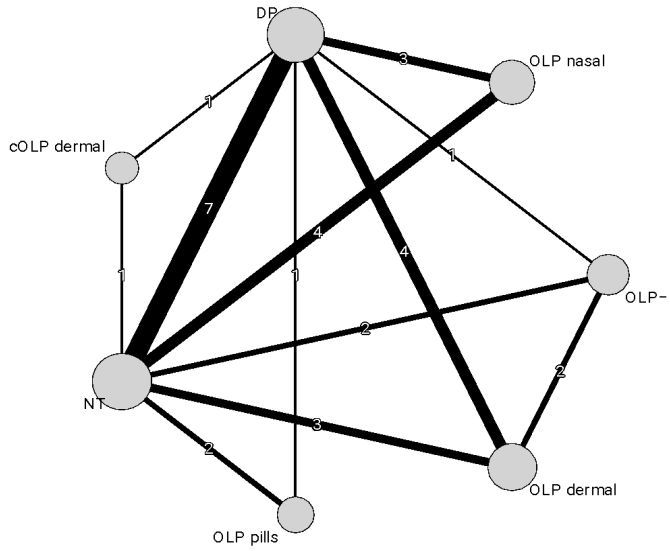
Competing interests' statement: The authors have no conflicts of interest to disclose.

Data availability statement: Data will be shared upon reasonable request addressed to sarah.buergler@unibas.ch.

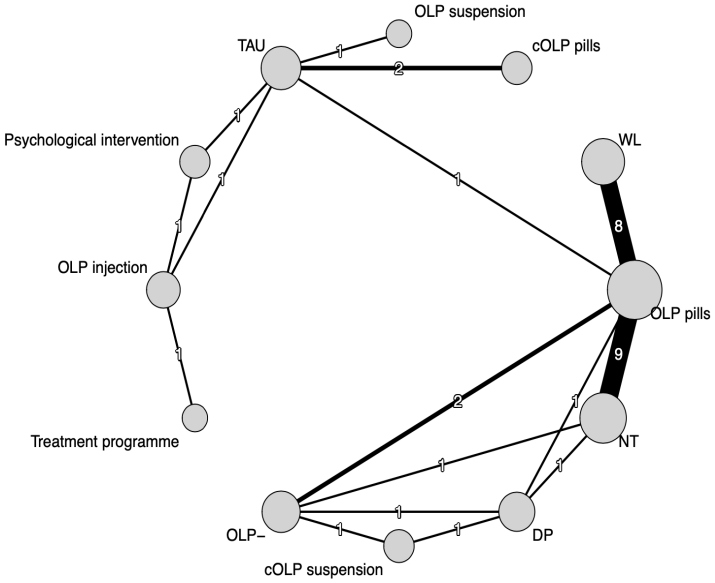
Funding: CL received funding from the Swiss National Science Foundation (SNSF): PZ00P1_201972**.**

Figures

Figure 1. Network Meta-analysis of eligible comparisons



A. Network meta-analysis of eligible comparisons for the nonclinical sample

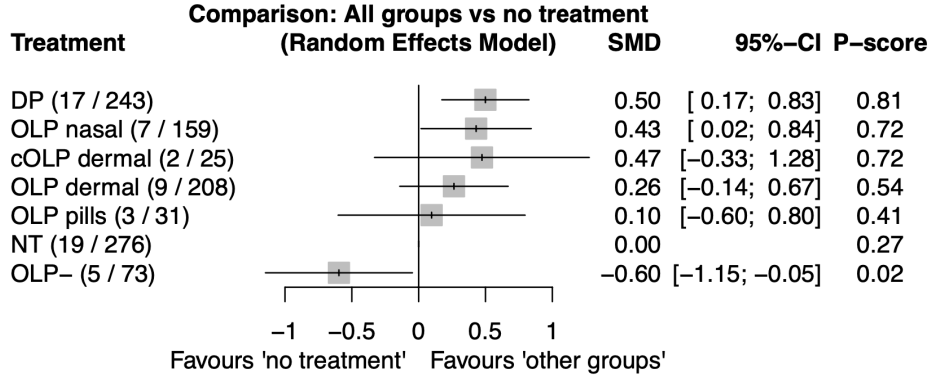


B. Network meta-analysis of eligible comparisons for the clinical sample

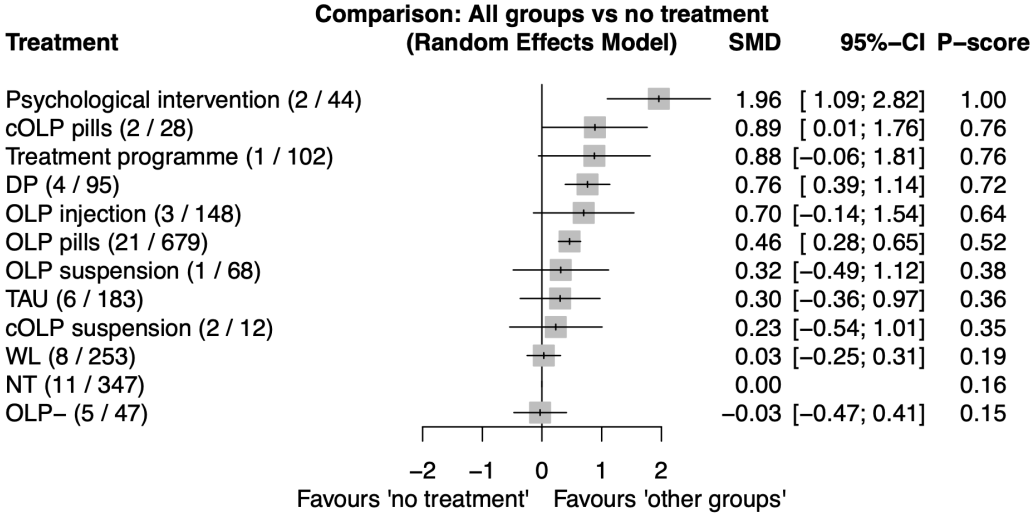
A, nonclinical. B, clinical. Width of the lines is proportional to the number of trials comparing every pair of treatments/groups.

Note. cOLP, conditioned Open-Label Placebo; DP, Deceptive Placebo; NT, No Treatment; OLP, Open-Label Placebo with rationale; OLP-, Open-Label Placebo without expectation induction; TAU, Treatment as Usual; WL, Wait List.

Figure 2. Forest plot of network meta-analysis of all trials



A. Forest plot of network meta-analysis of all trials for the nonclinical sample



B. Forest plot of network meta-analysis of all trials for the clinical sample

All groups were compared with no treatment (NT), which was the reference group. The brackets behind the group names indicate the following: number of direct comparisons with this group/number of patients in which the intervention/control was examined. SMD indicates standardized mean difference.

Note. cOLP, conditioned Open-Label Placebo; DP, Deceptive Placebo; NT, No Treatment; OLP, Open-Label Placebo with rationale; OLP-, Open-Label Placebo without expectation induction; TAU, Treatment as Usual; WL, Wait List.

References

1. Evers, A. W. M. *et al.* Implications of Placebo and Nocebo Effects for Clinical Practice: Expert Consensus. *Psychother Psychosom* **87**, 204–210 (2018).
2. Evers, A. W. M. *et al.* What Should Clinicians Tell Patients about Placebo and Nocebo Effects? Practical Considerations Based on Expert Consensus. *Psychother Psychosom* **90**, 49–56 (2021).
3. Blease, C., Colloca, L. & Kaptchuk, T. J. Are open-Label Placebos Ethical? Informed Consent and Ethical Equivocations. *Bioethics* **30**, 407–414 (2016).
4. Charlesworth, J. E. G. *et al.* Effects of placebos without deception compared with no treatment: A systematic review and meta-analysis. *J Evid Based Med* **10**, 97–107 (2017).
5. von Wernsdorff, M., Loef, M., Tuschen-Caffier, B. & Schmidt, S. Effects of open-label placebos in clinical trials: a systematic review and meta-analysis. *Sci. Rep.* **11**, 3855 (2021).
6. Spille, L., Fendel, J. C., Seuling, P., Göritz, A. S. & Schmidt, S. Open-label placebos: A systematic review and meta-analysis of experimental studies with non-clinical samples. (2022) doi:https://doi.org/10.21203/rs.3.rs-2093533/v1.
7. Sezer, D., Locher, C. & Gaab, J. Deceptive and open-label placebo effects in experimentally induced guilt: a randomized controlled trial in healthy subjects. *Sci. Rep.* **12**, 21219 (2022).
8. Locher, C. *et al.* Is the rationale more important than deception? A randomized controlled trial of open-label placebo analgesia. *Pain* **158**, 2320–2328 (2017).
9. Lembo, A. *et al.* Open-label placebo vs double-blind placebo for irritable bowel syndrome: a randomized clinical trial. *Pain* **162**, 2428–2435 (2021).
10. Mundt, J. M., Roditi, D. & Robinson, M. E. A comparison of deceptive and non-deceptive placebo analgesia: efficacy and ethical consequences. *Ann. Behav. Med.* **51**, 307–315 (2017).
11. Disley, N., Kola-Palmer, S. & Retzler, C. A comparison of open-label and deceptive placebo analgesia in a healthy sample. *J. Psychosom. Res.* **140**, 110298 (2021).
12. Kube, T. *et al.* Deceptive and Nondeceptive Placebos to Reduce Pain: An Experimental Study in Healthy Individuals. *Clin J Pain* **36**, 68–79 (2020).
13. Sandler, A. D. & Bodfish, J. W. Open-label use of placebos in the treatment of ADHD: a pilot study. *Child Care Health Dev* **34**, 104–10 (2008).
14. Forsberg, J. T., Martinussen, M. & Flaten, M. A. The Placebo Analgesic Effect in Healthy Individuals and Patients: A Meta-Analysis. *Psychosom Med* **79**, 388–394 (2017).
15. Charron, J., Rainville, P. & Marchand, S. Direct comparison of placebo effects on clinical and experimental pain. *Clin J Pain* **22**, 204–11 (2006).
16. Kaptchuk, T. J. *et al.* Sham device v inert pill: randomised controlled trial of two placebo treatments. *Bmj* **332**, 391–7 (2006).
17. de Craen, A. J., Tijssen, J. G., de Gans, J. & Kleijnen, J. Placebo effect in the acute treatment of migraine: subcutaneous placebos are better than oral placebos. *J Neurol* **247**, 183–8 (2000).
18. Macedo, A., Farré, M. & Baños, J. E. A meta-analysis of the placebo response in acute migraine and how this response may be influenced by some of the characteristics of clinical trials. *Eur J Clin Pharmacol* **62**, 161–72 (2006).
19. van Laarhoven, A. I. M. *et al.* Placebo effects on itch: a meta-analysis of clinical trials of patients with dermatological conditions. *J Invest Dermatol* **135**, 1234–1243 (2015).
20. Bannuru, R. R. *et al.* Effectiveness and Implications of Alternative Placebo Treatments: A Systematic Review and Network Meta-analysis of Osteoarthritis Trials. *Ann Intern Med* **163**, 365–72 (2015).
21. Meissner, K. *et al.* Differential effectiveness of placebo treatments: a systematic review of migraine prophylaxis. *JAMA Intern Med* **173**, 1941–51 (2013).
22. Colloca, L. & Miller, F. G. How placebo responses are formed: a learning perspective. *Philos Trans R Soc Lond B Biol Sci* **366**, 1859–69 (2011).

23. Kaptchuk, T. J. *et al.* Placebos without deception: a randomized controlled trial in irritable bowel syndrome. *PLoS One* **5**, e15591 (2010).

24. Haas, J. W., Winkler, A., Rheker, J., Doering, B. K. & Rief, W. No open-label placebo effect in insomnia? Lessons learned from an experimental trial. *J Psychosom Res* **158**, 110923 (2022).

25. Klinger, R., Kothe, R., Schmitz, J., Kamping, S. & Flor, H. Placebo effects of a sham opioid solution: a randomized controlled study in patients with chronic low back pain. *Pain* **158**, 1893–1902 (2017).

26. Leibowitz, K. A., Hardebeck, E. J., Goyer, J. P. & Crum, A. J. The role of patient beliefs in open-label placebo effects. *Health Psychol* **38**, 613–622 (2019).

27. 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* **13**, e0192758 (2018).

28. Michopoulos, I. *et al.* Different control conditions can produce different effect estimates in psychotherapy trials for depression. *J. Clin. Epidemiol.* **132**, 59–70 (2021).

29. Blease, C. R., Bernstein, M. H. & Locher, C. Open-label placebo clinical trials: is it the rationale, the interaction or the pill? *BMJ Evid Based Med* **25**, 159–165 (2020).

30. Watts, S. E., Turnell, A., Kladnitski, N., Newby, J. M. & Andrews, G. Treatment-as-usual (TAU) is anything but usual: a meta-analysis of CBT versus TAU for anxiety and depression. *J Affect Disord* **175**, 152–67 (2015).

31. Cuijpers, P., Quero, S., Papola, D., Cristea, I. A. & Karyotaki, E. Care-as-usual control groups across different settings in randomized trials on psychotherapy for adult depression: a meta-analysis. *Psychol. Med.* **51**, 634–644 (2021).

32. Liberati, A. *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* **6**, e1000100 (2009).

33. Bramer, W. M., Giustini, D., de Jonge, G. B., Holland, L. & Bekhuis, T. De-duplication of database search results for systematic reviews in EndNote. *J Med Libr Assoc* **104**, 240–3 (2016).

34. Higgins, J. P. *et al. Cochrane handbook for systematic reviews of interventions*. (John Wiley & Sons, 2019).

35. Gervasio, J., Zheng, S., Skrotzki, C. & Pachete, A. The effect of oral contraceptive use on cortisol reactivity to the Trier Social Stress Test: A meta-analysis. *Psychoneuroendocrinology* **136**, 105626 (2022).

36. Koechlin, H. *et al.* Nonpharmacological interventions for pediatric migraine: a network meta-analysis. *Pediatrics* **147**, (2021).

37. Shi, C., Westby, M., Norman, G., Dumville, J. C. & Cullum, N. Node-making processes in network meta-analysis of nonpharmacological interventions should be well planned and reported. *J. Clin. Epidemiol.* **101**, 124–125 (2018).

38. Friehs, T., Rief, W., Glombiewski, J. A., Haas, J. & Kube, T. Deceptive and non-deceptive placebos to reduce sadness: A five-armed experimental study. *J. Affect. Disord. Rep.* **9**, 100349 (2022).

39. Barnes, K., Yu, A., Josupeit, J. & Colagiuri, B. Deceptive but not open label placebos attenuate motion-induced nausea. *J. Psychosom. Res.* **125**, 109808 (2019).

40. Kube, T., Hofmann, V. E., Glombiewski, J. A. & Kirsch, I. Providing open-label placebos remotely—A randomized controlled trial in allergic rhinitis. *PLoS One* **16**, e0248367 (2021).

41. 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* **53**, 16–28 (2019).

42. Sterne, J. A. C. *et al.* RoB 2: a revised tool for assessing risk of bias in randomised trials. *Bmj* **366**, l4898 (2019).

43. Jelena Savovic. Personal communication. (2022).

44. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. vol. 2nd ed. (Routledge., 1988).

45. Rücker, G. Network meta-analysis, electrical networks and graph theory. *Res Synth Methods* **3**, 312–24 (2012).

46. Rücker, G., Krahn, U., König, J., Efthimiou, O. & Schwarzer, G. Netmeta: network meta-analysis using frequentist methods. *R Package Version 09–0* (2016).

47. Efthimiou, O. *et al.* GetReal in network meta-analysis: a review of the methodology. *Res Synth Methods* **7**, 236–63 (2016).

48. Dias, S., Welton, N. J., Caldwell, D. M. & Ades, A. E. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med* **29**, 932–44 (2010).

49. Higgins, J. P. *et al.* Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res Synth Methods* **3**, 98–110 (2012).

50. Salanti, G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res Synth Methods* **3**, 80–97 (2012).

51. Rücker, G. & Schwarzer, G. Resolve conflicting rankings of outcomes in network meta-analysis: Partial ordering of treatments. *Res Synth Methods* **8**, 526–536 (2017).

52. Duval, S. & Tweedie, R. A nonparametric “trim and fill” method of accounting for publication bias in meta-analysis. *J. Am. Stat. Assoc.* **95**, 89–98 (2000).

53. Duval, S. & Tweedie, R. Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* **56**, 455–463 (2000).

54. Chiocchia, V. *et al.* ROB-MEN: a tool to assess risk of bias due to missing evidence in network meta-analysis. *BMC Med.* **19**, 1–13 (2021).

55. Hyland, M. E. Motivation and placebos: do different mechanisms occur in different contexts? *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 1828–1837 (2011).

56. Peerdeman, K. J. *et al.* Expectations about the effectiveness of pain- and itch-relieving medication administered via different routes. *Eur J Pain* **22**, 774–783 (2018).

57. Locher, C. *et al.* Lay perspectives of the open-label placebo rationale: a qualitative study of participants in an experimental trial. *BMJ Open* **11**, e053346 (2021).

58. Sterne, J. A., Gavaghan, D. & Egger, M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol* **53**, 1119–29 (2000).

59. Sterling, T. D., Rosenbaum, W. L. & Weinkam, J. J. Publication Decisions Revisited: The Effect of the Outcome of Statistical Tests on the Decision to Publish and Vice Versa. *Am. Stat.* **49**, 108–112 (1995).

60. Chaimani, A., Salanti, G., Leucht, S., Geddes, J. R. & Cipriani, A. Common pitfalls and mistakes in the set-up, analysis and interpretation of results in network meta-analysis: what clinicians should look for in a published article. *Evid Based Ment Health* **20**, 88–94 (2017).

Supplementary Online Content

The role of population, expectation, modality and comparator on open-label placebo effects: A network meta-analysis

Sarah Buergler*, Dilan Sezer*, Jens Gaab & Cosima Locher

*shared first authorship

eAppendix 1. Search strategies and hits

Medline Ovid

(20210202; 956 hits)

(((placebo* or sham) adj2 (open-label* or told or nondecept* or non-decept* or nonconceal* or non conceal* or unconceal* or unblind* or nonblind* or non blind* or without decept* or without conceal* or without blind*)).ti,ab,kw,kf. or open placebo*.ti,ab,kw,kf. or ((placebos/ or Placebo Effect/) and (open-label* or told or nondecept* or non-decept* or nonconceal* or non conceal* or unconceal* or unblind* or nonblind* or non blind* or without decept* or without conceal* or without blind*).ti,ab.)) and (exp Random Allocation/ or exp Randomized Controlled Trial/ or exp Randomized Controlled Trials as Topic/ or RCT or (randomiz* or randomis*).ti,ab. or ((controlled clinical or non-inferiority or noninferiority or superiority or equivalence or pragmatic) ADJ2 trial$).ti,ab.)

Embase Ovid

(20210202; 5,487 hits)

(((placebo* or sham) adj2 (open-label* or told or nondecept* or non-decept* or nonconceal* or non conceal* or unconceal* or unblind* or nonblind* or non blind* or without decept* or without conceal* or without blind*)).ti,ab,kw. or open placebo*.ti,ab,kw. or ((placebo/ or Placebo Effect/ or sham procedure/) and (Open study/ or (open-label* or told or nondecept* or non-decept* or nonconceal* or non conceal* or unconceal* or unblind* or nonblind* or non blind* or without decept* or without conceal* or without blind*).ti,ab.))) and (randomization/ or exp randomized controlled trial/ or randomized controlled trial topic/ or "randomized controlled trial (topic)"/ or RCT or (randomiz* or randomis*).ti,ab. or ((controlled clinical or non-inferiority or noninferiority or superiority or equivalence or pragmatic) ADJ2 trial$).ti,ab.)

NOT (conference abstract or conference review).pt

CINAHL Ebsco

(20210202; 589 hits)

((((TI placebo* OR AB placebo*) OR (TI sham OR AB sham)) N2 ((TI open-label* OR AB open-label*) OR (TI told OR AB told) OR (TI nondecept* OR AB nondecept*) OR (TI non-decept* OR AB non-decept*) OR (TI nonconceal* OR AB nonconceal*) OR (TI "non conceal*" OR AB "non conceal*") OR (TI unconceal* OR AB unconceal*) OR (TI unblind* OR AB unblind*) OR (TI nonblind* OR AB nonblind*) OR (TI "non blind*" OR AB "non blind*") OR (TI "without decept*" OR AB "without decept*") OR (TI "without conceal*" OR AB "without conceal*") OR (TI "without blind*" OR AB "without blind*"))) OR (TI "open placebo*" OR AB "open placebo*") OR (((MH "placebos") OR (MH "Placebo Effect")) AND ((TI open-label* OR AB open-label*) OR (TI told OR AB told) OR (TI nondecept* OR AB nondecept*) OR (TI non-decept* OR AB non-decept*) OR (TI nonconceal* OR AB nonconceal*) OR (TI "non conceal*"

OR AB "non conceal*") OR (TI unconceal* OR AB unconceal*) OR (TI unblind* OR AB unblind*) OR (TI nonblind* OR AB nonblind*) OR (TI "non blind*" OR AB "non blind*") OR (TI "without decept*" OR AB "without decept*") OR (TI "without conceal*" OR AB "without conceal*") OR (TI "without blind*" OR AB "without blind*"))))

PsycINFO Ovid

(20210202; 406 hits)

(((placebo* or sham) adj2 (open-label* or told or nondecept* or non-decept* or nonconceal* or non conceal* or unconceal* or unblind* or nonblind* or non blind* or without decept* or without conceal* or without blind*)).ti,ab. or open placebo*.ti,ab. or (placebo/ and (open-label* or told or nondecept* or non-decept* or nonconceal* or non conceal* or unconceal* or unblind* or nonblind* or non blind* or without decept* or without conceal* or without blind*).ti,ab.))

eAppendix 2. Hits update

Medline Ovid

(20210201 bis 20220608; 66 hits)

limit *SEARCH* to dt=20210201-20220608

Embase Ovid

(20210201 bis 20220608; 640 hits)

limit *SEARCH* to dc=20210201-20220608

CINAHL Ebsco

(20210201 bis 20220608; 38 hits)

PsycINFO Ovid

(20210201 bis 20220608; 43 hits)

limit *SEARCH* to up=20210201-20220608

eAppendix 3. GRADE Ratings for each network

We used the Grading of Recommendations Assessment, Development, and Evaluation ratings (GRADE[1]) and the corresponding web application to apply this framework[2,3]. The certainty of evidence for each network estimate was assessed according to the following criteria:

Study limitations (Within study bias): The overall risk of bias of each study was categorized. According to the Cochrane Risk of Bias tool 2[4], we rated five risk of bias domains. We then used the contribution matrix to calculate the percentage of contribution from each study, and finally assessed the study limitation for each network estimate based on the weighted average risk of bias of the contributing studies. We selected the rule "Average Risk of Bias" in order to calculate the within study bias.

Reporting bias (Across studies bias): Since each of our comparisons had less than 10 comparisons, we could not use the ROB-MEN[5] tool to assess reporting bias. Therefore, a comparison-adjusted funnel plot with accompanying Egger test for asymmetry was conducted and used as a basis for the judgment.

Indirectness: We judged that there was no concern in this domain as the included studies matched our inclusion criteria and study questions.

Imprecision: In line with previous analyses[6], we considered a clinically meaningful threshold for standardized mean difference (SMD) to be 0.20.

Heterogeneity: We evaluated the degree of concerns through comparing the clinical inference based on the 95% confidence intervals (CI), the latter reflecting the degree of heterogeneity. Appling the same clinical inference framework as for imprecision, we saw no concerns in heterogeneity when the two judgements matched (e.g. no concern based on 95% CI and no concern based on 95% PI), some concerns when they differed by one degree (e.g. no concern based on 95% CI but some concerns based on 95% PI), and major concerns when they differed by two degrees (e.g. no concern based on 95% CI but major concerns based on 95% PI).

Incoherence (Inconsistency): For inconsistency, we looked at the results of side splitting and we saw major concerns when $p<0.05$ but no concern otherwise.

Nonclinical network

We found some concerns for *within-study bias* (i.e., study limitations) for all pairwise comparisons, due to the nature of the studies being unblind and most outcomes being self-reported. In terms of the *across-study bias* (i.e., reporting bias), the Egger test for funnel plot asymmetry was non-significant (p = .666) indicating that selection bias is not a big threat to the network meta-analysis. There was no concern for *indirectness*, since the included studies all matched our study questions. Evaluating *imprecision*, we found that all statistically significant comparisons revealed a clinically significant effect size. Furthermore, we examined *heterogeneity*, which is represented by the 95% prediction interval for each individual comparison. For all statistically significant comparisons there were at least some concerns regarding heterogeneity, indicating that there is a high variability of effects. Furthermore, we found no evidence for substantial and statistically significant heterogeneity in the network as a whole (within design Q = 2.27, p = .811, tau2 = 0.13; $I2$ = 66%). Finally, there was no evidence of incoherence between the direct and indirect evidence, i.e., all p-values were above 5%. For those comparisons where only indirect evidence was available incoherence was set to major concerns. Also, we identified evidence of inconsistency in the NMA when calculating the global design-by-treatment interaction test (between designs Q = 41.43, p < .001).

Comparison	Number of Studies	Within-study bias	Reporting bias	Indirectness	Imprecision	Heterogeneity	Incoherence
			Mixed evidence				
DP vs NT	7	Some concerns	Low risk	No concerns	No concerns	Major concerns	No concerns
DP vs OLP dermal	4	Some concerns	Low risk	No concerns	Some concerns	Some concerns	No concerns
DP vs OLP nasal	3	Some concerns	Low risk	No concerns	Major concerns	No concerns	No concerns
DP vs OLP pills	1	Some concerns	Low risk	No concerns	Major concerns	No concerns	No concerns
DP vs OLP-	1	Some concerns	Low risk	No concerns	No concerns	No concerns	No concerns
DP vs cOLP dermal	1	Some concerns	Low risk	No concerns	Major concerns	No concerns	No concerns
NT vs OLP dermal	3	Some concerns	Low risk	No concerns	Some concerns	Some concerns	No concerns
NT vs OLP nasal	4	Some concerns	Low risk	No concerns	No concerns	Major concerns	No concerns
NT vs OLP pills	2	Some concerns	Low risk	No concerns	Major concerns	No concerns	No concerns
NT vs OLP-	2	Some concerns	Low risk	No concerns	No concerns	Major concerns	No concerns
NT vs cOLP dermal	1	Some concerns	Low risk	No concerns	Major concerns	No concerns	No concerns
OLP dermal vs OLP-	2	Some concerns	Low risk	No concerns	No concerns	Some concerns	No concerns

		Indirect evidence					
OLP dermal vs OLP nasal	--	Some concerns ☐	Low risk	No concerns	Major concerns ☐	No concerns	Major concerns ☐
OLP dermal vs OLP pills	--	Some concerns ☐	Low risk	No concerns	Major concerns ☐	No concerns	Major concerns ☐
OLP dermal vs cOLP dermal	--	Some concerns ☐	Low risk	No concerns	Major concerns ☐	No concerns	Major concerns ☐
OLP nasal vs OLP pills	--	Some concerns ☐	Low risk	No concerns	Major concerns ☐	No concerns	Major concerns ☐
OLP nasal vs OLP-	--	Some concerns ☐	Low risk	No concerns	No concerns	Some concerns ☐	Major concerns ☐
OLP nasal vs cOLP dermal	--	Some concerns ☐	Low risk	No concerns	Major concerns ☐	No concerns	Major concerns ☐
OLP pills vs OLP-	--	Some concerns ☐	Low risk	No concerns	Some concerns ☐	Some concerns ☐	Major concerns ☐
OLP pills vs cOLP dermal	--	Some concerns ☐	Low risk	No concerns	Major concerns ☐	No concerns	Major concerns ☐
OLP- vs cOLP dermal	--	Some concerns ☐	Low risk	No concerns	No concerns	Major concerns ☐	Major concerns ☐

Clinical network

We found some concerns for *within-study bias* (i.e., study limitations) for most pairwise comparisons, due to the nature of the studies being unblind and most outcomes being self-reported. In terms of the *across-study bias* (i.e., reporting bias), the Egger test for funnel plot asymmetry was significant (p = .036) indicating that reporting bias is a threat to the network meta-analysis. There was no concern for *indirectness*, since the included studies all matched our study questions. Evaluating *imprecision*, we found that all statistically significant comparisons revealed a clinically significant effect size, except for two comparisons (cOLP suspension vs. DP, cOLP suspension vs. OLP-) where we found major concerns regarding the clinical significance of observed effects. Furthermore, we examine *heterogeneity*, which is represented by the 95% prediction interval for each individual comparison. For three statistically significant comparisons (TAU vs. cOLP pills, NT vs. cOLP pills, OLP- vs. OLP pills) there were some concerns regarding heterogeneity, indicating that there is some variability of effects. All other significant comparisons revealed no concerns. Furthermore, we found no evidence for substantial and statistically significant heterogeneity in the network as a whole (within design Q = 12.62, p = .557, tau2 = 0.024; $I2$ = 26.5%). Finally, there was evidence of incoherence between the direct and indirect evidence in three comparisons, i.e., cOLP suspension vs. OLP-, cOLP suspension vs. DP, DP vs. OLP-. For those comparisons where only indirect evidence was available incoherence was set to major concerns. Also, we identified evidence of inconsistency in the NMA when calculating the global design-by-treatment interaction test (between designs Q = 11.86, p = .018).

Comparison	Number of Studies	Within-study bias	Reporting bias	Indirectness	Imprecision	Heterogeneity	Incoherence
			Mixed evidence				
DP vs NT	1	Some concerns	Some concerns	No concerns	No concerns	No concerns	Some concerns
DP vs OLP pills	1	Some concerns	Some concerns	No concerns	Some concerns	Some concerns	No concerns
DP vs OLP-	1	Some concerns	Some concerns	No concerns	No concerns	No concerns	Major concerns
DP vs cOLP suspension	1	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
NT vs OLP pills	9	Some concerns	Some concerns	No concerns	No concerns	No concerns	No concerns
NT vs OLP-	1	Some concerns	Some concerns	No concerns	Major concerns	No concerns	No concerns
OLP injection vs Psychological intervention	1	No concerns	Some concerns	No concerns	No concerns	No concerns	Major concerns
OLP injection vs TAU	1	No concerns	Some concerns	No concerns	Some concerns	Some concerns	Major concerns
OLP injection vs Treatment programme	1	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
OLP pills vs OLP-	2	Some concerns	Some concerns	No concerns	No concerns	Some concerns	Major concerns
OLP pills vs TAU	1	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
OLP pills vs WL	8	Some concerns	Some concerns	No concerns	No concerns	No concerns	Major concerns
OLP suspension vs TAU	1	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
OLP- vs cOLP suspension	1	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
Psychological intervention vs TAU	1	No concerns	Some concerns	No concerns	No concerns	No concerns	Major concerns
TAU vs cOLP pills	2	Some concerns	Some concerns	No concerns	No concerns	Some concerns	Major concerns

		Indirect evidence					
DP vs OLP injection	--	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
DP vs OLP suspension	--	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
DP vs Psychological intervention	--	Some concerns	Some concerns	No concerns	No concerns	No concerns	Major concerns
DP vs TAU	--	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
DP vs Treatment programme	--	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
DP vs WL	--	Some concerns	Some concerns	No concerns	No concerns	No concerns	Major concerns
DP vs cOLP pills	--	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
NT vs OLP injection	--	Some concerns	Some concerns	No concerns	Some concerns	Some concerns	Major concerns
NT vs OLP suspension	--	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
NT vs Psychological intervention	--	Some concerns	Some concerns	No concerns	No concerns	No concerns	Major concerns
NT vs TAU	--	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
NT vs Treatment programme	--	Some concerns	Some concerns	No concerns	Some concerns	No concerns	Major concerns
NT vs WL	--	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
NT vs cOLP pills	--	Some concerns	Some concerns	No concerns	No concerns	Some concerns	Major concerns
NT vs cOLP suspension	--	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
OLP injection vs OLP pills	--	No concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
OLP injection vs OLP suspension	--	No concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
OLP injection vs OLP-	--	Some concerns	Some concerns	No concerns	Some concerns	Some concerns	Major concerns
OLP injection vs WL	--	Some concerns	Some concerns	No concerns	Some concerns	Some concerns	Major concerns
OLP injection vs cOLP pills	--	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns

OLP injection vs cOLP suspension	--	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
OLP pills vs OLP suspension	--	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
OLP pills vs Psychological intervention	--	No concerns	Some concerns	No concerns	No concerns	No concerns	Major concerns
OLP pills vs Treatment programme	--	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
OLP pills vs cOLP pills	--	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
OLP pills vs cOLP suspension	--	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
OLP suspension vs OLP-	--	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
OLP suspension vs Psychological intervention	--	No concerns	Some concerns	No concerns	No concerns	No concerns	Major concerns
OLP suspension vs Treatment programme	--	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
OLP suspension vs WL	--	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
OLP suspension vs cOLP pills	--	Some concerns	Some concerns	No concerns	Some concerns	Some concerns	Major concerns

OLP suspension vs cOLP suspension	--	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
OLP- vs Psychological intervention	--	Some concerns	Some concerns	No concerns	No concerns	No concerns	Major concerns
OLP- vs TAU	--	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
OLP- vs Treatment programme	--	Some concerns	Some concerns	No concerns	Some concerns	Some concerns	Major concerns
OLP- vs WL	--	Some concerns	Some concerns	No concerns	Major concerns	No concerns	Major concerns
OLP- vs cOLP pills	--	Some concerns	Some concerns	No concerns	Some concerns	No concerns	Major concerns
Psychological intervention vs Treatment programme	--	No concerns	Some concerns	No concerns	No concerns	No concerns	Major concerns
Psychological intervention vs WL	--	Some concerns	Some concerns	No concerns	No concerns	No concerns	Major concerns
Psychological intervention vs cOLP pills	--	Some concerns	Some concerns	No concerns	No concerns	No concerns	Major concerns
Psychological intervention vs cOLP suspension	--	Some concerns	Some concerns	No concerns	No concerns	No concerns	Major concerns
TAU vs Treatment programme	--	No concerns	Some concerns	No concerns	Some concerns	Some concerns	Major concerns

TAU vs WL	--	Some concerns ☐	Some concerns ☐	No concerns	Major concerns ☐	No concerns	Major concerns ☐
TAU vs cOLP suspension	--	Some concerns ☐	Some concerns ☐	No concerns	Major concerns ☐	No concerns	Major concerns ☐
Treatment programme vs WL	--	Some concerns ☐	Some concerns ☐	No concerns	Some concerns ☐	Some concerns ☐	Major concerns ☐
Treatment programme vs cOLP pills	--	Some concerns ☐	Some concerns ☐	No concerns	Major concerns ☐	No concerns	Major concerns ☐
Treatment programme vs cOLP suspension	--	Some concerns ☐	Some concerns ☐	No concerns	Major concerns ☐	No concerns	Major concerns ☐
WL vs cOLP pills	--	Some concerns ☐	Some concerns ☐	No concerns	Some concerns ☐	No concerns	Major concerns ☐
WL vs cOLP suspension	--	Some concerns ☐	Some concerns ☐	No concerns	Major concerns ☐	No concerns	Major concerns ☐
cOLP pills vs cOLP suspension	--	Some concerns ☐	Some concerns ☐	No concerns	Major concerns ☐	No concerns	Major concerns ☐

eAppendix 4. Details on inconsistency

Nonclinical network – local approach

```
Separate indirect from direct evidence (SIDE) using back-calculation method

Random effects model:

            comparison k prop   nma direct indir.  Diff     z p-value
        cOLP dermal:DP 1 0.78 -0.03  -0.09   0.20 -0.29 -0.29  0.7720
        cOLP dermal:NT 1 0.78  0.47   0.54   0.25  0.29  0.29  0.7720
 cOLP dermal:OLP dermal 0    0  0.21      .   0.21     .     .       .
  cOLP dermal:OLP nasal 0    0  0.04      .   0.04     .     .       .
  cOLP dermal:OLP pills 0    0  0.38      .   0.38     .     .       .
       cOLP dermal:OLP- 0    0  1.07      .   1.07     .     .       .
                 DP:NT 7 0.85  0.50   0.47   0.66 -0.19 -0.42  0.6773
          DP:OLP dermal 4 0.77  0.24   0.10   0.70 -0.61 -1.28  0.2011
           DP:OLP nasal 3 0.74  0.07   0.21  -0.33  0.55  1.09  0.2751
           DP:OLP pills 1 0.34  0.40  -0.01   0.62 -0.63 -0.79  0.4306
                DP:OLP- 1 0.43  1.10   1.44   0.84  0.60  1.02  0.3078
          OLP dermal:NT 3 0.70  0.26   0.20   0.41 -0.21 -0.47  0.6404
           OLP nasal:NT 4 0.87  0.43   0.50  -0.03  0.53  0.86  0.3890
           OLP pills:NT 2 0.91  0.10  -0.00   1.09 -1.09 -0.89  0.3747
                OLP-:NT 2 0.83 -0.60  -0.70  -0.12 -0.57 -0.77  0.4386
   OLP dermal:OLP nasal 0    0 -0.17      .  -0.17     .     .       .
   OLP dermal:OLP pills 0    0  0.17      .   0.17     .     .       .
        OLP dermal:OLP- 2 0.85  0.86   0.86   0.85  0.01  0.01  0.9882
    OLP nasal:OLP pills 0    0  0.33      .   0.33     .     .       .
         OLP nasal:OLP- 0    0  1.03      .   1.03     .     .       .
         OLP pills:OLP- 0    0  0.69      .   0.69     .     .       .

Legend:
 comparison - Treatment comparison
 k          - Number of studies providing direct evidence
 prop       - Direct evidence proportion
 nma        - Estimated treatment effect (SMD) in network meta-analysis
 direct     - Estimated treatment effect (SMD) derived from direct evidence
 indir.     - Estimated treatment effect (SMD) derived from indirect evidence
 Diff       - Difference between direct and indirect treatment estimates
 z          - z-value of test for disagreement (direct versus indirect)
 p-value    - p-value of test for disagreement (direct versus indirect)
```

Nonclinical network – global approach

```
Q statistics to assess homogeneity / consistency

                    Q df  p-value
Total           43.69 15   0.0001
Within designs   2.27  5   0.8112
Between designs 41.43 10 < 0.0001

Design-specific decomposition of within-designs Q statistic

          Design    Q df p-value
   DP:OLP dermal 0.12  1  0.7324
 NT:DP:OLP nasal 2.15  4  0.7083

Between-designs Q statistic after detaching of single designs

       Detached design     Q df  p-value
         DP:OLP dermal 40.30  9 < 0.0001
          NT:OLP nasal 37.82  9 < 0.0001
          NT:OLP pills 41.14  9 < 0.0001
      NT:DP:OLP dermal 40.17  8 < 0.0001
 NT:DP:OLP dermal:OLP-  6.98  7   0.4306
       NT:DP:OLP nasal 35.92  8 < 0.0001
       NT:DP:OLP pills 40.19  8 < 0.0001
    NT:OLP dermal:OLP- 14.51  8   0.0693

Q statistic to assess consistency under the assumption of
a full design-by-treatment interaction random effects model

                    Q df  p-value tau.within tau2.within
Between designs 41.43 10 < 0.0001          0           0
```

Clinical network – local approach

```
Separate indirect from direct evidence (SIDE) using back-calculation method

Random effects model:

                                  comparison k prop   nma direct indir.  Diff     z p-value
                    cOLP pills:cOLP suspension 0    0  0.65      .   0.65     .     .       .
                                cOLP pills:DP 0    0  0.12      .   0.12     .     .       .
                                cOLP pills:NT 0    0  0.89      .   0.89     .     .       .
                          cOLP pills:OLP injection 0    0  0.19      .   0.19     .     .       .
                              cOLP pills:OLP pills 0    0  0.42      .   0.42     .     .       .
                         cOLP pills:OLP suspension 0    0  0.57      .   0.57     .     .       .
                                cOLP pills:OLP- 0    0  0.92      .   0.92     .     .       .
          cOLP pills:Psychological intervention 0    0 -1.07      .  -1.07     .     .       .
                               cOLP pills:TAU 2 1.00  0.58   0.58      .     .     .       .
                 cOLP pills:Treatment programme 0    0  0.01      .   0.01     .     .       .
                                cOLP pills:WL 0    0  0.86      .   0.86     .     .       .
                            cOLP suspension:DP 1 0.89 -0.53  -0.93   2.58 -3.51 -2.98  0.0029
                            cOLP suspension:NT 0    0  0.23      .   0.23     .     .       .
                  cOLP suspension:OLP injection 0    0 -0.47      .  -0.47     .     .       .
                      cOLP suspension:OLP pills 0    0 -0.23      .  -0.23     .     .       .
                 cOLP suspension:OLP suspension 0    0 -0.08      .  -0.08     .     .       .
                           cOLP suspension:OLP- 1 0.78  0.26   0.89  -1.91  2.80  2.98  0.0029
      cOLP suspension:Psychological intervention 0    0 -1.72      .  -1.72     .     .       .
                             cOLP suspension:TAU 0    0 -0.07      .  -0.07     .     .       .
            cOLP suspension:Treatment programme 0    0 -0.65      .  -0.65     .     .       .
                             cOLP suspension:WL 0    0  0.20      .   0.20     .     .       .
                                       DP:NT 1 0.69  0.76   0.52   1.31 -0.79 -1.92  0.0546
                                 DP:OLP injection 0    0  0.06      .   0.06     .     .       .
                                     DP:OLP pills 1 0.68  0.30   0.11   0.72 -0.61 -1.51  0.1298
                                DP:OLP suspension 0    0  0.45      .   0.45     .     .       .
                                      DP:OLP- 1 0.35  0.79   1.82   0.24  1.58  2.98  0.0029
                    DP:Psychological intervention 0    0 -1.19      .  -1.19     .     .       .
                                       DP:TAU 0    0  0.46      .   0.46     .     .       .
                           DP:Treatment programme 0    0 -0.11      .  -0.11     .     .       .
                                       DP:WL 0    0  0.73      .   0.73     .     .       .
                                OLP injection:NT 0    0  0.70      .   0.70     .     .       .
                                    OLP pills:NT 9 0.99  0.46   0.47  -0.39  0.86  1.09  0.2749
                               OLP suspension:NT 0    0  0.32      .   0.32     .     .       .

                                      OLP-:NT 1 0.33 -0.03   0.42  -0.26  0.68  1.42  0.1564
                    Psychological intervention:NT 0    0  1.96      .   1.96     .     .       .
                                       TAU:NT 0    0  0.30      .   0.30     .     .       .
                           Treatment programme:NT 0    0  0.88      .   0.88     .     .       .
                                       WL:NT 0    0  0.03      .   0.03     .     .       .
                             OLP injection:OLP pills 0    0  0.24      .   0.24     .     .       .
                        OLP injection:OLP suspension 0    0  0.38      .   0.38     .     .       .
                               OLP injection:OLP- 0    0  0.73      .   0.73     .     .       .
       OLP injection:Psychological intervention 1 1.00 -1.26  -1.26      .     .     .       .
                             OLP injection:TAU 1 1.00  0.39   0.39      .     .     .       .
              OLP injection:Treatment programme 1 1.00 -0.18  -0.18      .     .     .       .
                             OLP injection:WL 0    0  0.67      .   0.67     .     .       .
                         OLP pills:OLP suspension 0    0  0.15      .   0.15     .     .       .
                               OLP pills:OLP- 2 0.69  0.49   0.23   1.07 -0.84 -1.77  0.0770
            OLP pills:Psychological intervention 0    0 -1.49      .  -1.49     .     .       .
                               OLP pills:TAU 1 1.00  0.16   0.16      .     .     .       .
                    OLP pills:Treatment programme 0    0 -0.41      .  -0.41     .     .       .
                               OLP pills:WL 8 1.00  0.43   0.43      .     .     .       .
                           OLP suspension:OLP- 0    0  0.35      .   0.35     .     .       .
      OLP suspension:Psychological intervention 0    0 -1.64      .  -1.64     .     .       .
                             OLP suspension:TAU 1 1.00  0.01   0.01      .     .     .       .
             OLP suspension:Treatment programme 0    0 -0.56      .  -0.56     .     .       .
                             OLP suspension:WL 0    0  0.28      .   0.28     .     .       .
                  OLP-:Psychological intervention 0    0 -1.99      .  -1.99     .     .       .
                                    OLP-:TAU 0    0 -0.34      .  -0.34     .     .       .
                          OLP-:Treatment programme 0    0 -0.91      .  -0.91     .     .       .
                                      OLP-:WL 0    0 -0.06      .  -0.06     .     .       .
                   Psychological intervention:TAU 1 1.00  1.65   1.65      .     .     .       .
 Psychological intervention:Treatment programme 0    0  1.08      .   1.08     .     .       .
                    Psychological intervention:WL 0    0  1.92      .   1.92     .     .       .
                          TAU:Treatment programme 0    0 -0.57      .  -0.57     .     .       .
                                       TAU:WL 0    0  0.27      .   0.27     .     .       .
                           Treatment programme:WL 0    0  0.85      .   0.85     .     .       .

Legend:
 comparison - Treatment comparison
 k          - Number of studies providing direct evidence
 prop       - Direct evidence proportion
 nma        - Estimated treatment effect (SMD) in network meta-analysis
 direct     - Estimated treatment effect (SMD) derived from direct evidence
 indir.     - Estimated treatment effect (SMD) derived from indirect evidence
 Diff       - Difference between direct and indirect treatment estimates
 z          - z-value of test for disagreement (direct versus indirect)
 p-value    - p-value of test for disagreement (direct versus indirect)
```

Clinical network – global approach

```
Q statistics to assess homogeneity / consistency

                    Q df p-value
Total           24.48 18  0.1400
Within designs  12.62 14  0.5569
Between designs 11.86  4  0.0184

Design-specific decomposition of within-designs Q statistic

         Design    Q df p-value
 cOLP pills:TAU 0.27  1  0.6041
   NT:OLP pills 4.40  6  0.6224
   OLP pills:WL 7.94  7  0.3375

Between-designs Q statistic after detaching of single designs

   Detached design     Q df p-value
      NT:OLP pills 11.86  3  0.0079
    OLP pills:OLP-  9.45  3  0.0239
   NT:DP:OLP pills  0.96  2  0.6199
 NT:OLP pills:OLP-  9.30  2  0.0096

Q statistic to assess consistency under the assumption of
a full design-by-treatment interaction random effects model

                    Q df p-value tau.within tau2.within
Between designs 11.86  4  0.0184          0           0
```

eAppendix 5. PRISMA checklist

Section/Topic	Item #	Checklist Item	Reported on Page #
TITLE			
Title	1	Identify the report as a systematic review *incorporating a network meta-analysis (or related form of meta-analysis).*	p.1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: **Background:** main objectives **Methods:** data sources; study eligibility criteria, participants, and interventions; study appraisal; and *synthesis methods, such as network meta-analysis.* **Results:** number of studies and participants identified; summary estimates with corresponding confidence/credible intervals; *treatment rankings may also be discussed. Authors may choose to summarize pairwise comparisons against a chosen treatment included in their analyses for brevity.* **Discussion/Conclusions:** limitations; conclusions and implications of findings. **Other:** primary source of funding; systematic review registration number with registry name.	p.2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known*, including mention of why a network meta-analysis has been conducted.*	p.3-4
Objectives	4	Provide an explicit statement of questions being addressed, with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	p.4
METHODS			
Protocol and registration	5	Indicate whether a review protocol exists and if and where it can be accessed (e.g., Web address); and, if available, provide registration information, including registration number.	p.6
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale. *Clearly describe eligible treatments included in the treatment network, and note whether any have been clustered or merged into the same node (with justification).*	p.5-6

Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	p.5
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	p.5 eAppendix 1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	p.5-6
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	p.6-7
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	p.5-7
Geometry of the network	**S1**	Describe methods used to explore the geometry of the treatment network under study and potential biases related to it. This should include how the evidence base has been graphically summarized for presentation, and what characteristics were compiled and used to describe the evidence base to readers.	p.7-9
Risk of bias within individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	p.7-8 eAppendix 3
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means). *Also describe the use of additional summary measures assessed, such as treatment rankings and surface under the cumulative ranking curve (SUCRA) values, as well as modified approaches used to present summary findings from meta-analyses.*	p.8-9
Planned methods of analysis	14	Describe the methods of handling data and combining results of studies for each network meta-analysis. This should include, but not be limited to: ● *Handling of multi-arm trials;* ● *Selection of variance structure;* ● *Selection of prior distributions in Bayesian analyses; and* ● *Assessment of model fit.*	p.8-9
Assessment of Inconsistency	**S2**	Describe the statistical methods used to evaluate the agreement of direct and indirect evidence in the treatment network(s) studied. Describe efforts taken to address its presence when found.	p.8-9 eAppendix 4
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	p.9 eAppendix 3
Additional analyses	16	Describe methods of additional analyses if done, indicating which were pre-specified. This may include, but not be limited to, the following:	p.9 eAppendix 6

		● Sensitivity or subgroup analyses; ● Meta-regression analyses; ● *Alternative formulations of the treatment network; and* ● *Use of alternative prior distributions for Bayesian analyses (if applicable).*	
RESULTS†			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	p.10 eFigure 1
Presentation of network structure	**S3**	Provide a network graph of the included studies to enable visualization of the geometry of the treatment network.	Figures 1A, 1B
Summary of network geometry	**S4**	Provide a brief overview of characteristics of the treatment network. This may include commentary on the abundance of trials and randomized patients for the different interventions and pairwise comparisons in the network, gaps of evidence in the treatment network, and potential biases reflected by the network structure.	eTable 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	p.10-11 eTable 1
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment.	eTable 1 eAppendix 3
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: 1) simple summary data for each intervention group, and 2) effect estimates and confidence intervals. *Modified approaches may be needed to deal with information from larger networks.*	Figures 2A, 2B eTable 2, 3
Synthesis of results	21	Present results of each meta-analysis done, including confidence/credible intervals. *In larger networks, authors may focus on comparisons versus a particular comparator (e.g. placebo or standard care), with full findings presented in an appendix. League tables and forest plots may be considered to summarize pairwise comparisons.* If additional summary measures were explored (such as treatment rankings), these should also be presented.	p.10-11 Figures 2A, 2B eTable 2, 3
Exploration for inconsistency	**S5**	Describe results from investigations of inconsistency. This may include such information as measures of model fit to compare consistency and inconsistency models, *P* values from statistical tests, or summary of inconsistency estimates from different parts of the treatment network.	eAppendix 4
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies for the evidence base being studied.	eTable 1 eAppendix 3

Results of additional analyses	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression analyses, *alternative network geometries studied, alternative choice of prior distributions for Bayesian analyses,* and so forth).	eAppendix 6
DISCUSSION			
Summary of evidence	24	Summarize the main findings, including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy-makers).	p.12-14
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review level (e.g., incomplete retrieval of identified research, reporting bias). *Comment on the validity of the assumptions, such as transitivity and consistency. Comment on any concerns regarding network geometry (e.g., avoidance of certain comparisons).*	p.14-15
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	p. 16
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review. This should also include information regarding whether funding has been received from manufacturers of treatments in the network and/or whether some of the authors are content experts with professional conflicts of interest that could affect use of treatments in the network.	p.17

PICOS = population, intervention, comparators, outcomes, study design.

* Text in italics indicate wording specific to reporting of network meta-analyses that has been added to guidance from the PRISMA statement.

† Authors may wish to plan for use of appendices to present all relevant information in full detail for items in this section.

eAppendix 6. Additional Results

Adverse events

Regarding adverse events, it is remarkable that few studies reported adverse events systematically or at all. In total, 15 of the 37 studies made a statement regarding adverse events. From these reports, it is apparent that relatively few adverse events occur in the context of OLP treatment. This suggests that OLP is a safe and mostly side effect free treatment. However, due to inconsistent or unreported adverse events, it is difficult to draw a conclusion.

Certainty of the evidence

The certainty of evidence for the network estimates of both samples was examined by using GRADE. The results for study limitations (within study bias), reporting bias (across-studies bias), indirectness, imprecision, heterogeneity, and incoherence can be found in the supplement (eAppendix 3-4, eFigure 2).

Sensitivity analysis

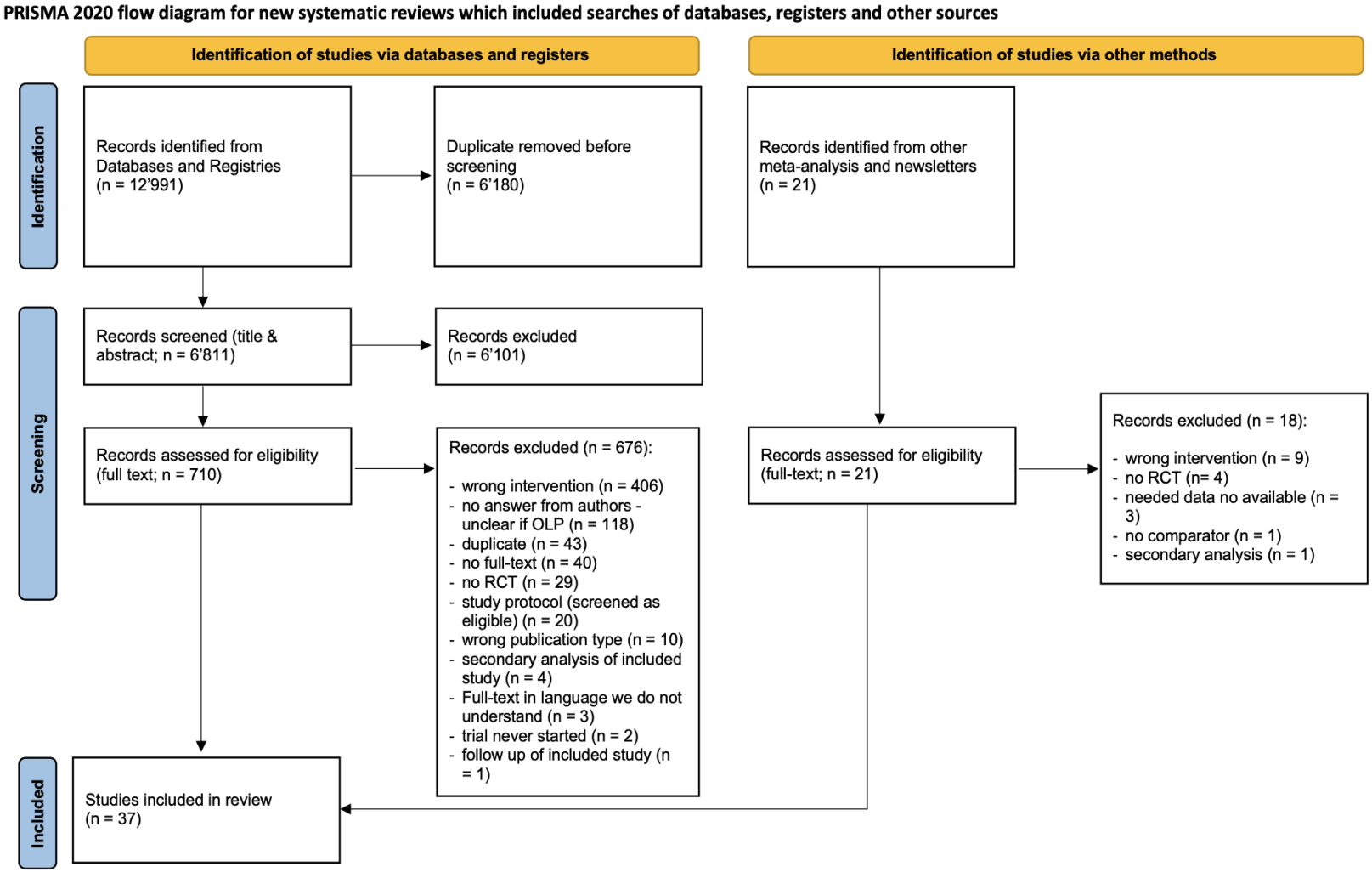
To investigate the impact of high risk studies, we conducted the analyses including only studies in which the risk of bias was low or moderate. In each sample, one study was high risk of bias and thus excluded and compared to the whole sample. The results in the nonclinical network remained unchanged in principal, solely OLP nasal changed from being marginally significant to insignificant. In the nonclinical sample, cOLP pills moved from being significant to non significant, as only one study with a cOLP pills group remained in the network. Otherwise results and heterogeneity measures remained comparable.

To investigate the impact of including studies with subclinical populations within the clinical sample, we conducted a sensitivity analysis by excluding studies with subclinical samples. In principle, the results remained unchanged with a trend for slightly bigger effect sizes when subclinical studies were excluded (see eFigure 3-6 in the supplement for the results of sensitivity analyses). Surprisingly, heterogeneity increased from I^2 = 26.5% (clinical all) to I^2 = 32.6% (clinical without subclinical).

Furthermore, owing to the great variance of included conditions within each of the two networks, we performed subgroup analysis for two broad areas: pain (i.e., chronic back pain, experimental pain, irritable bowel syndrome, knee ostheoarthritis) and psychological (i.e., depression, fatigue, conditions, well-being, insomnia, test anxiety, sadness, relaxation, stress). The results for the clinical pain network (11 studies) showed comparable results to the ones of the whole network, except the treatment programme changed to being significantly better than NT, whereas OLP- moved to being significantly worse than NT. Interestingly, heterogeneity was reduced from I^2 = 26.5% (clinical all) to I^2 = 0% (clinical pain). Within the nonclinical pain sample (N = 4), results did also change only marginally, with OLP nasal not being significantly better than NT anymore. Heterogeneity as well decreased from I^2 = 66% (nonclinical all) to I^2 = 51.7% (nonclinical pain). Within the psychological subsamples results could in general also be replicated (clinical psychological = 10 and nonclinical psychological = 3 studies), with the exception of DP being bigger

in the nonclinical sample and the effect size of OLP- changing from -0.03 to 0.30 in the clinical network. Heterogeneity decreased within the clinical sample from I^2 = 26.5% (clinical all) to I^2 = 0% (clinical psychological) and in the nonclinical network from I^2 = 66% (nonclinical all) to I^2 = 0% (nonclinical psychological). Overall, very few studies were included in the networks of these subgroup-analyses.

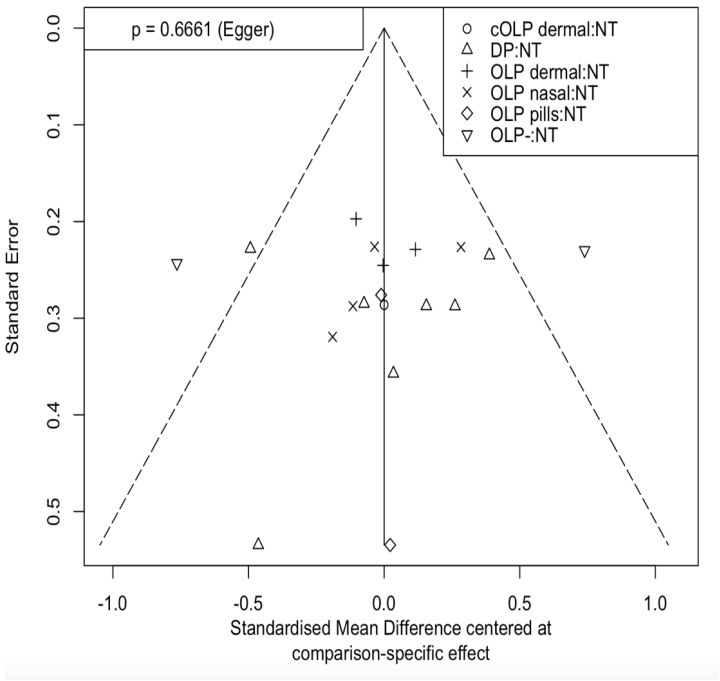
eFigure 1. Flowchart



From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71. For more information, visit: http://www.prisma-statement.org/

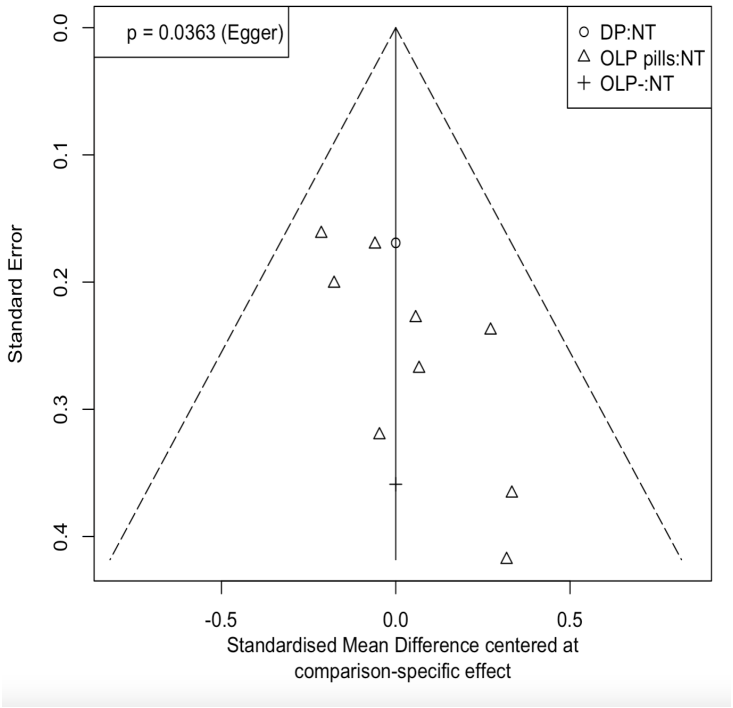
eFigure 2. Funnel plots with accompanying Egger test

Nonclinical network

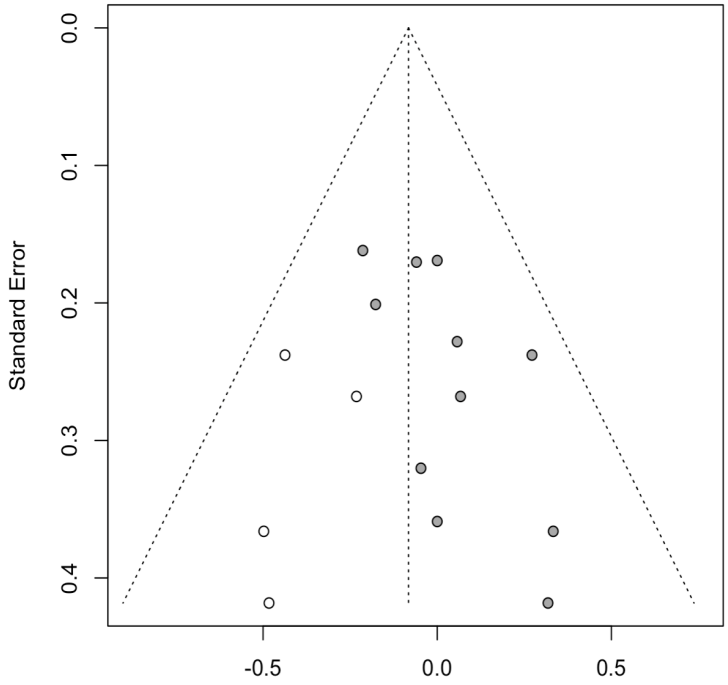


Note: Funnel plot with reference NT = , i.e. this plot only includes studies with NT as a control group depicting available comparisons with DP = deceptive placebo, OLP = open-label placebo, and cOLP = conditioned open-label placebo.

Clinical network



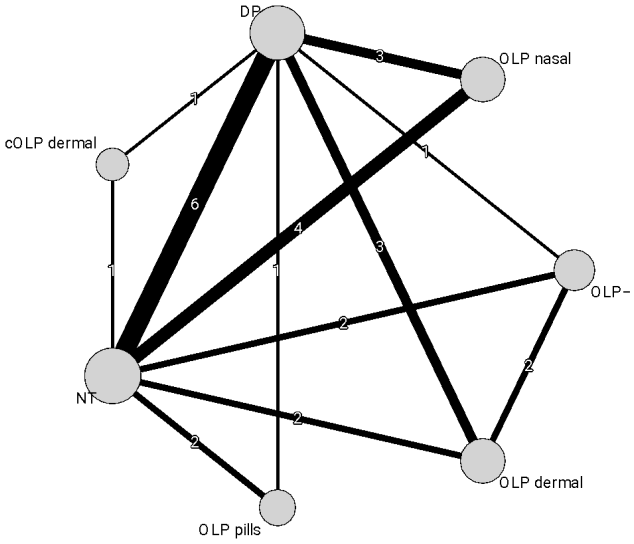
Note: Funnel plot with reference NT = no treatment, i.e. this plot only includes studies with NT as a control group depicting available comparisons with DP = deceptive placebo and OLP open-label placebo.



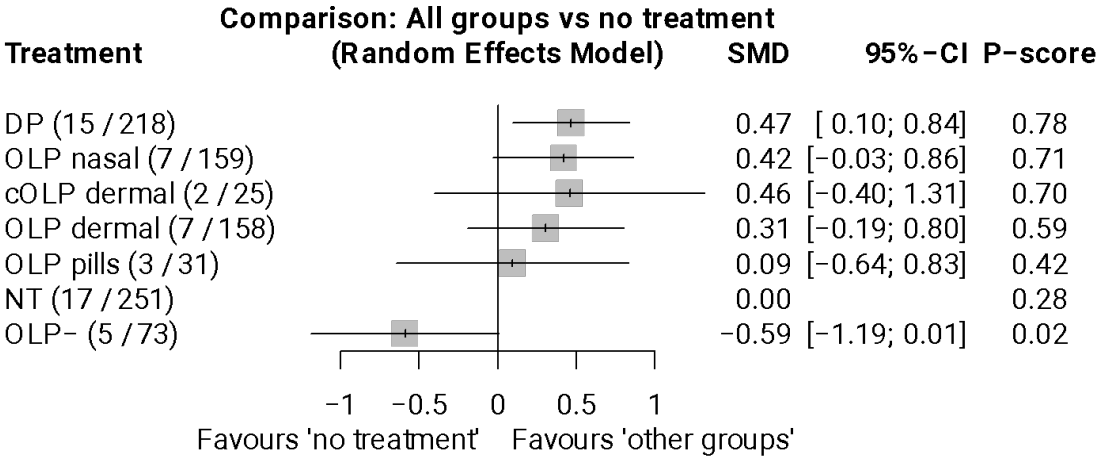
Note: Funnel plot with reference NT = no treatment after using the trim and fill method. The plot depicts the four comparisons (white dots) NT vs. OLP pills that are missing in order for the funnel to be symmetric.

eFigure 3. Plots of low and moderate risk of bias only (sensitivity analysis)

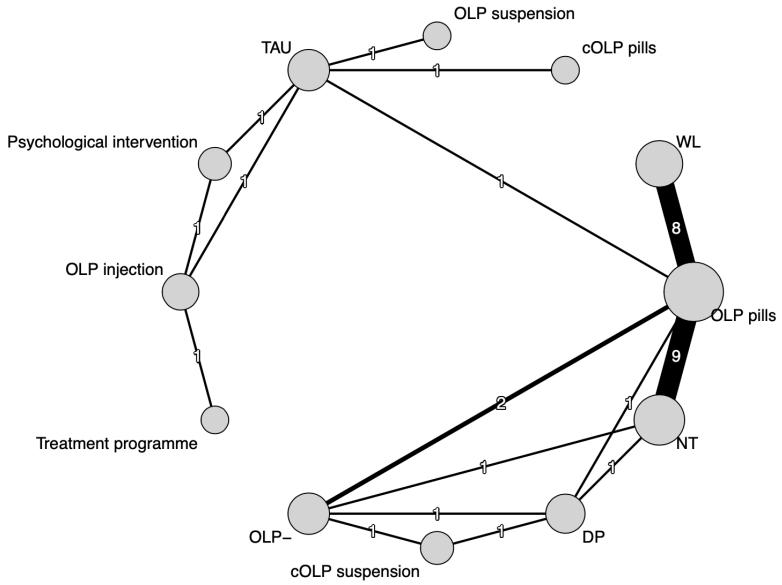
Netgraph of nonclinical network meta-analysis on low and moderate risk of bias only



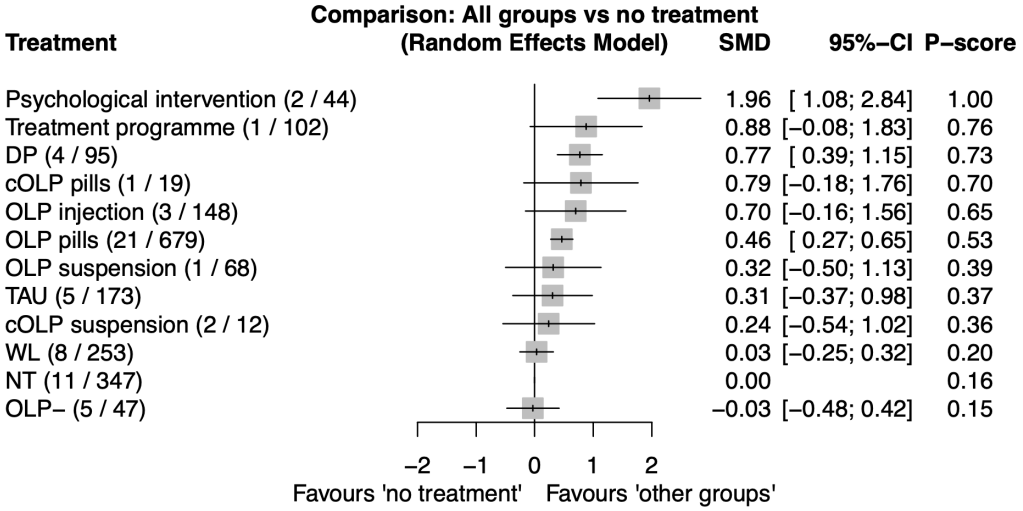
Forest plot of nonclinical network meta-analysis on low and moderate risk of bias only



Netgraph of clinical network meta-analysis on low and moderate risk of bias only

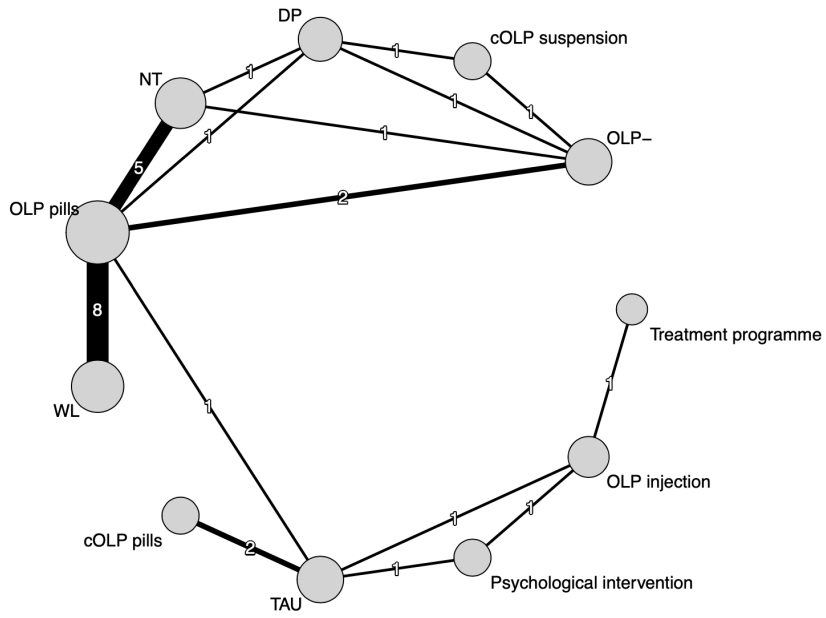


Forest plot of clinical network meta-analysis on low and moderate risk of bias only

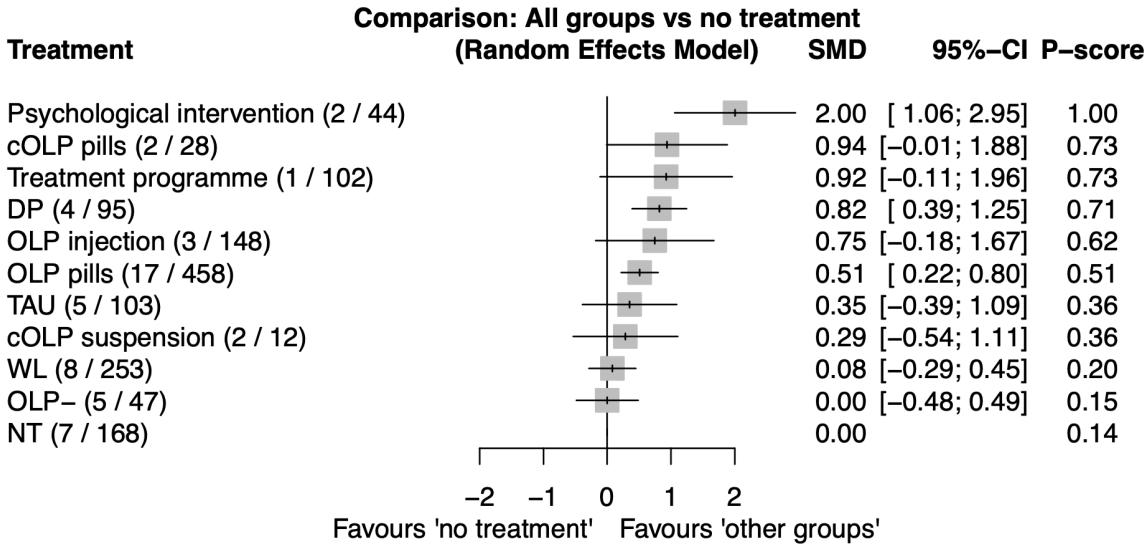


eFigure 4. Plots of clinical network without subclinical trials (sensitivity analysis)

Netgraph of network meta-analysis on clinical studies only

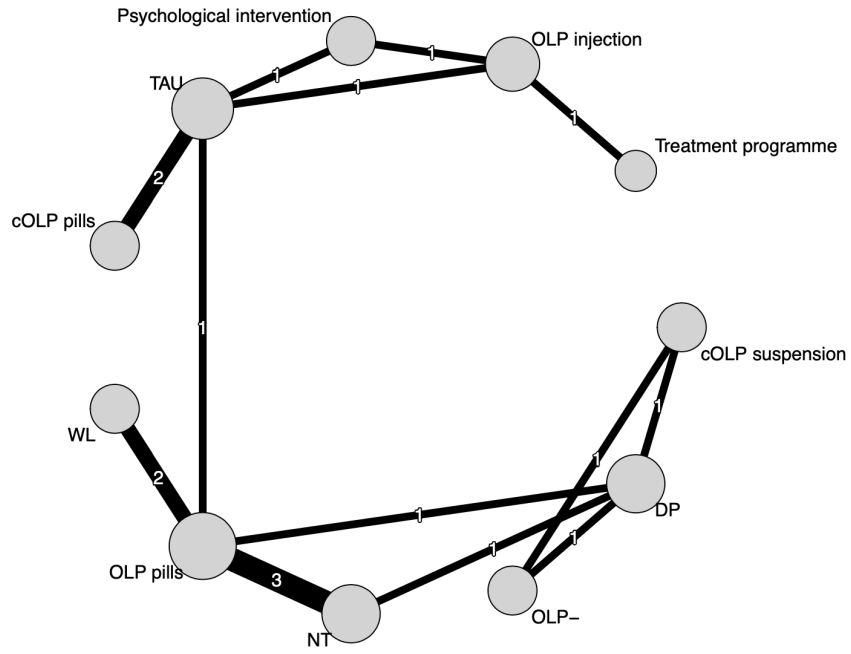


Forest plot of network-meta-analysis on clinical studies only

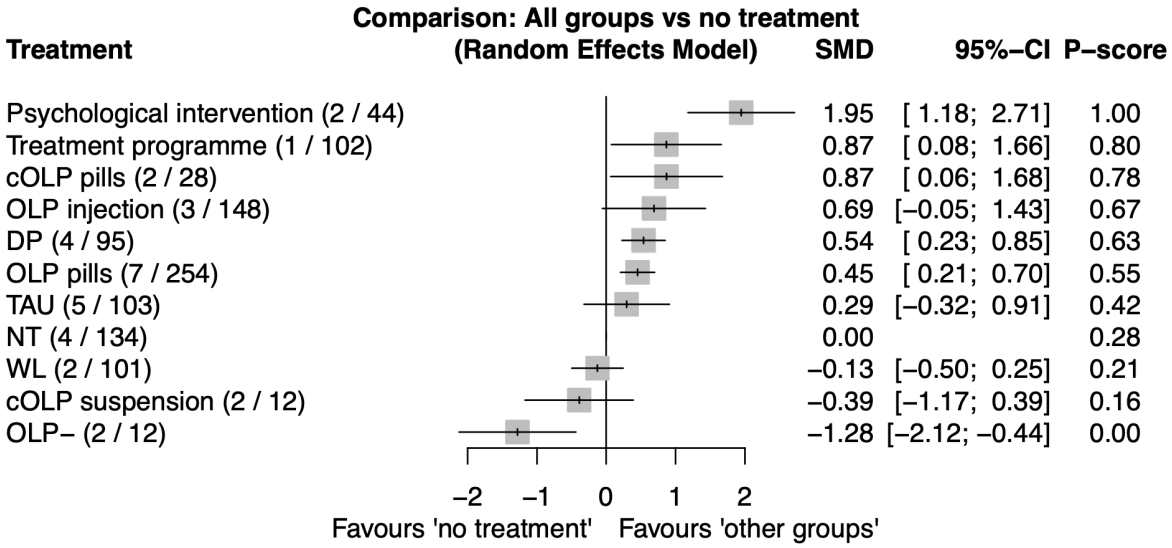


eFigure 5. Plots of pain trials (sensitivity analysis)

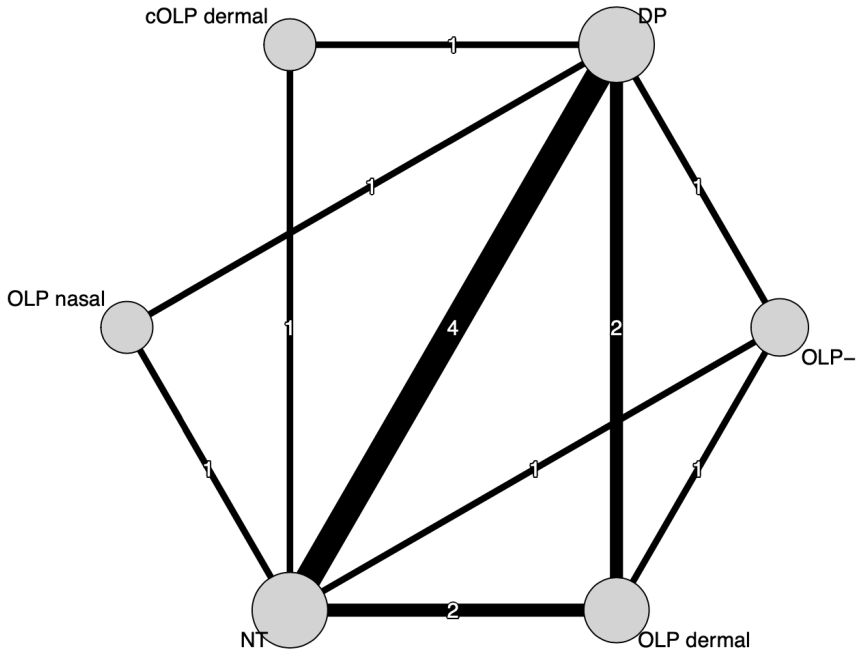
Netgraph of network meta-analysis on clinical pain studies only



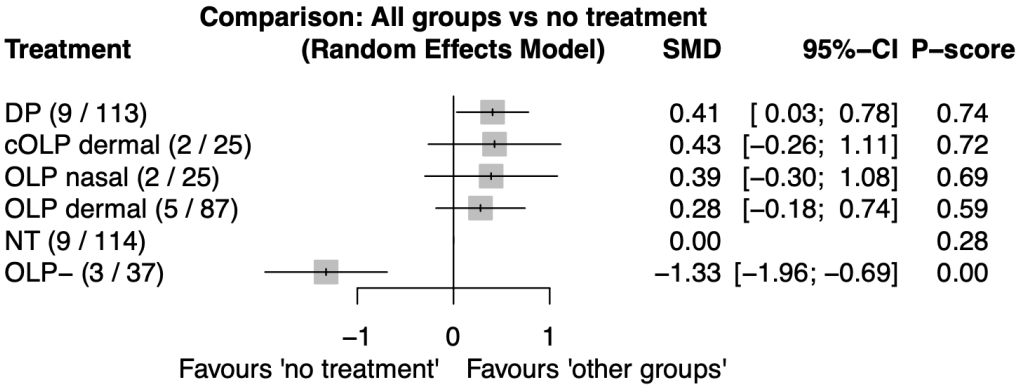
Forest plot of network meta-analysis on clinical pain studies only



Netgraph of network meta-analysis on nonclinical pain studies only

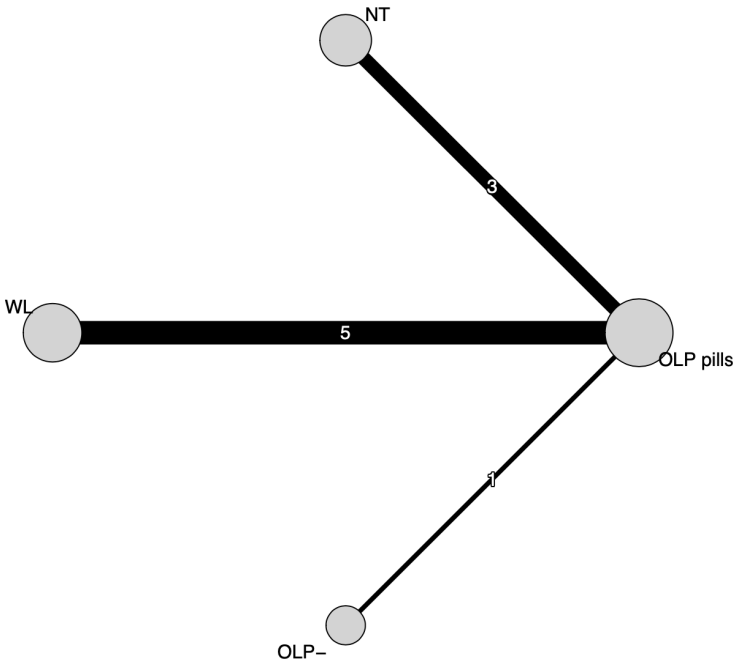


Forest plot of network meta-analysis on nonclinical pain studies only

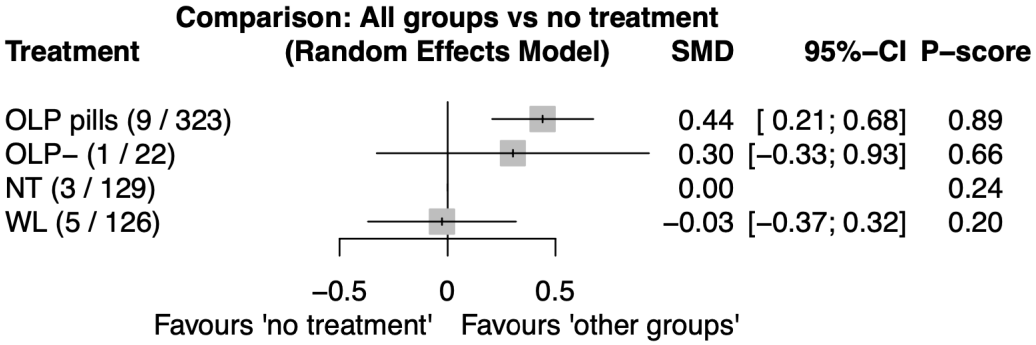


eFigure 6. Plots of psychological trials (sensitivity analysis)

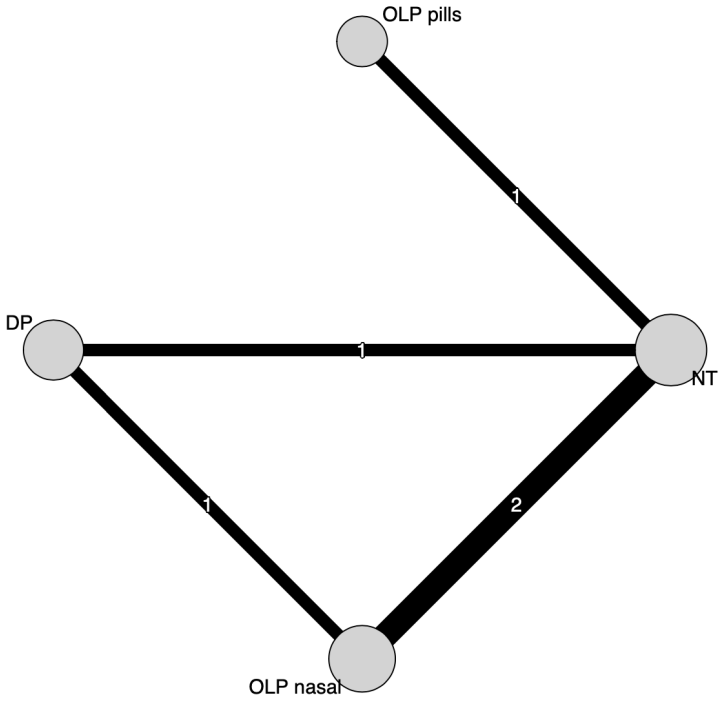
Netgraph of network meta-analysis on clinical psychological studies only



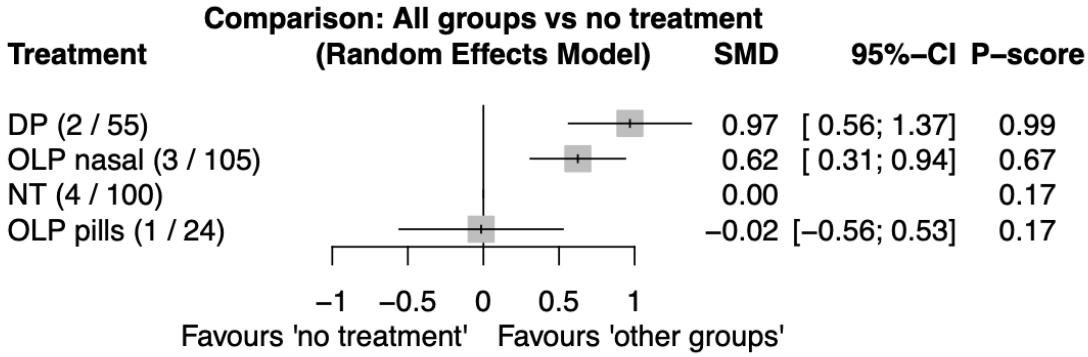
Forest plot of network meta-analysis on clinical psychological studies only



Netgraph of network meta-analysis on nonclinical psychological studies only



Forest plot of network meta-analysis on nonclinical psychological studies only



eTable 1. Demographics and study characteristic

Author, Year	Country	Condition/Diagnosis	Sample used in analysis	N (% female) per group	Mean age in years (SD) per group	Treatment duration in days	Intervention 1	Intervention 2	Intervention 3	Control	Outcome used for analysis	Rationale for choice of outcome	Risk of Bias
Ashar, 2021	USA	chronic low back pain	clinical	135 (53.67)	41.10 (15.67)	28	OLP injection (injection)	psychological intervention		TAU	pain intensity (NRS 0-10)	only PO	low
Bandak, 2022	Denmark	knee osteoarthritis	clinical	206 (45.65)	68.40 (8.25)	56	OLP injection (injection)			Treatment programme (exercise and education)	pain subscale (KOOS 0-100)	only PO	some concerns
Barnes, 2019	Australia	experimental nausea	nonclinical	61 (52.74)	21.50 (4.65)	2	OLP nasal (vapor) (semi + fully open)	DP (vapor)		NT	nausea (VAS 0-10)	only PO	some concerns
Carvalho, 2016	Portugal	chronic low back pain	clinical	83 (71.05)	44.25 (13.45)	21	OLP pills (pill)			WL	pain intensity (NRS 0-10)	most frequent	some concerns

Disley, 2021	UK	experimental pain	nonclinical	75 (86.67)	21.05 (5.04)	1	OLP nasal (spray)	DP (spray)		NT	pain intensity (VAS 0-100)	most frequent	some concerns
El Brihi, 2019	Australia	well-being	subclinical	88 (80.00)	19.00 (3.90)	7	OLP pills (capsule) (different doses merged)			NT	emotional distress (DASS)	most frequent	some concerns
Flowers, 2021	USA	acute pain (following spine surgery)	clinical	41 (NA)	60.15 (13.05)	17	cOLP pills (pill)			TAU	worst daily pain (mini-BP; NRS 0-10)	most frequent	some concerns
Friehs, 2022	Germany	experimental sadness	nonclinical	147 (70.26)	23.56 (4.25)	7	OLP nasal (spray sesame oil) (personal + scientific)	DP (spray)		NT (personal + scientific)	sadness subscale (PANAS-X)	only PO	some concerns

Haas, 2022	Germany	primary insomnia	clinical	45 (84.39)	30.07 (NA)	2	OLP pills (pill)			OLP-(pill)	subjective total sleep time in minutes	only PO	some concerns
Hahn, 2022	Germany	experimental sadness	nonclinical	84 (100.00)	24.74 (5.15)	1	OLP nasal (spray)			NT	sadness subscale (PANAS-X)	only PO	some concerns
Hoenemeyer, 2018	USA	cancer-related fatigue	clinical	73 (69.00)	57.20 (11.80)	21	OLP pills (pill)			WL	cancer related fatigue (FSI-14)	most frequent	some concerns
Ikemoto, 2020	Japan	chronic low back pain	clinical	48 (61.55)	66.75 (66.75)	84	OLP pills (pill)			TAU	pain intensity (NRS 0-10)	most frequent	some concerns
Kaptchuk, 2010	Israel	irritable bowel syndrome	clinical	80 (69.50)	46.50 (18.00)	21	OLP pills (pill)			NT	IBS symptom severity scale (IBS-SSS)	baseline available	some concerns
Kelley, 2012	USA	major depressive disorder	clinical	20 (70.00)	38.80 (12.60)	14	OLP pills (capsule)			WL	depression severity (HAM-D-17)	only PO	some concerns

Kleine-Borgmann, 2021	Germany	chronic low back pain	clinical	122 (NA)	59.33 (14.56)	21	OLP pills (capsule)			WL	pain intensity (NRS 0-10)	only PO	some concerns
Kleine-Borgmann, 2019	Germany	well-being & cognitive enhancement	subclinical	154 (67.50)	24.03 (2.79)	21	OLP pills (pill)			NT	stress (PSQ-20)	baseline available; most informative	some concerns
Klinger, 2017	Germany	chronic low back pain + experimental pain	clinical	48 (75.00)	50.89 (15.07)	1	cOLP suspension (saline cotton swab)	OLP- (saline cotton swab)		DP (conditioned + unconditioned), (saline cotton swab)	pain intensity (NRS 0-10)	only PO	some concerns
Kube, 2020	Germany	experimental pain	nonclinical	100 (49.50)	24.56 (5.66)	1	OLP dermal (cream) (expectancy + hope)	DP (cream)		NT	pain intensity (VAS 0-100)	only PO	high
Kube, 2021	Germany	allergic rhinitis	clinical	54 (68.68)	31.48 (12.67)	14	OLP pills (tablet) (augmented + limited)			WL (augmented + limited)	self-reported allergic symptoms (CSMS)	only PO	some concerns

Leibowitz, 2019	USA	experimental itch	nonclinical	NA (63.50)	24.55 (NA)	NA	OLP dermal (cream) (expectation + rationale)	OLP-(cream)		NT	physiological allergic reaction (size of the wheal)	only PO	some concerns
Lembo, 2021	USA	irritable bowel syndrome	clinical	211 (72.93)	42.00 (18.00)	42	OLP pills (pill)	DP (pill)		NT	IBS symptom severity scale (IBS-SSS)	only PO	some concerns
Locher, 2017	Switzerland	experimental pain	nonclinical	151 (68.00)	27.15 (9.51)	1	OLP dermal (cream)	OLP-(cream)	DP (cream)	NT	pain intensity (VAS 0-100)	most frequent	low
Meeuwis, 2021	Netherlands	experimental itch	nonclinical	55 (85.45)	21.89 (2.50)	1	OLP dermal (patch)			DP (patch)	mean itch (NRS 0-10)	only PO	some concerns
Meeuwis, 2019	Netherlands	experimental itch	nonclinical	45 (82.60)	21.80 (2.70)	7	OLP dermal (tonic)			DP (tonic)	AUC itch (NRS 0-10)	only PO	some concerns

Morales-Quezada, 2020	USA	acute pain (spinal cord injury and polytrauma)	clinical	19 (30.00)	47.30 (16.78)	6	cOLP pills (capsule)			TAU	opioid consumption (MEDC)	only PO	high
Mundt, 2017	USA	experimental pain	nonclinical	75 (57.33)	22.75 (5.89)	1	cOLP dermal (cream)	DP (cream)		NT	pain intensity (VAS 0-100)	only PO	some concerns
Nitzan, 2020	Israel	major depressive disorder	clinical	38 (NA)	49.91 (17.27)	56	OLP pills (capsule)			WL	depression severity (QIDS)	only PO	some concerns
Olliges, 2022	Germany	knee osteoarthritis	clinical	40 (60.15)	67.02 (9.47)	21	OLP pills (capsule)			NT	pain intensity (NRS 0-10)	most frequent	some concerns
Pan, 2020	Germany	menopausal hot flushes	subclinical	100 (100.00)	54.55 (NA)	28	OLP pills (pill)			NT	hot flushes composite score	most informative	some concerns
Schaefer, 2018	Germany	allergic rhinitis	clinical	46 (77.80)	24.67 (6.37)	14	OLP pills (pill)	OLP-(pill)		NT (with rationale + without rationale)	allergic symptoms composite score	only PO	some concerns

Schaefer, 2016	Germany	allergic rhinitis	clinical	25 (84.00)	26 (9.90)	14	OLP pills (pill)			NT	allergic symptoms composite score	only PO	some concerns
Schaefer, 2019	Germany	test anxiety	subclinical	58 (86.60)	22.90 (2.85)	14	OLP pills (pill)			NT	test anxiety (PAF)	most informative	some concerns
Schaefer, 2021	Germany	experimental acute stress	nonclinical	53 (53.31)	26.33 (8.77)	21	OLP pills (pill)			NT	acute stress (0-100)	most frequent	some concerns
Schienle, 2021	Austria	relaxation	subclinical	148 (71.00)	24.40 (2.70)	14	OLP suspension (sunflower oil)			TAU	PMR exercise quality: relaxation	baseline available; most informative	some concerns
Swafford, 2019*	USA	muscle strength	nonclinical	21 (47.60)	22.52 (3.00)	7	OLP pills (capsule)	DP (capsule)		NT	isometric peak torque	authors judgment	some concerns
Yennurajalingam, 2022	USA	cancer-related fatigue	clinical	84 (67.00)	56.00 (13.00)	7	OLP pills (tablet)			WL	cancer related fatigue (FACIT-F)	only PO	some concerns

Zhou, 2019	USA	cancer-related fatigue	clinical	40 (92.50)	47.30 (12.40)	22	OLP pills (tablet)			WL	cancer related fatigue (FACIT-F)	only PO	some concerns

Note. cOLP, conditioned Open-Label Placebo; DP, Deceptive Placebo; NT, No Treatment; OLP, Open-Label Placebo with rationale; OLP-, Open-Label Placebo without expectation induction; PO, Primary Outcome; TAU, Treatment as Usual; WL, Wait List; *, crossover study

eTable 2. Individual study data

Nonclinical network

author	year	merged groups	data from author	group	population	age mean	age sd	% female	country	continuous outcome	n	mean change	sd change
Barnes	2019	yes (fully & semi open)	yes	OLP nasal	experimental nausea	20.3	3.26	58.62	Australia	Self-report nausea, 6-item composite scale	29	6.14	9.78
Barnes	2019	no	yes	DP	experimental nausea	21.3	5.2	NA	Australia	Self-report nausea, 6-item composite scale	17	8.18	11.36
Barnes	2019	no	yes	NT	experimental nausea	22.9	5.5	NA	Australia	Self-report nausea, 6-item composite scale	15	2.86	11.07
Disley	2021	no	no	OLP nasal	experimental pain	21.05	5.04	86.666	UK	Pain Intensitiy, VAS	25	-0.12	20.35
Disley	2021	no	no	DP	experimental pain	21.05	5.04	86.666	UK	Pain Intensitiy, VAS	26	0.08	21.60
Disley	2021	no	no	NT	experimental pain	21.05	5.04	86.666	UK	Pain Intensitiy, VAS	24	-7.79	17.31
Friehs	2022	yes (personal & scientific)	no	OLP nasal	experimental sadness	24.56	6.55	69.79	Germany	Sadness subscale PANAS-X score total score 0-50	63	-2.20	9.60

Friehs	2022	yes (personal & scientific)	no	DP	experimental sadness	23.02	3.31	58.18	Germany	Sadness subscale PANAS-X score total score 0-50	55	1.00	6.06
Friehs	2022	no	no	NT	experimental sadness	23.1	2.9	82.8	Germany	Sadness subscale PANAS-X score total score 0-50	29	-6.00	8.22
Hahn	2022	no	no	OLP nasal	experimental sadness	23.67	3.31	100	Germany	Sadness subscale PANAS-X score total score 0-50	42	-4.27	8.76
Hahn	2022	no	no	NT	experimental sadness	25.81	6.98	100	Germany	Sadness subscale PANAS-X score total score 0-50	42	-12.01	10.87
Kube	2020	yes (Expectancy & Hope)	no	OLP dermal	experimental pain	25.16	6.41	62	Germany	Pain Intensitiy, VAS	50	-0.02	13.63
Kube	2020	no	no	DP	experimental pain	23.6	4.81	48	Germany	Pain Intensitiy, VAS	25	7.29	13.48
Kube	2020	no	no	NT	experimental pain	24.92	5.76	38.5	Germany	Pain Intensitiy, VAS	25	-2.70	13.92
Leibowitz	2019	no	yes	NT	experimental itch	24.55	NA	63.5	USA	Physiological allergic reaction (size of the wheal)	40	-1.65	1.09

Leibowitz	2019	no	yes	OLP-	experimental itch	24.55	NA	63.5	USA	Physiological allergic reaction (size of the wheal)	36	-1.61	0.83
Leibowitz	2019	yes (expectation & rationale)	yes	OLP dermal	experimental itch	24.55	NA	63.5	USA	Physiological allergic reaction (size of the wheal)	72	-1.56	0.88
Locher	2017	no	no	NT	experimental pain	27.9	8.52	73	Switzerland	Subjective heat pain intensity	40	1.89	3.33
Locher	2017	no	no	OLP-	experimental pain	28.27	11.34	65	Switzerland	Subjective heat pain intensity	37	-3.11	3.46
Locher	2017	no	no	OLP dermal	experimental pain	25.7	7.76	73	Switzerland	Subjective heat pain intensity	37	2.97	3.46
Locher	2017	no	no	DP	experimental pain	26.65	10.25	62	Switzerland	Subjective heat pain intensity	37	1.81	3.46
Meeuwis	2021	no	no	OLP dermal	experimental itch	21.67	2.6	85.19	Netherlands	Self reported mean itch, NRS	27	0.55	1.53
Meeuwis	2021	no	no	DP	experimental itch	22.11	2.39	85.71	Netherlands	Self reported mean itch, NRS	28	0.81	1.48
Meeuwis	2019	no	no	OLP dermal	experimental itch	21.8	2.7	82.6	Netherlands	AUC itch	22	49.71	223.04
Meeuwis	2019	no	no	DP	experimental itch	21.8	2.7	82.6	Netherlands	AUC itch	23	58.27	259.11

Mundt	2017	no	no	NT	experimental pain	22.75	5.89	57.33	USA	Mean pain intensity ratings, VAS	25	-6.14	11.33
Mundt	2017	no	no	DP	experimental pain	22.75	5.89	57.33	USA	Mean pain intensity ratings, VAS	25	1.25	12.46
Mundt	2017	no	no	cOLP dermal	experimental pain	22.75	5.89	57.33	USA	Mean pain intensity ratings, VAS	25	0.21	11.64
Schaefer	2021	no	no	OLP pills	experimental acute stress	25.25	7.28	58.33	Germany	Perceived stress, VAS	24	-31.29	35.07
Schaefer	2021	no	no	NT	experimental acute stress	27.41	10.25	48.28	Germany	Perceived stress, VAS	29	-30.76	34.60
Swafford	2019	no	yes	DP	muscle strength	22.52	3	47.6	USA	Peak torque of experiment 1	7	5.20	55.11
Swafford	2019	no	yes	OLP pills	muscle strength	22.52	3	47.6	USA	Peak torque of experiment 1	7	5.80	58.98
Swafford	2019	no	yes	NT	muscle strength	22.52	3	47.6	USA	Peak torque of experiment 1	7	4.90	32.06

Clinical network

author	year	merged groups	data from author	group	population	age mean	age sd	% female	country	continuous outcome	n	mean change	sd change
Ashar	2021	no	no	Psycho-logical intervention	chronic low back pain	42.6	16.2	58	USA	Pain intensity, VAS	44	3.04	1.23
Ashar	2021	no	no	OLP injection	chronic low back pain	39.4	14.9	49	USA	Pain intensity, VAS	44	1.32	1.51
Ashar	2021	no	no	TAU	chronic low back pain	41.3	15.9	54	USA	Pain intensity, VAS	47	0.78	1.36
Bandak	2022	no	no	OLP injection	knee osteoarthritis	66.7	8.2	47.2	Denmark	Pain score, KOOS (baseline-week 9)	104	7.30	15.22
Bandak	2022	no	no	Treatment program	knee osteoarthritis	70.1	8.3	44.1	Denmark	Pain score, KOOS (baseline-week 9)	102	10.00	15.07
Carvalho	2016	no	no	OLP pills	chronic low back pain	44.4	13.2	70.7	Portugal	Pain intensity, NRS	41	1.49	1.68
Carvalho	2016	no	no	WL	chronic low back pain	44.1	13.7	71.4	Portugal	Pain intensity, NRS	42	0.24	1.61
El Brihi	2019	yes (OLP 1/d & 4/d)	yes	OLP pills	well-being	19	3.9	80	Australia	Emotional distress (DASS)	61	7.30	9.16

El Brihi	2019	no	yes	NT	well-being	19	3.9	80	Australia	Emotional distress (DASS)	27	0.20	10.26
Flowers	2021	no	no	cOLP pills	acute pain (after spine surgery)	59.1	13.1	NA	USA	Worst daily pain (mini-BP; 0-10)	19	-0.60	2.26
Flowers	2021	no	no	TAU	acute pain (after spine surgery)	61.2	13	NA	USA	Worst daily pain (mini-BP; 0-10)	22	-1.50	1.44
Haas	2022	no	no	OLP pills	primary insomnia	31.04	NA	86.96	Germany	Subjective total sleep time in minutes	23	24.83	91.13
Haas	2022	no	no	OLP-	primary insomnia	29.09	NA	81.82	Germany	Subjective total sleep time in minutes	22	11.31	104.21
Hoenemeyer	2018	no	no	OLP pills	cancer-related fatigue	58.4	11.2	72	USA	FSI, Fatique Symptom Severity)	38	18.60	23.01
Hoenemeyer	2018	no	no	WL	cancer-related fatigue	56	12.4	66	USA	FSI, Fatique Symptom Severity)	35	6.10	22.75
Ikemoto	2020	no	no	OLP pills	chronic low back pain	68.2	68.2	65.4	Japan	Pain intensity, NRS	24	1.10	1.90
Ikemoto	2020	no	no	TAU	chronic low back pain	65.3	65.3	57.7	Japan	Pain intensity, NRS	24	0.80	1.90
Kaptchuk	2010	no	no	OLP pills	irritable bowel syndrome	47	18	65	Israel	IBS-SSS 0-500	37	92.00	99.00

Kaptchuk	2010	no	no	NT	irritable bowel syndrome	46	18	74	Israel	IBS-SSS 0-500	43	46.00	74.00
Kelley	2012	no	no	OLP pills	MDD	38.8	12.6	70	USA	Depression severity, HAM-D	11	1.64	4.52
Kelley	2012	no	no	WL	MDD	38.8	12.6	70	USA	Depression severity, HAM-D	9	-0.67	4.00
Kleine-Borgmann	2021	no	no	OLP pills	chronic low back pain	60.28	15.15	NA	Germany	Composite pain intensity score	63	0.62	1.81
Kleine-Borgmann	2021	no	no	WL	chronic low back pain	58.37	13.97	NA	Germany	Composite pain intensity score	59	-0.11	1.29
Kleine-Borgmann	2019	no	no	OLP pills	well-being & cognitive enhancement	23.97	2.83	68	Germany	Perceived Stress Questionnaire, PSQ20	79	-11.90	19.67
Kleine-Borgmann	2019	no	no	NT	well-being & cognitive enhancement	24.08	2.74	67	Germany	Perceived Stress Questionnaire, PSQ20	75	-16.74	17.22
Klinger	2017	no	no	OLP-	chronic low back pain + experimental pain	50.83	17.01	75	Germany	Back pain rating, NRS	12	-1.16	1.83
Klinger	2017	no	no	cOLP suspension	chronic low back pain + experimental pain	50.33	15.17	75	Germany	Back pain rating, NRS	12	0.67	2.12

Klinger	2017	yes (cond. & uncond. DP)	no	DP	chronic low back pain + experimental pain	51.52	13.05	75	Germany	Back pain rating, NRS	24	2.58	2.12
Kube	2021	yes (augmented & limited)	no	OLP pills	allergic rhinitis	26.95	10.56	64.3	Germany	Self-reported allergic symptoms, CSMS	28	2.20	3.81
Kube	2021	yes (augmented & limited)	no	WL	allergic rhinitis	36	14.77	73.05	Germany	Self-reported allergic symptoms, CSMS	26	2.90	3.76
Lembo	2021	no	no	OLP pills	irritable bowel syndrome	42.2	17.8	71.9	USA	IBS-SSS 0-500	68	90.60	89.50
Lembo	2021	no	no	NT	irritable bowel syndrome	40	17	73.3	USA	IBS-SSS 0-500	72	52.30	87.00
Lembo	2021	no	no	DP	irritable bowel syndrome	43.8	19.2	73.6	USA	IBS-SSS 0-500	71	100.30	99.60
Morales-Quezada	2020	no	yes	cOLP pills	acute pain (spinal cord injury and polytrauma)	44.9	16.93	30	USA	Opiod consumption, MDEC	9	66.00	99.55
Morales-Quezada	2020	no	yes	TAU	acute pain (spinal cord injury and polytrauma)	49.7	16.62	30	USA	Opiod consumption, MDEC	10	3.76	56.51

Nitzan	2020	no	no	OLP pills	major depressive disorder	48.17	16.86	NA	Israel	Depression severity,QIDS total score	18	1.95	5.06
Nitzan	2020	no	no	WL	major depressive disorder	51.65	17.68	NA	Israel	Depression severity, QIDS total score	20	0.45	4.12
Olliges	2022	no	yes	OLP pills	knee osteoarthritis	64.19	9.3	57.1	Germany	Pain intensity, NRS	21	0.44	1.35
Olliges	2022	no	yes	NT	knee osteoarthritis	69.84	9.63	63.2	Germany	Pain intensity, NRS	19	-0.28	1.99
Pan	2020	no	no	OLP pills	menopausal hot flushes	54.2	NA	100	Germany	Hot flush score, composite score	50	6.02	9.71
Pan	2020	no	no	NT	menopausal hot flushes	54.9	NA	100	Germany	Hot flush score, composite score	50	3.26	8.79
Schaefer	2018	no	no	OLP pills	allergic rhinitis	25	9	69.2	Germany	Allergic symptoms composite score	13	0.78	0.67
Schaefer	2018	no	no	OLP-	allergic rhinitis	23	3	69.2	Germany	Allergic symptoms composite score	13	0.43	0.90
Schaefer	2018	yes (with & without rationale)	no	NT	allergic rhinitis	26	7.11	95	Germany	Allergic symptoms composite score	20	0.05	1.03

Schaefer	2016	no	no	OLP pills	allergic rhinitis	26	9.9	84	Germany	Allergic symptoms composite score	11	0.88	0.93
Schaefer	2016	no	no	NT	allergic rhinitis	26	9.9	84	Germany	Allergic symptoms composite score	14	0.23	0.72
Schaefer	2019	no	no	OLP pills	test anxiety	22.3	2.3	80.6	Germany	Test anxiety, PAF	31	4.39	9.35
Schaefer	2019	no	no	NT	test anxiety	23.5	3.4	92.6	Germany	Test anxiety, PAF	27	0.07	6.00
Schienle	2021	no	no	OLP suspension	relaxation	24.4	2.7	71	Austria	Exercise quality relaxation	68	1.29	0.97
Schienle	2021	no	no	TAU	relaxation	24.4	2.7	71	Austria	Exercise quality relaxation	80	1.28	0.93
Yennurajalingam	2022	no	no	OLP pills	cancer-related fatigue	57	12	74	USA	Fatigue, FACIT-F	42	6.60	7.60
Yennurajalingam	2022	no	no	WL	cancer-related fatigue	55	14	60	USA	Fatigue, FACIT-F	42	2.10	9.40
Zhou	2019	no	no	OLP pills	cancer-related fatigue	47.3	12.4	92.5	USA	Fatigue, FACIT-F	20	4.30	10.43
Zhou	2019	no	no	WL	cancer-related fatigue	47.3	12.4	92.5	USA	Fatigue, FACIT-F	20	1.20	10.15

Note. cOLP, conditioned Open-Label Placebo; DP, Deceptive Placebo; NT, No Treatment; OLP, Open-Label Placebo with rationale; OLP-, Open-Label Placebo without expectation induction; TAU, Treatment as Usual; WL, Wait List.

eTable 3. Head to head comparisons

Nonclinical network

	cOLP dermal	DP	NT	OLP dermal	OLP nasal	OLP pills	OLP-
cOLP dermal		-0.09 [-1.00; 0.82]	0.54 [-0.37; 1.45]	.	.	.	.
DP	-0.03 [-0.83; 0.78]		**0.47 [0.12; 0.82]**	0.10 [-0.34; 0.54]	0.21 [-0.29; 0.72]	-0.01 [-1.28; 1.26]	**1.44 [0.57; 2.30]**
NT	0.47 [-0.33; 1.28]	**0.50 [0.17; 0.83]**		-0.20 [-0.69; 0.29]	**-0.50 [-0.94; -0.06]**	0.00 [-0.73; 0.74]	**0.70 [0.09; 1.30]**
OLP dermal	0.21 [-0.66; 1.08]	0.24 [-0.15; 0.62]	-0.26 [-0.67; 0.14]		.	.	**0.86 [0.26; 1.46]**
OLP nasal	0.04 [-0.83; 0.92]	0.07 [-0.36; 0.50]	**-0.43 [-0.84; -0.02]**	-0.17 [-0.70; 0.37]		.	.
OLP pills	0.38 [-0.68; 1.43]	0.40 [-0.34; 1.15]	-0.10 [-0.80; 0.60]	0.17 [-0.63; 0.96]	0.33 [-0.47; 1.13]		.
OLP-	**1.07 [0.12; 2.02]**	**1.10 [0.53; 1.66]**	**0.60 [0.05; 1.15]**	**0.86 [0.31; 1.41]**	**1.03 [0.37; 1.69]**	0.69 [-0.19; 1.57]	

Note. Column headers are identical to row headers. Cells contain the network estimates (SMDs) from network meta-analysis (direct and indirect evidence) in the lower triangle and the direct treatment estimates (SMDs) from pairwise comparisons in the upper triangle. Comparisons considered for RQ2 (modalities) are highlighted in yellow, for RQ3 (expectation) in green and for RQ4 (comparator) in blue. Legend: cOLP, conditioned Open-Label Placebo; DP, Deceptive Placebo; NT, No Treatment; OLP, Open-Label Placebo with rationale; OLP-, Open-Label Placebo without expectation induction; TAU, Treatment as Usual; WL, Wait List.

Clinical network

	cOLP pills	cOLP suspension	DP	NT	OLP injection	OLP pills	OLP suspension	OLP-	Psych. intervent.	TAU	Treatment programme	WL
cOLP pills		.	.	.	.	.	.	.	.	**0.58 [0.02; 1.15]**	.	.
cOLP suspension	0.65 [-0.50; 1.81]		**-0.93 [-1.71; -0.15]**	.	.	.	.	**0.89 [0.02; 1.76]**	.	.	.	.
DP	0.12 [-0.81; 1.06]	-0.53 [-1.26; 0.20]		**0.52 [0.07; 0.97]**	.	0.11 [-0.34; 0.55]	.	**1.82 [0.98; 2.66]**	.	.	.	.
NT	**0.89 [0.01; 1.76]**	0.23 [-0.54; 1.01]	**0.76 [0.39; 1.14]**		.	**-0.47 [-0.66; -0.29]**	.	-0.42 [-1.19; 0.34]	.	.	.	.
OLP injection	0.19 [-0.57; 0.95]	-0.47 [-1.59; 0.66]	0.06 [-0.84; 0.96]	-0.70 [-1.54; 0.14]		.	.	.	**-1.26 [-1.79; -0.72]**	0.39 [-0.12; 0.91]	-0.18 [-0.58; 0.23]	.
OLP pills	0.42 [-0.43; 1.28]	-0.23 [-1.00; 0.54]	0.30 [-0.07; 0.67]	**-0.46 [-0.65; -0.28**	0.24 [-0.58; 1.06]		.	0.23 [-0.28; 0.75]	.	0.16 [-0.48; 0.80]	.	**0.43 [0.22; 0.64]**
OLP suspension	0.57 [-0.15; 1.29]	-0.08 [-1.18; 1.01]	0.45 [-0.41; 1.31]	-0.32 [-1.12; 0.49]	0.38 [-0.29; 1.06]	0.15 [-0.63; 0.93]		.	.	0.01 [-0.43; 0.45]	.	.
OLP-	0.92 [-0.04; 1.87]	0.26 [-0.51; 1.03]	**0.79 [0.30; 1.29]**	0.03 [-0.41; 0.47]	0.73 [-0.20; 1.66]	**0.49 [0.07; 0.92]**	0.35 [-0.54; 1.23]		.	.	.	.
Psych. intervent.	**-1.07 [-1.85; -0.28]**	**-1.72 [-2.87; -0.58]**	**-1.19 [-2.11; -0.27]**	**-1.96 [-2.82; -1.09]**	**-1.26 [-1.79; -0.72]**	**-1.49 [-2.34; -0.65]**	**-1.64 [-2.34; -0.94]**	**-1.99 [-2.93; -1.04]**		**1.65 [1.11; 2.20]**	.	.
TAU	**0.58 [0.02; 1.15]**	-0.07 [-1.08; 0.93]	0.46 [-0.28; 1.20]	-0.30 [-0.97; 0.36]	0.39 [-0.12; 0.91]	0.16 [-0.48; 0.80]	0.01 [-0.43; 0.45]	-0.34 [-1.11; 0.44]	**1.65 [1.11; 2.20]**		.	.
Treatment programme	0.01 [-0.85; 0.87]	-0.65 [-1.84; 0.55]	-0.11 [-1.10; 0.87]	-0.88 [-1.81; 0.06]	-0.18 [-0.58; 0.23]	-0.41 [-1.33; 0.50]	-0.56 [-1.35; 0.23]	-0.91 [-1.92; 0.10]	**1.08 [0.41; 1.75]**	-0.57 [-1.23; 0.08]		.
WL	0.86 [-0.02; 1.74]	0.20 [-0.60; 1.00]	**0.73 [0.31; 1.16]**	-0.03 [-0.31; 0.25]	0.67 [-0.18; 1.51]	**0.43 [0.22; 0.64]**	0.28 [-0.52; 1.09]	-0.06 [-0.54; 0.41]	**1.92 [1.06; 2.79]**	0.27 [-0.40; 0.95]	0.85 [-0.09; 1.78]	

Note. Column headers are identical to row headers. Cells contain the network estimates (SMDs) from network meta-analysis (direct and indirect evidence) in the lower triangle and the direct treatment estimates (SMDs) from pairwise comparisons in the upper triangle. Comparisons considered for RQ2 (modalities) are highlighted in yellow, for RQ3 (expectation) in green and for RQ4 (comparator) in blue. Legend: cOLP, conditioned Open-Label Placebo; DP, Deceptive Placebo; NT, No Treatment; OLP, Open-Label Placebo with rationale; OLP-, Open-Label Placebo without expectation induction; TAU, Treatment as Usual; WL, Wait List.

eReferences

1. Salanti, G., Del Giovane, C., Chaimani, A., Caldwell, D. M. & Higgins, J. P. Evaluating the quality of evidence from a network meta-analysis. *PloS One* **9**, e99682 (2014).
2. Papakonstantinou, T., Nikolakopoulou, A., Higgins, J. P., Egger, M. & Salanti, G. Cinema: software for semiautomated assessment of the confidence in the results of network meta-analysis. *Campbell Syst. Rev.* **16**, e1080 (2020).
3. Nikolakopoulou, A. *et al.* CINeMA: an approach for assessing confidence in the results of a network meta-analysis. *PLoS Med.* **17**, e1003082 (2020).
4. Sterne, J. A. C. *et al.* RoB 2: a revised tool for assessing risk of bias in randomised trials. *Bmj* **366**, l4898 (2019).
5. Chiocchia, V. *et al.* ROB-MEN: a tool to assess risk of bias due to missing evidence in network meta-analysis. *BMC Med.* **19**, 1–13 (2021).
6. Cuijpers, P., Noma, H., Karyotaki, E., Cipriani, A. & Furukawa, T. A. Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: a network meta-analysis. *JAMA Psychiatry* **76**, 70075–707 (2019).

Universität
Basel

Fakultät für
Psychologie

Appendix D: Curriculum Vitae

The End