The Development and Evaluation of a new Experimental Model of an Active Placebo to Investigate how the Experience of Side Effects Influences the Placebo Effects

Christoph Patrick Werner MSc

A thesis submitted to fulfil requirements for the degree of

DOCTOR OF PHILOSOPHY

School of Psychology

Faculty of Science

The University of Sydney

New South Wales, Australia

August 2021

Statement of Originality

This is to certify that to the best of my knowledge the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Signature: 

Name: Christoph Patrick Werner

Date: 31/08/2021

Contributions of the Candidate

The research presented in this thesis represents original work undertaken by the candidate, in conjunction with the School of Psychology, Faculty of Science at the University of Sydney. Ethics approval was granted by the University of Sydney Human Research Ethics Committee (Project Number: 2018/107, see Appendices).

The candidate was responsible for coordinating the research under the supervision of A/Prof Ben Colagiuri, Prof Louise Sharpe, and Dr Kate Fassee. The candidate took primary responsibility for all aspects of the research presented in this thesis, including data collection, analysis, and writing, under the guidance and supervision of the aforementioned supervisors. The candidate wrote this thesis and maintains chief responsibility for the thesis.

Signature: 

Name: Christoph Patrick Werner

Date: 31/08/2021

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Signature: 

Name: Ben Colagiuri

Date: 31/08/2021

Acknowledgements

It is with deep gratitude that I extend my thanks and special acknowledgment to the many people who have contributed, directly or indirectly, to the completion of this thesis.

First, I would like to say a big thanks the people who participated in this research, either as healthy participant or as someone who suffers from insomnia. All volunteers showed incredible generosity with their commitment to advance the current state of knowledge in the field of placebo research.

To my primary supervisor, A/Prof Ben Colagiuri, it has been a pleasure and privilege to have completed my PhD under your guidance. You welcomed me warmly to Australia and the rest of the lab. I knew I could always rely on you supporting me on my journey to become an independent researcher. You always supported my ideas and improved them with your expertise in the area. Thanks Ben! To my secondary supervisors, Prof Louise Sharpe and Dr Kate Fassee, thank you both for providing your ongoing guidance and contributing your immense expertise. Louise, I am particularly thankful for your incredible experience and advice.

My gratitude goes out to my other collaborators, Prof Maree Hackett, Prof Nick Glozier, and to the brilliant interns I had the pleasure to mentor, Lea Fischer, Lea Schäfer, Steffen Holsteg, Olivia Fischer, and Laura Petri.

I want to express my great gratitude to the Australian Commonwealth who made all my research possible with the generous support I received being awarded an International Research and Training Scholarship and Stipend Offset. Without this support, it would not have been a viable choice to move to Australia and commence a PhD candidacy at the University of Sydney.

I would also like to extend my thanks to all the other members and colleagues of the Colagiuri lab. Kirsten, Jonno, Biya, and Amelia you truly made my journey to Australia and the

adventure of doing a PhD on the other side of the world a true pleasure and I will keep you forever in fond memory. Biya I will always be indebted to you for your hands-on help during the SARS-CoV-2 pandemic while I was stuck in Europe.

To my better half (and soon to be wife) Doreen, I could not have wished for a better partner on my side in the entire world. You supported me in any way possible. Be it from your true and heartfelt joy when I told you that I was accepted to the PhD program, during the nearly four years of living in opposite corners of the world with limited possibilities to spend time together, the countless obstacles life put in front of me while conducting my studies, to finally helping keep my spirits up finishing everything and writing this thesis. You are my true hero.

To my parents, Peter and Traudi, and siblings Ralph, Jasmin, and Beat, thank you always for your ongoing support and love, even if it meant skyping at inconvenient times to share a laugh and a glass of wine together. And finally, to my amazing old friends from Switzerland and all the newly found ones in Australia, thank you all for keeping me motivated and making life outside the PhD even more joyful.

Contents

List of Tables

List of Figures

List of Abbreviations

AE	Adverse Event
ANZCTR	Australian and New Zealand Clinical Trial Registry
CSD-C	Consensus Sleep Diary – Core Version
DASS-21	Depression Anxiety Stress Scale 21-item short version
RCT	Double-blind Randomised Placebo-Controlled Trial
GP	General Practitioner
ISI	Insomnia Severity Index
oAwake	Objective Number of Nightly Awakenings
oSOL	Objective Sleep Onset Latency
oSQ	Objective Sleep Quality
oTST	Objective Total Sleep Time
PSG	Polysomnography
PSQI	Pittsburgh Sleep Quality Index
PTA / cPTA	Perceived Treatment Allocation / Participants' certainty of their choice regarding PTA
sAwake	Subjective Number of Nightly Awakenings
sSOL	Subjective Sleep Onset Latency
sSQ	Subjective Sleep Quality
sTST	Subjective Total Sleep Time
TSQM-II	Treatment Satisfaction Questionnaire for Medications – Version 2
WHOQOL-BREF	World Health Organisation Quality of Life Questionnaire – Brief Version

Abstract

Most conventional placebos, e.g., lactose pills, have no perceptible features during or after the process of intake and therefore are not likely to mimic the perceptible features of a real medication administered during a clinical trial or in clinical practice. This is because most drugs have distinctive side effects, like vitamins influencing urine colour, or other perceptible effects, like cough pills having an unpleasant taste. This means that using conventional placebos to understand the placebo effect might not accurately reflect the conditions under which placebo effects occur outside the laboratory. To my knowledge, only two approaches have been mentioned in the literature so far aimed to account for perceptible features of medications. An early approach, labelled as impure placebo, involved participants receiving a real medication, but importantly without direct effects on the condition being treated, like antibiotics for viral infections. A more recent approach focuses on using so-called active placebos, that do not contain an established drug, but rather a specific ingredient with the ability to create a perceptible sensation without any other effects on the participant, e.g., capsaicin to induce nose tingling.

The current project sought to extend this research by developing and evaluated a new experimental model of an active placebo in the form of capsules to investigate how perceptible features of a placebo influences its effect. The active placebo contained a benign food colouring, i.e., beetroot extract (E162), that caused a slight red colouration of urine and had a distinct noticeable taste during ingestion, but otherwise had no known biological effects. A systematic review and meta-analysis first showed that adverse event rates differed statistically significantly between placebo groups of clinical trials investigating different medications and that the adverse events experienced in the placebo groups statistically significantly correlated with the adverse events reported by the corresponding medications group, while there was no overall association

between adverse event rates reported by placebo groups and their placebo response. Three studies evaluated the new model for its capacity to elicit the targeted perceptible feature compared to a conventional placebo (Study 1), whether the active placebo elicited a higher placebo and nocebo effect than a conventional placebo (Studies 1 and 2), whether the active placebo influenced beliefs about treatment allocation (Study 2), and how differently framed information affected the efficacy of active placebos (Study 3).

Study 1 demonstrated that the new active placebo model successfully elicited the target side effect in ~50% of participants, which was what was suggested as a priori as a minimum to be effective, while being mostly well tolerated, but no placebo effect was observed on sleep for either active or conventional placebo relative to no treatment. Study 2 showed that most participants receiving a conventional placebo correctly guessed that they had been allocated to the placebo group, while active placebo participants were split evenly between indicating having received a placebo and active medication, suggesting differential blinding as a result of the active placebo. However, there was no overall placebo effect on sleep, but active placebo group showed greater improvement on the Insomnia Severity Scale than conventional placebo participants. Results from Study 3 indicated no statistically significant effect of message framing on sleep following active placebo administration, nor an overall placebo effect. Pooled analysis of data from all three studies confirmed the absence of a statistically significant placebo effect for both a conventional and active placebo relative to no treatment.

The findings from this thesis indicated no statistically significant placebo effect on sleep for either conventional or active placebo. Interestingly, however, the results of Study 2 indicate that active placebos can facilitate the maintenance of participant blinding in randomised controlled trials and that in such contexts, a lack of side effects might have diminished the

placebo response. As such, the active placebo model developed here sheds light on how side effects may influence outcomes in randomised controlled trials and could help inform the development of techniques to improve blinding in practice.

Keywords: active placebo, nocebo, insomnia, adverse event rates, perceptible features, randomised controlled trials

Chapter 1: Literature Review and Research Aims

This chapter gives an overview about the background of the research conducted for this thesis, highlights the existing gaps in knowledge in the current scientific literature, and introduces the aims of the research and the methodological approaches chosen to achieve these aims.

The placebo effect is a fascinating phenomenon that has captured the attention of clinicians and researchers because placebos seemingly improve health problems without having any active ingredients. At first sight, the fact that a mere capsule filled with lactose can improve a person's health problems, let alone trigger objectively measurable changes in the body and brain seems impossible. This chapter commences with defining what constitutes a placebo, comprehensively outlines how the placebo and nocebo effects are defined and how these definitions differ from placebo and nocebo responses, gives an overview about the origins of placebos, and describes the shift in perspective from being a nuisance that needs to be controlled for in clinical trials to a powerful mechanism that might be utilised to enhance clinical outcomes in medical practice. Following this, an overview of the factors that make it possible that an otherwise inert capsule has the potential to elicit changes in the health status of patients, and models of the current scientific literature explaining any such effect are presented. The following section then reviews how randomised controlled trials (RCTs) became the gold standard for the evaluation of new medicines and why RCTs are considered the pinnacle of evidence-based medicine with the highest quality of evidence and the lowest amount of bias. After these remarks, the focus is set on exploring why side effects or better said the lack thereof might jeopardise the validity of RCTs. Before concluding the literature review, the last section explains

why a further investigation of placebo effects in sleep research with participants suffering from insomnia symptoms is important and outlines the detailed research aims.

Defining Placebo

This section is aimed at introducing some of the most important definitions and conceptualisations used throughout this dissertation. The meaning of the term 'placebo' and its use has changed over time from being used to please patients when no real medicine was available before modern medicine's scientific breakthroughs until today's understanding where placebos are mostly seen as inert sugar pills used as a comparator in clinical trials.

What is a Placebo?

The term placebo was first introduced into the medical jargon in the 18th century by the English physician Alexander Sutherland (Jütte, 2013). The Latin word "placebo" is best translated as "I shall please.".

With advances in the medical sciences and the associated methodology of testing new medications a change in practice regarding placebos happened from an undistinguished treatment to satisfy patients to a necessary control intervention in trials of new medications. In 1899 placebos were already described as harmless substances with the example of bread pills (The New Sydenham Society's Lexicon, 1899). It took until the year 1937 that a medical dictionary had defined a placebo as an inert substance (Taber, 1937). Since then, placebos have generally been defined as an inert intervention that is indistinguishable from a real treatment in its appearance.

Placebos come in many shapes, colours, intensity, and cover a plethora of different treatment modalities (Meissner & Linde, 2018). The most common form of placebo application are capsules, pills, or tablets containing lactose in RCTs evaluating new pharmacological medications or used in experimental research (Ashar et al., 2017). Fässler et al. (2015) reviewed the placebo literature with the aim of investigating if different levels of invasiveness had an impact on the extents of a placebo response, thereby elegantly showing the variety of placebos. Placebo can relate to anything ranging from the most prominent sugar pill all the way to sophisticated surgeries where patients are put under general anaesthesia, actually cut open, but without the actual surgical procedure applied. Some of the less invasive placebos include the classical pill, tablet, or capsule containing lactose or placebo creams that are administered externally. Some of the more invasive placebos listed include placebo inhalation or sham acupuncture needles. Tough et al. (2009) specifically developed and evaluated spring-loaded acupuncture needles that are unable to puncture the skin and retract into the handle making them indistinguishable from real acupuncture needles, even for people with prior experience in acupuncture. Finally, the spectrum of placebos ranges all the way to intravenous saline injections or placebo surgery as conducted in the famous hallmark study about arthroscopic surgery for osteoarthritis of the knee. Moseley et al. (2002) went all the way to trick patients into believing that they had undergone the actual knee surgery for their placebo-controlled trial. Patients underwent local anaesthetic procedures, were cut open by surgeons that were then informed to perform the actual debridement or to simulate the actual surgery step by step asking assistants for all the needed tools and splashing saline to simulate the sound of lavage.

Defining Placebo and Nocebo Effects vs. Responses

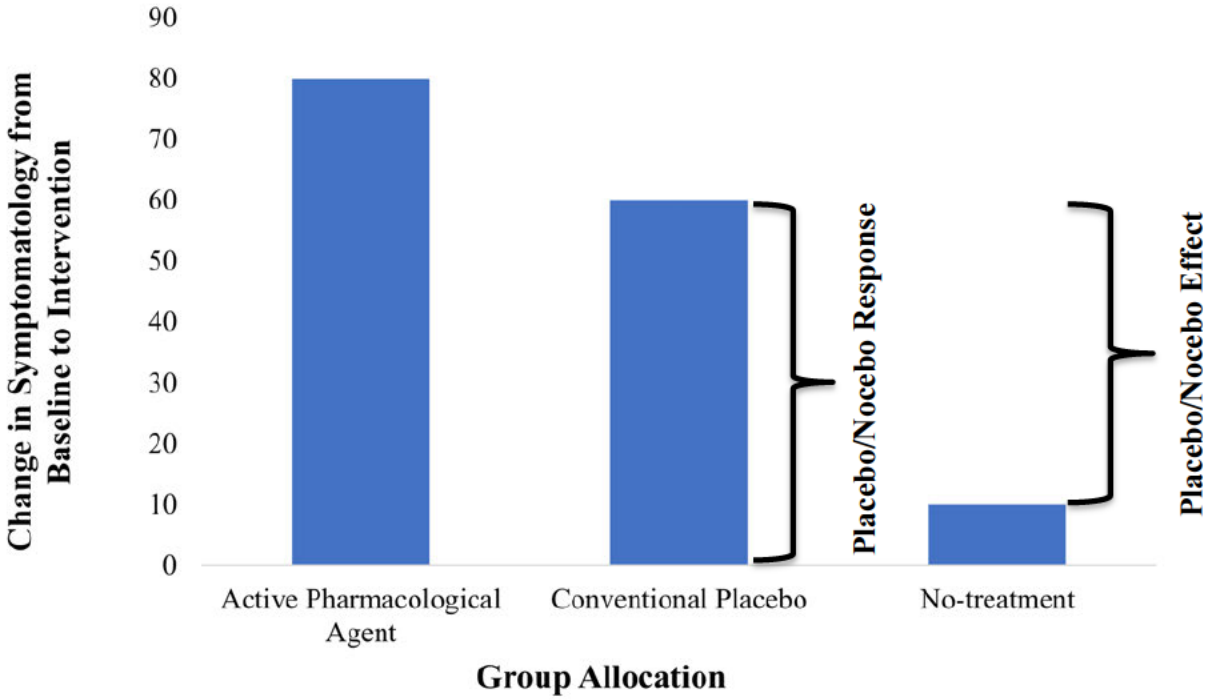
The placebo effect has been defined in many different ways depending on the clinical context or the researchers' perspective. Placebo effect and placebo response is often used interchangeably, but the two terms describe two rather different phenomena. For this thesis, the following definitions were chosen because they ideally suit a research perspective in the context of pharmacological RCTs. Following Kirsch (2013), the placebo response was defined as the observed within-group change from a baseline to a follow-up measure after the application of a placebo intervention. While the placebo response is a simple within-group difference over time, the placebo effect is a more complex difference of differences, meaning that the placebo effect solely accounts for the change caused by a placebo and is controlled for other factors that might be responsible for causing a change in a no-treatment group. Therefore, it is only possible to talk about placebo effects when a no-treatment group is included in a study or trial. This implies that the placebo effect represents the change that is elicited by the application of a placebo alone, while the placebo response includes changes in symptomatology that are caused by the application of a placebo and factors like spontaneous remission, regression to the mean, and natural symptom fluctuations of a disease state (Colloca, 2017; Colloca & Miller, 2011a; Miller & Brody, 2011).

While the concept of placebo effect generally concerns the improvement after an application of a placebo, negative consequences associated with a placebo application are generally summarised with the term nocebo effect. Compared to the wide-spread knowledge about the placebo effect, even in the general population, most people including researchers and clinicians are less familiar with the nocebo phenomenon. Blasini et al. (2017) nicely defined a

nocebo and the nocebo effect. A nocebo generally defines any intervention or procedure resulting in negative expectations about a health outcome. This might happen for example when disclosing potential side effects of a medication or when a patient experiences a therapeutic encounter that was earlier associated with adverse events or negative health outcomes. Parallel to Kirsch (2013) differentiating between the placebo effect and response, it is important to distinguish between the nocebo effect and the nocebo response. The nocebo response is defined by any negative health outcomes e.g., a worsening of symptomatology compared to the baseline or the experience of adverse events after the application of a placebo. As Colloca and Miller (2011b) have stated, the detection of a nocebo effect always required a no-treatment control group, or alternatively a group that is not disclosed about potential side effects. Without the comparison to a no-treatment control group a reported side effect after the onset of a placebo treatment might have been misattributed as a side effect when in fact it might have just been caused by changes in the persons quality of life, personal distress levels, normal physiological processes, or just as part of natural history (Barsky et al., 2002; Rief et al., 2006). Figure 1.1 illustrates the importance of a no-treatment group to differentiate between the placebo or nocebo effect and the placebo or nocebo response.

Figure 1.1

Difference Between Placebo/Nocebo Response and Placebo/Nocebo Effect



Historical Background

Origins of Placebo

Before the scientific advances of modern medicine in western culture, physicians often lacked effective treatments and so they often administered placebos to fulfil patients' expectations and guarantee they would leave satisfied. In the 18th and 19th century, practitioners did not perceive or prescribe placebos as inert sugar pills. If early practitioners had nothing specific ready for a patient's illness, they just prescribed any ointment they had so the patient would not be disappointed and leave empty-handed (Jütte, 2013). At that time the term placebo was better defined as a common-place treatment or method to satisfy the patient rather than having actual effects for the patient's underlying suffering (Shapiro, 1964). The third president of the United States of America described the state-of-the-art medicine, or rather the use of placebos in a letter to Caspar Wistar.

> One of the most successful physicians I have ever known, has assured me, that he used more bread pills, drops of colored water, and powders of hickory ashes, than of all other medicines put together. It was certainly a pious fraud. (Jefferson, 1807)

Emerging interest in the placebo effect beyond a control condition

In the middle of the 20th century, placebos were established as a control condition in clinical trials and the placebo effect was primarily seen as a nuisance that needed to be accounted for in trials (Langer, 1987). In 1955 Henry K. Beecher, an anaesthesiologist and medical ethicist was the first attempting to quantify the placebo effect. His milestone publication "The Powerful

Placebo" can arguably be identified as the article that started the modern era of placebo research and is still the most cited placebo publication (Beecher, 1955). Beecher was interested in the rate of placebo responders across different trials. He therefore assessed how many placebo participants showed a satisfactory improvement (e.g., in postoperative pain) between two times. Satisfactory improvement in a patient's pain was therefore defined as an improvement by at least 50% between the two time. Across 15 studies ranging from headache to severe post-operative wound pain he summarised that 35.2% of patients given a placebo had been satisfactorily relieved. This demonstrated how powerful the placebo effect was and how important it was that future pharmacological trials and experimental studies used a blinded design using placebos as a control group.

Dispute about the Existence of the Placebo Effect

Beecher's publication started a long and intensive scientific discourse about the existence of the placebo effect, its magnitude, and how important it really was to control for placebo effects. Kienle and Kiene (1997) fundamentally questioned Beecher's results and did their own analysis for 14 of the 15 studies Beecher had originally included – the other study did not report sufficient data for their reanalysis. The authors concluded that none of these 14 trials had demonstrated any reason to assume the existence of a placebo effect. Although the list of methodological issues regarding the placebo phenomenon are important when evaluating placebo research it has to be stated that the approach chosen by Kienle and Kiene (1997) to re-evaluate Beecher's findings has to be classified as scientifically unsystematic and purely relied on qualitative point-by-point personal judgements rather than coherently defined criteria and sound data analysis. The authors rated each of the included studies using two subjectively answered

questions: *"1. Is the existence of the placebo effect demonstrated in those 15 trials that Beecher had surveyed in "The Powerful Placebo"? 2. If not, what are the factors that can create the false impression of a placebo effect?".* Because Kienle and Kiene (1997) based their conclusions only on these subjective ratings, their strong conclusion regarding the non-existence of a placebo effect should be evaluated cautiously.

Shortly after Kienle and Kiene, another study was interested in evaluating if there was any merit to the placebo effect. Hróbjartsson and Gøtzsche (2001) were interested in comparing no-treatment groups to placebo groups across clinical trials investigating all different kinds of diseases to estimate the size of the placebo effect. They identified 32 clinical trials including 3,795 patients that reported binary outcomes and did not find a statistically significant placebo effect. For the 82 clinical trials involving 4,370 patients reporting continuous outcomes they found a beneficial effect of placebos compared to no treatment, but the placebo effect was only statistically significant for subjective outcomes. The median sample size of the included studies was rather small being 51 for binary outcomes and 27 for continuous outcomes and the authors found a significant effect of sample size on the placebo effect, with larger studies reporting smaller placebo effects. Based on the limited number of available trials and their analysis with sparse data across many different conditions the authors concluded that there is generally little evidence that placebos have powerful clinical effects, although mentioning that there were significant effects on subjective continuous outcome and for the treatment of pain.

Hrobjartsson and Gotzsche's paper received some criticism, primarily focussing on the methodological validity of their meta-analysis and the distinct conclusions and implications they had drawn from the rather scarce and heterogeneous evidence base (Ader, 2001; Greene et al., 2001; Kirsch & Scoboria, 2001). To follow up on the surprisingly low effect size for the placebo

analgesia in clinical trials where placebo was only used as a control condition, Vase et al. (2002) decided to conduct a meta-analysis that compared the latter trials with studies that investigated placebo mechanisms as the primary focus. Interestingly, their direct comparison resulted in a large placebo effect with a standardised mean change of 0.95 for studies investigating placebo as their primary focus compared to a small effect size of 0.15 for the trials that were included in the meta-analysis by Hróbjartsson and Gøtzsche (2001). A potential explanation for this large difference in effect sizes might be due to different instructions in double-blind trials and clinical practice instructions used in studies directly investigating placebos. In RCTs participants typically receive the information that they will either receive a placebo or an active medication, introducing uncertainty in participants. Contrary to this uncertainty studies directly investigating the placebo effect most often deceivingly inform participants that they are receiving an active treatment, causing no uncertainty.

The scientific discussion slowly changed over the years from questioning the existence of the placebo effect to a discussion about the extent of the placebo effect. First, Hróbjartsson and Gøtzsche (2003) argued in a commentary about a lack of methodological rigor, stating that Vase's meta-analysis did not weigh in the sample size of the included studies when computing the overall effect size for the placebo effect. This issue was rectified re-analysing Vase's meta-analysis adjusting for the methodological shortcomings, which resulted in an even larger estimate of the placebo effect with a standardised mean change of 1.14 (Price et al., 2003). The reporting of even larger placebo effects by Price and colleagues motivated Hrobjartsson and Gotzsche to update their meta-analysis with new clinical trials that had been published while the debate was ongoing. They again came to the same conclusion that there was no powerful placebo effect detectable for binary outcomes and following their earlier work they again found a small

but significant placebo effect on continuous measures for pain. In line with the significant placebo effect for pain they now detected a statistically significant, but small placebo effect for phobias (Hróbjartsson & Gøtzsche, 2004).

Miller and Rosenstein (2006) criticised the approach Hrobjartsson and Gotzsche had chosen to investigate the extent of the placebo effect and its usefulness in clinical practice. Their main argument was that the source of data in the form of clinical trials was poorly chosen to make conclusions about how powerful the placebo effect might be in clinical practice. Due to the blinded nature of controlled clinical trials participants experience uncertainty regarding their treatment allocation (Miller & Rosenstein, 2006). Arguably, this does not reflect clinical practice well where patients in fact do receive a real treatment. Concluding, the authors stated that there is sufficient evidence supporting the reality of the placebo effect, and that further research is needed to investigate the placebo effect's clinical significance across different conditions.

Quantifying the Placebo Response and the Placebo Effect

The above-mentioned dispute about the power of a placebo over time resulted in a convergence of opinions that there is evidence for placebo effects at least under certain conditions. Therefore, the focus of placebo research shifted towards the question of whether there are differences in the magnitude of the placebo response in different contexts, like different health conditions. To compare patients' placebo response across different psychiatric conditions Khan et al. (2005) compared patients with psychosis, obsessive-compulsive disorder, generalised anxiety disorder, depression, post-traumatic stress disorder, and panic disorder. Interestingly, they found that the mean percentage symptom improvements within the placebo groups over the course of the studies differed strongly between the conditions. While psychotic patients receiving

placebos only showed a negligible near null improvement, anxiety, depression, and traumatised patients showed improvements around 30%, and panic patients halved their symptoms with placebos. In line with the responses between different health conditions to placebo pills, a systematic review and meta-analysis investigated to what extent patients with different conditions responded to placebo surgery. Comparing placebo surgery across pain-related conditions, obesity, gastroesophageal reflux disorder, and others, Jonas et al. (2015) summarised that patients showed large effect sizes for sham surgery. The placebo overall accounted for 65 percent of the improvements observed with the real surgical interventions, ranging from 57 percent for gastroesophageal reflux disorder and 71 percent for obesity. Placebo surgery for pain-related conditions accounted for nearly 80 percent of the improvement with real surgery and were so large that the effects were statistically indistinguishable from bona fide surgeries. It is noteworthy to mention that these surgery studies did not contain no-treatment groups, so it is not possible to make statements about the placebo effect, only the placebo response as other non-specific factors influence the outcomes.

To fill this knowledge gap, Hróbjartsson and Gøtzsche (2010) conducted a Cochrane review including randomised trials that included a comparison between placebo interventions and no-treatment groups for all clinical conditions. Their results were in line with their previous work about placebo effects in the context of randomised clinical trials, showing evidence for small placebo effects in certain areas that are most likely not of clinical relevance, except for pain, where the effect size for placebo interventions ranged from nearly negligible all the way up to clinically important. From the perspective of placebo research the most relevant finding stemmed from the meta-regression analyses Hróbjartsson and Gøtzsche (2010) conducted. They reported that clinical trials using physical placebo interventions like sham acupuncture, trials

specifically investigating the placebo phenomenon, and trials in which participants were not aware of the existence of a placebo control condition were associated with larger placebo effects.

The insight that the extent of the placebo effect might heavily depend on the information given to participants offered an opportunity to explain the discrepancy in findings between the large placebo effects proclaimed by Vase and colleagues and the arguably non-existent to small placebo effects observed by the Hróbjartsson group. That information might influence participants expectations and their response to a placebo was the topic of a contemporary review discussing the validity of double-blind placebo-controlled trials (Colagiuri, 2010). After reviewing the literature, the author exemplified three different ways how participant expectancies could limit the validity of RCTs. First, when blinding of participants fails any effects in the treatment group receiving active medication could be caused by participants expectations rather than the medication itself. Second, strong placebo effects caused by participant expectancies could create ceiling effects in both groups. This in term could hinder the detection of a beneficial drug effect. Third, participants' uncertainty about the treatment allocation in dbRCTs could lower their expectations, leading to weaker treatment responses, compared to clinical practice where patients do not have to wonder if they are receiving a real medication.

After the time period spanning from around 1990 to 2010 that was primarily aimed at quantifying the placebo effect, the focus shifted slowly towards its usefulness. This sparked an increasing interest in both meta-analytical investigations and experimental research into the underlying processes and factors associated with (larger) placebo effects as researchers and clinicians started to understand the placebo phenomenon's importance for the validity of placebo-controlled randomised trials and its potential to increase patients' health outcomes in clinical practice (Geers & Miller, 2014). The next section of this chapter introduces the most

recent neurobiological and psychological knowledge gains underlying the placebo phenomenon and outlines the most important definitions and conceptualisations within placebo research necessary for a comprehensive understanding of this dissertation.

A Biopsychosocial Approach explaining Placebo and Nocebo Effects

A complex and multidimensional phenomenon as the placebo or nocebo effect is best described using a biopsychosocial approach. Therefore, this section first gives an overview about the biological factors hypothesised to promote the placebo and nocebo phenomenon. Then a detailed discussion about psychological theories and models is presented. At the end of this section social and contextual factors contributing to placebo and nocebo effects are discussed.

Biological Underpinnings of the Placebo and Nocebo Effect

Several neuroscientific studies have shown that placebo and nocebo effects are associated with autonomic, neuroendocrine, and immune responses (Wager & Atlas, 2015). Most of the neurobiological research regarding the placebo effect was conducted in the area of pain because it offers elegant experimental procedures allowing the examination of neurobiological processes involved in placebo effects. Nakamura et al. (2012) were interested whether there were autonomic responses associated with a dose-dependent placebo analgesia. They used a common procedure to condition placebo responses. Here participants first received a strong painful stimulus. Then participants were conditioned to surreptitiously lowered pain stimuli to associate analgesia with the two placebo creams. This conditioning phase took place under the guise of

three analgesic creams that were presented with three different strengths, a strong analgesic, a weak analgesic, and a control cream. In the final test phase, all participants received the same painful stimulus together with either the strong, weak, or control cream. The results showed a gradual reduction in noxious autonomic responses in stimulus-evoked skin conductance, electro encephalogram activity in the N1-P2 electrodes, and pupil diameter that proportionally corresponded to the strength of the placebo creams. These results are in line with other research demonstrating a close link between the autonomic nervous system and placebo responses (Meissner, 2011).

Placebo and nocebo effects have also been associated with neuroendocrine systems involved in the human stress response and appetite regulation. Nocebo suggestions lead to increased peripheral cortisol levels (Johansen et al., 2003). Cortisol is typically seen as a biological marker of stress with higher levels indicating more stress. Interestingly, this response of the hypothalamic-pituitary-adrenocortical axis was elicited by verbal instructions alone. The verbal instructions that an already applied (placebo) treatment would further increase the pain participants were in was sufficient to increase the level of cortisol release (Johansen et al., 2003).

Placebos do not only influence the autonomic nervous system and the hormone excretion but also the immune system. In a pharmacological taste conditioning study, Goebel et al. (2002) had demonstrated for the first time ever that the human immune system can be conditioned by pairing a distinctly tasting drink with the application of the immunosuppressive drug cyclosporin A. First the drug was paired together with the drink in four sessions over three consecutive days. After one week, participants were re-exposed to the drink, but this time with placebo capsules. Giving the drink together with placebo capsules showed and induced suppression of the immune system in multiple factors.

The above-mentioned studies reviewed mechanisms and processes associated with placebo and nocebo effects. It is important to mention that the biological underpinnings are activated by psychological processes. The next section reviews the most influential psychological models and theories used in the current literature to explain placebo and nocebo effects.

Psychological Models

The prior section explaining the biological underpinnings of placebo and nocebo effects has already touched on the two most prominent psychological theories responsible, the expectancy theory and the conditioning theory. For a long time, proponents of the placebo expectancy and placebo conditioning theory argued about which one of the theories is better suited to explain placebo effects. Over the course of the debate, it became apparent that both theories are validly explaining separate parts responsible for the placebo and nocebo effect. Once it became clear that none of the two original theories exclusively explained the placebo or nocebo effect researchers discovered that the two theories in fact are entangled and best used in combination. With time researchers introduced the mindset model to explain the placebo and nocebo phenomenon. The next three subsections introduce the three theories in chronological order starting with the oldest one, the conditioning theory of placebo and nocebo effects.

The Conditioning Theory.

The conditioning theory of placebo and nocebo effects originates in the early work on conditioned reflexes conducted by Ivan Pavlov. In his famous experiments he trained dogs to learn the association between an auditory stimulus and the presentation of food (Pavlov, 1927).

The first study that reportedly used Pavlovian conditioning to elicit placebo responses in rats was published in the year 1962. Herrnstein (1962) demonstrated that it was possible to create a similar behavioural suppression with the mere application of a saline injection after a rat had been condition with scopolamine. The author argued that he did not see a reason why this simple paradigm based on Pavlovian conditioning should not elicit similar responses in humans. Some years later Wickramasekera (1980) came to the same conclusions arguing that the placebo phenomenon in humans could be described as a conditioned response. Humans could learn to associate the practitioner, the environment, or the method of administration when receiving actual active treatments. If there are sufficient repetitions to learn this relationship, the associated stimuli could elicit a conditioned response even without the active medication.

Conditioned placebo and nocebo responses have been observed by multiple studies relating to immunosuppression and analgesia to nociceptive stimuli, and nocebo hyperalgesia via reinforcement strategies (Ader & Cohen, 1982; Colagiuri et al., 2015; Colloca & Benedetti, 2006; Janssens et al., 2019; Voudouris et al., 1990).

Conditioning placebo and nocebo responses as one of the oldest paradigms of studying the psychological mechanisms has been studied across a large variety of health conditions, including itch (Bartels et al., 2014), nausea (Quinn & Colagiuri, 2016), asthma (Castes et al., 1998), immunosuppression (Ader, 2003; Goebel et al., 2002), Parkinson's Disease (Benedetti et al., 2004; de la Fuente-Fernandez et al., 2001), and most prominently in pain (Colloca & Benedetti, 2005). Placebo conditioning has also been used to support the treatment of attention deficit hyperactivity disorder and psoriasis, a skin condition where skin cells pathologically multiply and form build-ups of bumpy red patches (Ader et al., 2010; Sandler et al., 2010). The two researchers made clever use of conditioning processes involved when taking medications

and substituted the medications with placebos on some days of treatment. As discussed in great detail by Doering and Rief (2012) this procedure was labelled placebo-controlled dose reduction and offers the opportunity to reduce the overall applied dosage of a medication, hence reducing problems with side effects while upholding the effectiveness making use of a conditioned placebo response.

Learning processes like conditioning are clearly important for the acquisition of placebo responses and are closely associated with the formation of expectancies. Dependent on the prior experiences a patient forms expectations, both negative and positive according to the valence of the experience with medications, contexts, and practitioners. While the conditioning model can explain important aspects in the formation of placebo effects there have been studies conducted that elicited placebo and nocebo effects using verbal suggestions alone. The next subsection about the expectancy theory discusses these cases in more detail.

The Expectancy Theory.

As mentioned in the prior section about conditioning mechanisms, expectancies play an important role in eliciting placebo and nocebo responses. This section introduces the reader in more detail into the research that built the foundation of the expectancy theory of placebo and nocebo effects, what role expectancies play in eliciting placebo and nocebo effects and how expectancies can be used to increase placebo effects and decrease the extent of nocebo effects.

The terms expectancy and expectations are mostly used synonymously. Although, they are connected, it is useful to differentiate between the two terms in the context of placebo and nocebo effects. Expectations have typically been described as a cognition in the form of

conscious constructs or beliefs that can be verbally expressed by a patient (Corsi & Colloca, 2017), while expectancies have been specified as sub-consciously driven psychophysiological entities that do not have to be fully aware (Colloca, 2017; Laferton et al., 2017).

In a milestone publication for the response expectancy theory, Kirsch (1985) hypothesised that a person receiving a placebo experiences an effect because he or she expects an effect to occur. While most of the expected outcomes in learning theory about classical conditioning can be described as stimulus expectancies that are occurrences of external events like money or grades (Rescorla, 1988, 1991), Kirsch (1997) defined response expectancy as the anticipation of an automatic reaction to situations or behaviours. As examples, response expectancy for coffee consumption would be the expectation of feeling more alert or feeling less pain after taking a pain medication. The difference between stimulus expectancies and response expectancies lies in the fact that response expectancies are directly self-confirming as they are the result of one person's own course of action, stable, and less prone to extinction compared to stimulus expectancies in the context of placebo effects (Kirsch, 2018).

There is robust evidence that verbal instructions are capable of introducing strong nocebo responses. Benedetti et al. (2007) dedicated a whole review about the potential of instructions eliciting nocebo effects where he unravelled how negative verbal instructions alone triggered the activation of cholecystokinin which is associated with anticipatory anxiety and facilitates pain transmission. The power of verbal instructions alone eliciting placebo effects on the other hand does not have an equally robust evidence base at least in the context of clinical trials, as is discussed in more details later. A publication using an elegant methodology and with high clinical relevance demonstrated the power of different verbal suggestions with patients who had just undergone thoracic surgery for lung cancer (Pollo et al., 2001). Patients were all treated with

the powerful analgesic buprenorphine at request and received a basal intravenous infusion of saline. Pollo et al. (2001) randomised the patients into three groups receiving different instructions regarding the infusion. The natural history group did not receive any analgesic information, the classic double-blind administration group was either told that they would receive a painful painkiller or a placebo, and the last (deceptive) group was informed that the infusion was a potent painkiller. The results showed clear placebo effects and how powerful and clinically relevant these instructions can be. Over the first three days after the surgery the patients in the natural history group on average had requested 11.55 mg buprenorphine, the double-blind group only 9.15 mg, and the deceptive group had only requested 7.65 mg buprenorphine. Controlling for the level of analgesia the authors have demonstrated that instructions influence explicit expectations and can nearly cut the use of pain medication in half.

Investigations using open-hidden paradigms showed how important patients' explicit expectations are for the treatment outcomes. In open-hidden paradigms the application of a medication is either done hidden without any indication to the patient or clearly stated by a doctor or nurse in the open condition. In an experimental study with patients who had just undergone surgery the analgesic medication was either applied by a doctor in full visibility with the instruction that a painful painkiller had just been applied (open administration), or the medication was automatically applied by a machine without the patient's knowledge. Amanzio et al. (2001) demonstrated that the hidden administration of non-steroidal anti-inflammatory pain medication was only half as efficacious as the open administration. These results were later replicated by Bingel et al. (2011) who showed that even powerful opioid medications lose nearly a third of their efficacy when applied in a hidden compared to an open way. These investigations not only show how important expectations are for the placebo response, but also that the placebo

effect makes up a good proportion of normal interventions. Further, these studies show how easily clinicians can make use of simple instructions to boost the efficacy of treatment by making use of the placebo effect.

The open-hidden paradigm was crucial to showcase the importance of patients' expectations and showcased the importance of contextual factors in the environment as well as the role of the practitioner applying a treatment. A more comprehensive review of the importance of social and contextual factors is provided further down, but first, the next section focuses on the role of expectations in the broader mindfulness model and computational model of placebo and nocebo effects.

The Mindset Model.

Zion and Crum (2018) have defined mindsets as a lens or frame of mind that orients a person to a set of beliefs, associations, and expectations to guide attentional and motivational processes. This means that mindsets simplify complex information so individuals can make sense about themselves and their surroundings. In a health context, mindsets provide a framework, so a patient better understands a disease and the associated treatments. The authors argue that mindsets and expectations are closely connected, but that they are two distinct constructs. According to Zion and Crum (2018) expectations are best described as specific cognitions or beliefs about a future event, like expecting pain relief from a morphine injection. On the other hand, the authors explain that mindsets are more general and comprehensive psychological constructs that inform a patient about various mindset-consistent expectations. As an example, they state that a person might have the mindset that getting cancer represents a

personal catastrophe and might be associated with multiple different negative expectations like the expectation that the treatment will be painful or more complex expectations that the patient might not be able to cope with cancer treatment. The authors make a distinction between arguably simple expectations encompassing only simple cognitions that a treatment might work or not and the way more complex constructs of mindsets that act as a scaffolding helping patients understand the broader impact of treatments and illnesses.

Compared to the conditioning and expectancy theory, the mindset model has only recently been proposed. The effect of mindsets has been studied in a variety of conditions within the medical context and can be specifically targeted to help improve health outcomes across a variety of conditions ranging from blood pressure, weight loss, cortisol response, hormone secretion, immune function, and cognitive performance (Crum & Zuckerman, 2017; Crum et al., 2011; Crum & Langer, 2007; Crum et al., 2013). Mindsets are best understood as a more comprehensive model that originated from the expectancy theory that more holistically encompasses multiple specific expectations. Proponents of the mindset model are not just limiting mindsets to an aggregation of specific expectations, but also include a patient's belief about oneself and the beliefs how someone might change their mindset with interventions to alter expectations to try to improve outcomes, like coping with an illness for example (Yeager & Dweck, 2020). It is therefore most useful in the modern medical context where patients often have already existing mindsets based on prior experience with health problems, treatments, contexts, and health personnel.

Social and Contextual Factors influencing Placebo and Nocebo Effects

As we have learned from the prior section covering conditioning, expectancy, and mindsets all three models to some degree depend on the prior experience people have made that trigger placebo and nocebo responses. As humans we have evolved as a social species and our specially developed social cognitive skills have allowed us to flourish as a species (Herrmann et al., 2007). It is therefore not surprising how important observational learning and social influences, patient-provider interactions, and contextual factors can be in the general context of placebo and nocebo effects. This section introduces the most relevant aspects of social and contextual factors that are relevant to the context of this thesis.

A plethora of characteristics of a treatment can influence its effectiveness. Generally speaking, humans are affected by colour, mood might change according to colours, colours influence what products we buy, and might even change well-being and health (Küller et al., 2006; Lengen, 2015; Spies et al., 1997). A systematic review conducted by de Craen et al. (1996) found 12 studies that investigated the effect of colour on drugs' perceived effect and effectiveness. They found that red, yellow, and orange drugs were mostly associated with stimulant effects. Blue and green drugs were mainly associated with tranquillising effects. The evidence regarding the effect of colour on drugs effectiveness was not conclusive, but they reviewed some studies that indicated that colours might influence the effectiveness of drugs targeting the central nervous system. Similarly to the colour of drugs being associated with specific mechanisms of action, college students were asked to indicate the perceived strength of capsules and tablets and indicated that they perceived capsules to be more effective (Buckalew & Coffield, 1982).

Further characteristics of treatments that have been studied are a drug's price and if it is branded or generic. Espay et al. (2015) specifically conducted a study with the aim to test the effects of price on the placebo response in Parkinson patients. Using a cross-over design, patients either first received a cheap or expensive subcutaneous novel injectable dopamine agonist, that in fact both were placebo injections only containing saline solution. They observed improvement in both placebo conditions, but patients receiving the expensive placebo improved significantly more than the cheap one. Interestingly, the improvement with the expensive placebo was halfway between the cheap placebo and what is usually observed with levodopa, the actual dopamine agonist used for treatment of Parkinson's disease. Furthermore, the brain activity recorded during the functional magnetic resonance imaging showed that the brain activity of the expensive, but not cheap placebo mimicked the brain activity of levodopa.

These results convincingly demonstrated how important the perception of cost influences patients' placebo response. It is therefore not surprising that studies comparing branded versus generic treatments resulted in the same results. A cross-sectional survey with participants from the general population showed that people do not have the same level of trust in generic medications as in branded ones (Figueiras et al., 2008). Especially when medical conditions were perceived as more severe people reported to believe less in generic compared to the branded medication. An experimental examination of the effects of branded versus generic labelling used a clever design, where participants first took a placebo disguised as "Betaprol", a fast-acting beta-blocker for the treatment of pre-exam anxiety in a first session (Faasse et al., 2013). In session two participants were randomised to either a no-change condition again receiving "Betaprol", a branded change condition called "Novaprol", or a generic change condition. The no-change placebo condition staying on "Betaprol" demonstrated the greatest decreases in

systolic blood pressure and state anxiety compared to the branded and generic change. Interestingly, the change to generic group reported more side effects than the no-change group, or the branded change group.

This section has highlighted important factors influencing the placebo and nocebo effect. The next section introduces randomised controlled trials, and the specific aspects placebos play when it comes to the validity of the current gold standard of evidence-based medicine.

Randomised Controlled Trials – Failed Participant Blinding as Achilles' Heel

One of the known first clinical trials in the history of medicine was conducted by the Scottish naval physician James Lind (Lind, 1753). In the year 1747, he treated sailors suffering from similarly severe scurvy with six different treatments. The six conditions consisted of two sailors each and received either cider, vinegar, seawater, diluted sulphuric acid, or two different mixtures of food that contained nutmeg and garlic, or oranges and lemons. He thought it was important to standardise the sailors living conditions and diets to make sure the change in their condition could only be attributed to the interventions. After one week of treatment the two sailors in the fruit conditions had nearly recovered, while the other five groups' scurvy symptoms remained mostly unchanged. Unknowingly, Lind's trial already incorporated two key features of modern clinical trials, the comparison between at least two treatments and the attempt to control for participant characteristics. Over the last two and a half centuries clinical trials have become more sophisticated. The main improvements in methodology were the addition of randomisation to guarantee that participants or experimenters cannot influence the treatment allocation and the introduction of blinding to rule out that expectations of participants or

researchers bias the efficacy and safety estimations (Hackshaw, 2009). The current gold standard to evaluate the benefits and risks of drugs are randomised controlled trials that are most often double-blind, meaning participants and experimenters are both blinded and use a placebo as a control group (Hariton & Locascio, 2018). According to a guide about clinical trials by Hackshaw (2009), RCTs consist of four key design features, inclusion and exclusion criteria, control conditions, randomisation, and blinding. The goal of these design features is to guarantee the highest possible scientific validity.

RCTs are the highest form of evidence that can be empirically acquired in therapeutic studies (Oxford Centre for Evidence-Based Medicine, 2021). Only systematic reviews summarising well-conducted RCTs forming a homogeneous body of evidence trump single RCTs in the hierarchy of evidence-based medicine. RCTs being the gold standard to evaluate the effectiveness of new medications makes sense compared to the other lower-level sources of evidence like cohort studies or expert opinions at the end of the list. Although double-blind placebo-controlled clinical trials are the top of the evidence hierarchy, that does not mean that they are free of any limitations, as had already been discussed by Kaptchuk (2001).

So, what are the problems with the current gold standard of drug evaluation? One of the core underlying paradigms of randomised clinical trials is blinding of participants and experimenters to avoid any biases due to expectations for an accurate assessment of the drug's effectiveness and side effects (Hackshaw, 2009). When discussing the role of blinding in clinical trials it is always beneficial to imagine being a participant in one of these trials. Participants in clinical trials typically have to fulfil some clinical diagnosis or at least report that they are suffering from certain symptoms associated with a disease. In most cases clinical trials target specific ICD-10 diagnoses that by definition nearly always contain some aspects of suffering.

This implies that patients in RCTs hope to be allocated to the drug group and to get some relief for their condition. Naturally, being allocated to a (placebo) control group is often associated with disappointment and participants in control groups are more likely to withdraw from studies (Bell et al., 2013; Kemmler et al., 2005; Lindström et al., 2010).

The debate about problems with blinding in clinical trials nearly goes back to the inception of double-blind placebo-controlled randomised trials as the standard for pharmacological trials in the 1950s (Straus & von Ammon Cavanaugh, 1996). Researchers as early as the 1980s conducted literature reviews summarising the problem of unblinding due to side effects in pharmacological trials (Thomson, 1982). Their literature review found some publications dating back as early as 1959 mentioning that several researchers, trial workers, as well as participants had made comments about the potential of side effects unblinding investigators (Letemendia & Harris, 1959; Nash, 1962; Stallone et al., 1975). This criticism was especially important because RCTs about psychotropic interventions back then primarily focused on the study of lithium. Lithium is known to have relatively high rates of side effects, with individual side effects being experienced in as much as 70 percent of people taking lithium (Gitlin, 2016).

The first article about the success of blinding in modern pharmacological double-blind placebo-controlled randomised clinical trials included a random sample of 200 RCTs published in major general medical and psychiatric journals between the year 1998 and 2001. Fergusson et al. (2004) found that only seven out of 97 general medicine trials had provided evidence on the success of blinding and five of the seven trials reported problems with blinding. In the psychiatric trials only eight of 94 trials reported blinding success and four of them reported problems with blinding. Hróbjartsson et al. (2007) conducted a similar systematic review with a

much larger sample of trials randomly chosen from the Cochrane Central Register of Controlled Trials published in the year 2001. Only 31 out of the 1599 RCTs reported on the success of blinding. Out of the 31 trials, blinding was considered successful in 14, unclear in 10, and broken in seven. While these findings replicated the earlier review about blinding, the authors had additionally contacted 200 of the trialists that did not include any information regarding blinding in their publication to check whether they did not check the success of blinding while conducting the trial or if it was only omitted in the publication. From the 130 responses only 15 indicated that they had assessed the success of blinding. The authors concluded that blinding is rarely assessed at best, when trialists assessed blinding, they were uncertain how to proceed and analyse the success of blinding, and finally that the reporting of blinding was often incomplete. The authors urged that there is an important need for methods to assess blinding and improved reporting.

As such, although RCTs represent the current highest quality source of evidence, they do not come without problems. Evidence identifying problems with failed participant blinding possibly jeopardising the validity of double-blind randomised placebo-controlled trials dates back as long as RCTs existed.

Active Placebos

The most common example of impure placebos involves the use of a potent medication that is applied to a health condition without any indication or any reasonable biomedical mechanism causing an improvement. Linde et al. (2011) depicted a real-world example of a doctor prescribing an impure placebo. In their example, a patient with a suspected viral upper

respiratory tract infection is asking for an antibiotic that has helped when the patient had experienced infections in the past and the doctor complies with the demand and prescribes the antibiotic. Obviously, antibiotics are effective medications for the treatment of bacterial infections but are not indicated or effective for the treatment of viral infections.

The term impure placebo is primarily used to describe real world clinical settings. Using qualitative semi-structured interviews with primary care physicians, Fent et al. (2011) found that physicians mostly used pure and impure placebos in cases of non-severe diseases for which there were no satisfactorily somatic explanations. According to the doctors they mostly used complementary or alternative therapies.

While the concept of pure and impure placebos at first glance seems reasonable, there has been considerable conceptual criticism. Grünbaum (1986) primarily criticised that the vocabulary used to define the placebo was confusing and obscure, and that Shapiro and Morris' definition would constitute that the current scientific knowledge had to be sufficient to rule out that a placebo was pure and did not hold any unknown elements that in fact elicited a real reaction making it de facto a medication.

While criticism towards differentiating between pure and impure placebos has validity, for the scope of this dissertation the focus was set on the concept of active placebos. The terminology of active placebo has received increasing interest over the last decade, especially in the field of pharmacological trials. Jensen et al. (2017) defined active placebos as a control intervention that mimics the side effects of the pharmacological agent that is being studied without having any effects on the studied outcome variables. Unfortunately, there is no clear-cut difference between the definition of active and impure placebos.

From reviewing the scientific literature, the difference between the two concepts seems to stem primarily from the context that the different nomenclatures have been used, the purpose for using the placebos, and what was used as placebo. Impure placebos are mostly associated with being prescribed by GPs in the form of alternative medicines (e.g., homeopathic remedies) in everyday clinical settings to make use of psychological effects when there are no clear somatic causes and satisfy patients' demand to receive a treatment (Linde et al., 2011). Active placebos on the other hand are mainly medications used in the context of pharmacological RCTs with the distinct purpose to introduce side effects as closely matching the side effects caused by the drug under investigation with the purpose to increase the validity of RCTs (Jensen et al., 2017). The argumentation that conventional pure placebos being completely inert might jeopardize the blinding in RCTs has been discussed in the literature for some time (Bystritsky & Waikar, 1994). The first RCT using an active placebo as a control group was already published in 1961 and investigated a tricyclic-antidepressant called imipramine (Daneman, 1961). The authors used atropine, an anticholinergic drug to account for the side effects caused by the antidepressant medication under investigation.

Jensen et al. (2017) conducted a comprehensive systematic literature review of RCTs to investigate how often active placebos have been used in RCTs, characterised the different medications used as active placebos, and summarising methodological articles reviewing the benefits and problem associated with use of active placebos. To estimate how many trials used active placebos they randomly sampled 200 RCTs published in October 2013 and found that merely one single trial had used an active placebo as a control group. The systematic review of the past literature conducted by Jensen et al. (2017) resulted in a total of 89 RCTs with an active placebo group. The trials were published between 1961 and 2014 and they mainly investigated

pain conditions with 31 trials, depression with 22 trials, allergies with 12 publications. The most common drugs used as active placebos were antihistamines (e.g., diphenhydramine) contributing 35 percent of all active placebo uses. Anticholinergics (e.g., atropine), benzodiazepines (e.g., midazolam or diazepam), histamines, and benztropines were nearly equally frequently used as active placebos. Most active placebo medications were used to account for sedative and anticholinergic side effects like dry mouth, drowsiness, dizziness, nausea, sedation, and constipation.

Research Aims

Background

Placebos are one of the most widely used control methods in RCTs. In fact, double-blind randomised placebo-controlled trials (dbRCTs) are the current gold standard to evaluate new medications and are regarded as the highest-quality of evidence currently available in evidence-based medicine (Murad et al., 2016). Even though most placebo-controlled RCTs use placebos as a comparator, there is an argument to be made that the placebo effect has not yet been sufficiently researched to live up to the expectation that RCTs are the strongest possible source of evidence for the evaluation of new pharmacological treatments. One of the most prominent criticisms that have been mentioned in the context of the validity of dbRCTs it that there are often problems with failed participant blinding. Hróbjartsson et al. (2007) conducted a review of dbRCTs published within one year in the Cochrane Central Register of Controlled Trials. Out of 1,599 blinded trials only a minority of 31 even mentioned tests for the success of blinding and out of these only 45 percent considered blinding successful.

The efficacy of the active treatment and the experience of adverse events, or the lack thereof is the cause most often discussed as being responsible for failed participant blinding in RCTs. Findings from reviews and meta-analyses across different fields of pharmacological RCTs ranging from antidepressants, pain medications, to statins have concluded that an integral reason associated with failed participant blinding in dbRCTs is an imbalance in the experience of side effects between the drug and placebo groups (Colagiuri et al., 2019; Gupta et al., 2017; Kirsch, 2014).

As an early approach to rectify the discrepancy in experiencing side effects between the drug and placebo groups, researchers started using another prescription medication that caused the participants in the control group to experience distinctive side effects, but supposedly had no direct effect on the condition to be treated, labelled “active placebos”. Different types of active placebos used in past pharmacological RCTs have been summarised and evaluated by Jensen et al. (2017). They found that mostly antihistaminic, anticholinergic, and sedative drugs had been used as active placebos in past RCTs. However, the authors concluded that active placebos were rarely used (e.g., 0.5% of RCTs published in 2013) despite their promising merit as a methodological tool, and that they should be considered more often, especially when the expected effects of the pharmacological agent are modest and the risk of bias due to failed participant blinding is high.

Arguably, the above-mentioned approach to use a medication as active placebo is semantically misleading, if not inappropriate and comes with ethical as well as methodological problems. Methodological problems with medications as active placebos include unwanted indirect therapeutic effects because it is generally just assumed that they have no beneficial effects for the condition under investigation (Jensen et al., 2017). Further, most patients suffer

from comorbidities that might be alleviated by the medication used as placebo. This in turn might cause a general improvement in symptomatology, including the condition under investigation (Salamone, 2000). The most prominent ethical problem with the use of medications as active placebos are serious side effects, or even potentially lethal unwanted events.

To my knowledge, Rief and Glombiewski (2012) were the first to develop an experimental approach of an active placebo that did not use an established medication. They used a nasal spray containing capsaicin to create a tingling sensation in the nose, but otherwise no active pharmacological contents. Their milestone invention is an important advance in active placebo methodology that started a new way of thinking about active placebos and created a wide variety of possibilities for experimental investigations. However, while their active placebo elicits a noticeable sensation during treatment onset when the nasal spray is applied, one could argue that there is a need for an active placebo that can be applied orally (i.e., as pill, tablet, or capsule) given that oral medications are one of if not the most common route of administration for pharmacological interventions.

The general introduction of this thesis introduced the most common models and theories presented in the scientific literature to explaining placebo and nocebo effect. While most models or theories are intertwined to a certain degree, this thesis mainly relies on the expectancy model. What participants expect and believe is deemed the most important aspect when explaining placebo and nocebo effects in the context of double-blind placebo-controlled randomised studies.

Research Aims

The general aim of this thesis was to develop and evaluate a much-needed new model of an active placebo in the form of an orally administered capsule than can be widely administered in experimental research to gain insights into the effects of side effects on the placebo effect generally and in clinical trials specifically as a more accurate control condition. An overview about the central research aims of this thesis, the respective research approach that was chosen and key methodological features to achieve the aims is provided in Table 1.1.

Table 1.1

Overview of Detailed Research Aims and Methodological Approach

Aim	Chapter	Approach	Sample	Intervention & Comparator	Primary Outcomes	Setting
Investigate if the adverse event (AE) reporting in placebo groups mimic the drug, and if more AEs lead to larger placebo responses.	Chapter 2	Systematic Review and Meta-analysis	Adult patients with insomnia	Any pharmacological treatment that was compared to a placebo group	Subjective and objective outcomes for: • SOL • TST • Awake • SQ	Double-blind parallel randomised placebo-controlled trials
Proof of concept of the new active placebo model consisting of beetroot extract and oxalic acid.	Chapter 3	Experimental study	Adult and healthy students	• Active placebo • Conventional placebo • No-treatment control	• ISI • Urine colouration	Single-blind RCT with one week of baseline measures and one week of placebo treatment
Test if the active placebo elicits a larger placebo effect compared to a conventional placebo and if the experience of the target side effect influences participants' perceived treatment allocation	Chapter 4	Experimental study	Adults with insomnia	• Active placebo • Conventional placebo • No-treatment control	• PTA • ISI	Double-blind RCT with one week of baseline measures and one week of placebo treatment

Find out if giving participant either a positive, negative, or no framing information about the meaning of the experience of the target side effect (urine colouration) influences the placebo effect and the bothersomeness of experiencing given side effect	Chapter 5	Experimental study	Adults with insomnia	• Active placebo plus positive framing • Active placebo without framing • Active placebo plus negative framing • No-treatment control	• ISI • Bothersomeness of the target side effect	Single-blind RCT with one week of baseline measures and one week of placebo treatment
Calculate pooled analysis across the data collected from all three experimental studies	Chapter 6	Post-hoc data analysis	Adult and healthy students and adults with insomnia	• Active placebos • Conventional placebos • No-treatment control	• ISI	Single-blind and double-blind RCTs with one week of baseline measures and one week of placebo treatment

Note. ISI, Insomnia Severity Index; PTA, perceived treatment allocation; SOL, sleep onset latency; TST, total sleep time; Awake, number of nightly awakenings; SQ, sleep quality.

Thesis Structure

This thesis is structured in seven distinct chapters. As a first step a systematic review and meta-analysis was conducted to compare adverse event rates and profiles between placebo and drug groups of pharmacological RCTs and investigating the association between adverse event rates and the placebo response (Chapter 2). In a first experimental proof-of-concept study (Study 1) the new model of an active placebo was tested about its reliability to elicit the targeted side effect (Chapter 3). Study 2 compared the active placebo against a conventional (lactose) placebo to evaluate its influence on the placebo effect and participants' perceived treatment allocation (Chapter 4). Study 3 focused on investigating how the information given to participants receiving the active placebo influences the efficacy and experience of side effects (Chapter 5). Following the three experimental chapters a pooled data analysis consisting of participant data from all three experimental studies was carried out (Chapter 6). Concluding this thesis, a general discussion (Chapter 7) reviews the main findings from the original empirical research conducted for this thesis, puts the research conducted in context regarding its strengths and weaknesses, and finishes with implications for pharmacological RCT, clinical practice, and future directions for experimental research.

Chapter 2: Meta-Analysis Side Effects in Insomnia RCTs and the Placebo Response

This chapter reports the findings of a systematic literature review and meta-analysis that was conducted to compare adverse event rates and profiles between placebo and drug groups of pharmacological RCTs and investigated if there was an association between adverse event rates and the placebo response.

Introduction

Pharmacological interventions for insomnia are widely used with approximately 13 million people in the US, 2.3 million in Australia, and 2.5 million inhabitants of the German adult population report using prescription sleep medication in the past month (Adams et al., 2017; Chong et al., 2013; Techniker Krankenkasse, 2017). Despite their widespread use and potential benefits, many people with insomnia report concerns about pharmacological treatments, in particular about the possible side effects (Stinson et al., 2006). Patients frequently cease pharmacological treatment for insomnia due to side effects or fear of consequences associated with long-term use (Barter & Cormack, 1996; Iliffe et al., 2004; Siriwardena et al., 2008). Patients most often complain about daytime drowsiness, headache, dizziness, and falls in elderly populations. Cheung et al. (2018) found that most patients expressed concerns about side effects and long-term worries about dependence.

However, there is increasing evidence indicating that not all side effects are attributable to the medication itself. Instead, negative information and expectancies can contribute to side effects via the nocebo effect. In relation to side effects, multiple studies show that simply warning patients about side effects can increase their occurrence (Colagiuri et al., 2012; Mondaini et al., 2007; Neukirch & Colagiuri, 2015). For example, Mondaini et al. (2007)

compared side effects in a RCT of finasteride for benign prostatic hyperplasia, where one group was informed about the possibility of erectile dysfunction, decreased libido, and problems with ejaculation as side effects and another group was not. Patients informed about the potential, but uncommon side effects had a significantly higher (44%) proportion of one or more sexual side effects compared to those not warned (15%). Nocebo-induced side effects have also been demonstrated in experimental models of sleep difficulty where warning participants about side effects has increased their occurrence when they were given placebos but told that it was a medication to treat their insomnia (Colagiuri et al., 2012; Neukirch & Colagiuri, 2015).

The existence of nocebo-induced side effects creates a paradox whereby warning individuals about potential side effects during the informed consent process may actually cause them harm via the nocebo effect. While there have been some attempts to develop communication strategies to minimise nocebo effects caused by side effect information (e.g., positive framing, (Barnes et al., 2019)), what constitutes sufficient evidence for a side effect to be listed for a medication in the first place has received less attention. In clinical trials, the efficacy of a medication is taken as the difference between the active drug and the placebo. However, in terms of side effects, it is common for any side effect reported by those receiving the active drug to be listed as a potential side effect, irrespective of whether or not they occurred at an equivalent rate in those receiving the placebo. Meta-analysis indicates very little difference in side effects reported between patients receiving statins versus placebo in double-blind RCTs (Finegold et al., 2014), particularly with regard to muscle pain, one of the most frequently reported side effects. Yet, statin drug information leaflets commonly contain over 50 listed side effects (e.g. simvastatin (Merck & Co., 2021)) with this negative information believed to directly contribute to statin side effects via the nocebo effect. This is consistent with Tan et al. (2014)

analysis of 15 commonly prescribed drugs indicating that side effect information in medication leaflets was inconsistent, excessive, and frequently overlapped with nonspecific general daily symptoms, e.g. fatigue, headache, that were unlikely to actually be attributable to the drugs themselves.

In order to develop more accurate side effect profiles of drugs and to reduce nocebo-induced side effects caused by drug information leaflets, it is critical to understand side effects that occur within placebo groups, which, by definition, cannot be attributable to the drug. Meta-analyses of placebo-controlled RCTs in depression and migraines that show that side effects reported by placebo groups differ depending on the class of drug administered and typically mimic the active drug side effects (Amanzio et al., 2009; Rief et al., 2009). This suggests that side effects reported by placebo participants are influenced by the contextual factors surrounding the clinical trial, consistent with the nocebo effect. However, to date, there has been no systematic analysis of side effects in placebo groups of pharmacological trials for insomnia and it is unclear whether similar processes apply as were demonstrated in migraine and depression RCTs.

To address this gap, the main aim of this systematic literature review and meta-analysis was to build on the findings from the experimental research about the nocebo effect and the meta-analyses about depression and migraine RCTs (Amanzio et al., 2009; Rief et al., 2009) to investigate whether the adverse event reporting in placebo groups mimic the drug in the context of double-blind parallel placebo-controlled trials for pharmacological treatments of insomnia in adult patients. Therefore, two main hypotheses were formulated. First, we expected that the adverse event rate of placebo participants reporting at least one adverse event will differ depending on the class of drug being investigated (hypothesis 1.1). Second, we predicted that

adverse event profiles, measured as the percentage of placebo participants experiencing a specific adverse event in placebo groups of these trials will mimic the adverse event profile of their corresponding drug group (hypothesis 1.2). As an additional exploratory analysis, it was also investigated if larger rates of placebo participants reporting at least one adverse event were associated with higher placebo responses (hypothesis 2).

Methods

The protocol for this systematic literature review and meta-analysis was pre-registered via PROSPERO, where full details about the search and analysis plan is available (CRD42018097395; Werner et al., 2018). This meta-analysis is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher et al., 2009).

Selection Criteria

This analysis focused on results from published parallel double-blind RCTs testing pharmacological interventions for adult patients with insomnia. Published peer-reviewed articles and dissertations were included in any language as long as one of the authors clearly understood the article's content, this included English, German, French, Italian, and Dutch. All articles classified as reviews, theoretical articles, meta-analyses, conference posters or abstracts were excluded. The following criteria for inclusion and exclusion were defined before the literature search according to a guideline for meta-analyses in mental health research (Cuijpers, 2016).

Sample:

Samples were restricted to adults with physician diagnosed insomnia. The majority of participants (≥80%) had to meet at least one of the following criteria: (1) have a predominant complaint of dissatisfaction with sleep quality or quantity associated with difficulty initiating or maintaining sleep, or early-morning awakening of at least four weeks. The four-week cut point for insomnia is considered long enough to eliminate studies involving transient insomnia and short enough to include studies involving persistent insomnia. (2) the authors of the RCT stated that the participants had insomnia. This definition combines the approach by Buscemi et al. (2007) and the criteria for insomnia disorder (307.42) according to the Diagnostic and Statistical Manual of Mental Disorders fifth edition (DSM-5; American Psychiatric Association, 2013).

Interventions:

Based on a systematic review of the most commonly used pharmacological treatments for insomnia, we restricted interventions to prescription drugs approved by the Food and Drug Administration (FDA), over-the-counter drugs, off-label treatments, and herbal therapies (Lie et al., 2015). Following Lie and colleagues, treatments were categorised into drug classes as benzodiazepines, nonbenzodiazepines, melatonin agonists, tricyclic antidepressants, barbiturates, orexin receptor antagonists, atypical antipsychotics, antihistamines, other antidepressants, or herbal remedies.

Study design:

The study had to employ a double-blind randomised placebo-controlled parallel design.

Type of placebo:

To guarantee standardisation the placebo comparator had to be a conventional, inert placebo, i.e., with no active ingredients. This means that active placebos, such as atropine, vitamins, or other medications, that are not intended to treat insomnia but have been used as a comparator were excluded because they have properties themselves that are able to elicit adverse events.

Search Strategy

To identify all relevant articles EMBASE, PsycINFO, MEDLINE, CENTRAL, and Web of Science were searched with combinations of key words and text words regarding pharmacological treatments, insomnia, and RCTs. To search for grey literature and to decrease search bias ClinicalTrials.gov, International Clinical Trials Registry Platform, FDA Center for Drug Evaluation and Research, LILIACS, Australian New Zealand Clinical Trials Registry, African Index Medicus, and ProQuest Dissertation and Theses were searched. Additionally, one author (CW) searched reference lists of relevant systematic reviews and meta-analyses (Belanger et al., 2007; Brzezinski et al., 2005; Buscemi et al., 2007; Dundar et al., 2004; Ferracioli-Oda et al., 2013; Glass et al., 2005; Holbrook et al., 2000; McCall et al., 2003; Smith et al., 2002) for further eligible studies. The literature search included the timeframe from inception until March 6th, 2020. For the detailed search terms see eTable 1 in Appendix A

Outcomes Measures

The primary outcome of interest in this meta-analysis was the rate of placebo participants reporting at least one adverse event and the rate of specific adverse events, defined as any symptoms that have first occurred or worsened in severity after initiation of treatment. Sleep related outcomes defined as sleep onset latency, total sleep time, number of awakenings and quality of sleep were extracted for subjective and objective measures as secondary outcomes. If a study reported more than one type of measure for any of the above-mentioned sleep outcomes (i.e., subjective sleep quality as assessed by the Insomnia Severity Index and the Pittsburgh Sleep Quality Inventory), then the measure with higher test-retest reliability ratings according to the literature were extracted. The pre-registration of the analyses on PROSPERO listed the assessment of drop-out rates between placebo groups in trials about different drug classes as a secondary outcome, but there were insufficient data to perform any meaningful analyses.

Risk of Bias and Quality of Evidence Assessment

Risk of bias (RoB) was assessed using the Cochrane risk of bias tool (Cumpston et al., 2019). Selective reporting was coded according to whether authors specified main outcomes a priori and whether they reported measures for all mentioned measures. Blinding of assessors concerned whether studies reported observer-rated outcomes and if outcome assessors were blinded. Selective attrition was coded using information on whether studies reported missing or complete outcome data. To assess the quality of generation of allocation sequence the implementation strategy had to be adequately implemented and the concealment had to be independent. Randomization concerned whether authors stated somewhere that assignment of participants to treatments was randomized or if they used any synonymous description. The last

item assessing risk of bias was whether authors calculated intention-to-treat analyses or only completer analyses. Total scores of these ratings were used to assess overall risk of bias and individual study quality ratings.

The quality of evidence (QoE) assessment was conducted using the Cochrane GRADE approach, which assesses grades of recommendation, assessment, development, and evaluation (Schünemann et al., 2013). Limitations in study design or execution were based on allocation concealment, lack of blinding, incomplete accounting of patients and outcome events, selective outcome reporting. Indirectness of evidence was based on differences in population, intervention, outcomes, comparisons. Inconsistency of results was defined as present if the study had diverging results on different outcome categories (e.g., inconsistency was rated as low QoE for an individual study if total sleep time significantly improved and sleep onset latency worsened). Imprecision was rated as low QoE if studies include relatively few patients and few events and thus had a wide confidence interval or large standard deviation compared to body of studies. Publication bias for the GRADE approach was rated as low quality of evidence if the study reported preliminary findings, was labelled as a pilot study, or if it was a relatively old publication or the journal was not international where the authors could not confirm if the journals already used a peer-review process with at least two independent reviewers. As this was not a meta-analytical investigation in the conventional way investigating the effect of an interventions on different outcomes, QoE was not assessed for different types of outcome domains.

Both RoB and QoE were independently assessed by two authors (CW & SH). Disagreements were resolved through discussion between the two authors.

Screening and Data Extraction

Screening and data extraction were undertaken independently by two researchers (CW and SH) and discrepancies were resolved through discussion. For details for the screening process see Figure 2.1.

The aggregation of specific adverse event data followed the MedDRA (Medical Dictionary for Regulatory Activities) hierarchy for classifying different adverse events and their suggested terminology (Brown et al., 1999).

Meta-analytic Strategy

Data extraction.

For the purpose of effect size calculation, group means (*M*), corresponding standard deviations (*SD*) and group sizes (*N*) were extracted primarily. In cases where one of these values were missing, other statistical data was extracted if they could be converted into means and standard deviations according to formulas provided by Borenstein et al. (2011), Cumpston et al. (2019), and Weir et al. (2018).

When studies reported only standard errors, these were directly converted to standard deviations. For the transformation of medians into means and interquartile range into standard deviation, a normal distribution was assumed, unless authors of an RCT stated otherwise. If studies reported only adjusted outcome values, then these data were extracted, but sensitivity analyses were calculated without these studies to check for possible bias. If only the total included sample, but not individual group sizes were reported, the total sample was divided by

the number of groups to receive individual group sizes, because of the random allocation of participants to study groups. If studies did not report measures of uncertainty like standard deviations, standard errors, or confidence intervals, then they were excluded from analyses.

Effect size calculation.

To calculate meta-analyses and meta-regressions the package metafor (Viechtbauer, 2010) and the open-source software environment R 3.5.3 (R Core Team, 2019a) was used. The adverse event rate was included as the percentage of participants in a given placebo group reporting at least one adverse event and the specific adverse event profiles as the percentage of participants in a group that reported a specific adverse event (i.e., 20 percent of placebo participants in a study reported headaches). In terms of changes in sleep, within-group standardised mean changes (SMC) were calculated for all placebo groups between baseline and post-intervention measures. SMC were calculated using the escalc() function with the standardised mean change using change score standardisation (SMCC) method, to account for the fact that the groups are dependent. This dependency was included in the calculation using the correlation coefficient of the pre- and postscore. These correlations were planned to be extracted based on coefficients reported within papers, inferred from given test statistics from repeated measure tests, and from test-retest reliability of studies evaluating the scales. However, there was insufficient data to do this for any study. Therefore, a simple correlation coefficient was calculated based on the baseline and post-intervention means of the sleep measures that had been extracted. Then sensitivity analyses were calculated using varying correlation coefficients ranging from zero to one to test the robustness of the resulting effect size estimations. For the purpose of this meta-analysis all within-group comparisons over different measure points are

presented in the same way with positive effect sizes implying an improvement in the respective sleep outcome.

The rma.uni() and rma.mv() functions were used to run multivariate/multilevel (mixed-effect) linear models using restricted maximum-likelihood estimation. Multi-level random-effects-type meta-analytical models are specifically designed to account for non-independence among effect sizes by allowing for the addition of random terms in the model (Nakagawa & Santos, 2012; Viechtbauer, 2010).

For the first analysis (hypothesis 1.1) it was predicted that the adverse event rate of placebo participants reporting at least one adverse event in the placebo groups would be predicted by the factor drug category using the rma.uni() function as each included study only reported one adverse event data point. For the specific adverse event profiles (hypothesis 1.2), the specific adverse event rates of the placebo groups were predicted using the factor drug category with the rma.mv() function. For this analysis, the adverse event type was introduced as an inner and the study as the outer random term in each model to account for the potential lack of independence among adverse event rates derived from the same article and to account for the fact that not every article reported all possible specific adverse events. To analyse the possible relationship of the rate of placebo participants at least reporting one adverse event on the placebo response (hypothesis 2), the within-group effect size was predicted based on the adverse event rate using the rma.mv() function. Here the effect size type (i.e., subjective sleep onset latency, objective total sleep time etc.) was added as the inner random term to account for possible associations between the sleep outcome measures and the study as the outer random term to account for the possibility of a lack of independence among multiple effect sizes from the same study.

Moderators.

Moderator analyses for seven continuous (drug dosage, treatment duration, publication year, baseline severity, and participant age) and four categorical moderators (adverse event assessment strategy, comorbidity, region, and funding source) were analysed using meta-regression models for both adverse event outcomes and sleep related outcome measures.

Based on Fu et al. (2010), meta-regressive analyses were only performed when a minimum of four studies for categorical and a minimum of 10 studies for continuous variables reported outcomes.

Publication bias.

Risk of publication bias was visually evaluated using funnel plots and statistically by adding the standard deviation of the effect sizes as a moderator to the multivariate/multilevel linear (mixed-effects) models, since neither leave-one-out nor trim-and-fill apply for these models. This procedure follows the same logic as Egger's regression test (Sterne & Egger, 2005) and was already implemented in a meta-analysis by Habeck and Schultz (2015). The Funnel plots and analyses regarding the publication bias can be found in Appendix A. Although some asymmetry was observed indicating a publication bias, the pattern of the main analyses remained unchanged. For the ease of readability, I decided to report the main analyses that are not adjusted for the publication bias in the Result section and added all R outputs in Appendix A.

Outliers and influential data.

Data were defined as outliers if standardized residuals were greater as three or as influential points if hat values were larger as two times the average hat value (Aguinis et al.,

2013; Stevens, 1984; Viechtbauer, 2010). It was planned to calculate statistical models with and without data points that fulfilled both criteria and to compare them using a Wald-type test (Viechtbauer, 2010). No single data point in the three main analyses fulfilled both criteria. Scatter plots depicting hat values and standardized residuals are in Appendix A.

Sensitivity analysis.

Sensitivity analyses were conducted to account for the possibility of systematic differences in study characteristics. Thus, the influence of studies reporting type of study (journal article vs. dissertation), risk of bias, and quality of evidence were assessed using meta-regression in accordance with the procedures which were previously described (Cumpston et al., 2019; Viechtbauer, 2010).

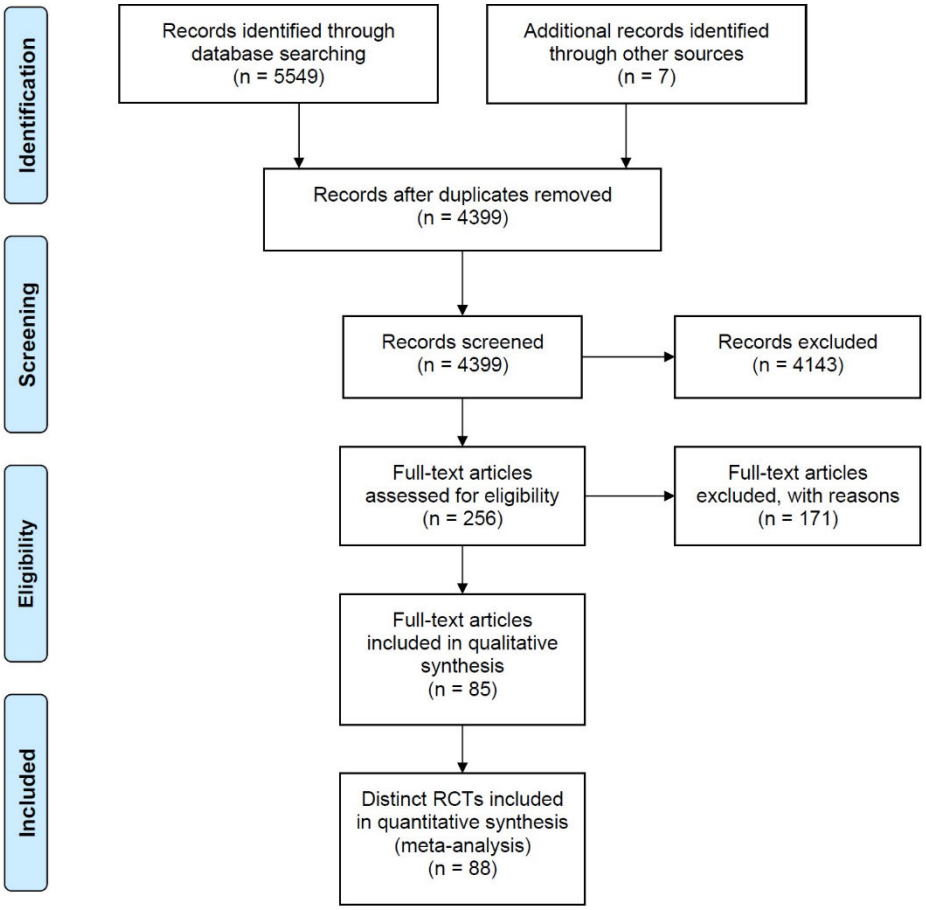
Results

Study Selection and Characteristics

Details concerning the screening process are illustrated in Figure 2.1. The systematic literature search revealed 4,399 studies. After the title and abstract screening 256 full-text publications were potentially relevant. After the full-text review 85 articles reporting 88 distinct RCTs were included in the data extraction. The retrieved articles were published between 1973 and 2018. The total number of participants included in this meta-analysis is 27,885 (61.49% women) ranging from N = 24 to N = 1,155 participants in each study. Across the studies participants' mean age was 50.69 years with study means ranging from 29.55 to 81.00 years of age.

Figure 2.1

Consort Flow Chart of Study Inclusion Process

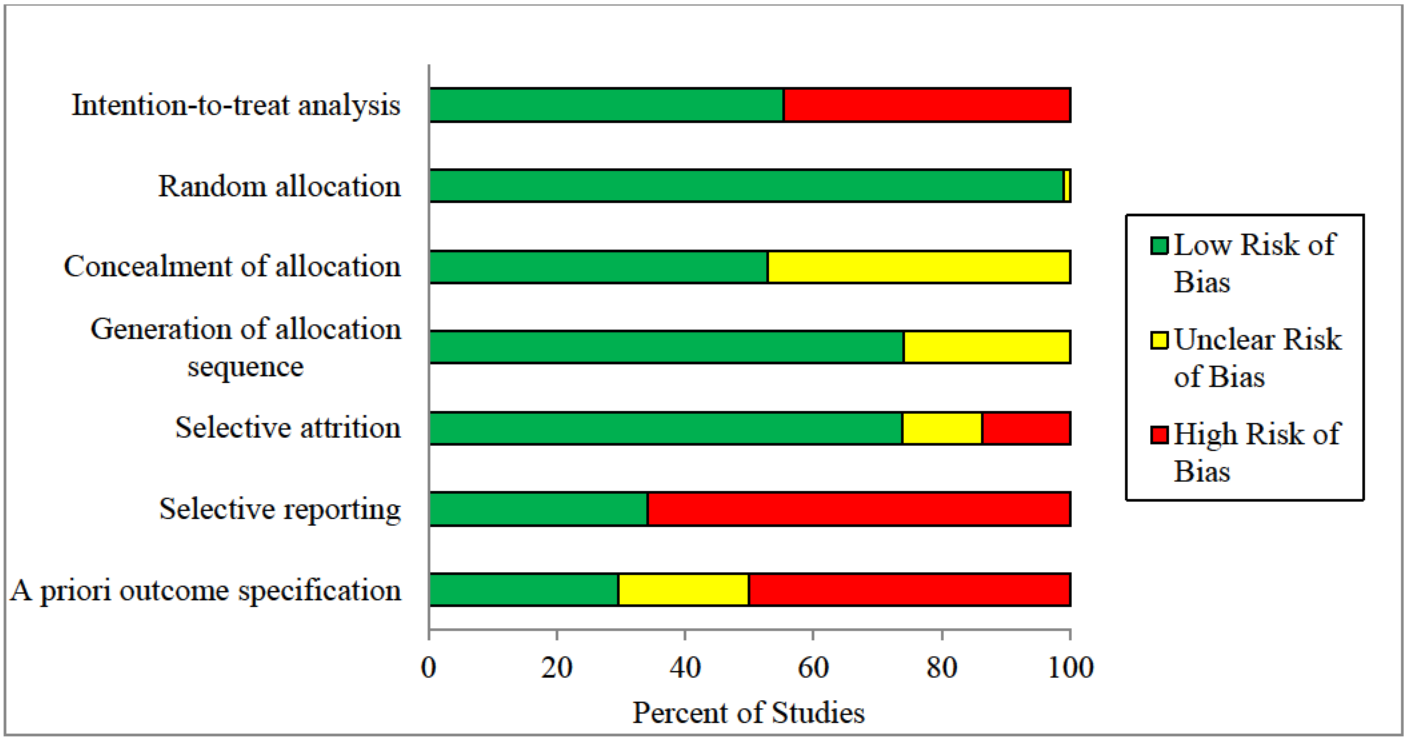


The included RCTs reported a total of 88 placebo groups, 42 benzodiazepine groups, 42 nonbenzodiazepine groups, 16 melatonin agonist groups, 12 gaboxadol groups, 10 orexin receptor antagonist groups, seven tricyclic antidepressant groups, four herbal remedy groups, three SSRI/SNRI antidepressant groups, two barbiturate groups, and one antihistamine group. The mean treatment duration was 39.44 days and ranged from one day to a maximum of one year. For further details see eTable 2 in Appendix A.

Figure 2.2 shows how many percent of the included studies were rated as low, moderate, or high risk of bias for each domain. Risk of bias assessment revealed that the majority of assessed domains showed low risk of bias. The domains raising concerns for bias were the selective reporting of outcomes, the a priori specifications of outcomes, and the intention-to-treat analysis.

Figure 2.2

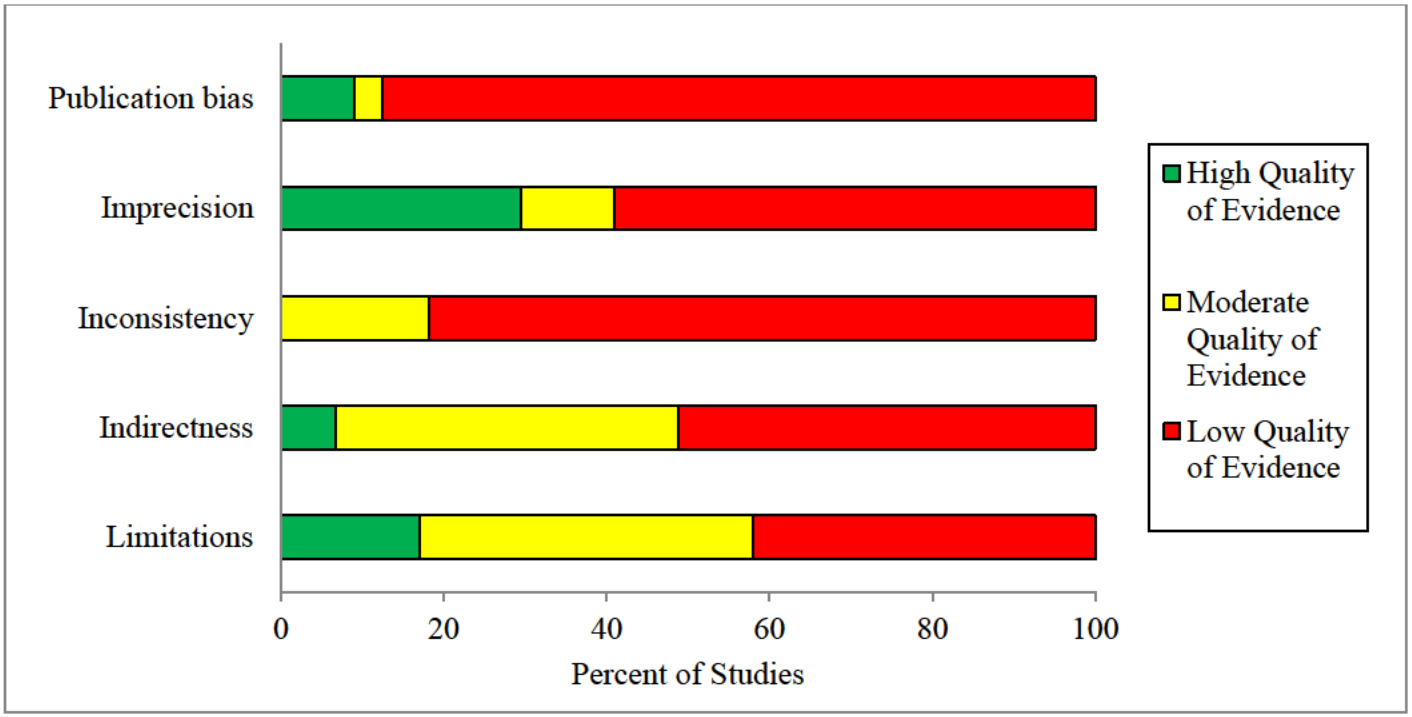
Risk of Bias Assessment



The quality of evidence assessment, shown in Figure 2.3 revealed that most of the evidence has to be rated as low quality while only a small proportion of the evidence can be rated as high quality. It was observed that most issues with QoE stemmed from older studies and that more recent studies generally achieved higher QoE ratings.

Figure 2.3

Quality of Evidence Assessment according to the GRADE approach

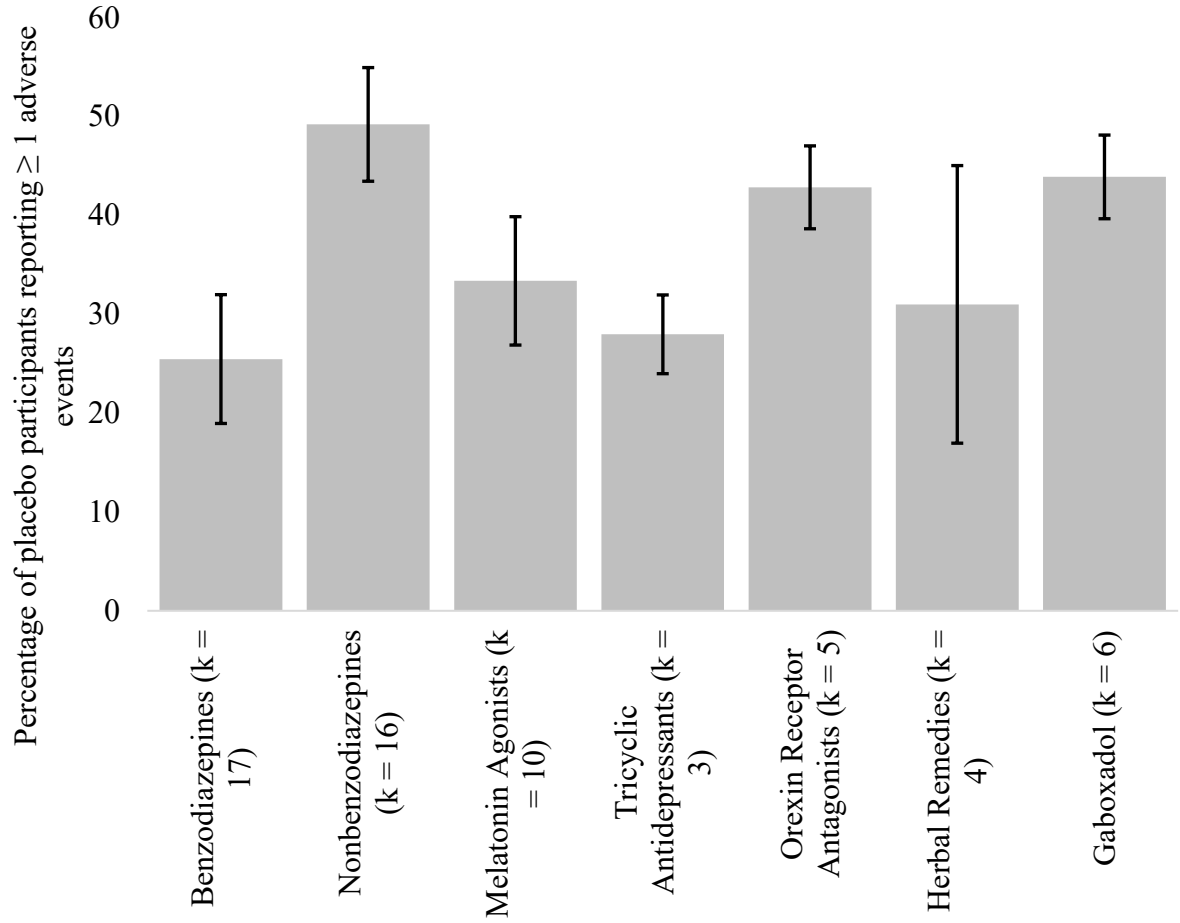


Adverse Events Across Placebo Groups of Different Drug Trials

To analyse how the rate of placebo participants that reported one or more adverse events differed across trials investigating different drug classes data from seven drug classes including a total of 61 placebo groups was available. The mean rate of placebo participants who reported at least one adverse event is depicted in Figure 2.4 and ranged from ~27% for benzodiazepines hypnotics to ~50% for non-benzodiazepine hypnotics. The Q statistic was calculated for the primary omnibus test comparing the adverse event rates. When statistically significant this indicates that the between group variation is greater than would be expected by chance alone. The difference between the adverse event rates between the different drug groups was statistically significant with $Q(6,61) = 17.33$, $p = .0081$.

Figure 2.4

Rate of Placebo Participants Reporting One or More Adverse Events Across Different Drug Classes



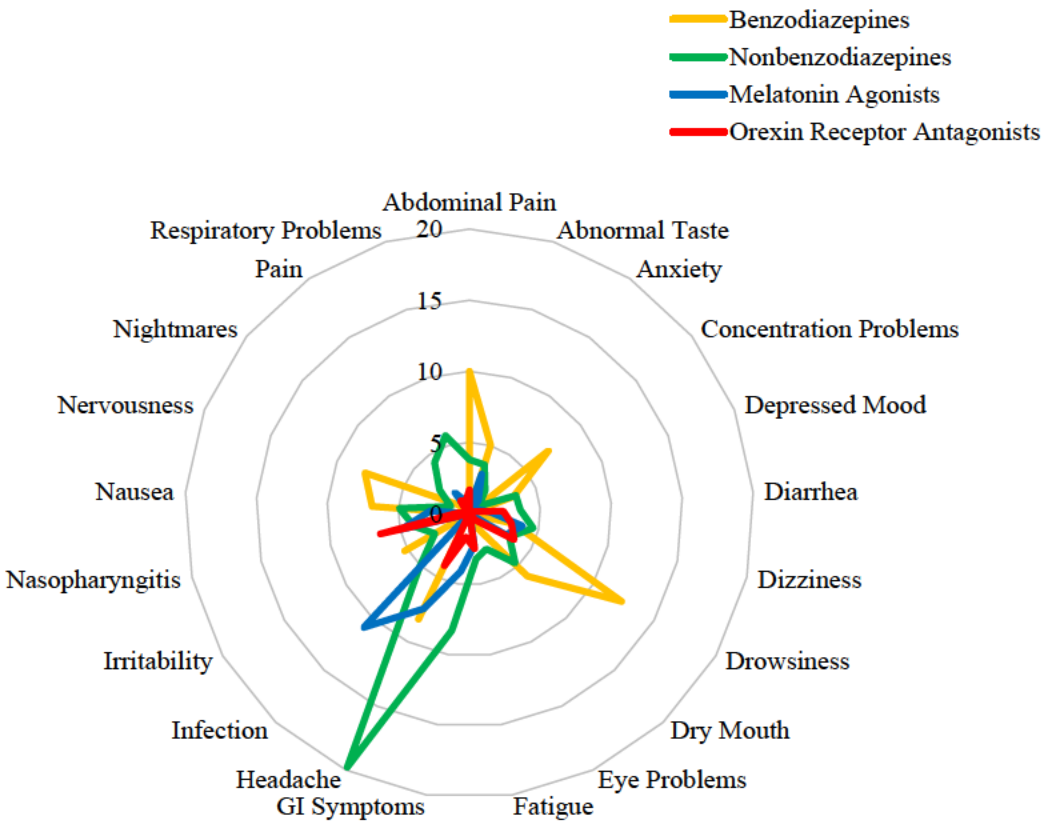
Note. Error bars depict the standard error of the mean. k refers to the number of placebo groups in each drug class.

The specific adverse event profiles of placebo participants of all different drug classes in this review included 33 distinct adverse event types. Figure 2.5 shows an illustrative, not comprehensive example of the most commonly reported adverse events across drug classes. From the 70 papers reporting specific adverse events it was possible to include a total of 582

specific AE reports in this analysis. For the statistical comparison of the specific adverse events between groups a multivariate model was used predicting the percentage of a specific adverse event in the different placebo groups based on the drug class investigated. An inner random term was included for the type of adverse event (i.e., nausea, pain etc.) and an outer random term on the study level to account for the fact that some studies contributed more adverse events than others. The reported Q statistic, when statistically significant indicates that the between group variation in the adverse events is greater than would be expected by chance alone. The percentage of specific adverse events reported differed significantly between placebo groups in trials of different drug classes, $Q(7,697) = 66.10$, $p < .0001$.

Figure 2.5

Profiles of Specific Adverse Event Rates of Placebo Participants Across Different Drug Classes



Note. This spider chart illustrates the specific adverse event profiles of placebo participants across drug classes. For illustrative purposes I decided to show adverse events for the four drug classes that made up the majority of adverse event reporting and that included sufficient data on specific adverse events, i.e., at least 75 percent of specific adverse events were reported.

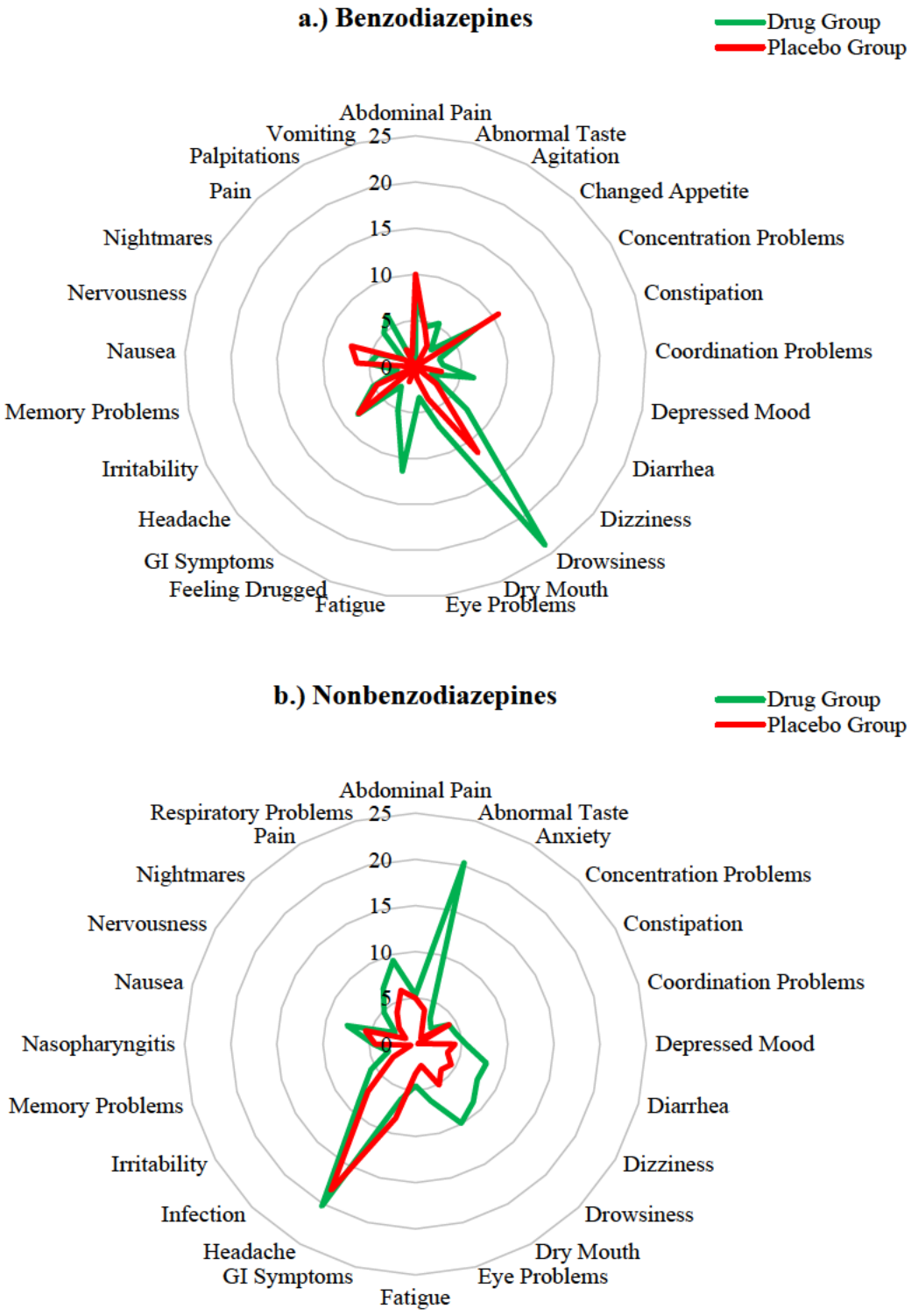
Adverse Event Profiles between Placebo and their Respective Drug Groups

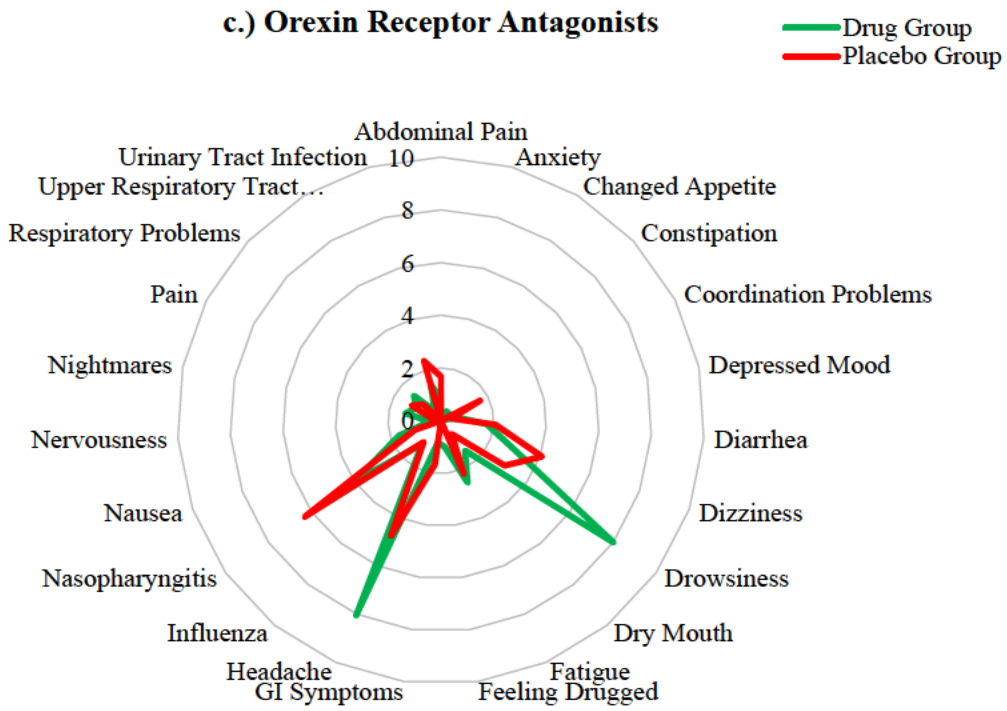
Figure 2.6 illustrates the close association of specific adverse event rates for the head-to-head comparison between the placebo and corresponding drug groups. To analyse the adverse event profiles between the placebo and their corresponding drug groups across all drug

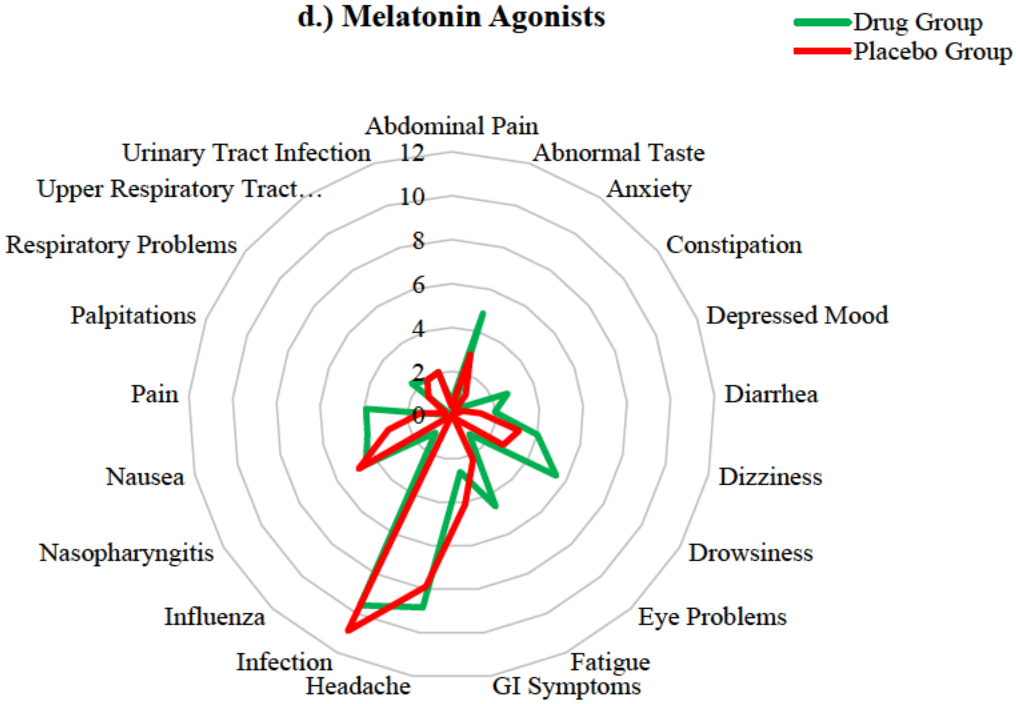
classes a multilevel Pearson's correlation was calculated with a random term on the study level to adjust for the fact that not every study contributed the same level of adverse event information. A total sample of 100 specific adverse event reports where information was available for the placebo and the corresponding drug group was available. There was a statistically significant and strong association with the specific adverse event reports in the placebo groups closely following the rates in the respective drug groups, $r(98) = .44$ $p < .001$, with 95% confidence intervals for the correlation coefficient ranging from 0.26 to 0.58.

Figure 2.6

Comparison of Specific Adverse Event Profiles between Placebo and the Corresponding Drug Groups







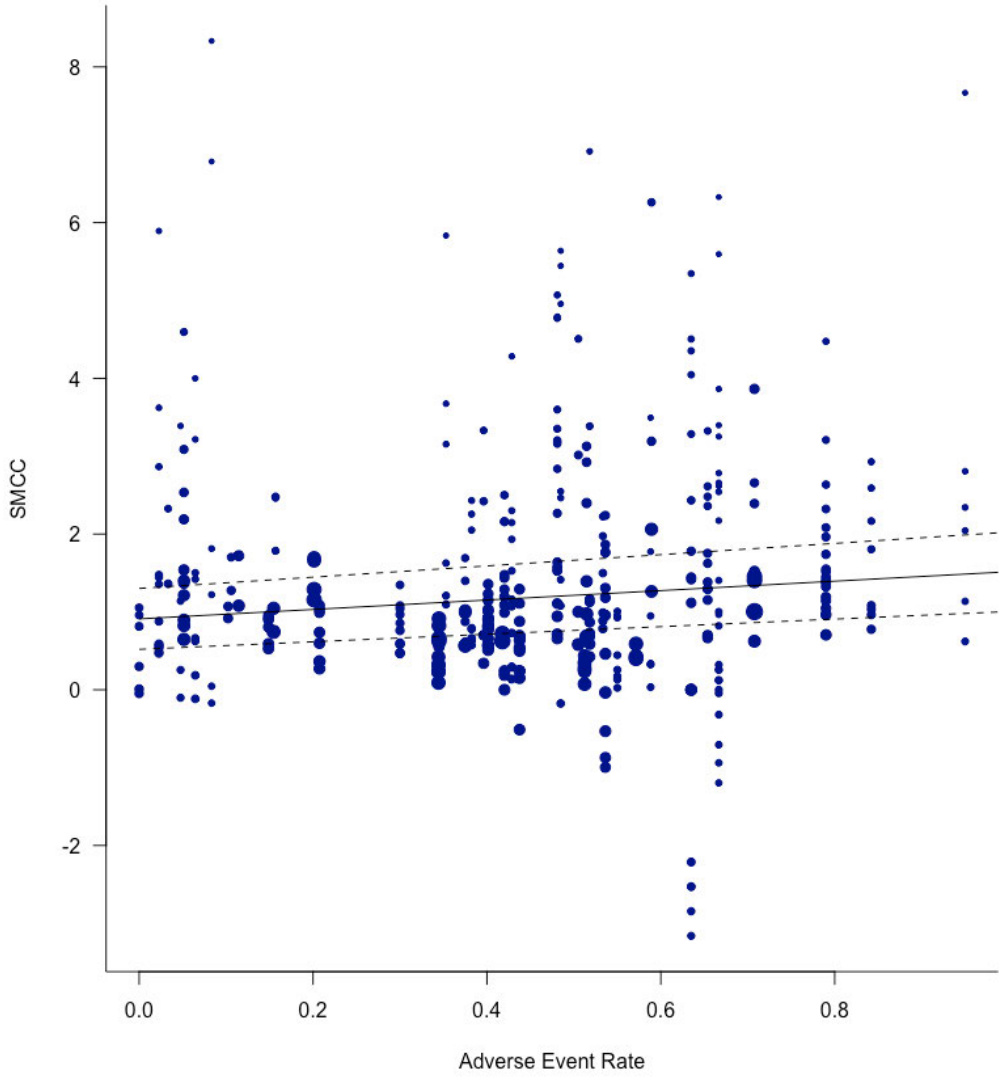
Note. These spider charts illustrate the specific adverse event profiles (in percentage) between the placebo and their relative drug group for the four drug classes providing most adverse event data.

Association between Adverse Event Rate and the Placebo Response

The relationship between the rate of placebo participants reporting at least one adverse event and the placebo response, i.e., the improvement in sleep is shown in Figure 2.7. To analyse if the adverse event rate of placebo groups predicted the placebo response, a total number of 382 effect sizes were available (comprised of the eight outcome measures defined in the method section) and the corresponding adverse event rates within the placebo groups from a total of 49 RCTs. A multivariate meta-analysis model was used predicting the within-group placebo response from baseline to post-intervention using the adverse event rate. The advantage of the multivariate model is that it is possible adjust the model for dependencies using additional random terms. The inner random term was used to adjust for the multiple effect size types included in the model, while the outer random term was chosen on the study level to adjust for the fact that some studies contributed more effect sizes than others. There was no statistically significant overall association between the rate of placebo participants reporting at least one adverse event and the placebo response, $Q(1, 382) = 2.11$, $p = .146$. The statistically non-significant Q statistic from the multivariate model indicates that the association between the within-group placebo response and the adverse event is not greater than would be expected by chance.

Figure 2.7

The Relationship between the Adverse Event Rate of Placebo Groups and the Placebo Response



Note. This figure shows the association between the adverse event rate of placebo groups as percentage of participants experiencing at least one adverse event and the placebo response as within-group standardised mean change scores from baseline to post-intervention. The size of the dots depicts the weight of the study in the statistical model, with larger dots representing a bigger weight. The dotted lines represent the 95% confidence intervals around the regression line.

Moderator Analyses

The moderation analysis was calculated with mixed-effect regression models. Therefore, a statistical models from the main analyses was used with the addition of the seven continuous (drug dosage, treatment duration, publication year, baseline severity, and participant age) and four categorical moderators (adverse event assessment method, comorbidity, geographical region, and funding source) to see if the added variables might explain more of the statistical variability or if they even were the underlying drivers of the observed effects.

The moderation analysis is reported in all statistical details in Table 2.1. When including the moderators in the model for the comparison between the adverse event rate between placebo groups of trials investigating different drug classes the predictor drug category remained statistically significant (p = .0217), confirming the originally observed effect. Further, participants' mean age (p = .0074) and the geographical region (p < .0001) were identified as statistically significant moderators, whereas drug dosage (p = .0590), treatment duration (p = .1666), publication year (p = .5786), method of adverse event assessment (p = .7158), comorbidity (p = .8387), and funding source (p = .1319) had no statistically significant effect on the adverse event rate of placebo groups. Baseline severity was dropped from the statistical model due to insufficient data. The heterogeneity in the statistical model with the moderators was no longer statistically significant, $Q(1, 23) = 0.84$, p = .3591.

Table 2.1

Moderation Analysis Adverse Event Rates Between Placebo Groups of Different Drug Classes

Predictor	*Estimate*	*SE*	z	*p*	*95% CI* (lower, upper)	Omnibus test for categoric predictors (with more than two levels)
Intercept	-28.35	52.00	-0.55	.5856	-130.26, 73.56	
Factor: Drug category						***Q*(4,23) = 11.47, *p* = .0217**
Level: Nonbenzodiazepines	0.13	0.50	0.25	.7998	-0.85, 1.11	
Level: Melatonin agonists	0.17	0.53	0.31	.7555	-0.87, 1.21	
Level: Orexin receptor antagonists	0.09	0.67	0.13	.8949	-1.21, 1.39	
Level: Herbal remedies	5.50	3.00	1.83	.0669	-0.38, 11.38	
Factor: AE assessment						*Q*(2,23) = 0.67, *p* = .7158
Level: Unstructured clinician interview	-0.02	0.13	-0.13	.8950	-0.28, 0.25	
Level: Systematic assessment via diaries & questionnaires	0.10	0.13	0.79	.4294	-0.14, 0.35	
Factor: Comorbidity						
Level: Comorbidity present	0.04	0.20	0.20	.8387	-0.36, 0.44	
Factor: Geographical region						***Q*(9,23) = 178.29, *p* < .0001**
Level: Belgium	0.58	0.11	5.14	< .0001	0.36, 0.80	

Level: France	0.05	0.15	0.37	.7130	-0.24, 0.35
Level: France & Israel	0.18	0.14	1.32	.1873	-0.09, 0.46
Level: Germany	0.30	0.13	2.26	.0241	0.04, 0.55
Level: International	0.57	0.16	3.59	.0003	0.26, 0.88
Level: Europe	0.47	0.16	2.99	.0028	0.16, 0.78
Level: UK	0.27	0.20	1.34	.1782	-0.13, 0.67
Level: USA	0.60	0.12	4.91	< .0001	0.36, 0.84
Level: North America	0.59	0.15	3.89	< .0001	0.29, 0.89
Factor: Funding source					
Level: Mixed (Industry & Government)	-0.20	0.13	-1.51	.1319	-0.46, 0.06
Drug dosage	-0.01	0.01	-1.89	.0590	-0.02, 0.00
Treatment duration	-0.00	0.00	-1.38	.1666	-0.01, 0.00
Publication year	0.02	0.03	0.56	.5786	-0.04, 0.07
Participants' mean age	**-0.02**	**0.01**	**-2.68**	**.0074**	**-0.03, -0.00**

Note. The standard procedure to accurately represent categorical moderators in (mixed-effect) meta-regression analyses is to code dummy variables representing each level of the predictor. The first (alphabetical) level of each categorical predictor is represented in the intercept. Here the intercept represents drug category level "Benzodiazepines", AE assessment level "unsystematic spontaneous or voluntary reports by patient", comorbidity level "no comorbidity", geographical region level "Australia", and funding source level "Industry". Statistically significant moderators are bolded.

The moderation analysis for the effect of the adverse event rate of placebo participants on the placebo response did not result in a single statistically significant predictor and still showed large heterogeneity, $Q(1, 180) = 6125.40$, $p < .0001$. See Table 2.2 for detailed information.

Table 2.2

Moderation Analysis Effect of Adverse Event Rates of Placebo Groups on the Placebo Response

Predictor	*Estimate*	*SE*	z	*p*	*95% CI* (lower, upper)	Omnibus test for categoric predictors (with more than two levels)
Intercept	-42.64	360.37	-0.12	.9058	-748.95, 663.66	
Adverse event rate	-3.00	7.68	-0.39	.6958	-18.05, 12.04	
Factor: Drug category						$Q(4,180) = 0.27$, $p = .9917$
Level: Nonbenzodiazepines	-0.21	1.47	-0.14	.8855	-3.08, 2.66	
Level: Melatonin agonists	-6.28	13.50	-0.47	.6416	-32.73, 20.17	
Level: Tricyclic antidepressants	-4.30	13.30	-0.32	.7466	-30.36, 21.77	
Level: Herbal remedies	-58.62	167.49	-0.35	.7292	-386.89, 269.65	
Factor: AE assessment						$Q(2,180) = 0.28$, $p = .8708$

Level: Unstructured clinician interview	-6.76	13.01	-0.52	.6035	-32.26, 18.74	
Level: Systematic assessment via diaries & questionnaires	-0.67	1.70	-0.39	.6943	-4.01, 2.67	
Factor: Comorbidity						
Level: Comorbidity present	-8.09	16.09	-0.50	.6153	-39.62, 23.45	
Factor: Geographical region						$Q(7,180) = 2.41$, $p = .9339$
Level: Belgium	2.17	3.01	0.72	.4715	-3.74, 8.07	
Level: France	-4.01	7.91	-0.51	.6118	-19.52, 11.49	
Level: France & Israel	5.60	11.35	0.49	.6219	-16.65, 27.85	
Germany	4.27	10.41	0.41	.6816	-16.13, 24.67	
International	0.29	1.74	0.16	.8702	-3.13, 3.70	
Europe	-1.43	1.72	-0.83	.4063	-4.79, 1.94	
UK	6.80	15.05	0.45	.6516	-22.70, 36.29	
Drug dosage	0.10	0.28	0.35	.7292	-0.45, 0.64	
Treatment duration	0.05	0.07	0.74	.4589	-0.08, 0.18	
Publication year	0.02	0.18	0.11	.9167	-0.33, 0.36	
Participants' mean age	0.19	0.37	0.52	.6048	-0.54, 0.92	

Note. The standard procedure to accurately represent categorical moderators in (mixed-effect) meta-regression analyses is to code dummy variables representing each level of the predictor. The first (alphabetical) level of each categorical predictor is represented in the intercept. Here the intercept represents drug category level "Benzodiazepines", AE assessment level "unsystematic spontaneous or voluntary reports by patient", comorbidity level "no comorbidity", and geographical region level "Australia".

Sensitivity Analyses

Like the moderation analysis, the sensitivity analysis was calculated using mixed-effect regression models. The statistical models from the main analyses were used plus the categorical predictors type of study (journal article vs. dissertation) and the continuous predictors risk of bias and quality of evidence. This was done in order to test the robustness of the results. Should the original predictor drug category not remain statistically significant when the additional predictors of the sensitivity analyses were added this would mean that some other factor was responsible for the different adverse event rates.

In the sensitivity model for the comparison between the adverse event rate between placebo groups of trials investigating different drug classes the predictor drug category was still statistically significant ($p < .0001$), confirming the robustness from the main analysis. Further, risk of bias ($p < .0001$), quality of evidence ($p < .0001$), and reporting type ($p < .0001$) was identified as statistically significant predictors influencing the adverse event rate. The heterogeneity in the statistical model with the moderators was still statistically significant, $Q(1, 61) = 934.47$, $p < .0001$. See Table 2.3 for full information about the results of the sensitivity analysis.

Table 2.3

Sensitivity Analysis Adverse Event Rates Between Placebo Groups of Different Drug Classes

Predictor	*Estimate*	*SE*	z	*p*	*95% CI* (lower, upper)	Omnibus test for categoric predictors (with more than two levels)
Intercept	-0.19	0.03	-6.83	< .0001	-0.25, -0.14	
Factor: Drug category						***Q*(6,61) = 387.79, *p* < .0001**
Level: Nonbenzodiazepines	0.24	0.02	12.51	< .0001	0.20, 0.28	
Level: Melatonin agonists	0.25	0.02	12.61	< .0001	0.21, 0.28	
Level: Tricyclic antidepressants	0.29	0.02	9.00	< .0001	0.23, 0.36	
Level: Orexin receptor antagonists	0.41	0.03	13.72	< .0001	0.35, 0.46	
Level: Herbal remedies	0.12	0.03	4.37	< .0001	0.06, 0.17	
Level: Gaboxadol	0.42	0.03	16.27	< .0001	0.37, 0.48	
Total risk of bias	**0.03**	**0.00**	**9.11**	**< .0001**	**0.03, 0.06**	
Factor: Quality of evidence						***Q*(2,61) = 142.61, *p* < .0001**
Level: Low	0.03	0.02	2.22	.0263	0.00, 0.06	
Level: Moderate	0.25	0.02	11.05	< .0001	0.21, 0.30	
Factor: Reporting type						
Level: Registry	**-0.20**	**0.02**	**-10.15**	**< .0001**	**-0.23, -0.16**	

Note. The standard procedure to accurately represent categorical moderators in (mixed-effect) meta-regression analyses is to code dummy variables representing each level of the predictor. The first (alphabetical) level of each categorical predictor is represented in the intercept. Here the intercept represents drug category level “Benzodiazepines”, Quality of evidence level “high”, and Reporting type level “Peer-reviewed journal”.

Like the main analysis, the sensitivity analysis for the effect of the adverse event rate of placebo participants on the placebo response resulted in the adverse event rate being a statistically non-significant predictor for the placebo response (p = .2448). Risk of bias (p = .6406), quality of evidence (p = .4218), and reporting type (p = .3960) had no statistically significant effect on the placebo response and there was still large heterogeneity, $Q(1, 382) = 15364.58$, $p < .0001$. See Table 2.4 for all statistical information.

Table 2.4

Sensitivity Analysis Effect of Adverse Event Rates of Placebo Groups on the Placebo Response

Predictor	*Estimate*	*SE*	z	*p*	*95% CI* (lower, upper)	Omnibus test for categoric predictors (with more than two levels)
Intercept	0.96	0.39	2.47	.0137	0.20, 1.72	
Adverse event rate	0.50	0.43	1.16	.2448	-0.35, 1.35	
Total risk of bias	-0.03	0.06	-0.47	.6406	-0.14, 0.09	
Factor: Quality of evidence						$Q(2,382) = 1.73$, $p = .4218$
Level: Low	0.41	0.32	1.28	.1993	-0.22, 1.03	
Level: Moderate	0.17	0.24	0.69	.4920	-0.31, 0.64	
Factor: Reporting type						
Level: Registry	**-0.36**	**0.42**	**-0.85**	**.3960**	**-1.18, 0.47**	

Note. The standard procedure to accurately represent categorical moderators in (mixed-effect) meta-regression analyses is to code dummy variables representing each level of the predictor. The first (alphabetical) level of each categorical predictor is represented in the intercept. Here the intercept represents Quality of evidence level “high” and Reporting type level “Peer-reviewed journal”.

Discussion

A systematic review and meta-analysis was conducted to investigate the rate of placebo participants reporting adverse events and the profile of specific adverse events of placebo groups across trials investigating different drug classes. Although one would expect participants receiving inert placebo pills should always experience the same rate of adverse events, the analyses showed that placebo participants experienced statistically significantly different rates of adverse effects across different drug classes investigated. Further, when comparing the specific adverse event profiles head-to-head between the placebo groups and their corresponding drug group, a statistically highly significant positive correlation was found. Meaning that participants in placebo groups frequently report the same adverse events as their counterparts receiving the actual drug. In both cases, these findings remained robust when adjusting the statistical analysis in the moderator and sensitivity analyses for the most typical methodological and clinical characteristics. This suggests that contextual factors, like participating in a trial about of certain drug influences side effect reporting or possibly even participants' experiencing of side effects even when allocated to placebo.

These findings have important implications for the evaluation of medication side effects. Informing participants about potential side effects using information leaflets or verbally during the informed consent process seems to perpetuate side effects via the nocebo effect. As was reviewed by Barnes et al. (2019) there have been attempts to use different communication strategies to reduce nocebo effects caused by side effect information. While these approaches might be particularly important for clinicians informing patients about side effects to reduce their occurrence and improve patient satisfaction, different communication strategies will not solve

the problem of what constitutes a side effect that has to be put on the information leaflet in the first place.

To create medication leaflets that report side effect information that is based on robust evidence there need to be important changes in clinical trial methodology and reporting standards. Conducting this systematic review, it was often not even possible to determine if the researchers had used unsystematic adverse event assessments like voluntary patient reports or if they made use of more systematic assessments like daily diaries. There are many reasons for the lack of high-quality evidence for side effect information. The most important source for side effect information are RCTs as they are the current gold standard in evaluating new medications (Hackshaw, 2009). According to the U.S. Food and Drug Administration, a new medication must demonstrate its effectiveness and safety (U.S. Food and Drug Administration, 2020). The results of this review showed that there is a lack of scrutiny devoted towards assessing and reporting adverse events in RCTs compared to the efficacy outcomes. There is a pronounced difference in the requirements between the highly specified and often well-validated effectiveness outcome measures needed to register a RCT online compared to the assessment strategies for adverse events. Further, when studies investigating medications are published they often only describe the most common adverse events, not all information is presented, or they do not even mention adverse events at all (Schroll et al., 2016).

Although a rigorous inclusion and exclusion criteria were chosen to guarantee a homogenous sample of only the highest quality trials in this review, there were still many RCTs that did not fulfil all the domains that were assessed using the Cochrane risk of bias tool and the GRADE approach for the assessment of the quality of evidence. In fact, missing, unreported, or only partial data on adverse events was a problem that limits the conclusions of this meta-

analysis. For example, only data about the adverse event rate of 61 of the total 88 placebo groups could be retrieved.

The moderation and sensitivity analyses further revealed that the adverse event rate of placebo groups is statistically significantly different depending on the geographic region the trial was conducted, but it was not possible to find a clear tendency pointing to one area reporting larger adverse event rates. The participants' mean age was negatively associated with the adverse event with older placebo groups reporting fewer adverse events. A compelling argument for high-quality trials and comprehensive reporting is the finding that trials with a lower risk of bias, meaning higher quality seem to report higher adverse events in the placebo groups. Furthermore, RCTs that were published in peer-reviewed journals reported higher rates of adverse event rates in placebo groups compared to studies that reported their findings in registries without a peer-review process. Higher reported adverse event rates in higher quality trials are in line with prior research. The underlying reason is most likely that higher quality investigations more often use systematic strategies to assess adverse events.

Several limitations should be considered when interpreting the results of this meta-analysis. The first is the low overall quality of evidence and the risk of bias of included studies, which most likely resulted from the poor reporting standards of adverse events in RCTs. Articles included in this review often did not report all of the adverse event information they had collected, especially when it comes to less frequent adverse events. Many authors only reported adverse events that occurred in at least three percent of participants. While it is understandable that authors are often limited in the amount of information that can be presented in a paper, it is suggested that future studies make all their data available using online supplements or open-science repositories.

Although an attempt was undertaken to minimise heterogeneity in the study sample already by defining strict inclusion and exclusion criteria and later using moderator analyses some of the analyses still suffered from unexplained heterogeneity. The results should therefore be interpreted cautiously, and other researchers are encouraged to replicate these meta-analytic findings in experimental and clinical studies.

Conclusions

Placebo participants' adverse event rates and specific adverse event profiles suggest that participants expectations, most likely developed during the informed consent process, influences their reporting of adverse events. This finding is further evidence for the nocebo effect that information in verbal or written form alone are sufficient for people to at least report, if not experience side effects. Currently, the rates of adverse events reported in the medication arm are used to populate the patient information provided with the medication. However, these results suggest that the true rates of side-effects could be significantly lower. As a result, the available information might introduce unnecessary burden for patients in that they might experience side effects caused by the nocebo effect rather than the medication per se. Having accurate evidence-based side effect information is particularly important for insomnia for two reasons. First, as Cheung et al. (2018) showed many people are hesitant to start pharmaceutical treatment for insomnia because they fear side effects and negative long-term consequences. This fear might be reduced if medication leaflets contained a more accurate estimate of the rate of side-effects. This could be achieved by using the rate of adverse events in the medication group and subtracting the rate of adverse events in the placebo group. Second, many patients cease their pharmacological insomnia treatment because they experience side effects. These results suggest that if fewer side

effects were described in patient information leaflets, and those that were described were less frequent, patients may experience fewer nocebo-induced side effects because less negative information was presented. This might improve adherence rates, which could in turn significantly improve treatment outcomes.

It became clear from this meta-analysis that research does not give enough thoughts to how adverse events are assessed, reported, and how this data is then used to inform communications about side effects. To reduce the risk of nocebo effects in the general population and increase the validity of RCTs it is suggest that adverse event assessments in clinical trials should undergo the same scrutiny as efficacy outcomes. Therefore, future clinical trials should assess adverse events at least using systematic surveys and report them with as much detail as they report the efficacy outcomes. Ideally, a structured approach using a validated questionnaire like the General Assessment of Side Effects (GASE; Rief et al., 2011) should be used to assess general symptoms during the baseline and intervention phase of the trial. Once the assessment and registration of said assessment strategy have been pre-registered it is important to transparently make this data public. A potential solution to ensure an adequate assessment and reporting of adverse events in clinical trials might be that clinical trial registries specifically ask researchers to define the adverse event assessment strategy in as much detail as is already required for efficacy outcomes. Once reliable and complete side effect information is reported it is important that this information is appropriately presented to patients, so that one can minimise consequences of the nocebo effect as good as possible.

This chapter has reviewed the literature about side effects of placebo groups in RCTs. The retrospective analysis of side effect data has shown that there is a need to better understand the role side effects play in clinical trials. While this meta-analysis did not find a statistically

significant association between placebo groups' adverse event rates and the placebo response, the analysis was based on sparse and heterogenous data. Therefore, the next three chapters report the findings of three experimental chapters that specifically focus on the influence side effects play in the context of placebo effects.

Chapter 3: Study 1 - Proof of Concept New Model of Active Placebo

This chapter introduces the new model of an active placebo and reports the findings of Study 1 of this thesis that evaluated how well the new active placebo model elicited the target side effect. Study 1 additionally served the purpose of a proof-of-concept study trialling the study design, outcome measures, and procedures to inform the following studies.

Introduction

Jensen et al. (2017) searched the literature for publications discussing the advantages and disadvantages of using active placebos in RCTs. The search resulted in only one publication with the main aim of discussing active placebos. Salamone (2000) reviewed active placebos in the context of pharmacological depression trials and found methodological problems associated with the drugs that were used as active placebos. Most of the drugs used as active placebos had some efficacy for the treatment of depression or disorders that are often observed comorbidly with depression, and several of the drugs were clinically used to augment the antidepressant treatment. Jensen et al. (2017) summarised the literature arguing that practical problems and unintended therapeutic effects of active placebos speak against the use of placebos, while the reduced risk of failed participant blinding, and ethical arguments favour the use of active placebos in clinical trials. The main argument in favour of using active placebos in RCTs is the reduced risk of unblinding. A total of five articles stated the advantage or necessity of using active placebos to uphold blinding, which is one of the core underlying principles of RCTs (e.g., Cohen & Jacobs, 2007; Colagiuri, 2010; Howick, 2009; Kirsch & Sapirstein, 1999; Moncrieff, 2001). Unintended therapeutic effects of drugs used as active placebos was the main criticism

identified in their systematic review, with three more articles stating this concern (e.g., Colagiuri, 2010; Cooney, 1998; Kirsch & Sapirstein, 1999).

As reviewed above, pharmacological clinical trials referred to other active medications as active placebos to mimic the side effects of the drug under investigation. Although this approach arguably solves the problem of upholding blinding throughout the trial, the risk of unintended therapeutic effects remains. In a more recent approach to active placebos, Rief and Glombiewski (2012) created an active placebo that does not rely on another active medication to mimic side effects. Their model of an active placebo was tested in an experimental setting involving 144 healthy participants under the guise of a new analgesic nasal spray using experimental heat pain. Half of the nasal sprays used in their study were inert placebos containing sesame oil, while the active placebos contained sesame oil together with 0.014 percent capsaicin to induce a prickling sensation in the nose. Compared to the inert placebos, participants in the active placebo group demonstrated a higher pain threshold after placebo application. The active placebo's increase in pain threshold was significantly larger than the inert placebo and was of medium effect size (Cohen's $d = 0.68$).

A major strength of Rief and Glombiewski (2012)'s nasal capsaicin active placebo was that it alleviated the main concern that the active placebo might have unintended therapeutic effects of its own. However, it could be argued that the prickling sensation caused by the capsaicin nasal spray differs from common side effects in the sense that the sensation was associated with the delivery of the treatment (like pain at the site of an injection) as opposed to the side effects that develop as a result of the pharmacology of a treatment (like nausea following chemotherapy). Further, use of a nasal spray could also be argued to be quite different to the

most common routes of treatment administration in clinical trials of pharmacological medications, i.e., pills, tablets, or capsules.

To my knowledge, only one investigation trying to use something similar to an active placebo in the form of a pill could be identified in the literature. Szabo et al. (2018) compared an inert lactose placebo labelled as mood enhancing "super pill" against a commercially available Tic Tac mint as the active placebo. After baseline measures young healthy athletes received either the "super pill" or the Tic Tac and were asked to observe any effects for the three minutes after intake, before they were asked to fill out the post-intervention questionnaires. Szabo et al. (2018) observed that the active placebo outperformed the inert placebo group on short-term physical wellbeing and positive affect. The authors therefore concluded that the experience of onset sensations of an "active" placebo elicits larger changes than expectation enhancing labelling of an inert placebo. While the authors use the term active placebo, the same problem applies to their model of an active placebo as with the nasal spray used in Rief and Glombiewski (2012) study, that it only elicited a sensation during onset, but did not actually induce a side effect.

So, why is there a need for a new model of an active placebo for clinical trials and experimental research? The goal of RCTs is to evaluate the benefits and risks that are associated with a new drug. As long as clinical trials are based on the comparison between an inert placebo not eliciting any sensations during onset or side effects and an active medication that elicits side effects and for example has a distinct taste after intake, then there may never be an accurate estimate of the beneficial effect of a pharmacological agent itself. As long as only the medication elicits active sensations and side effects the comparison to an inert placebo is always going to be made up of the pharmacological agent and the psychological effects associated. The factors

underpinning placebo effects have been discussed to a large extent in this introduction. The most important reasons hypothesised to cause the amplification of efficacy between inert and active placebos are an increase in expectations. Although the experience of onset sensations or side effects might generally enhance placebo effects, it is especially important to account for these factors in double-blind RCTs where blinding is a key underlying paradigm and participants are desperately looking for an improvement from their suffering. In this situation, the experience of side effects or any other distinct sensations may reassure patients that they were allocated to the actual treatment group. Besides the reasons just mentioned, research has clearly demonstrated that active placebos outperform inert placebos for experimentally induced conditions, on a limited time period, and in healthy adults. While the capsaicin nasal spray (Rief & Glombiewski, 2012) and the Tic Tac mint model (Szabo et al., 2018) of active placebos managed to induce treatment onset sensations, they importantly failed to induce real noticeable side effects on a longer time scale as is typically the case in the context of clinical trials or in clinical practice. Further, they were only investigated in single-session, short-term experiments using healthy adults.

The new active placebo model had to fulfil many requirements to be considered a true active placebo eliciting a side effect while not having any effects on the condition to be treated. To be most useful in experimental studies and in the context of clinical trial research the new active placebo had to be conceptualised as capsules, pills, or tablets because this is the dominant method of application of pharmaceutical medication. Further the active placebo was ideally required to be based on non-prescription substances to make it a viable experimental model that can be applied on multiple occasions and by non-medical professionals. Not being based on pharmaceutical ingredients that require a prescription further offers the benefits that it can be

used in many more health conditions without causing any unwanted complications due to potential medication interactions or unwanted therapeutic effects.

Using an active placebo eliciting a target side effect and potentially other unwanted effects naturally increases the ethical hurdles compared to an inert placebo only containing lactose that had already been used in many of studies or trials. Therefore, the requirements for the ingredients had to be chosen in a way that enabled eliciting a noticeable side effect while being as safe as possible. In addition to only containing safe ingredients that can be applied on multiple occasions, there had to be evidence that the placebo stayed true to its name and did not have any unwanted therapeutical effects on the condition to be treated.

To create an active placebo that is safe and can be widely applied in experimental research beetroot extract was chosen to elicit beeturia, an unusual red colouration of urine or stool. The idea was based on a study published in 1995 that investigated pigment-decolourising systems in the intestinal tract (Eastwood & Nyhlin, 1995). Beetroot is a vegetable that is most often consumed after cooking in boiling water. Its red pigment, called betanin, is widely used in the food industry as a colourant labelled E162 (Timberlake & Henry, 1988). The pigment is a pH and redox indicator and only approximately 14 percent of the general population experience beeturia after consuming beetroot because it is decolourized by most people's digestive system (Watson et al., 1963). To protect beetroot extract or more precisely its red pigment betanin from decolourisation, Eastwood and Nyhlin (1995) gave patients beetroot orally together with 1 g of oxalic acid. Oxalic acid is a reducing agent that naturally occurs in beetroot and protects it from being decolourised. One hundred gram of beetroot naturally contains about 700 to 800 mg of oxalic acid (Holland et al., 1991). Eastwood and Nyhlin (1995) not only managed to increase the rate of beeturia experienced in patients by using oxalic acid, but they also demonstrated that

beeturia resulted from colonic absorption of beetroot extract. They studied the location of beetroot absorption using patients who had undergone an ileostomy. The ileum is a part of the small intestine before the colon. Ileostomy is a surgical procedure where a stoma (hole) on the surface of the belly lets stool exit the body without going through the colon. Ileostomy typically needs to be performed when a part or the whole large intestine, colon, rectum, or anus needs to be removed due to cancer and can either be temporary or permanent (Kock et al., 1977).

In applying the beetroot extract and oxalic acid formulation orally to patients who had undergone such a drastic surgery without any noticeable adverse events, Eastwood and Nyhlin (1995) had delivered the first evidence that applying beetroot extract together with 1 g of oxalic acid can be considered to be safe. To study if high-calcium diets have the potential to abolish hyperoxaluria, a condition often believed to be associated with the formation of kidney stones, Hess et al. (1998) even applied 2.2 grams of oxalic acid per day without any negative consequences for participants.

The daily dose for the new active placebo groups in this thesis consisted of a total of four capsules per day containing each 600 mg beetroot extract, totalling to 2.4 grams of beetroot extract per day. Beetroot extract (E162) has no known adverse effect on the human body and is certified for use in food and medicines by international food and drug agencies without any restrictions on the dose (for further details see: EFSA Panel on Food Additives & Nutrient Sources added to Food, 2015; Therapeutic Goods Administration, 2011). In line with Eastwood and Nyhlin (1995), the active placebo capsules also contained a minimal amount of 250 mg oxalic acid to avoid decolourisation and guarantee the absorption of the beetroot extract in the colon. Oxalic acid is a naturally occurring part of our daily diet and can be found in many common foods like spinach, beetroot, rhubarb, or star fruits. This amount is equal to a daily

consumption of approximately 120 grams spinach or 150 grams of beetroot (Holland et al., 1991; Riemenschneider & Tanifuji, 2011). Further, only a minimal amount (2-5%) of oxalic acid is actually absorbed into the bloodstream and this has been shown to have no effects on the human body (Eastwood & Nyhlin, 1995; Hodgkinson & Zarembski, 1968; Riemenschneider & Tanifuji, 2011). Nevertheless, to account for the non-maleficence principle and to guarantee safety and well-being of participants, the daily application of beetroot extract was limited to only 2.4 grams and one gram of oxalic acid, and the length of the application was restricted to seven days.

Besides evidence of the safety of the new model of the active placebo is safe to take, it was also important to evaluate evidence that it is an actual placebo and does not elicit any changes on the condition to be treated, sleep in the case of this thesis. There have been multiple reviews conducted about the biological effects of betanin, beetroot extract, and beetroot juice. Some of them found small effects of large quantities of beetroot juice (~500 ml/d) on blood pressure but the authors argued that these effects are associated with nitrate and nitrite that are found in beetroot juice in sufficient quantities to elicit changes on blood pressure (Esatbeyoglu et al., 2015; Kapadia & Rao, 2013; Siervo et al., 2013). Importantly, none of the literature reviews had found any effects of betanin, beetroot extract, or beetroot juice on sleep. Only just recently a small-scale study consisting of three groups with 10 participants in each condition (beetroot juice supplement vs. placebo vs. no-treatment) was conducted with athletes that found that the beetroot juice supplement had significantly improved sleep quality compared to the placebo and no-treatment condition, as measured by the PSQI (Shamloo et al., 2019). However, that study has to be interpreted with caution due to the many methodological problems and missing details that were not reported in their manuscript. For example, the study only reported on a very small sample, has methodological problems including a clearly distinguishable placebo and the

reporting of their methodology and results does not allow to infer what statistical tests were used or how they were calculated. Further, the authors stated that they had tested male athletes in their study, but the exclusion criteria stated that participants with a regular background of exercise were excluded.

Even if the finding of Shamloo et al. (2019) study turned out to be accurate, this would still not question the status of the new model of an active placebo as their interventional supplement consisted of 300 ml beetroot juice containing a total of 900 mg nitrate on a daily basis, whereas the new model of an active placebo only consists of 2.4 g beetroot extract per day being the equivalent to 30 mg nitrate. According to Gallardo and Coggan (2019) who conducted a systematic review about the effects of nitrate on physiological processes defined the minimum required dosage of nitrate with at least 5 mmol (the equivalent of 310 mg) of nitrate per serving. They concluded that every dose below that 310 mg nitrate would not be sufficient to even detect any physiological effects in humans. This reassures the fact that the new model of an active placebo is correctly labelled as a placebo and is not capable of eliciting any detectable physiological effects.

To test if the active placebo induced the desired side effect a similar approach was chosen as Neukirch and Colagiuri (2015), who studied the placebo effect for sleep using a longitudinal parallel group design with one week of baseline sleep measures followed by one week of either no-treatment or placebo treatment. To increase the ecological validity the study was conducted as parallel, double-blind, randomised, placebo-controlled trial under the guise of a new treatment for sleep problems. Sleep problems as the condition to be treated offered many advantages for the evaluation of the new active beetroot placebo. Sleep problems are a prevalent health concern

for many students with up to 60 percent of university students experiencing problems with sleep due to sharing a room with someone else, structural changes in their lives after having moved out, and the challenges associated with studying (Schlarb et al., 2017). Moreover, sleep problems are a useful health concern to investigate placebo effects and compare an inert to an active placebo because Yeung et al. (2018) has summarised that placebo effects in sleep are robust and range from small to moderate effect sizes when compared to no-treatment. The sample for Study 1 was primarily recruited from first- and second-year undergraduate science students enrolled in at the University of Sydney.

As the primary aim was to test how well the active placebo group elicits the target side effects the active placebo was considered to be successful when at least half (i.e., 50%) of all participants receiving the active placebo reported experiencing an unusual colouration of their urine colouration during the treatment week. As a comparator for the success rate of inducing side effects the new active beetroot placebo was compared to the most commonly use pharmacological active placebo. Berna et al. (2017) investigated if side effects enhance the treatment response using healthy participants and induced heat pain. Half of participants received atropine, an anticholinergic drug that was most commonly used as active placebo in RCTs to induce dry mouth. Fortunately, Berna et al. (2017) included the number of participants reporting dry mouth after receiving atropine in the supplementary material. Out of the 50 participants that had received 1.2 mg atropine 29 (58 percent) reported dry mouth as a side effect.

Based on the argumentation provided above, two additional secondary a priori hypotheses were formulated. First, the presence of a statistically significant placebo effect for the two placebo groups (inert and active) compared to the no-treatment group on the outcome measures was hypothesised. Second, the active placebo group was hypothesised to elicit

statistically larger placebo effects compared to the inert placebo groups on the outcome measures.

Method

The University of Sydney's Human Research Ethics Committee had reviewed and approved all ethical aspects regarding the recruitment, materials, and procedures for this study (Project Number: 2018/107). Detailed documentation of the ethical approvals and the necessary modifications can be found in Appendix B.

The full methodology of this study was registered on the Australian and New Zealand Clinical Trial Registry (ANZCTR, identifier ACTRN12618001493235; Australian and New Zealand Clinical Trials Registry, 2018a).

Participants

The desired participant sample for this study consisted of healthy adults recruited in the greater Sydney area. Participants were recruited via the University of Sydney's Psychology Research Participation System (SONA-PSYCH) that offers first and second year undergraduate students enrolled in psychology courses to participate in research studies to receive course credit. Further, participants receiving cash payment as remuneration were recruited via the University of Sydney's paid Research Participation System (SONA-PAID) that allows students enrolled across all faculties and seniorities to participate in research studies. Additionally, paid participants were recruited from the general population in the greater Sydney area via posting the advertisement

flyer in the area surrounding the University of Sydney's Camperdown campus. Recruitment lasted from May to October 2018. The advertisement flyer can be found in Appendix B.

Participants had to be older than 18 years of age and because the primary aim of this study was to evaluate how well the active placebo elicited the target side effect participants were not required to fulfil a specific threshold of sleep difficulty. Participants were excluded from participation if they (1) were currently taking prescription medication with the exception of contraceptive pills, (2) were pregnant, breastfeeding, or trying to conceive, (3) had received professional treatment for sleep problems in the last three months, (4) had any intolerances for antihistamines, lactose, or beetroot extract, (5) suffered from an abnormal or deficient kidney functioning or any other medical condition. Because the placebo capsules consisted of gelatine, participants were further made aware that study participation was only possible if they were okay with taking capsules that were not vegan or halal. Participants either received two hours' worth of course credit or AUD $50 for their participation.

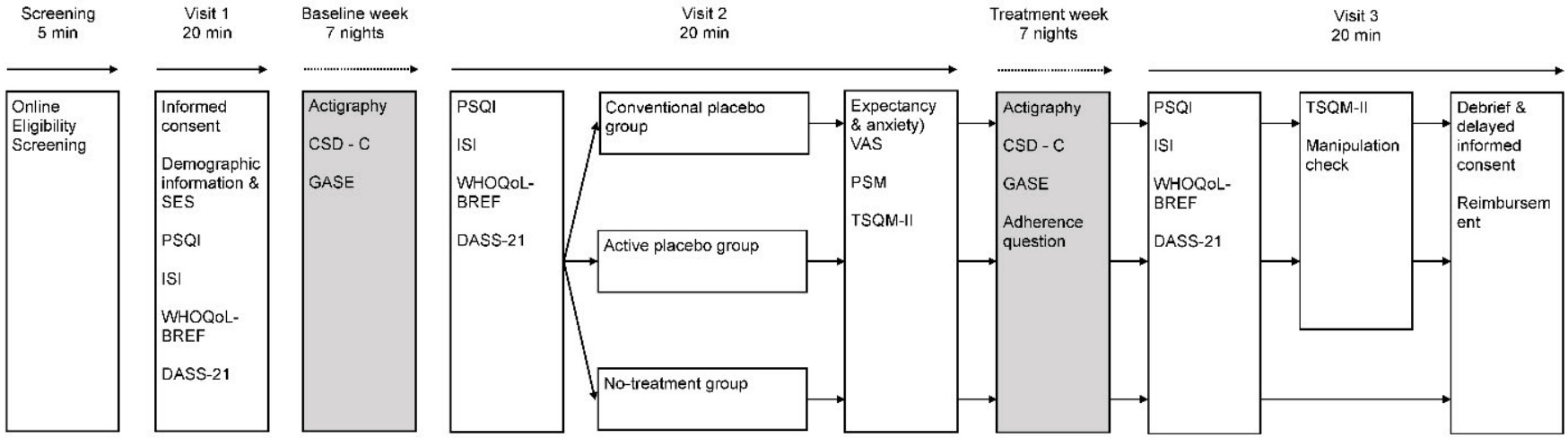
Design

Figure 3.1. shows the study design with the randomisation to the three study groups and the two time periods. The experiment used a 3 x 2-mixed design with treatment (inert vs. active vs. no-treatment) and time (baseline vs. intervention) as factors. The factor time was divided in a baseline and a treatment period, each covering seven nights. In visit two, participants were randomly allocated to one of the groups and received either a placebo treatment or no treatment. Participants were proportionally allocated to the three groups using simple randomisation with a 1:1:1 ration. The placebo treatment unbeknownst to the participants either consisted of an active beetroot placebo or an inert lactose placebo.

The independent variable was group allocation to one of the three experimental groups and the independent variables were participants report of an unusually redder urine colouration and the Insomnia Severity Index.

Figure 3.1

Study Design and Procedure with all Measures



Materials

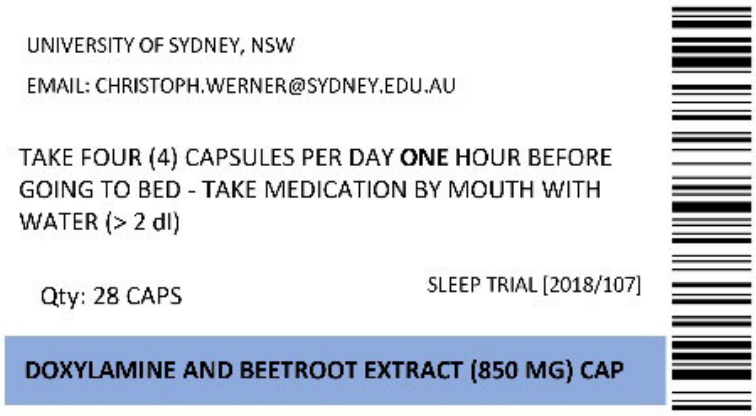
Placebo capsules.

All capsules used for this study were manufactured by a fully licenced and registered compounding pharmacy called Pharmacy@UNSW located at the University of New South Wales. The capsules were blue/white in colour, consisted of gelatine, were about 20 mm long (size 00), and depending on the density of the ingredients held approximately 735 to 750 mg powder. The conventionally used inert placebos as most often encountered in RCTs just contained 750 mg lactose powder as filler material.

The active placebo capsules for this study contained 500 mg of the food colour E162 (i.e., beetroot extract) and 250 mg of oxalic acid. Following Eastwood and Nyhlin (1995) adding oxalic acid to the capsules protected the discolouration of the red pigment in beetroot extract, ensuring its absorption in the colon. The successful absorption of beetroot extract in theory should then induce beeturia, an unusually red coloured of urine. Each participant received a clear capsule container with 28 capsules, four capsules to be taken each evening approximately one hour before going to sleep. The capsule containers were labelled in order to further support the cover story of an active medication. The label, which is displayed in Figure 3.2., was specifically designed for this study to imitate typical prescribed medication labels.

Figure 3.2

Label on Capsule Container



Cover story.

Participants were deceived about the actual goal of the study and received the information that they were participating in a RCT that tested the effect of a new compound consisting of an antihistamine (doxylamine) and an antioxidant (beetroot extract) for the treatment of insomnia symptoms. Participants received the information that they would be randomly allocated to either the active medication group, or to a no-treatment control group. To create a credible cover story, participants were informed that beetroot extract was added as a protective layer around doxylamine, an antihistamine that is already available over-the-counter and is often used by people suffering from insomnia due to its ability to induce sleepiness. The argument was made that beetroot extract was added for two reasons. First, that a preliminary study had already shown that beetroot extract in itself had sleep promoting characteristics and had helped participants improve their sleep. Secondly, that the beetroot extract was layered around the doxylamine as an additional protective coating to protect the doxylamine from being

degraded in the intestinal system by microbiota and enzymes. Therefore, the doxylamine will be absorbed to a larger degree into the blood stream as conventional doxylamine available over-the-counter. To increase standardisation of information given to participants during participants' visits an experimenter manual was created and experimenters were trained to deliver the information according to the manual and in a standardised fashion. The full experimenter manual for the three study visits can be found in Appendix B. Both, the inert and active placebo groups received the same explanations and instructions about the treatment under investigation and potentially associated side effects.

Measures and outcomes

Table 3.1 defines the primary and secondary outcome measures and when these measures were assessed over the course of the study. The hierarchy of outcome measures and definition of assessment time points was pre-defined and registered within the ANZCTR (Identifier ACTRN12618001493235; Australian and New Zealand Clinical Trials Registry, 2018a).

Table 3.1

Definitions of Outcomes

Outcome Hierarchy	Outcome Measure and Definition of Assessment Time Point for Statistical Analyses
Primary outcome (1)	Insomnia severity assessed using the Insomnia Severity Index (ISI)
Timepoint (1)	Baseline (assessed at a single timepoint on the day starting treatment, for the 7 nights period prior to receiving treatment) Treatment (assessed at a single timepoint on the day after finishing 7 nights of treatment, for the 7 nights period receiving treatment)
Primary outcome (2)	Self-reports of urine colouration, using the same 4-point scale as the GASE (0=not present, 1=mild, 2=moderate, 3=severe)
Timepoint (2)	Baseline (assessed daily during the 7 nights prior to randomisation) Treatment (assessed daily during the 7 nights receiving treatment after randomisation)
Secondary outcome (1)	Sleep quality assessed using the Pittsburgh Sleep Quality Index (PSQI)
Timepoint (1)	Baseline (assessed at a single timepoint on the day starting treatment, for the 7 nights period prior to receiving treatment) Treatment (assessed at a single timepoint on the day after finishing 7 nights of treatment, for the 7 nights period receiving treatment)
Secondary outcome (2)	Sleep duration (self-report) using the Consensus Sleep Diary Version C (CSD-C)
Timepoint (2)	Baseline (assessed daily during the 7 nights prior to randomisation) Treatment (assessed daily during the 7 nights receiving treatment after randomisation)
Secondary outcome (3)	Reports of daily symptoms using the General Assessment of Side Effects (GASE)
Timepoint (3)	Baseline (assessed daily during the 7 nights prior to randomisation) Treatment (assessed daily during the 7 nights receiving treatment after randomisation)

Secondary outcome (4)	Depression, anxiety, and stress using the Depression Anxiety Stress Scale (DASS-21)
Timepoint (4)	Baseline (assessed at a single timepoint on the day starting treatment, for the 7 nights period prior to receiving treatment) Treatment (assessed at a single timepoint on the day after finishing 7 nights of treatment, for the 7 nights period receiving treatment)
Secondary outcome (5)	Quality of life using the World Health Organisation's quality of life assessment (WHOQOL-BREF)
Timepoint (5)	Baseline (assessed at a single timepoint on the day starting treatment, for the 7 nights period prior to receiving treatment) Treatment (assessed at a single timepoint on the day after finishing 7 nights of treatment, for the 7 nights period receiving treatment)
Secondary outcome (6)	Treatment Satisfaction Questionnaire for Medication version II (TSQM-II)
Timepoint (6)	Post-treatment (at the end of the seven-night period receiving treatment)
Secondary outcome (7)	Sleep duration (objective) using Actigraphy (Philips Actiwatch 2)
Timepoint (7)	Baseline (assessed daily during the 7 nights prior to randomisation) Treatment (assessed daily during the 7 nights receiving treatment after randomisation)

Note. The hierarchy and time points of these measures have been registered on ANZCTR. Please note that the measures in some cases were assessed more often over the course of the study as defined here for the purpose of the statistical analyses.

Insomnia Severity Index (ISI; Amended)

This short self-report questionnaire, consisting of seven items is a brief screening measure of self-rated insomnia and a validated outcome measure for insomnia research (Bastien et al., 2001). The period of time assessed in this study was a one-week interval. The seven items cover the following criteria: severity of SOL, sleep maintenance difficulties, satisfaction with sleeping pattern, interference with daily functioning and noticeability attributed to sleep problems. Each item is rated on a 0 - 4 scale, the total score ranges from 0 to 28 with higher scores indicating more severe insomnia symptoms. Bastien et al. (2001) categorised ISI scores from zero to seven as no clinically significant insomnia, eight to 14 as subthreshold insomnia, 15 to 21 as moderate clinical insomnia, and scores from 22 to 28 as severe clinical insomnia. Generally, a score of 10 or above is considered as treatment-worthy insomnia. The internal consistency is moderately high (Cronbach's alpha = .74). The ISI is a reliable and sensitive measure to detect changes in self-reported insomnia severity over the course of a treatment (Bastien et al., 2001). In the current study this questionnaire was used alongside the PSQI to quantify sleep problems across the three visits of the study.

General Assessment of Side Effects (GASE; Amended)

This self-reporting questionnaire appraises 36 symptoms on a weekly basis, asking for the experience of each symptom (e.g., headache, nausea, dizziness, dry mouth, etc.) and rating the severity on a 0 - 3 scale (Rief et al., 2011). After rating a symptom as present (every score other than 0), the participant had to state whether the symptom is related to the current medication taken. A GASE total score (sum of all symptom ratings) indicates general symptom load. This questionnaire has high internal consistency (Cronbach's alpha = .89) and high

discriminate validity (Rief et al., 2011). The GASE was used in this study to evaluate the experience of side effects on a daily basis. The period of time was therefore changed from a one-week interval to a one-day interval. In order to account for the specific side effect of the active placebo and other symptoms that seem to generally occur in the human population (Petrie et al., 2014), additional items were added, asking about red coloured urine and faeces, back pain, fatigue, runny or stuffy nose, and joint pain creating a GASE version asking about a total of 47 symptoms.

Pittsburgh Sleep Quality Index (PSQI-A; Amended)

The PSQI is a self-reporting questionnaire where participants rate their sleep quality and disturbances over a one-month-interval (Buysse et al., 1989). It consists of 19 self-rated questions and five questions answered by the person sharing the bed or room. The PSQI assesses sleep duration and the frequency of sleep-related problems. The items are rated on a discontinuous 0-4 scale, except for item six asking for sleep quality which is rated on a 5-point scale.

The sum of these 19 items creates a global score with a range from 0 to 21 points, higher scores indicating poorer sleep quality. A PSQI score of more than five indicated that a patient is having severe difficulty in at least two areas or moderate difficulties in more than three areas. Further, a score of more than five provides a sensitive and specific measure to differentiate between good sleepers and people with sleep problems. The test-retest reliability is considered good ($r = 0.85$) and the internal consistency is acceptable with Cronbach's alpha = .83 (Buysse et al., 1989). For this study, the PSQI was adjusted to ask about a one-week time period, compared to the original one-month period in accordance with the study design. Participants answered the

PSQI on three different occasions during the three in-lab visits over the course of their participation.

Consensus Sleep Diary – Core (CSD-C)

The CSD-C is a self-rated questionnaire incorporating nine items evaluating sleep quality (e.g., time of getting into bed, time when attempt to fall asleep, sleep onset latency, number and duration of awakenings during the night, time of the final awakening, final rise time and perceived quality of sleep (Carney et al., 2012). Each item is rated on a 5-point Likert scale. The correlations between the CSD-C, the ISI and an actigraphy measure, as an objective measure of sleep, were shown to be significant (for all indices excluding sleep efficiency), thus leading to the conclusion of high validity of the CSD-C (Maich et al., 2018). Participants were instructed to fill out the CSD-C each morning as soon as possible after their final awakening. This was done across the complete two-week study duration.

Treatment Satisfaction Questionnaire for Medication Version II (TSQM-II; Amended)

This questionnaire measures the self-perceived effectiveness, side effects, convenience, and global satisfaction of a medical treatment, using 10 items assessing how satisfied or dissatisfied the participant was, considering different aspects of the medical treatment (Atkinson et al., 2005). Each item is rated on a scale ranging from one to seven, except for three items on side effects, which are answered binary or on a scale ranging from one to five. The scores for the TSQM-II sub-scales and the total score are calculated as percentages from 0 to 100, with higher scores indicating more satisfaction with medical treatment. The internal consistency is high (Cronbach's alpha = .81- .94) and the TSQM-II was shown to have similar dimensions compared to the full version of the TSQM (Atkinson et al., 2005). In this study, the TSQM-II was used as a

pre- and post-treatment assessment of expectation on treatment outcome and experience of side effects. In order to address the assessment of expectation on treatment outcome in the pre-treatment phase (during visit two, after randomisation), we changed the tense in the questions to future tense, whereas for the post-treatment assessment (at visit three), items were used as in the original version.

World Health Organization Quality of Life Questionnaire (WHOQOL; Amended)

The WHOQOL-BREF questionnaire is the short version of the WHOQL-100 and assesses quality of life as defined by the World Health Organisation (Skevington et al., 2004). It consists of 26 items, all answered on a 1-5 Likert scale, the WHOQOL-BREF has an underlying four factor structure, with the factors: physical, psychological, social, and environment quality of life. The WHOQL-BREF has acceptable internal consistency (Cronbach's alpha > .7) and good discriminating validity in distinguishing healthy from ill responders (Skevington et al., 2004). The assessed time period is a two-week interval, which was amended for this study to a one-week interval and was used to assess general quality of life at all three study visits.

Depression Anxiety Stress Scale 21 (DASS-21)

The DASS-21, a short version of the 42-items (DASS; Lovibond & Lovibond, 1995), is a self-report questionnaire, consisting of three scales measuring depression, anxiety and stress symptoms. Each scale contains seven items and all of the items are rated on a 1-4 scale. (Henry & Crawford, 2005) showed that the DASS-21 has good internal consistency (Cronbach's alpha = .93 for total scale) and good convergent and discriminant validity. The DASS-21 was used in the current study to measure depression, anxiety and stress and was administered in all three visits.

Perceived Sensitivity to Medicines (PSM) Scale

This short five-item self-report questionnaire assesses the perceived sensitivity to medicines (Horne et al., 2012). Responses are scored on a 1-5 Likert scale leading to a total score of 5-25 with higher scores indicating more perceived sensitivity. The PSM showed good internal consistency (Cronbach's alpha = .8) and high Retest – Reliability (r = .89), as well as good construct and predictive validity (Horne et al., 2012). The PSM was assessed during visit two after participants had been randomly allocated to one of the three study arms.

Expectation and Anxiety Visual Analogue Scale (VAS)

Visual analogue scales are a way of measuring a construct on an interval scale by asking the participant to move a slider over a continuous scale, depicted as a straight horizontal line of 100 mm length with the ends defined as the extreme limits of the construct. As such it is possible with a VAS to assess single constructs with many perceptible gradations and account for the inherent interval-scaled structure of a construct. Although originally designed and predominately used in the assessment of subjective pain, VAS has also been shown to have good reliability and validity in detecting anxiety (Williams et al., 2010).

In this study, two VASs were used as measures of expectation and anxiety after instruction on treatment. Two horizontal lines were chosen with a length of 100 mm and named the two extremes (0 = not at all and 100 = absolutely). For the expectation VAS we asked to what extent the participants expected an improvement for their sleep problems due to the treatment. The VAS measuring anxiety asked how anxious participants were about experiencing side effects associated with the treatment.

Adherence Questions

Adherence was assessed on a daily basis using the daily surveys participants had to fill out as a self-report item asking them how many capsules they had taken. Additionally, participants were instructed to return the capsule containers with all capsules not taken for a manual capsule count as a double-check for the self-reports.

Actiwatch 2 (Respironics Inc., USA).

Actigraphy measurement of movements is a non-invasive method of estimating sleep time by comparing times during the day where the participant moved (wake) or did not move (sleep). The Actiwatch 2 (Respironics Inc., USA) used for this study is a watch-like device worn at the wrist, that records movement via a solid-state "Piezo-electric" accelerometer with a sampling rate of 32 Hz and a sensitivity of 0.025 g. The watch has a size of 43 mm x 23 mm x 10 mm and weighs 16 g (with band). As logging interval, we chose a 30-second sampling rate, meaning that the watch recorded 2880 epochs per day. The Actiwatch 2 sampling rate was chosen in accordance with a previous placebo study assessing sleep problems (Neukirch & Colagiuri, 2015).

Sleep data obtained from the Actiwatch 2 were calculated using Respironics' Actiware 5 software (Respironics Inc, USA), estimating TST and SOL. As a measure of sleep-wake-pattern, Actigraphy can be regarded as a valid and sensitive method for detecting sleep epochs, whereas the specificity (proportion of non-sleep epochs correctly identified) is considered to be low (Ancoli-Israel et al., 2003; de Souza et al., 2003).

The Actiwatch 2 devices were used in the current study to assess objective TST in addition to the self-reported TST using the CSD-C. Due to the limited availability of only six

Actiwatch 2 devices, only a subsample of participants was equipped with actigraphy devices. Participants were instructed to wear the actigraphy devices for the whole two-week period of the study and were only allowed to take the devices of to shower or during sport activities that prohibited participants from wearing the devices.

REDCap Survey System

Study data were collected and managed using REDCap (Research Electronic Data Capture) tools hosted at The University of Sydney (Harris et al., 2019; Harris et al., 2009). REDCap is a secure, web-based software platform designed to support data capture for research studies, providing 1) an intuitive interface for validated data capture; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for data integration and interoperability with external sources. REDCap was chosen for data collection for this study because it was specifically designed to support data collection in RCTs, because it is directly hosted by the University of Sydney guaranteeing the security of data, and lastly because it offers log files showing every single change that was done to data guaranteeing the collected data's integrity.

Randomisation and Blinding

To guarantee the double-blind nature of this RCT regarding the two placebo conditions, an independent co-worker at the lab chose three numbers for each group, drawing from a pool of numbers from 1 to 9. Three numbers (e.g., 1, 3, 8) indicated group one, three numbers (e.g., 2, 6, 7) indicated group two and the same for group three (e.g., 4, 5, 9). The experimenters were semi-blind to the numbers, knowing the three numbers that indicated no-treatment, but not knowing

what numbers corresponded to which placebo group. Each of the numbers that was associated with participants receiving placebo capsules belonged to an individual drawer. To avoid an experimenter from breaking blind completely in case a participant unintentionally talked about the experience of the target side effect nine numbers compared to only three were used for participants' group allocation.

Procedure

The study design and procedure are depicted in Figure 3.1, in the Design section above. The total duration of the study covered a two-week period, the first week was used for baseline assessment and the second week was the treatment period. Participants were asked to come to the University of Sydney's Camperdown campus three times, each visit lasting about 20 minutes.

To avoid any coercion because students enrolled in undergraduate courses at the University of Sydney could participate for course credits, participants had to make initial contact with researchers. Once participants had made the initial contact, they were sent the participation information statement and a link to an online eligibility screening questionnaire. Eligible participants were sent a list with available time slots to choose from. The three individual in-lab visits had to be scheduled one week apart to guarantee that both the seven-night baseline and seven-night treatment week matched. All three visits were conducted in the Brain and Behaviour Lab at the Faculty of Science at the University of Sydney.

Visit 1

In visit one, participants were informed about the aims and the procedures of the study and completed written informed consent. The aims of the study were described and explained

according to the cover story to make participant believe they would participate in a RCT about a new sleep remedy.

After all potential questions had been answered to participant's satisfaction, they were asked to provide demographic information and filled in questionnaires determining their sleep during the last week (PSQI and ISI), the WHOQOL-BREF and the DASS-21. After completion of the questionnaires, participants received the actigraphy devices alongside the instruction to continuously wear it for the next two weeks. At the end of the first visit, all participants were familiarised with the CSD-C and instructed to fill out the CSD-C every morning after their final awakening and the GASE every evening before going to bed.

Visit 2

After one week of baseline measures participants spent at home, they came for the second visit (See the grey box in Figure 3.1.). After participants had been guided to the laboratory facilities, they filled out the PSQI, ISI, WHOQOL-BREF, and the DASS-2. Participants were instructed to fill out the questionnaires regarding the first seven nights of the baseline week. Participants were then randomly allocated to one of the three groups (no-treatment, conventional, placebo, or active placebo) by drawing a number from an envelope (simple randomisation).

After the randomisation participants in the treatment group received the labelled capsule container with 28 blue-white capsules. Participants in the placebo groups were informed to take four capsules each day, one hour before their intended sleep onset. After this verbal instruction, participants were handed an information sheet that looked like a typical medication leaflet informing them about how to take the capsules, what to consider while taking the capsules, and potential side effects.

According to the cover story participants were asked to participate in a study trialling a new sleep medication consisting of an antihistamine (doxylamine) and a natural antioxidant (beetroot extract). Therefore, the information sheet mostly contained information present on the real doxylamine leaflet with the addition of the target side effect of the active placebo (unusually dark to red colouration of urine or stool). Potential side effects listed on the information sheet included side effects of the musculoskeletal, nervous, respiratory, and renal system. These side effects on the leaflet were indicated as "frequency not known". Additionally, participants in both placebo groups were told that at least 50 percent of participants were to experience an unusually dark or red colouration of their urine in the morning due to the capsules containing the red beetroot extract. Participants in the no-treatment group were honestly told that they were not receiving any medical treatment and that they will act as a control group for the natural course of sleep problems, general health, and well-being.

Subsequently, participants of the placebo groups rated their expectation regarding the effectiveness of the treatment and their anxiety about the occurrence of treatment associated side effects on visual-analogue scales. Further, they filled out the TSMQ-II and the PSM.

For the next week, all of the participants were asked to continue with their daily online surveys as they did during the baseline week. The placebo groups were instructed to fill out the additional adherence questions at the end of the GASE in the evening. At the end of the visit participants in the treatment condition got an official participation statement that they are taking part in a research study to increase the credibility of the study and to avoid participants from getting into any inconveniences should they be questioned by anyone (i.e., police or suspicious parents) about the capsules they were carrying around.

Visit 3

One week later, participants were invited to come to the lab for the third and last visit. First, all participants were asked to return the actigraphy device and the capsule container if they had received one during visit two. Then they started to fill out the PSQI, ISI, WHOQOL-BREF, and DASS-21. Participants in the placebo groups were additionally asked to fill out the TSQM-II and a manipulation check, asking to what group they believe they had been allocated. Eventually, participants were fully debriefed on the real aims of the study and reimbursed with two course credits or AUD$50 cash.

Sample Size Calculation

Following a power analysis, the required sample size was estimated as 24 participants finishing the study in each of the three groups, totalling 72 participants. The analysis was conducted with the software environment R version 3.4.2 (R Core Team, 2018) and the pwr package version 1.2-2 (Champely, 2018) using three groups, an effect size of $f = 0.4$, a significance level of $\alpha = .05$, and a power of $1\text{-}\beta = .85$. For the power analysis, outcome data from an earlier study about the placebo effect in sleep was used to calculate an appropriate effect size estimate for the placebo effect between the no-treatment group and the placebo groups (Colagiuri et al., 2012). Although experimental studies typically use a desired power of $1\text{-}\beta = .80$, a higher power level was chosen to further reduce the risk of type II errors (Baker et al., 2021). Due to expected attrition, recruitment was continued until the necessary number of participants in each group had finished the study.

Data Analysis

Participant data were excluded if participants did not answer daily measures at least four nights out of seven for the baseline and intervention week, or if they did not honestly fill out questionnaires (i.e., biological men reporting menstruation pain, contradicting side effects, or sleep values that are impossible). Adherence was defined as taking at least two capsules out of the four instructed on at least five out of the seven treatment nights. Nonadherence led to exclusion from statistical analyses. Participants were excluded from the analyses that reportedly experienced the flu, common cold, or gastroenteritis for more than one day because this would have biased participants symptom reporting and treatment outcomes.

All statistical analyses for the baseline characteristics and the outcomes were carried out using R version 4.0.4 (R Core Team, 2019b). Results were considered to be statistically significant when $p < .05$. Chi-squared tests and one-way analysis of variance (ANOVA) were used to compare demographic information and baseline characteristics across the three groups. Baseline variables that differed between the three groups with a p-value of $< .1$ were included in the primary and secondary analyses as covariates.

Changes in primary and secondary outcome measures (ISI, PSQI, CSD-C, actigraphy, GASE, DASS-21, WHOQOL-BREF, TSQM-II) between the baseline and intervention week were analysed using analysis of co-variance (ANCOVAs) with orthogonal planned contrasts to analyse differences 1) between the placebo groups and the no-treatment group, 2) between the active and conventional placebo group. According to Twisk et al. (2018), longitudinal analysis of covariance is advised when it comes to the analysis of treatment effects in repeated-measures designs in RCTs as they adjust for baseline differences and therefore avoid regression to the

mean as a confounding variable. For the between group analysis, two-way ANCOVA were calculated to examine the effects of treatment condition (no treatment, conventional, and active placebo) on primary and secondary outcomes while controlling for baseline scores as defined in the prior section.

Mediation analyses were conducted to examine the influence of expectations regarding (1) treatment efficacy, and (2) anxiety about side effects, (3) treatment satisfaction using the TSQM- II, and (4) the PSM scale on primary and secondary outcomes, if there was a statistically significant difference between the two placebo groups. The mediation analyses were calculated using the mediation package version 4.5.0 (Tingley et al., 2014) within the software environment R version 4.0.4 (R Core Team, 2019b).

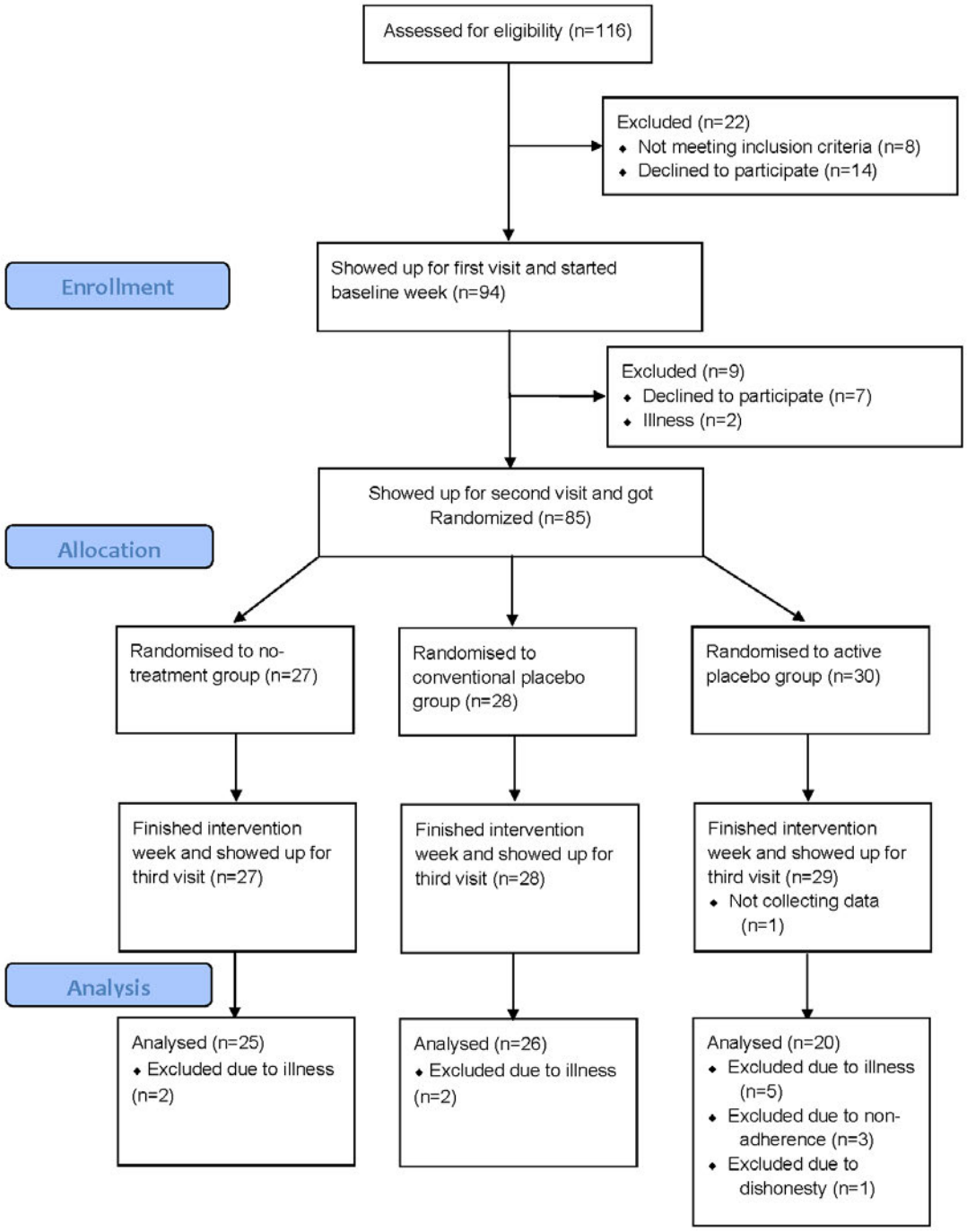
Results

An illustration of the participant flow from the eligibility screening to the final analyses is presented in Figure 3.3. A total of 116 participants expressed interest in this study and completed the online eligibility screening. After the screening, eight participants had to be excluded because they did not fulfil all inclusion criteria and an additional 14 participants declined to participate or never showed up for their first appointment. From the 94 participants showing up to the first appointment seven declined participation during the informed consent and two participants withdrew during the baseline week of the study due to illness. Out of the 85 participants who were randomly allocated to the three groups 84 finished the treatment week and showed up for the last visits. As for the predefined exclusion criteria, some participants needed to be excluded from analyses to uphold the quality of data. Most participants had to be excluded because due to

concomitant illnesses (n = 9; e.g., gastroenteritis, common cold, or influenza). Less common reasons for exclusion from data analyses included non-adherence (n = 3) and dishonesty (n = 1). While the individual reasons for exclusion of participants from the analyses did not differ statistically significantly across groups, the overall numbers of participants that had to be excluded from the analysis did, X^2 (2, 84) = 8.20, p = .017, with more dropouts in the active placebo arm than the other two groups.

Figure 3.3

CONSORT Flow Diagram



Note. CONSORT Flow diagram amended from Moher et al. (2001).

Demographics and Descriptive Data

Specific demographic information and sample characteristics obtained at the first visit for each of the three groups is presented in Table 3.2. The analysis sample for this study comprised 71 participants (44 women), between the ages of 18 and 41 years (M = 20.86, SD = 4.04), of which a large majority (63 participants) were studying at the University of Sydney participating for course credits. The two predominant nationalities present in this study were people from China (n = 26) and Australia (n = 21). As the study was conducted and advertised at the University of Sydney participants' general education level was high. All participants were either enrolled in an undergraduate course (n = 65), or in postgraduate degrees (n = 6). Sixty participants reported to be single and only ten indicated that they are in a relationship. Demographics and other characteristics at the first visit were similar across groups, only the age differed statistically significantly between groups.

Table 3.2

Demographic Information and Characteristics at First Visit

Variable	No-treatment (n = 25)	Conventional placebo (n = 26)	Active Placebo (n = 20)	Omnibus tests of statistically significant between group differences
	M (SD)	*M (SD)*	*M (SD)*	
Age (years)	**22.56 (5.90)**	***19.92 (2.12)***	**19.95 (2.14)**	***F* (2, 68) = 3.68, *p* = .031**
Gender				X^2 (2, 84) = 1.04, p = .596
Women	15	18	11	
Men	10	8	8	
Education level[a]				X^2 (8, 84) = 8.43, p = .393
Undergraduates	21	25	19	
Postgraduates	4	1	1	
Relationship status				X^2 (2, 84) = 0.01, p = .996
Single	21	22	17	
In a relationship	4	4	3	
Height (cm)	168.34 (10.46)	169.69 (9.15)	170.75 (11.23)	F (2, 68) = 0.32, p = .731
Weight (kg)	63.26 (11.26)	61.08 (12.83)	64.80 (14.38)	F (2, 68) = 0.50, p = .607
ISI	10.52 (5.82)	9.69 (5.88)	9.70 (5.34)	F (2, 68) = 0.17, p = .845
PSQI	6.24 (1.76)	6.50 (1.70)	6.05 (1.88)	F (2, 68) = 0.37, p = .689

DASS-21	13.2 (10.63)	14.23 (12.19)	16.65 (13.75)	F (2, 68) = 0.46, p = .631
WHOQOL-BREF	3.56 (0.57)	3.62 (0.52)	3.49 (0.58)	F (2, 68) = 0.34, p = .712
VAS Prior experience with medications	74.24 (16.85)	67.35 (19.85)	70.90 (17.38)	F (2, 68) = 0.92, p = .403
VAS Prior experience with side effects of medications	75.24 (18.29)	74.04 (24.39)	78.65 (18.45)	F (2, 68) = 0.29, p = .749

Note. The information provided in this table represents the analysis sample. [a] Education level was assessed using nine different levels ranging from "less than year 12 or equivalent" all the way to "doctorate". As only four levels were used representing undergraduate and postgraduate diplomas and degrees levels were condensed to undergraduate and postgraduate students for presentation this table. Statistically significant differences between the three groups as indicated by omnibus ANOVA, and Chi-square tests are bolded.

Primary Outcomes

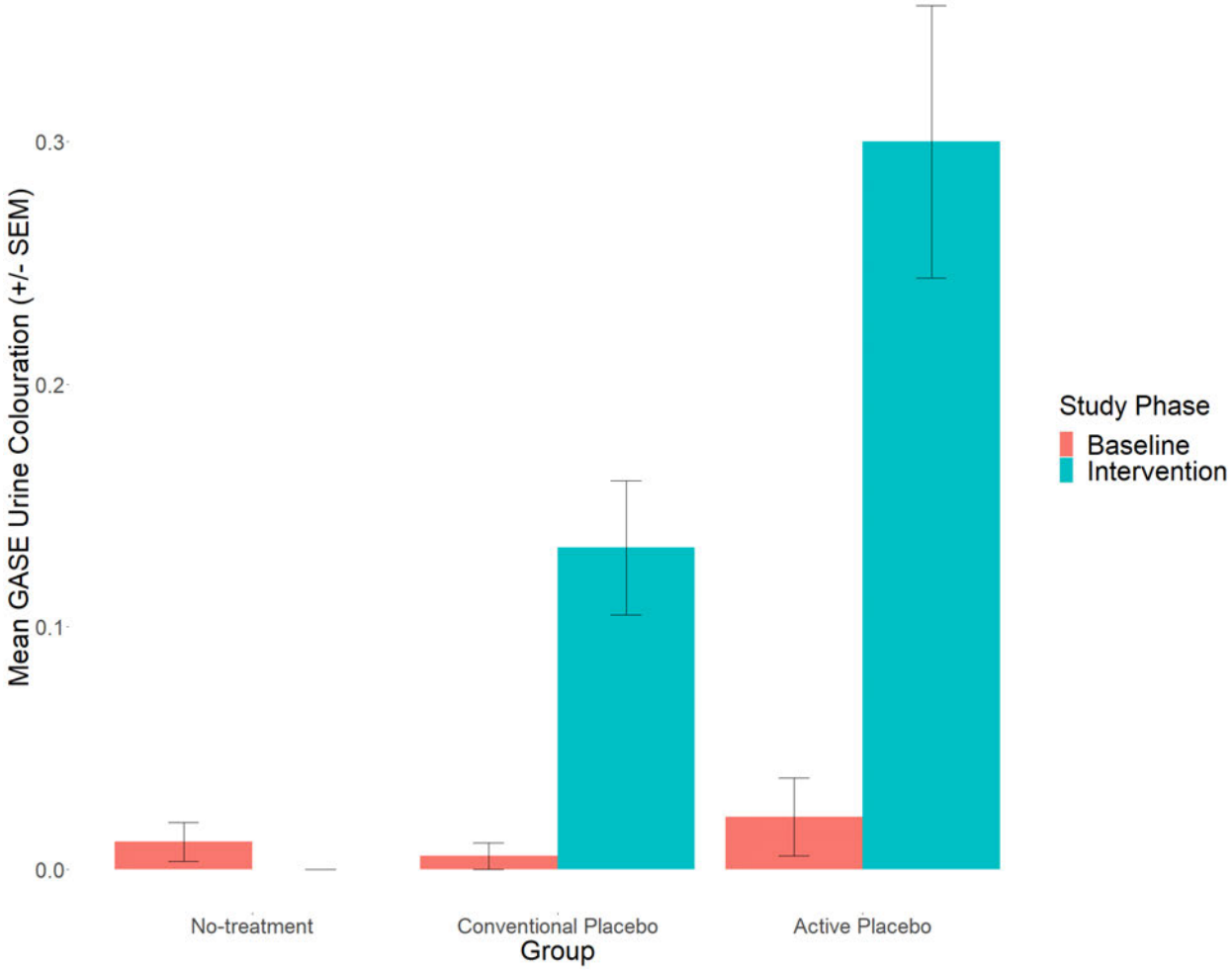
GASE Urine Colouration

In the no-treatment group not a single participant reported having experienced an unusual urine colouration during the second week of the study. Despite having received pure lactose placebos, 23 percent of participants in the conventional placebo group reported a urine colouration during the treatment week. In the active (beetroot) placebo group 50 percent of participants reported having experienced an unusual urine colouration. This difference between the groups was statistically significant, X^2 (2,71) = 15.92, p = .001.

Figure 3.4 shows how differently the three groups rated the severity of their urine colouration between the baseline and intervention week. Adjusting for the urine colouration during the baseline week, the ANCOVA analysis showed that the three groups differed statistically significantly in their reports of the severity of urine colouration during the treatment week, F (2,67) = 9.49, p < .001. The three baseline-adjusted urine colouration severity scores fitted for the ANCOVA model showed a small negative urine colouration severity for the no-treatment group (M = -0.002, SE = 0.452), a small increase in urine colouration for the conventional placebo group (M = 0.903, SE = 0.445), and the largest increase for the active placebo group (M = 2.129, SE = 0.508). The first planned orthogonal contrast comparing the no-treatment group to the two placebo groups showed that the two placebo groups reported a statistically significantly more discoloured urine during the treatment week as the no-treatment group (F (1,69) = 3.85, p < .001). The second planned orthogonal contrast compared the two placebo groups. The active placebo group indicated a statistically significantly more severe urine colouration as the conventional placebo group (F (1,44) = 2.33, p = .023).

Figure 3.4

Mean Baseline and Treatment GASE Severity of Urine Colouration across Groups



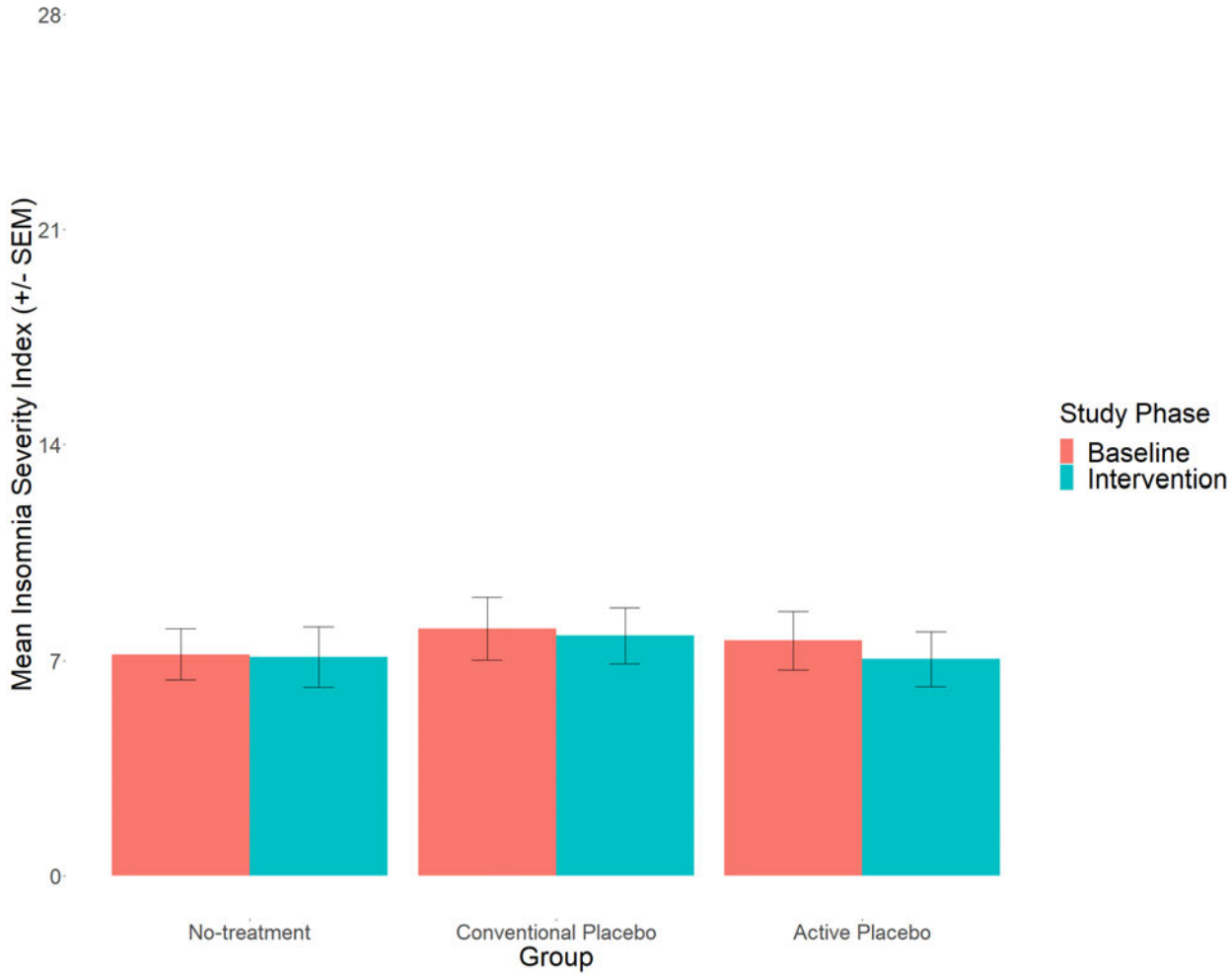
ISI

Figure 3.5 shows how the three groups' ISI score for the baseline and intervention week. Adjusting for the ISI during the baseline week, the ANCOVA analysis did not show that the three groups differed statistically significantly on the ISI during the treatment week, F (2,67) = 0.19, p = .826. The baseline-adjusted ISI scores fitted for the ANCOVA model were M = 7.47

(*SE* = 0.538) for the no-treatment group, *M* = 7.480 (*SE* = 0.527) for the conventional placebo group, and *M* = 7.037 (*SE* = 0.601) for the active placebo group. The first planned orthogonal contrast comparing the no-treatment group to the two placebo groups showed that the two placebo groups did not differ statistically significantly in terms of improvement on the ISI from baseline to intervention compared to the no-treatment group (F (1,69) = 0.32, p = .752). The second planned orthogonal contrast comparing the two placebo groups similarly to the first contrast did not indicate a statistically significant difference on the ISI between the conventional and active placebo group (F (1,44) = 0.56, p = .581).

Figure 3.5

Mean Baseline and Treatment Insomnia Severity Index across Groups



Secondary Outcomes

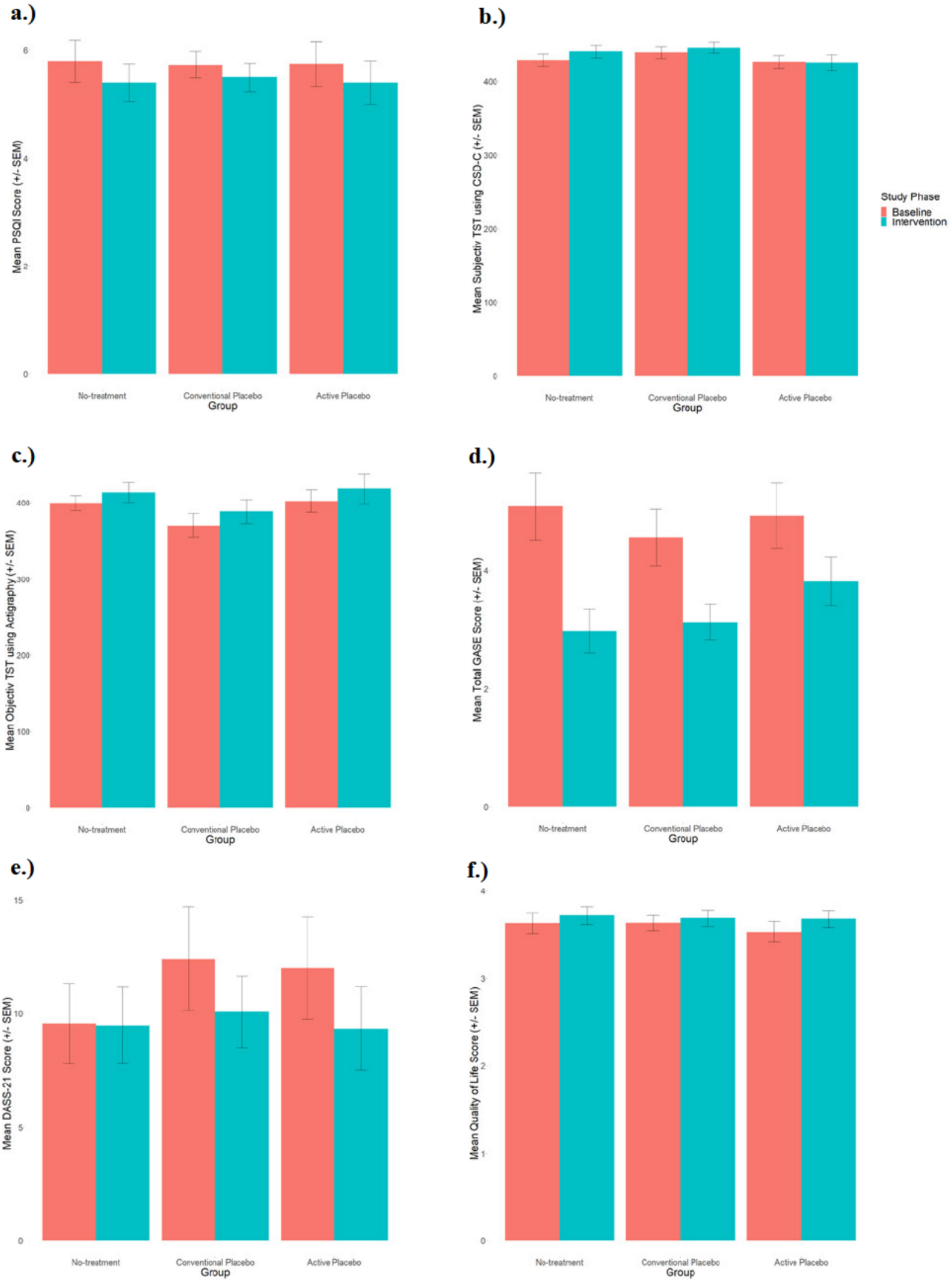
PSQI

Figure 3.6a depicts the PSQI scores of all three groups for the baseline and intervention week. As the PSQI and the ISI are similar measures the results for the PSQI were in line with the earlier reported ISI. Adjusting for the PSQI during the baseline week, the ANCOVA analysis did

not show that the three groups differed statistically significantly on the PSQI during the treatment week, F (2,67) = 0.07, p = .935. The baseline-adjusted PSQI scores fitted for the ANCOVA model were M = 5.380 (SE = 0.279) for the no-treatment group, M = 5.515 (SE = 0.273) for the conventional placebo group, and M = 5.406 (SE = 0.312) for the active placebo group. The first planned orthogonal contrast comparing the no-treatment group to the two placebo groups showed that the two placebo groups did not differ statistically significantly in terms of improvement on the PSQI from baseline to intervention compared to the no-treatment group (F (1,69) = 0.23, p = .818). The second planned orthogonal contrast comparing the two placebo groups similarly to the first contrast did not indicate a statistically significant difference on the PSQI between the conventional and active placebo group (F (1,44) = 0.27, p = .792).

Figure 3.6

Mean Baseline and Treatment Outcomes across Groups for Secondary Outcomes



Subjective TST as assessed with the CSD – C

Figure 3.6b illustrates the sTST of all three groups for the baseline and intervention week. The findings with the CSD – C regarding the subjective daily TST were in line with the ISI and the PSQI. Adjusting for the sTST during the baseline week, the ANCOVA analysis did not show that the three groups differed statistically significantly on the sTST during the treatment week, F (2,67) = 0.60, p = .550. The baseline-adjusted sTST in minutes fitted for the ANCOVA model were M = 442.197 (SE = 9.896) for the no-treatment group, M = 443.246 (SE = 9.725) for the conventional placebo group, and M = 428.394 (SE = 11.073) for the active placebo group. The first planned orthogonal contrast comparing the no-treatment group to the two placebo groups showed that the two placebo groups did not differ statistically significantly in terms of improvement on sTST from baseline to intervention compared to the no-treatment group (F (1,69) = 0.52, p = .607). The second planned orthogonal contrast comparing the two placebo groups similarly to the first contrast did not indicate a statistically significant difference on sTST between the conventional and active placebo group (F (1,44) = 1.01, p = .318).

Objective TST as assessed by Actigraphy

Figure 3.6d depicts the three groups' objective TST in minutes based on actigraphy assessed using the Actiwatch 2 devices between the baseline and intervention week. Compared to the analysis sample of the subjective outcomes that consist of 71 participants only a sub-sample of 18 participants were equipped with an Actiwatch 2 device and recorded Actigraphy across the two study weeks. This small sub-sample was unevenly distributed across the three groups with 11 participants in the no-treatment group, two participants in the conventional

placebo, and five participants in the active placebo group providing objective actigraphy data. The objective TST outcomes fell in line with the subjective assessments of the ISI, PSQI, and CSD – C not showing any statistically significant differences between the groups. Adjusting for objective TST at baseline, the ANCOVA analysis showed that the three groups did not differ statistically significantly in their objective total sleep time as assessed by actigraphy for the treatment week, F (2,14) = 0.29, p = .751. The baseline-adjusted objective TST in minutes fitted for the ANCOVA model were M = 414.436 (SE = 13.913) for the no-treatment group, M = 398.461 (SE = 16.492) for the conventional placebo group, and M = 413.719 (SE = 19.554) for the active placebo group. The first planned orthogonal contrast comparing the no-treatment group to the two placebo groups showed that the two placebo groups did not statistically significantly differ in their objective TST from baseline to intervention compared to the no-treatment group (F (1,16) = 0.44, p = .665). In line with the first contrast, the second planned orthogonal contrast comparing the two placebo groups did not indicate a statistically significant difference between the conventional and active placebo group (F (1,5) = 0.59, p = .567).

Total Side Effects as assessed by the GASE

Figure 3.6d shows the differences in total side effect reporting assessed using the GASE total score of all three groups for the baseline and intervention week. Adjusting for the GASE score during the baseline week, the ANCOVA analysis showed that the three groups statistically significantly differed in their total GASE score during the treatment week, F (2,67) = 3.293, p = .043. The baseline-adjusted total GASE scores fitted for the ANCOVA model were M = 19.843 (SE = 2.263) for the no-treatment group, M = 22.975 (SE = 2.573) for the conventional placebo

group, and M = 26.329 (SE = 2.932) for the active placebo group. The first planned orthogonal contrast comparing the no-treatment group to the two placebo groups showed that the two placebo groups statistically significantly differed in terms of totally experienced side effects from baseline to intervention compared to the no-treatment group (F (1,69) = 2.47, p = .016). The second planned orthogonal contrast comparing the two placebo groups did not indicate a statistically significant difference on total GASE score between the conventional and active placebo group (F (1,44) = 0.50, p = .622).

DASS-21

Figure 3.6e shows how differently the three groups rated their depression, anxiety, and stress score between the baseline and intervention week. Adjusting for the DASS-21 score at baseline, the ANCOVA analysis showed that the three groups did not differ statistically significantly in their DASS-21 scores for the treatment week, F (2,67) = 0.78, p = .468. The baseline-adjusted DASS-21 scores fitted for the ANCOVA model were M = 10.607 (SE = 0.970) for the no-treatment group, M = 9.345 (SE = 0.949) for the conventional placebo group, and M = 8.893 (SE = 1.080) for the active placebo group. The first planned orthogonal contrast comparing the no-treatment group to the two placebo groups showed that the two placebo groups did not statistically significantly differ in DASS-21 scores from baseline to intervention compared to the no-treatment group (F (1,69) = 1.23, p = .224). In line with the first contrast, the second planned orthogonal contrast comparing the two placebo groups did not indicate a statistically significant difference on DASS-21 scores between the conventional and active placebo group (F (1,44) = 0.32, p = .754).

WHOQOL-BREF

Figure 3.6f depicts the three groups' quality-of-life ratings between the baseline and intervention week. Adjusting for the WHOQOL-BREF rating at baseline, the ANCOVA analysis showed that the three groups did not differ statistically significantly in their quality-of-life ratings for the treatment week, F (2,67) = 0.41, p = .668. The baseline-adjusted WHOQOL-BREF scores fitted for the ANCOVA model were M = 3.695 (SE = 0.053) for the no-treatment group, M = 3.665 (SE = 0.052) for the conventional placebo group, and M = 3.736 (SE = 0.060) for the active placebo group. The first planned orthogonal contrast comparing the no-treatment group to the two placebo groups showed that the two placebo groups did not statistically significantly differ in their quality-of-life ratings from baseline to intervention compared to the no-treatment group (F (1,69) = 0.08, p = .938). In line with the first contrast, the second planned orthogonal contrast comparing the two placebo groups did not indicate a statistically significant difference on WHOQOL-BREF ratings between the conventional and active placebo group (F (1,44) = 0.901, p = .371).

TSQM-II

Treatment satisfaction was only assessed after the treatment week by the two placebo groups. The two-sided Welch t-test showed that the two placebo groups did not rate the treatments differently regarding their satisfaction as measured by the TSQM-II, t (42.64) = 1.29, p = .205. The conventional placebo group rated the treatment satisfaction as M = 59.936 (SE = 3.551) and the active placebo as M = 53.333 (SE = 3.697).

Discussion

The primary goal of this study was to evaluate how well the new active placebo model using beetroot elicited the desired target side effect in the form of an unusually red urine colouration. This study showed that the active beetroot placebo induced the desired target side effect in the form of an unusual red urine colouration in half (i.e., 50%) of all participants. Compared to the historically most often used pharmacological active placebo atropine eliciting dry mouth as a side effect in 58% of participants (Berna et al., 2017), the new active placebo using beetroot eliciting beeturia was considered sufficiently successful in terms of eliciting side effects. Interestingly, nearly a quarter of participants in the conventional placebo group reported an unusual urine colouration. The baseline-adjusted analysis showed an overall statistically significant effect that placebo participants overall rated their urine colouration as more severe as the no-treatment group that did not receive any placebos or the information that they might experience a urine colouration as a side effect. In line with the count data of participants reporting beeturia, participants in the active placebo group rated their daily red urine colouration statistically significantly more severe than the conventional placebo group.

While the new active placebo model successfully elicited the target side effect, none of the planned orthogonal contrast regarding the health-related outcomes between the two placebo groups and the no-treatment group, or between the conventional and active placebo groups were statistically significant. This means that applying an inert or active placebo did not elicit a statistically significant placebo effect on any of the sleep outcomes or other secondary health-related outcomes. The only exception was an observed difference in experiencing daily symptoms as assessed with the GASE between the no-treatment group and the two placebo

groups. Adjusting for the baseline week, the two placebo groups reported a larger daily symptom load as the no-treatment group.

The finding of an increased symptom load of participants receiving a placebo compared to the no-treatment group adds more evidence to the robustness of the nocebo effect. The absence of a placebo effect while observing a nocebo effect could be explained by the healthy participant sample recruited for this study. As discussed in the general literature review of this dissertation and specifically reviewed by multiple articles, there is strong evidence supporting that information presented to patients in the form of information leaflets or mere verbal suggestion is sufficient to introduce nocebo effects (Benedetti et al., 2007; Blasini et al., 2017; Petrie & Rief, 2019). It is therefore not surprising that participants in the conventional and active placebo conditions that both received written and verbal information about potential side effects when taking the capsules reported experiencing more daily symptoms than the no-treatment group that was not informed about any such effects and was not asked to take any capsules.

The absence of a statistically significant placebo effect on any of the sleep outcomes might be explained best by participant characteristics. As an illustration the current participant sample is best compared with the study by Neukirch and Colagiuri (2015) that observed a statistically significant placebo effect. They used a very similar design with one week of baseline and intervention measurements comparing a conventional placebo to no treatment. On average, their participants reported scores of approximately 10 points on the PSQI at baseline. According to Buysse et al. (1989) who validated the PSQI using normal sleepers and people with sleep problems' global PSQI scores above five can be indicative of poor sleep and their sample of people with sleep problems on average reported nine to ten points on the PSQI. The baseline sleep problems observed in Neukirch and Colagiuri (2015) demonstrated similar PSQI scores as

the sample of people with sleep problems in the original validation study (Buysse et al., 1989). The fact that the participant sample of the current study on average only reported scores <6 on the global PSQI at baseline, it is reasonable to argue that most participants were already fairly good sleepers at baseline. This argument is further supported by the mean ISI scores for the baseline week that ranged between seven and eight points for the groups at baseline. Bastien et al. (2001) categorised ISI scores of less or equal than seven as no clinically significant insomnia, eight to 14 as subthreshold, and generally an ISI of 10 or more is considered treatment-worthy. Arguably, participants in this current study were already better sleepers at baseline than the placebo group in Neukirch's study during the treatment week. It stands to reason that the current study was object to a floor effect where participants health related outcome measures were well enough from the beginning that there was no room for any significant improvements.

Besides the study sample's characteristics, a further possible explanation for the absence of an overall statistically significant placebo effect might be found in the experimental procedures used in the current study. As discussed in the general introduction, creating expectations that a treatment will result in a beneficial health outcome is an important aspect in eliciting placebo effects. While the procedures, including the cover story were implemented to resemble an actual RCT in as much detail as possible, some participants might have realised the underlying purpose of the current study. Because the principal investigator of the current study is being increasingly known as a placebo researcher, giving lectures about placebo effects at the University of Sydney and most of the participants were students enrolled at the University of Sydney participants might have realised that they were participating in placebo research instead of a pharmacological RCT as the cover story suggested.

An additional potential explanation for the statistically non-significant overall placebo effect in the current study might be related to placebo effects not being as strong or reliable for sleep problems as other conditions. Although the most recent systematic review and meta-analysis concluded that there is strong evidence for a reliable placebo effect for sleep problems Yeung et al., 2018), placebo effects have not been investigated as often and thoroughly in sleep problems as in other conditions like pain and depression (Ashar et al., 2017). The small number of published studies about placebo effects in sleep, therefore, might have led to a biased overestimation of the placebo effect in terms of its reliability and effect size.

Although the two placebo treatments were not rated statistically significantly different regarding participants' treatment satisfaction, there were some limitations associated with this study, specifically with the new active placebo model. The first and likely most important limitation identified while conducting this study had to do with the oxalic acid content in the active placebo capsules. In accordance with Eastwood and Nyhlin (1995), a large majority of participants receiving the active placebo did not report any adverse experiences. Nevertheless, two participants in the current study experienced gastroenteritis or acid reflux, most likely as a result of the oxalic acid contained in the capsules that were taken one hour before bed. Both participants decided to withdraw from the study due to these problems. This contributed to the imbalance of attrition rates between the three groups with more people being lost from baseline to follow up in the active placebo compared to the other groups. A further limitation of this study was that the three groups differed statistically significantly in their age, with the no-treatment group being approximately 2.5 years older than the two placebo groups. Although this difference was statistically significant, it did not significantly affect any of the primary or secondary outcomes. As another limitation of this study, it has to be mentioned that the objective sleep

outcome in the form of actigraphy was only assessed in a subsample of merely 18 participants, that were not evenly distributed across the three groups.

Although this study had some limitations, it importantly demonstrated that the new active placebo model successfully elicited the target side effect in at least 50% of participants – the primary aim – and that participants were similarly satisfied with the active placebo compared to a conventional placebo. While no statistically significant difference could be observed in the two placebo groups' improvements for any of the health-related outcome measures, there is at least a coherent trend observable in the expected direction with the active placebo at least numerically outperforming the conventional placebo group. Therefore, the methodological issues were addressed in Study 2 of this dissertation.

The next chapter presents the findings from a double-blind investigation that only included participants with treatment-worthy sleep problems to evaluate if the active placebo elicits larger improvements than a conventional placebo and if the active placebo influences participants' experienced treatment allocation.

Chapter 4: Study 2 - Double-blind Active vs. Conventional Placebo Efficacy and Perceived Treatment Allocation

This study built on Study 1 to test whether the active placebo produced larger placebo effects than a conventional placebo in the context of a (fake) double-blind randomised controlled trial, including assessing participants' perceived treatment allocation as a possible mediator of this effect.

Introduction

Study 1 was successful regarding the primary aims of the active placebo successfully eliciting the target side effect, being mostly well tolerated, and participants rating the active placebo treatment as satisfactory as the conventional placebo. Because the primary aim of Study 1 was to test the new active placebo on its ability to elicit the target side effect the recruitment strategy did not require participants to have sleep problems. Participants enrolled in Study 1 therefore were predominantly good sleepers. This meant that floor effects on sleep outcomes could have obscured evidence of a placebo effect.

Building on Study 1, the aims of Study 2 were twofold. The primary aim was to assess if the experience of side effects caused by the new active placebo model caused participants to be more likely to believe that they had actually received a medication compared to a conventional placebo group. The secondary aim was to investigate whether the active placebo group experienced a larger placebo effect for sleep as the conventional placebo group.

Identifying the potential reasons why patient blinding fails in RCTs that are supposed to be double-blind is important, because successful blinding minimises bias and maximises the internal validity of trial data (Karanicolas et al., 2010). When taking a trial participant's

perspective an improvement in symptomatology or lack thereof, or the experience of side effects or lack thereof might inform participants about the group they had been allocated to (Altman et al., 2004; Hróbjartsson et al., 2007; Sackett, 2004). When a patient that had suffered from a disease for a long time prior to participation in a trial suddenly experiences a noticeable improvement in their condition they might conclude that they had been allocated to the medication group. If there is no improvement during the trial a patient might either assume that the medication was not working, or more likely that they had been allocated to the inert placebo group. While the improvement or lack thereof might lead to failed participant blinding, Bello et al. (2017) argued that perceptible differences between the experimental and control intervention were mostly responsible for failed participant blinding. Again, when looking at blinding from a patient's perspective, experiencing a side effect might lead a patient to believe that they had been allocated to the active medication group. Because every single participant in a RCT must be informed about potential side effects, it is likely that a patient not experiencing any side effects will believe that they received the inert placebo.

As was reviewed in the general introduction there has been much attention directed towards evaluating failed participant blinding and perceived treatment allocation in clinical trials. Experimental work has highlighted how important participants' perception about treatment allocation can be. Across two experiments Colagiuri and Boakes (2010) studied the impact of false feedback given to participants about their cognitive performance after they had been given a treatment that participants were informed would either be caffeine or a placebo. In fact, all participants received a placebo and they had unbeknownst to them been randomised to either receive the information that their cognitive performance had improved after treatment, or that there was little change. The findings from their first experiment demonstrated that participants

led to believe their performance had improved were statistically significantly more likely to believe that they had been allocated to the treatment group compared to the no-change information group. Even more importantly, the second experiment showed that participants' actual cognitive ability improved more when they had prior received positive feedback and therefore believed that they had been allocated to the treatment group, compared to the group made to believe that they had shown little change. The finding that perceived treatment allocation can influence the actual treatment outcome was further confirmed by Laferton et al. (2018) who reanalysed the findings of a double-blind depression RCT and found that perceived treatment allocation at week six, but not actual treatment assignment predicted improvements in depressive symptoms that occurred over the next six weeks.

While failed participant blinding and perceived treatment allocation already received scientific attention, not so much research has been conducted investigating the role of side effects on perceived treatment allocation and how the experience of side effects might further affect treatment effects. Being interested in the effects of side effects on the placebo effect Thomson (1982) conducted what would now be known as a systematic literature review or even a meta-analysis. The author collected a total of 75 double-blind trials testing tricyclic drug therapies for the treatment of depression. The comparison was simply made between trials using an inert placebo and trials using atropine as an active placebo as it mimics the side effects of tricyclic medications. The results showed that drugs in 43 out of 68 trials outperformed the inert placebo, while only one out of seven trials showed a superior effect of the drug over atropine. The authors suggested three potential explanations for the observed findings. First, that anticholinergic drugs like atropine might have antidepressant effects themselves. Second, that researchers using atropine as an active comparison are generally more motivated to conduct

thorough and unbiased research. And third, the authors hypothesised a placebo amplification phenomenon, where side effects further increased the efficacy of the drug compared to the inert placebo. The additional amplification of the drug's effect was hypothesised via psychological aspects caused by the experience of side effects in the drug group compared to the placebo group that did not experience any side effects.

In the year 2008 a meta-analysis published by Irving Kirsch and colleagues made headlines all around the world. They had analysed RCTs submitted to the U.S. Food and Drug Administration that were submitted by pharmaceutical companies in the process of approval for six newer generation antidepressants (Kirsch et al., 2008). They had found that the placebo response accounted for 75 percent of the improvement observed in the drug groups, a difference that was statistically significant, but nowhere near what was considered to be of clinical relevance. The debate that ensued rejuvenated placebo research as it highlighted the importance of the placebo phenomenon. Following the publication of Kirsch's meta-analysis in 2008 a heated discussion ensued about the validity of the meta-analytical finding. Some researchers and clinicians could not understand that an inert placebo accounted for three quarters of the effects of approved antidepressant drugs that were prescribed to millions of patients. Fountoulakis and Möller (2011) for example reanalysed Kirsch et al's data and concluded that their analysis suffered flaws and that they had not reported all results, meaning their conclusions were partially unjustified and overemphasised. Although Fountoulakis and Möller (2011) found some problems, they stated that most results were verified and that expectancy plays an important role for the placebo response. Kirsch (2014) argued that the serotonin hypothesis that is supposed to cause depression is nothing but a myth. His argument is that no matter what type of drug had been prescribed to depressed patients (i.e., selective serotonin reuptake inhibitors, tricyclic

antidepressants, or even tranquilisers or thyroid medications) the drug response on depressive symptoms was always similar and the placebo response always accounted for approximately 75 percent of the drug response. Based on these findings the authors concluded that the antidepressant drugs in fact did not have a specific antidepressant effect, but the placebo response within the drug groups was enhanced due to the experience of side effects (Kirsch, 2014).

In arguing that the experience of side effects amplify the placebo response within the drug group of double-blind RCTs, Kirsch (2014) came to the same conclusions as Thomson (1982) had hypothesised. Although these two authors came independently to the same conclusions, they did not report any actual data to support their suggestions. Actual evidence about the influence of side effects on perceived treatment allocation and the influence of experiencing side effects on a treatments efficacy was reported in a systematic review and meta-analysis about pharmacological RCTs for the treatment of chronic pain. Colagiuri et al. (2019) included trials published between 2006 and 2016 and found that only 23 out of 408 trials had reported blinding data. When analysing the success of blinding, their analysis showed that blinding across trials was broken. Due to their meta-analytical approach the authors were able to calculate moderator analyses to investigate factors that are associated with blinding. Their results indicated that higher rates of adverse events and larger treatment effect sizes were associated with higher rates of failed participant blinding. As was also discussed in great detail in the general introduction, the role of side effects influencing perceived treatment allocation and the placebo effect did not get enough attention in the form of experimental investigations specifically targeting the role of participants experiencing side effects.

The current study therefore aimed to investigate how the new model of an active placebo eliciting beeturia as the target side effect affected participants' perceived treatment allocation and how the experience of this side effect influenced the placebo effect regarding their sleep problems. Building on Study 1, the current study's methodology needed to be adjusted in two core aspects. First, the study was presented as a double-blind RCT evaluating a new sleep remedy against a placebo control group so that participants' perceived treatment allocation could be assessed. Second, the advertisement strategy and inclusion criteria were adjusted as to only include participants with insomnia symptoms. This change was necessary to guarantee that there was enough room for an improvement in participants' insomnia symptoms that a placebo effect could be observed in the first place.

Following the literature provided above two primary a priori hypotheses were formulated. First, it was hypothesised that participants in the active placebo group were more likely to believe that they had received an actual medication compared to the conventional placebo. Second, the active placebo group was expected to elicit a larger placebo effect compared to the inert placebo group on the primary and secondary outcome measures.

Methods

The University of Sydney's Human Research Ethics Committee reviewed and approved all ethical aspects regarding the recruitment, materials, and procedures for this study (Project Number: 2018/107). This included all necessary modifications following the insights gained from Study 1. Detailed documentation of the ethical approvals and the necessary modifications can be found in Appendix C.

The detailed methodology of this study was registered before enrolment of the first participant on ANZCTR (Identifier ACTRN12618002048268; Australian and New Zealand Clinical Trials Registry, 2018b). Because the methodology of the current study built on Study 1, only methodological aspects that are specific to the current study are presented to avoid unnecessary redundancies. While the design of the study resembles Study 1, the cover story differs in an important aspect, in that the current study was disguised as a double-blind placebo-controlled trial including three groups (no-treatment, placebo, and medication), while participants in Study 1 were instructed that they were either allocated to no-treatment or to the active medication.

Participants

The participant sample for this study consisted of sleep-impaired adults recruited in the greater Sydney area. Contrary to Study 1, the advertisements for the current study specifically mentioned that volunteers for the evaluation of a new sleep medication were wanted that regularly suffered from sleep problems, including difficulty initiating sleep, waking up frequently during the night and having difficulties falling back asleep, or waking up earlier than desired without being able to fall back asleep. The advertisement flyer can be found in Appendix C.

Participant recruitment of the current study used the same recruitment opportunities that were used in Study 1. In addition, the current study made use of the University's casual work webpage, Gumtree, Google Advertisement, and Facebook Advertisement. Because volunteers had to attend campus for participation online advertisement was limited to the greater Sydney area.

To be eligible for the current study an inclusion criteria was formulated that potential participants had to score at least 10 or more points on the ISI, meaning they suffered from treatment-worthy sleep problems (Bastien et al., 2001). Because of the insights gained from Study 1 and to further minimise the potential harm to participants caused by the oxalic acid contained in the active placebo, participants with a sensitive stomach or any stomach problems, especially acid reflux were excluded from participation. The full list of exclusion criteria were: (1) If they were currently taking prescription medication with the exception of contraceptive pills. (2) If they were pregnant, breastfeeding, or trying to conceive. (3) If they had received professional treatment for sleep problems in the last three months. (4) If they had an antihistamines, lactose, or beetroot allergy, or any other intolerances. (5) If they suffered from an abnormal or deficient kidney functioning or any other medical condition. (6) If they had a sensitive stomach or gastric problems, especially acid reflux. In addition to the listed exclusion criteria participants that had already participated in Study 1 were not allowed to participate again in the current study. As in Study 1 participants were made aware that the capsules were not vegan or halal. If a participant withdrew from the study, remuneration was proportionate to the time they had invested up until their withdrawal. The recruitment period for this study started in January 2019 and was finished in October 2019.

Design

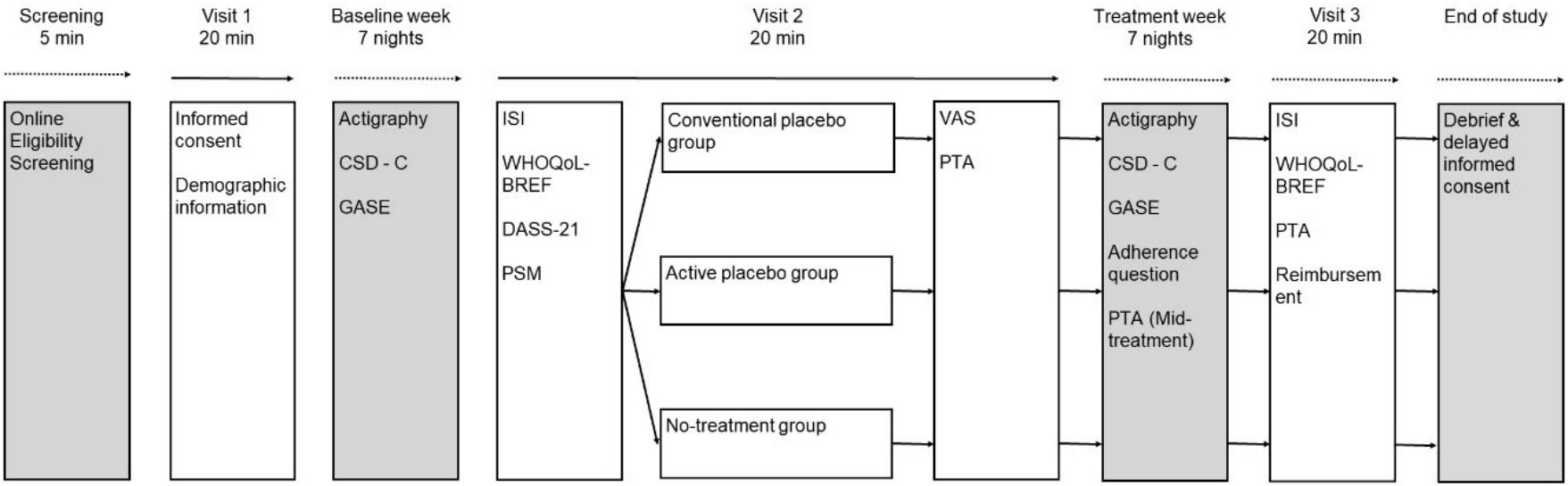
Figure 4.1. shows study design with the randomisation to the three study groups and the two time periods. The experiment used a 3 x 2-mixed design with treatment (inert vs. active vs. no-treatment) and time (baseline vs. intervention) as factors. The factor time was divided in a baseline and a treatment period, each covering seven nights. In visit two, participants were

randomly allocated to one of the groups and received either a placebo treatment or no treatment. The placebo treatment unbeknownst to the participants either consisted of an active beetroot placebo or an inert lactose placebo.

The independent variable was group allocation to one of the three experimental groups and the independent variables were participants perceived treatment allocation and the Insomnia Severity Index.

Figure 4.1

Study Design and Procedure with all Measures



Materials

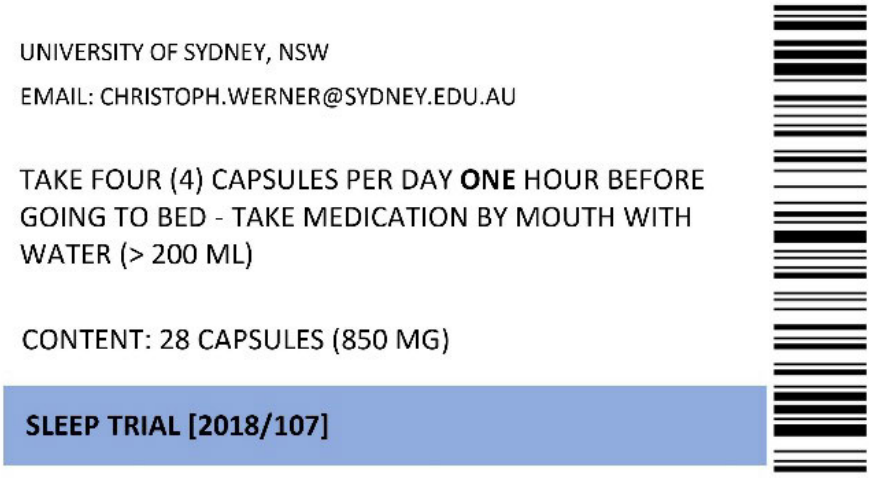
Placebo capsules.

Placebo capsules were identical to Study 1 with the following exception. The active placebo was adjusted to only contain half the amount of oxalic acid and therefore contained more beetroot extract. The active placebo capsules for the current study each contained 625 mg of the food colour E162 (i.e., beetroot extract) and 125 mg of oxalic acid.

Contrary to Study 1 where capsule containers were labelled to contain a mix of beetroot extract and doxylamine, the capsule containers in the current study were labelled neutrally according to the cover story that a double-blind placebo-controlled randomised trial was conducted for sleep problems. The label that was used in this study is displayed in Figure 4.2.

Figure 4.2

Capsule Containers Label



Cover story.

The current study used a similar cover story as Study 1 making participants believe a new medication was tested, but the current study was presented as a double-blind RCT according to the new aims. As per Study 1, participants were deceivingly informed that they were participating in an open-label RCT that tested the effect of a new compound against no treatment. However, critically in Study 2, participants were informed that they would participate in a double-blind placebo-controlled randomised trial that consisted of three different groups, an active medication group, a placebo group, and a no-treatment group.

As in Study 1 information given to participants were standardised using an experimenter manual that can be found in Appendix C. Both, the inert and active placebo groups received the same explanations and instructions about the treatment under investigation and potentially associated side effects. Experimenters were blinded towards participants' allocation to the conventional or active placebo group.

Measures

The hierarchy of outcome measures and definition of assessment time points was pre-defined and registered within the ANZCTR (Identifier ACTRN12618002048268; Australian and New Zealand Clinical Trials Registry, 2018b). Table 4.1 defines the primary and secondary outcome measures and when these measures were assessed over the course of the study.

Measures were identical to Study 1 with the exception of additional measures and changes listed below.

Table 4.1

Definitions of Outcome Measures and Assessment Time Points

Outcome Hierarchy	Outcome Measure and Definition of Assessment Time Point for Statistical Analyses
Primary outcome (1)	Perceived treatment allocation (PTA), using a forced-choice question with the options (0=no-treatment group, 1=placebo group, 2=doxylamine group)
Timepoint (1)	Baseline (assessed at a single timepoint post-randomisation, but prior to receiving treatment) Mid-treatment (assessed at a single timepoint on the day after 3 nights of treatment) Post-treatment (assessed at a single timepoint at the end of the seven-night period receiving treatment)
Primary outcome (2)	Insomnia severity assessed using the Insomnia Severity Index (ISI)
Timepoint (2)	Baseline (assessed at a single timepoint on the day starting treatment, for the 7 nights period prior to receiving treatment) Treatment (assessed at a single timepoint on the day after finishing 7 nights of treatment, for the 7 nights period receiving treatment)
Secondary outcome (1)	Total sleep time (self-report) using the Consensus Sleep Diary Version C (CSD-C)
Timepoint (1)	Baseline (assessed daily during the 7 nights prior to randomisation) Treatment (assessed daily during the 7 nights receiving treatment after randomisation)
Secondary outcome (2)	Sleep onset latency (self-report) using the Consensus Sleep Diary Version C (CSD-C)
Timepoint (2)	Baseline (assessed daily during the 7 nights prior to randomisation) Treatment (assessed daily during the 7 nights receiving treatment after randomisation)
Secondary outcome (3)	Reports of daily symptoms using an amended 10-item version of the General Assessment of Side Effects (GASE). We shortened the GASE to only include the most relevant items and added a question regarding urine colouration. Additionally, we changed the items from a 4-point multiple choice about the severity of the symptoms to a visual analogue scale (VAS) ranging from 0 (= “not present”) to 100 (= “severe”), with
Timepoint (3)	

	33 (= “mild”) and 66 (= “moderate”)
	Baseline (assessed daily during the 7 nights prior to randomisation) Treatment (assessed daily during the 7 nights receiving treatment after randomisation)
Secondary outcome (4)	Quality of life using the World Health Organisation’s quality of life assessment (WHOQOL-BREF)
Timepoint (4)	Baseline (assessed at a single timepoint on the day starting treatment, for the 7 nights period prior to receiving treatment) Treatment (assessed at a single timepoint on the day after finishing 7 nights of treatment, for the 7 nights period receiving treatment)
Secondary outcome (5)	Total sleep time (objective) using Actigraphy (Activinsights, GENEActiv Original)
Timepoint (5)	Baseline (assessed daily during the 7 nights prior to randomisation) Treatment (assessed daily during the 7 nights receiving treatment after randomisation)
Secondary outcome (6)	Sleep onset latency (objective) using Actigraphy (Activinsights, GENEActiv Original)
Timepoint (6)	Baseline (assessed daily during the 7 nights prior to randomisation) Treatment (assessed daily during the 7 nights receiving treatment after randomisation)
Secondary outcome (7)	Participants’ certainty of their choice regarding perceived treatment allocation (cPTA), operationalised as VAS ranging from 0 (= “not certain at all”) to 100 (= “absolutely certain”)
Timepoint (7)	Baseline (assessed at a single timepoint post-randomisation, but prior to receiving treatment) Mid-treatment (assessed at a single timepoint on the day after 3 nights of treatment Post-treatment (assessed at a single timepoint at the end of the seven-night period receiving treatment)

Note. The hierarchy and time points of these measures have been registered on ANZCTR. Please note that the measures in some cases were assessed more often over the course of the study as defined here for the purpose of the statistical analyses.

Perceived Treatment Allocation (PTA)

Participants' Perceived Treatment Allocation was assessed with a study-specific forced choice question with three options ("No-treatment group (no capsules given)", "Placebo group", and "Doxylamine group"). On three occasions participants were asked to what group they thought that they had been allocated. Right after indicating the group to which participants thought, they had been allocated, they were asked to indicate how certain they were about their choice (cPTA). The question "How certain are you about your group allocation?" was answered using a VAS ranging from 0 (= "not certain at all") to 100 (= "absolutely certain").

General Assessment of Side Effects (GASE; Amended)

It became apparent through multiple participant comments during Study 1 that filling out the complete 47-item GASE each morning caused unnecessary participant inconveniences. This was especially problematic because participants rarely experienced some of the symptoms included on the GASE (i.e., palpitations, difficulty urinating, or convulsions) hence not adding any real benefits to the study while causing unnecessary time commitments for participants. Therefore, the GASE for the current study was reduced to the 10 most important symptoms based on Study 1. The 10-item GASE for the current study included the following symptoms: headache, dry mouth, dizziness, nausea, fatigue / loss or energy, muscle weakness, upset stomach / indigestion, yellow to reddish colouration of urine or stool, drowsiness, and difficulty concentrating.

The original format of the GASE as used in Study 1 presented participants with a 0 - 3 Likert scale to assess single items. To increase the sensitivity and to implement participants' feedback from Study 1 complaining about having troubles deciding between two of the four

crude options, the GASE was amendment to a continuous VAS. To keep a resemblance and guide participants in rating the severity of the experienced symptoms the VAS ranged from 0 (= "not present") to 100 (= "severe"), with two ankers chosen as 33 (= "mild") and 66 (= "moderate"). Similar to Study 1, if participants rated symptoms as present, meaning any score other than zero, they were asked to indicate if they thought that the experience of this symptom was associated with the medication they had taken.

GENEActiv Original (Activinsights, UK).

Actigraphy measurement of movements is a non-invasive method of estimating sleep parameters by comparing times during the day where participants were active during the day or resting during the night. The GENEActiv (Activinsights, UK) used for this study is a watch-like wrist-worn device, that records movement using a tri-axial accelerometer, light intensity, and skin temperature. All data was assessed at a sampling rate of 50 Hz, meaning 50 data points each second for all parameters. The GENEActiv is 40 mm wide and 13 mm deep weighing 27 g (with band). The GENEActiv Original was chosen over the previously used Actiwatch 2, because it offers some important advantages. First of all, Activinsights supports open-source offering the opportunity to get all data in raw SI units, compared to the Actiwatch 2 used Study 1 that only allowed to get aggregate data outputs from their native analysis program, without the chance of knowing what was calculated due to their closed-source programs. Further, the newer technology of the GENEActiv Original allows to collect data at a higher sampling rate creating more accurate data analysis while even being able to collect data over longer time periods. Lastly, and most importantly an advantage of the GENEActiv Original over the Actiwatch 2 for participants

is the general wearing comfort that is higher due to the materials used for manufacturing. Further increasing the ease of usability is the fact that the GENEActiv is more robust allowing participants to wear it during any physical activity. The GENEActiv is also waterproof which means that participants do not have to take it off to shower or while spending time at the beach.

Sleep data obtained from the GENEActiv Original were calculated using open-source R package GGIR version 2.3-0 (van Hees et al., 2021), estimating TST and SOL. The GGIR package's autocalibration algorithm has been validated to calculate sleep outcomes based on actigraphy data for healthy participants as well as clinical populations from multiple devices with a focus on data collected with Activinsights devices like the GENEActiv Original (van Hees et al., 2014; van Hees et al., 2015).

The GENEActiv devices were used in the current study to assess objective SOL and TST. Compared to the shortfall of available devices in Study 1, meaning that only a subsample was equipped with actigraphy devices, all participants in the current study received a GENEActiv Original. Participants were instructed to wear the actigraphy devices for the whole two-week period of the study and were only allowed to take the devices off during sport activities that prohibited participants from wearing the devices.

Randomisation and Blinding

As REDCap was specifically designed for RCTs its built-in solution to randomise participants to different groups was used for the current study because it supports customised stratified randomisations. Therefore, REDCap was used to randomly allocate participants to the three groups during the second visits taking participants' gender (factorised with three levels: female, male, and diverse) and baseline ISI score (factorised with three levels: ISI < 10, ISI 10 –

15, and ISI > 15) into account creating three groups that are similar in their baseline characteristics. The stratification strategy incorporated a 1:2:2 randomisation ratio, so that twice the number of participants were allocated to each placebo group compared to the no-treatment group. The strategy to keeping the experimenter blinded consisted of the same approach as was described in Study 1.

Procedure

To avoid unnecessary redundancies only changes in procedures that deviate from Study 1 are mentioned. Changes were also made regarding the cover story to further support the credibility of the cover story. The study design and procedure are depicted in Figure 4.1 in the Design section above.

Visit 1

In visit one, participants were informed about the aims and the procedures of the study and completed written informed consent. The aims of the study were described and explained according to the cover story to make participant believe they would participate in a double-blind RCT about a new sleep remedy, therefore participants had to specifically agree to the collection of biological material in the form of a mouth swab carried out during the last visit that would allegedly be used to measure traces of the new compound under investigation. After all potential questions had been answered to participant's satisfaction, they were asked to provide demographic information. After completion of the questionnaire, participants received the actigraphy devices alongside the instruction to continuously wear it for the next two weeks. At the end of the first visit, participants were familiarised with the daily online survey (CSD-C and GASE) that they were asked to fill out each morning as soon as possible after their final awakening.

Visit 2

After one week of baseline measures at home, participants came to the lab for the second visit. There, participants were asked to fill out the first block of retrospective baseline questionnaires, namely the ISI, WHOQOL-BREF, DASS-21, and the PSM. Participants were then randomly allocated to one of the three groups (no-treatment, conventional placebo, or active placebo).

After the randomisation participants in the two placebo conditions received a capsule container with 28 blue-white capsules. Participants were informed to take four capsules each evening, one hour before their intended sleep onset. After this verbal instruction, participants were handed an information sheet that looked like a typical medication leaflet informing them

about how to take the capsules, what to consider while taking the capsules, and potential side effects. Participants in the no-treatment group were told that they were not receiving any medical treatment and that they will act as a control group for the natural course of sleep problems, general health, and well-being.

Subsequently, participants of the placebo groups rated their expectations regarding the effectiveness of the treatment, their anxiety about the occurrence of treatment associated side effects on VAS, and to guess to what group they had been allocated (PTA).

For the intervention week, all of the participants were asked to continue with their daily online surveys as they did during the baseline week. The placebo groups were instructed to fill out an additional adherence question after the GASE on a daily basis. At the end of the visit participants in the treatment condition got an official participation statement that they were taking part in a research study to increase the credibility of the study and to avoid participants from getting into any inconveniences should they be questioned about the capsules by police, suspicious parents, or anyone else. After the first three nights of the treatment week, participants were asked to guess their allocation to the study groups (PTA) as a mid-treatment assessment.

Visit 3

One week later, participants were invited to come to the lab for the third and last visit. First, all participants were asked to return the actigraphy device and the capsule container if they had received one during the second visit. Participants that had received capsules then underwent a mouth swap collecting saliva under the guise that their sample would be tested for traces of the new compound by an independent laboratory to objectively test their adherence. Then they started to fill out the retrospective intervention week questionnaires, namely the ISI, WHOQOL-

BREF, PSM, and PTA. Eventually, participants were fully debriefed on the real aims of the study and reimbursed with two course credits or 50 AUD cash.

Sample Size Calculation

For the sample size estimation regarding perceived treatment allocation, an earlier study about the influence of feedback on placebo effects in double-blind RCTs was referenced to (Colagiuri & Boakes, 2010). Based on these results and adjusting for the fact that the active placebo elicited the desired side effect in 50% of participants compared to the 100% of feedback participants got in Colagiuri and Boakes (2010), the required sample size would be 13 per group (f = 0.895, α = .05, and 1-β = .90). For the sample size estimation regarding the placebo effect on the sleep outcome, a study about influences of active placebos compared to benign placebos on pain was used as basis (Rief & Glombiewski, 2012). The power analysis resulted in a required completer sample of at least 46 participants in each of the placebo groups (f = 0.34, α = .05, and 1-β = .90). The power analyses were conducted with the software environment R version 3.4.2 (R Core Team, 2018) and the pwr package version 1.2-2 (Champely, 2018) for general linear models using two groups. Based on these power analyses the allocation of participants to the three groups used a 1:2:2 randomisation ratio for no-treatment: active placebo: benign placebo, until 46 participants in the two respective placebo groups and 23 participants in the no-treatment group were reached. This approach was chosen because the primary interest was the difference in perceived treatment allocation and sleep outcomes between the two placebo groups, with the no treatment group serving only as a comparison for the overall placebo effect. Due to expected attrition, recruitment was continued until the necessary numbers in each group were realised.

Data Analysis

Exclusion of participants from data analysis, assessment of baseline characteristics, the statistical program, and definition of statistical significance in the current study was identical to Study 1.

Changes in primary and secondary outcome measures (PTA, ISI, CSD-C, GASE, actigraphy, WHOQOL-BREF, and cPTA) between the baseline and intervention week were again analysed using ANCOVAs with orthogonal planned contrasts to analyse differences 1) between the placebo groups and the no-treatment group, 2) between the active and conventional placebo group.

Mediation analyses were calculated according to the methodology of Study 1.

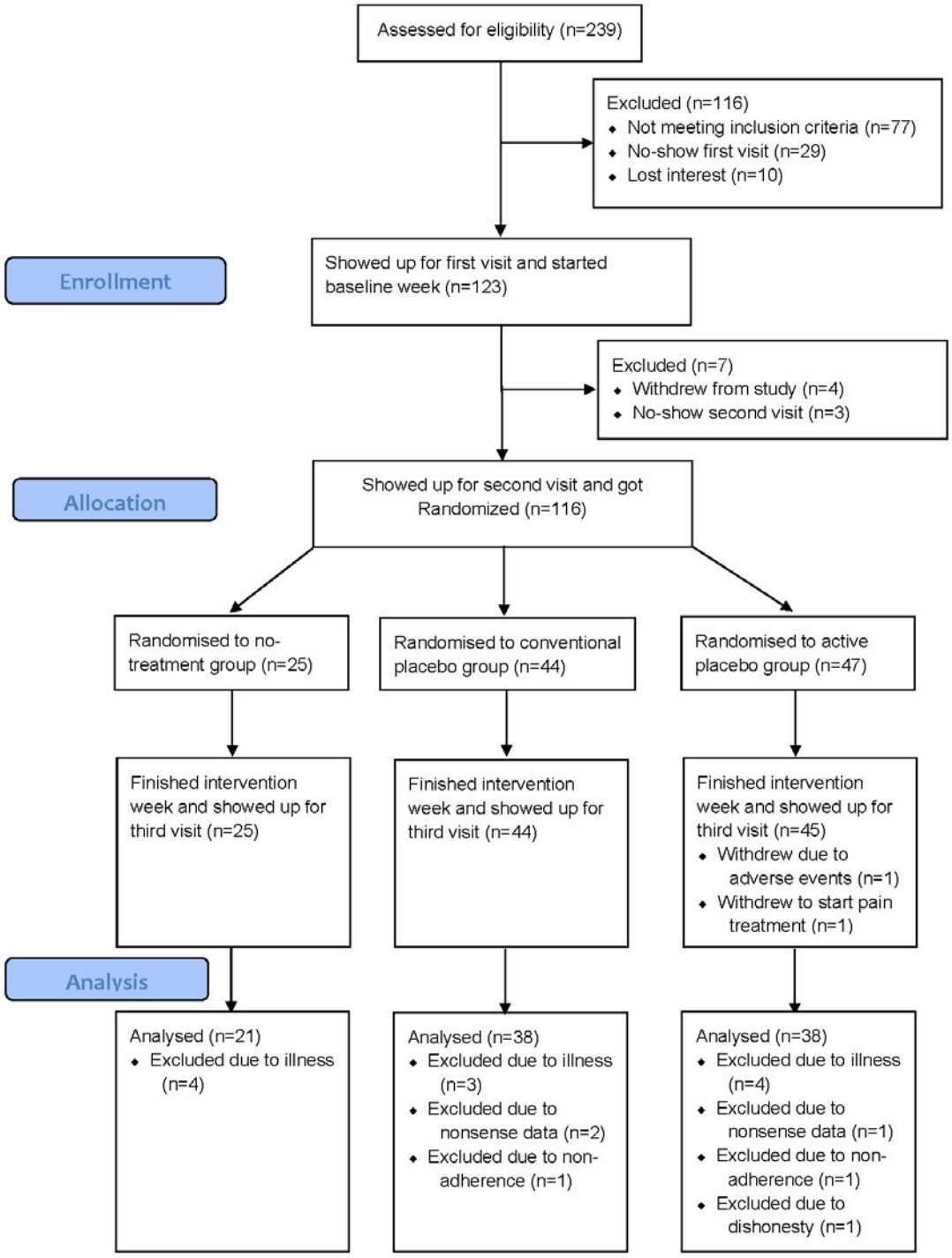
Results

The participant flow from the online eligibility screening to the final analysis sample is presented in Figure 4.3. A total of 239 people completed the online eligibility screening. After the screening, 77 participants had to be excluded because they did not fulfil all inclusion criteria, 10 lost interest and cancelled the first appointment, and an additional 29 participants never showed up for their first appointment without giving any reason. From the 124 participants showing up to the first appointment four declined participation during the informed consent and three participants withdrew during the baseline week of the study before the second appointment. Out of the 116 participants who were randomly allocated to the three groups 114 finished the treatment week and attended the final visits. Two participants in the active placebo group

withdrew from the study, one to start pain treatment for a sustained injury, and one experienced a mild adverse event and opted out of the study. As for the predefined exclusion criteria, some participants needed to be excluded from analyses to uphold the quality of data. Most participants had to be excluded because they got ill (n = 11; e.g., gastroenteritis, common cold, or influenza). Less common reasons for exclusion from data analyses included data provided by participants that was incoherent (n = 3), non-adherence (n = 2), and dishonesty (n = 1), meaning one student participant enrolling for course credits was answering all questionnaires with the option most to the left on display. The attrition rate from randomisation during the second appointment and the analysis sample were not statistically significantly different between the three groups, X^2 (2, 116) = 0.51, p = .776.

Figure 4.3

CONSORT Flow Diagram



Note. CONSORT Flow diagram amended from Moher et al. (2001).

Demographics and Descriptive Data

Specific demographic information and sample characteristics for each of the three groups is presented in Table 4.2. The analysis sample for this study comprised 97 participants (63 women), between the ages of 18 and 50 years (M = 21.65, SD = 5.55). A majority of participants (n = 78) were studying at the University of Sydney participating for course credits (n = 54) or cash reimbursement through the paid participant pool (n = 24). Only 19 participants could be recruited form the general population. Similar to Study 1, the two predominant nationalities present in this study were people from China (n = 34) and Australia (n = 32). As the study was conducted and mostly advertised at the University of Sydney the participants' general education level was high. Most participants were either enrolled in an undergraduate course (n = 86), or were enrolled in postgraduate degrees (n = 4). Only five participants reported having a vocational qualification and two indicated that they did not finish year 12 (equivalent to high school). Seventy-four participants reported to be single, 21 indicated that they are in a relationship, and two reported that they were married but living separate from their partner. Demographics and other characteristics as assessed at the first visit or as baseline measures during the second visit were similar across groups for most characteristics. The exceptions were relationship status, with more people being in a relationship in the no-treatment group compared to the two placebo groups, and participants prior experience with medications that was rated as more satisfying in the no-treatment group compared to the two placebo groups. Although statistical omnibus tests were significant, relationship status and prior experience with medications did not differ between the two placebo groups.

Table 4.2

Demographic Information and Characteristics at First Visit

Variable	No-treatment (n = 21)	Conventional placebo (n = 38)	Active Placebo (n = 38)	Omnibus tests of statistically significant between group differences
	M (SD)	*M (SD)*	*M (SD)*	
Age (years)	22.38 (5.31)	20.53 (5.41)	22.37 (5.77)	F (2, 94) = 1.29, p = .281
Gender				X^2 (2, 97) = 0.73, p = .696
Women	15	23	25	
Men	6	15	13	
Education level[a]				X^2 (12, 97) = 15.49, p = .216
Undergraduates	16	35	35	
Postgraduates	3	0	1	
Vocational training	2	3	2	
Relationship status				**X^2 (4, 97) = 10.48, p = .033**
Single	**12**	**33**	**29**	
In relationship	**9**	**5**	**7**	
Live separated	**0**	**0**	**2**	
Height (cm)	170.29 (12.07)	168.39 (9.65)	167.05 (9.22)	F (2, 94) = 0.70, p = .497
Weight (kg)	63.29 (11.24)	63.58 (13.93)	65.21 (12.90)	F (2, 94) = 0.21, p = .811

ISI	13.95 (4.26)	12.11 (3.53)	13.79 (3.84)	F (2, 94) = 2.42, p = .095
DASS-21	15.95 (8.63)	14.61 (11.45)	16.50 (11.30)	F (2, 94) = 0.30, p = .741
WHOQOL-BREF	3.58 (0.42)	3.47 (0.45)	3.38 (0.46)	F (2, 94) = 1.41, p = .250
VAS Prior experience with medications	**77.91 (16.18)**	**65.76 (15.39)**	**64.74 (18.96)**	***F* (2, 94) = 4.57, *p* = .013**
VAS Prior experience with side effects of medications	70.62 (23.44)	66.05 (24.25)	70.67 (23.96)	F (2, 94) = 0.44, p = .647

Note. The information provided in this table represents the analysis sample. [a] Education level was assessed using nine different levels ranging from "less than year 12 or equivalent" all the way to "doctorate". Levels were condensed less than year 12, vocational qualification, undergraduate and postgraduate degrees. Statistically significant differences between the three groups as indicated by omnibus ANOVA, and Chi-square tests are bolded.

Primary Outcomes

Perceived Treatment Allocation

Participants perceived treatment allocation across the three assessment time points is illustrated in Figure 4.4. The PTA assessment directly after the randomisation showed that the three groups differed statistically significantly in their guesses (X^2 (4, 97) = 92.81, p < .001), and as expected there was no statistically significant difference between the two placebo groups (X^2 (2, 76) = 2.08, p = .353). All participants in the no-treatment group correctly indicated their allocation, whereas placebo participants in both groups where nearly evenly split between indicating that they had been allocated to the active medication or placebo group. In the conventional placebo group 47% indicated they had been allocated to the medication group and in the active placebo group 60% indicated they had been assigned to the medication group.

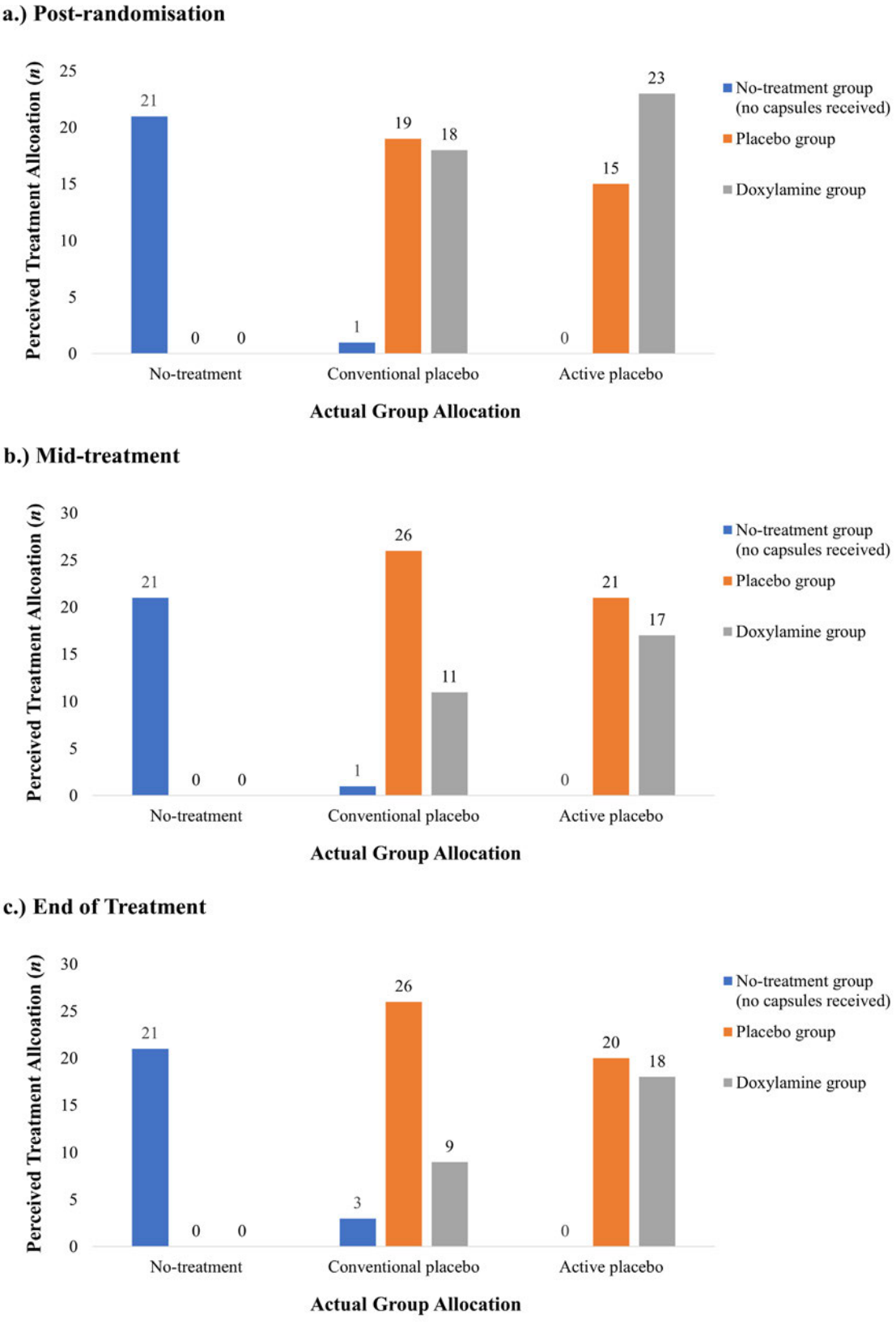
The PTA assessment during the intervention week showed again that there was a statistically significant overall difference between the three groups, with all participants in the no-treatment group correctly guessing their allocation (X^2 (4, 97) = 93.75, p < .001). As observed, the difference between the two placebo groups was not statistically significant (X^2 (2, 76) = 2.82, p = .244), although a trend started to be observable with only a minority of participant in the conventional placebo group (29%) indicating that they had been allocated to the medication group, while guesses in the active placebo group were still nearly evenly distributed with 45% indicating they had received the active medication.

The last assessment of participants' perceived treatment allocation took place after the intervention week, at the start of the last appointment. This time not only did the overall results indicate statistically significant differences between the three groups (X^2 (4, 97) = 86.83, p <

.001), but the difference between the two placebo groups was also statistically significant (X^2 (2, 76) = 6.78, p = .034). While all no-treatment participants again correctly indicated their allocation, a small minority in the conventional placebo group (24%) indicated that they had been allocated to the medication group. Participants in the active placebo group still showed a nearly even distribution between placebo and medication choices, with 47% believing they had received a medication.

Figure 4.4

Participants Perceived Treatment Allocation across the Three Assessment

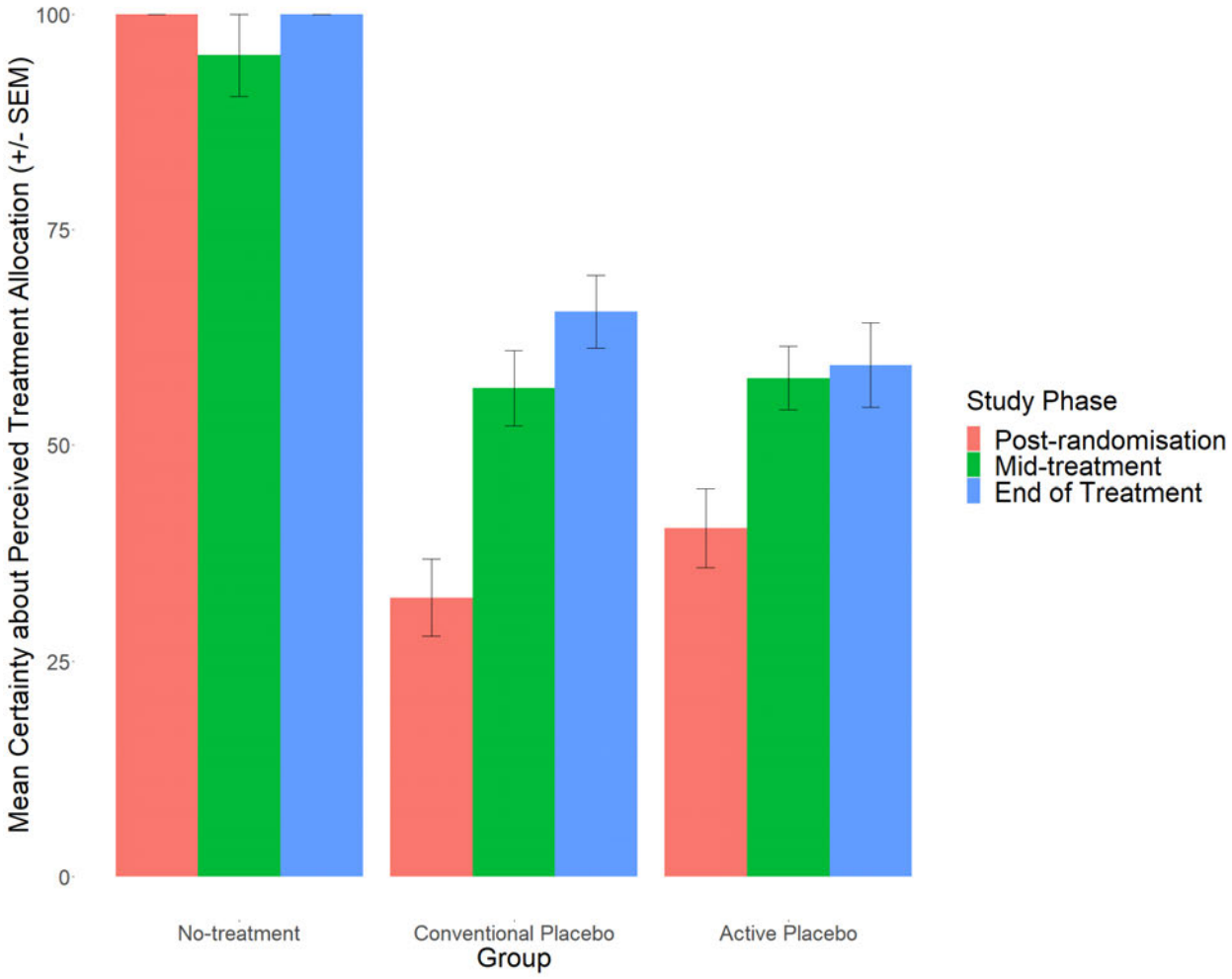


Certainty of Participants' Choice regarding PTA

After participants indicated their perceived treatment allocation, they were directly asked to indicate how certain they were about their choice. The results for participants' certainty about their PTA are plotted in Figure 4.5. Adjusting for the cPTA ratings after the randomisation, the ANCOVA analysis did not show a statistically significant main effect for the factor group on certainty ratings after the treatment week, F (2,93) = 2.69, p = .074. The baseline-adjusted certainty ratings for participants' perceived treatment allocation fitted for the ANCOVA model were M = 78.299 (SE = 6.853) for the no-treatment group, M = 73.250 (SE = 4.058) for the conventional placebo group, and M = 63.533 (SE = 3.804) for the active placebo group. The first planned orthogonal contrast comparing the no-treatment group to the two placebo groups showed that the two placebo groups did not differ statistically significantly in terms of change in cPTA from post-randomisation to the end of treatment compared to the no-treatment group (F (1,95) = 1.20, p = .232). The second planned orthogonal contrast comparing the two placebo groups did not indicate a statistically significant change in participants' certainty about treatment allocation for the active placebo group compared to the conventional placebo group (F (1,74) = 1.84, p = .069).

Figure 4.5

Mean Certainty about PTA Choice across Groups

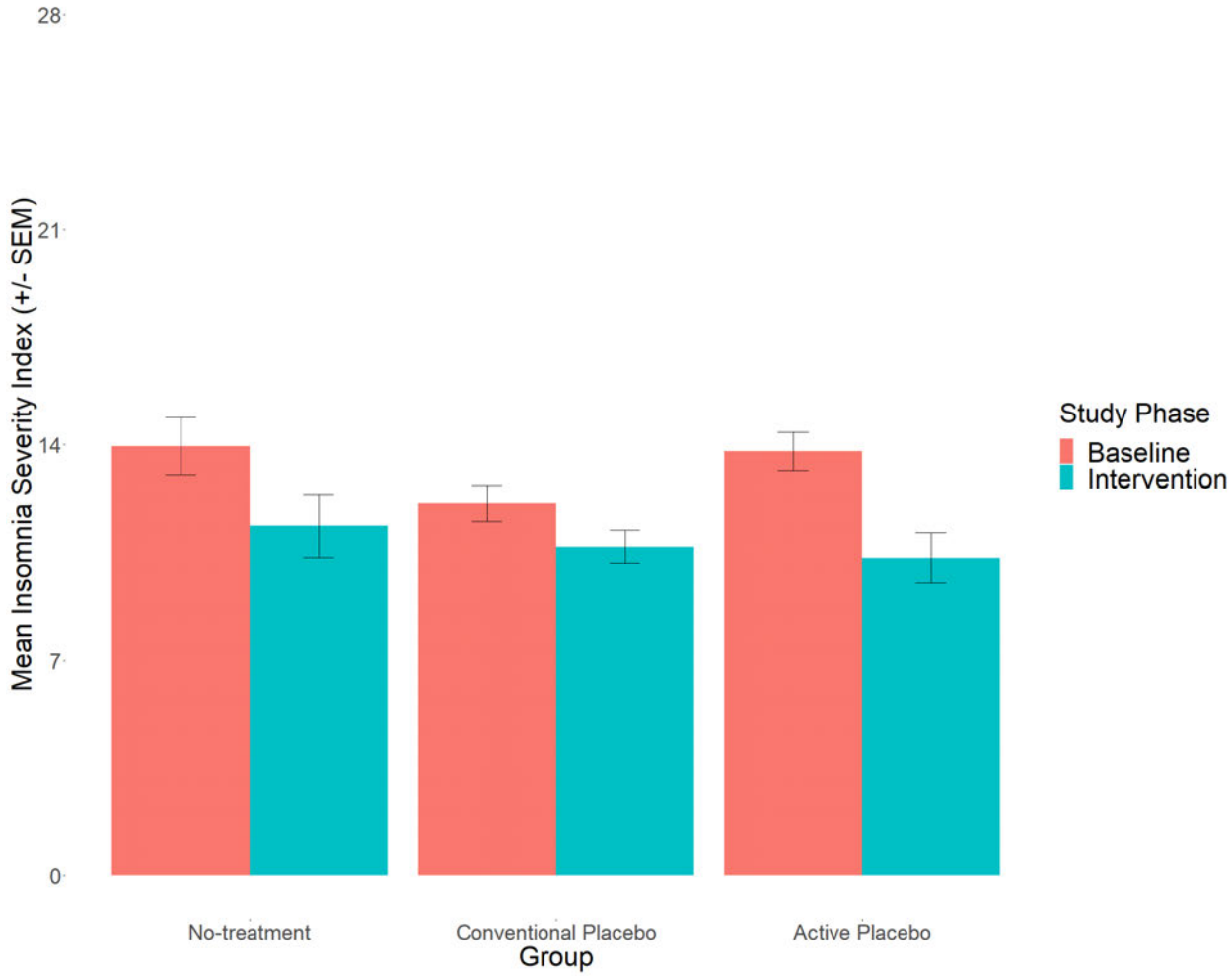


Insomnia Severity Index

Figure 4.6 shows how the three groups' ISI score for the baseline and intervention week. Adjusting for the ISI during the baseline week, the ANCOVA analysis did not show a statistically significant main effect for the factor group on the ISI during the treatment week, F (2,93) = 2.08, p = .131. The baseline-adjusted ISI scores fitted for the ANCOVA model were M = 10.809 (SE = 0.736) for the no-treatment group, M = 11.480 (SE = 0.553) for the conventional placebo group, and M = 9.889 (SE = 0.547) for the active placebo group. The first planned orthogonal contrast comparing the no-treatment group to the two placebo groups showed that the two placebo groups did not differ statistically significantly in terms of improvement on the ISI from baseline to intervention compared to the no-treatment group (F (1,95) = 0.15, p = .881. The second planned orthogonal contrast comparing the two placebo groups indicated a statistically significantly larger improvement on the ISI for the active placebo group compared to the conventional placebo group (F (1,74) = 2.03, p = .045).

Figure 4.6

Mean Baseline and Treatment Insomnia Severity Index across Groups



Secondary Outcomes

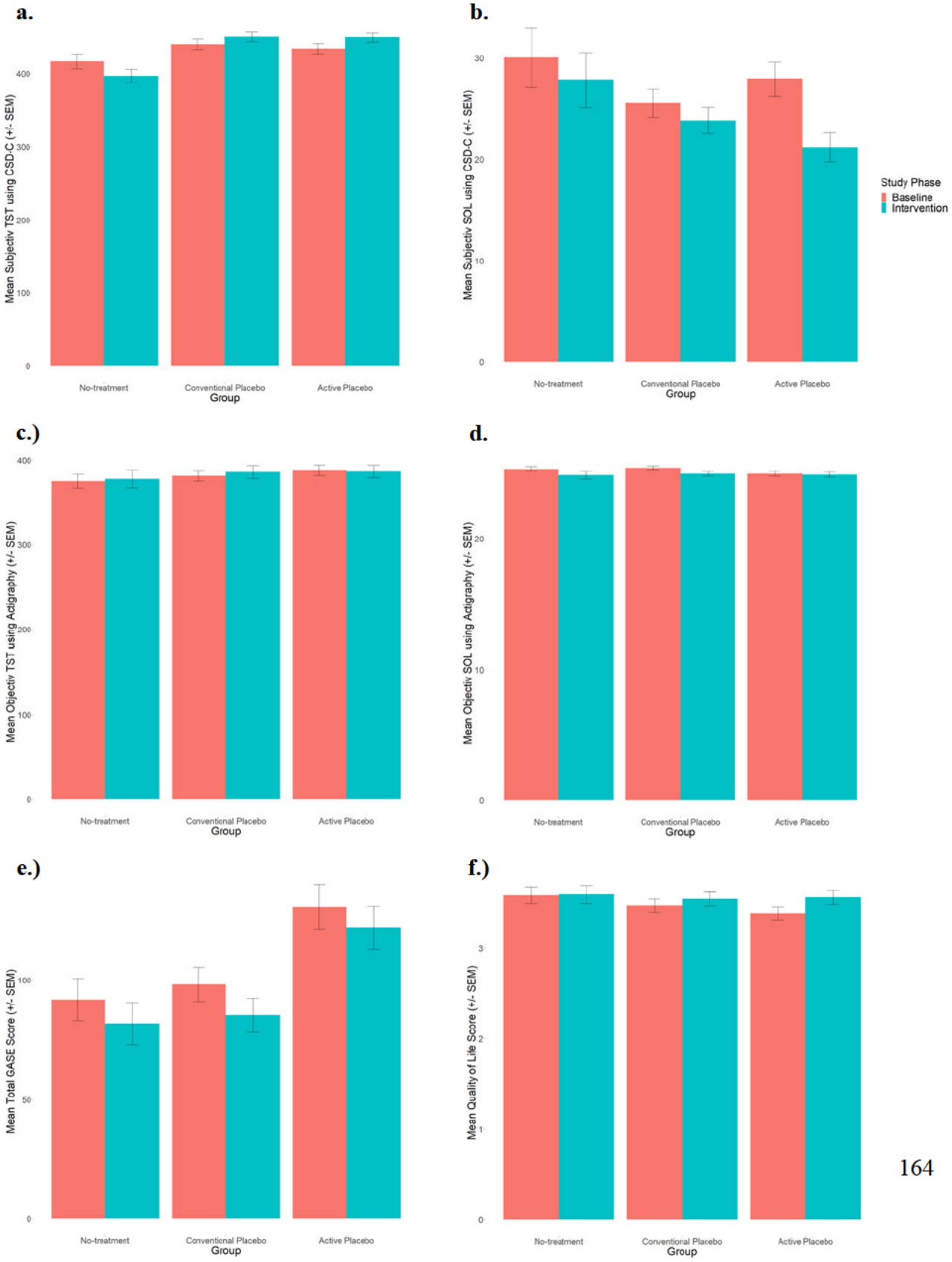
Subjective Total Sleep Time

Figure 4.7a shows the three groups' subjective total sleep time in minutes as assessed using daily sleep diaries (CSD – C) for the baseline and intervention week. Adjusting for the sTST during the baseline week, the ANCOVA analysis showed a statistically significant main

effect for the factor group on sTST during the treatment week, F (2,93) = 6.08, p = .003. The baseline-adjusted sTST fitted for the ANCOVA model were M = 407.270 (SE = 10.145) for the no-treatment group, M = 445.730 (SE = 7.506) for the conventional placebo group, and M = 448.631 (SE = 7.484) for the active placebo group. The first planned orthogonal contrast comparing the no-treatment group to the two placebo groups showed that the two placebo groups statistically significantly differed in subjective total sleep time from baseline to intervention compared to the no-treatment group (F (1,95) = 3.48, p < .001), because the no-treatment groups sleep duration declined and the placebo groups improved. The second planned orthogonal contrast comparing the two placebo groups did not indicate a statistically significant difference for sTST between the two placebo groups (F (1,74) = 0.27, p = .785).

Figure 4.7

Mean Baseline and Treatment Outcomes across Groups for Secondary Outcomes



Onset Latency

Figure 4.7b shows the three groups' subjective sleep onset latency in minutes as assessed using daily sleep diaries (CSD – C) for the baseline and intervention week. Adjusting for the sSOL during the baseline week, the ANCOVA analysis did not show a statistically significant main effect for the factor group on sSOL during the treatment week, F (2,93) = 1.70, p = .189. The baseline-adjusted sSOL fitted for the ANCOVA model were M = 26.240 (SE = 2.625) for the no-treatment group, M = 24.973 (SE = 1.951) for the conventional placebo group, and M = 20.929 (SE = 1.947) for the active placebo group. The first planned orthogonal contrast comparing the no-treatment group to the two placebo groups showed that the two placebo groups did not differ statistically significantly in subjective sleep onset latency from baseline to intervention compared to the no-treatment group (F (1,95) = 1.11, p = .270). Similarly, the second planned orthogonal contrast comparing the two placebo groups did not indicate a statistically significant difference for sSOL between the two placebo groups (F (1,74) = 1.47, p = .146).

Objective Total Sleep Time

Figure 4.7c shows the three groups' objective total sleep time in hours as assessed using actigraphy for the baseline and intervention week. Adjusting for the oTST during the baseline week, the ANCOVA analysis showed no statistically significant main effect for the factor group on oTST during the treatment week, F (2,90) = 0.16, p = .852. The baseline-adjusted oTST fitted for the ANCOVA model were M = 391.95 (SE = 10.31) for the no-treatment group, M = 389.16 (SE = 7.67) for the conventional placebo group, and M = 385.07 (SE = 7.47) for the active

placebo group. The first planned orthogonal contrast comparing the no-treatment group to the two placebo groups showed that the two placebo groups did not differ statistically significantly in objective total sleep time from baseline to intervention compared to the no-treatment group (F (1,92) = 0.42, p = .679). In agreement with the first contrast, the second planned orthogonal contrast comparing the two placebo groups did not indicate a statistically significant difference for oTST between the two placebo groups (F (1,72) = 0.38, p = .703).

Objective Sleep Onset Latency

Figure 4.7d shows the three groups' objective sleep onset latency in minutes as assessed using actigraphy for the baseline and intervention week. Adjusting for the oSOL during the baseline week, the ANCOVA analysis showed no statistically significant main effect for the factor group on oSOL during the treatment week, F (2,90) = 0.56, p = .575. The baseline-adjusted oSOL fitted for the ANCOVA model were M = 24.747 (SE = 0.283) for the no-treatment group, M = 24.915 (SE = 0.211) for the conventional placebo group, and M = 25.106 (SE = 0.206) for the active placebo group. The first planned orthogonal contrast comparing the no-treatment group to the two placebo groups showed that the two placebo groups did not differ statistically significantly in objective sleep onset latency from baseline to intervention compared to the no-treatment group (F (1,92) = 0.83, p = .411). Following the first contrast, the second planned orthogonal contrast comparing the two placebo groups did not indicate a statistically significant difference for oSOL between the two placebo groups (F (1,72) = 0.65, p = .518).

Side Effects

Figure 4.7e shows the three groups' total symptom load as assessed by the GASE for the baseline and intervention week. Adjusting for the GASE score during the baseline week, the ANCOVA analysis showed no statistically significant main effect for the factor group on GASE during the treatment week, F (2,93) = 0.19, p = .826. The baseline-adjusted GASE score fitted for the ANCOVA model were M = 675.001 (SE = 110.021) for the no-treatment group, M = 662.874 (SE = 81.733) for the conventional placebo group, and M = 732.073 (SE = 82.144) for the active placebo group. The first planned orthogonal contrast comparing the no-treatment group to the two placebo groups showed that the two placebo groups did not differ statistically significantly in total GASE from baseline to intervention compared to the no-treatment group (F (1,95) = 0.18, p = .857). Following the first contrast, the second planned orthogonal contrast comparing the two placebo groups did not indicate a statistically significant difference on the GASE between the two placebo groups (F (1,74) = 0.60, p = .553).

Quality of Life

Figure 4.7f shows the three groups' quality of life score as assessed by the WHOQOL-BREF for the baseline and intervention week. Adjusting for the QOL score during the baseline week, the ANCOVA analysis showed no statistically significant main effect for the factor group on QOL during the treatment week, F (2,93) = 2.12, p = .126. The baseline-adjusted QOL score fitted for the ANCOVA model were M = 3.489 (SE = 0.059) for the no-treatment group, M = 3.534 (SE = 0.044) for the conventional placebo group, and M = 3.630 (SE = 0.044) for the active placebo group. The first planned orthogonal contrast comparing the no-treatment group to

the two placebo groups showed that the two placebo groups did not differ statistically significantly in quality-of-life scores from baseline to intervention compared to the no-treatment group (F (1,95) = 1.38, p = .170). Following the first contrast, the second planned orthogonal contrast comparing the two placebo groups did not indicate a statistically significant difference on QOL between the two placebo groups (F (1,74) = 1.54, p = .127).

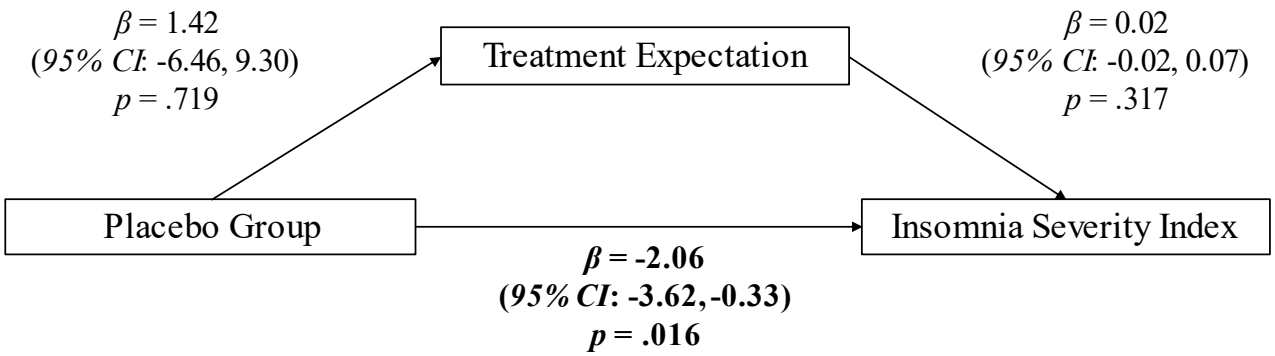
Mediation Analysis

Because the contrast testing the placebo responses on the Insomnia Severity Index was statistically significant between the two placebo groups a causal mediation analysis was performed. Figure 4.8 shows how expectations regarding treatment effectiveness mediated the impact of group on the improvements on the ISI.

The effect of group (active vs. conventional placebo) on the improvement on the ISI was not statistically significantly mediated via treatment expectation. The direct effect of group on ISI improvement was statistically significant. The indirect effect was not statistically significant. The significance of the complete indirect effect was tested using bootstrapping procedures. Unstandardized indirect effects were computed for each of 1’000 bootstrapped samples, and the 95% confidence interval was computed by determining the indirect effects at the 2.5th and 97.5th percentiles. The bootstrapped unstandardized indirect effect was β = 0.035, and the 95% confidence interval ranged from β = – 0.228 to β = 0.370. Thus, the indirect effect was statistically not significant (p = .780).

Figure 4.8

Causal Mediation Analysis



Note. This figure shows the results from the causal mediation analysis. Statistically significant regression coefficients are bolded.

Exploratory Analyses

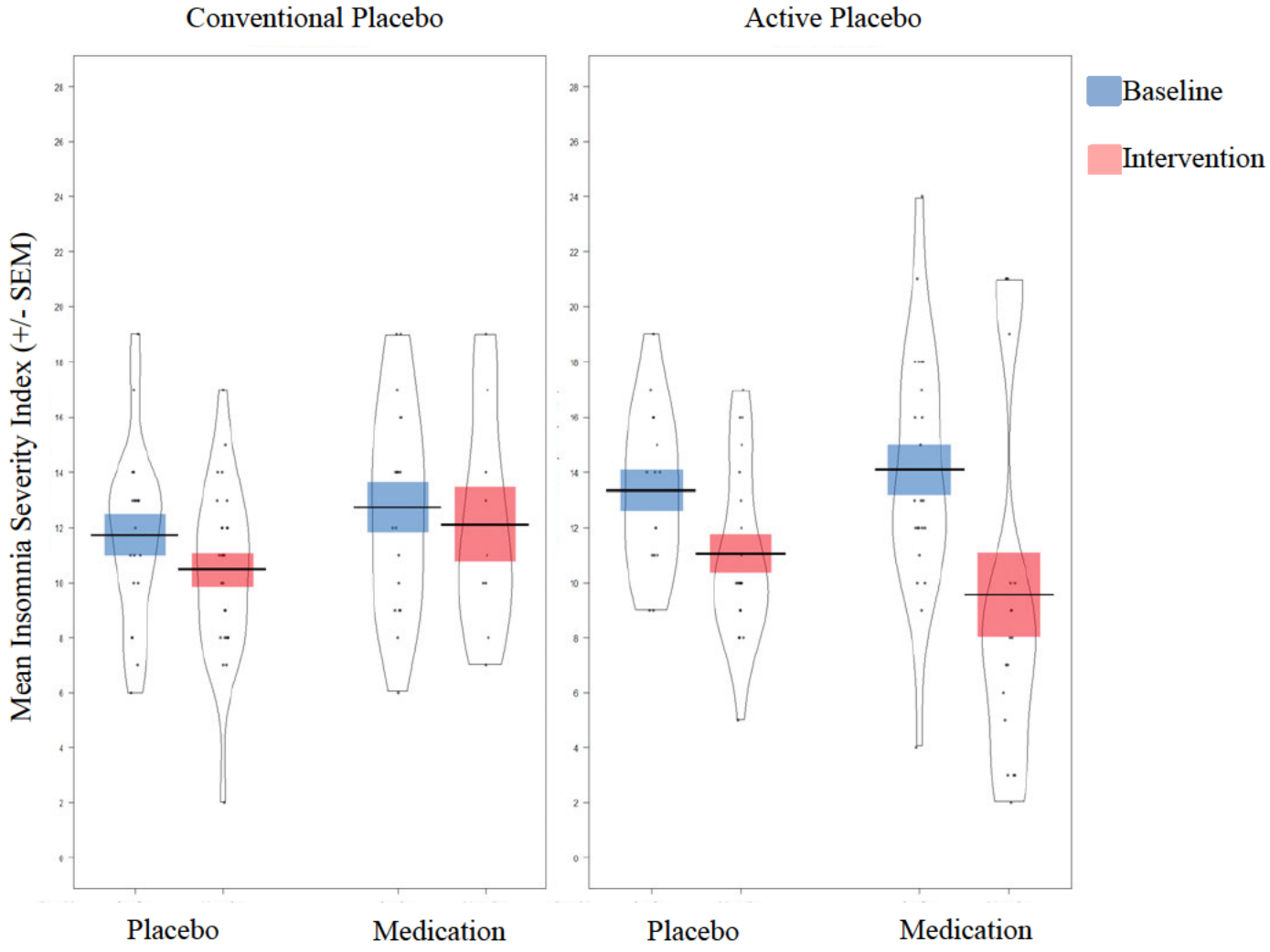
Exploratory analyses were calculated to investigate potential reasons explaining the absence of an overall statistically significant placebo effect on the ISI. Therefore, an analysis was conducted predicting the dependent variable ISI by the factor group with the two placebo groups and participants' perceived treatment allocation as a covariate. The result of this analysis is plotted in Figure 4.9. There was no statistically significant main effect for perceived treatment allocation on the ISI (F (1,71) = 0.75, p = .599) and the interaction between perceived treatment allocation and the actual group allocation was also not statistically significant (F (1,71) = 2.47, p = .121).

Because the exploratory analysis with participants perceived treatment allocation did not result in any meaningful insight further exploratory analysis were conducted splitting the placebo

groups if they had actually reported an unusual urine colouration. An additional exploratory analysis was conducted investigating why some participants in the no-treatment group reported unusually large improvements. Because both analyses were statistically non-significant, they are reported in detail in Appendix C.

Figure 4.9

Mean Baseline and Treatment Insomnia Severity by Perceived Treatment Allocation



Note. The figure above shows participants ISI on the y-axis. The conventional placebo group is depicted in the left panel and the active placebo in the right panel. The results are split by participants perceived treatment allocation, indicated on the x-axis as "Placebo" or "Medication". The horizontal lines are the means and *SEMs* are plotted as the boxes around the mean. The points in the density distributions are participants' individual ISI scores at baseline (blue) and intervention (red).

Discussion

The primary aim of this study was to assess if the experience of side effects caused by the new active placebo model caused participants to be more likely to believe that they had actually received a medication compared to a conventional placebo group. The secondary aim was to investigate whether the active placebo group experienced a larger placebo effect as the conventional placebo group.

The observed findings regarding the primary aims of the study were mostly confirmed by the primary outcome measures. While most participants in the conventional placebo group correctly indicated that they had been allocated to the placebo group, participants in the active placebo group were nearly evenly split in their beliefs that they had either received an actual medication or a placebo. This result clearly supports the findings from retrospective meta-analyses that argued that participants blinding often failed in RCTs (Colagiuri et al., 2019; Fergusson et al., 2004; Hróbjartsson et al., 2007). This study adds support to the hypothesis that the experience or lack thereof contributes to this methodological problem in RCTs. Interestingly, the current study, against the findings from Study 1, did not find an overall nocebo effect of placebo participants reporting an increased daily symptom load during the intervention week compared to the no-treatment group. This study further found experimental evidence that a placebo eliciting a target side effect caused a larger placebo response as a conventional lactose placebo. This adds further experimental support to the earlier retrospective observations from meta-analyses that hypothesised that the lack of side effects in the placebo group leading to failed participant blinding might increase the drug-placebo difference, hence overestimating the

drugs superiority over the placebo that is actually caused by the direct beneficial effect of a pharmacological agent.

Even though the active placebo, successfully upheld blinding, while participant blinding failed in the inert placebo group, and the active placebo group showed a larger placebo response compared to the inert placebo on the ISI, this study has some limitations. First of all, it has to be stated that there was no statistically significant overall placebo effect of the two placebo groups over the no-treatment group on the ISI. This might have to do with the unusually large improvement of some of the no-treatment participants, even showing improvements above the threshold of six points on the ISI defined as clinically meaningful changes of insomnia treatments (Yang et al., 2009). The no-treatment group in the current study improved by around four points on the ISI, which is more than twice as much of an improvement as observed by Yeung et al. (2020) who similarly conducted an experimental placebo study with people suffering from sleep problems. On the other hand, another experimental study about placebo effects in sleep similarly did not observe an overall placebo effect on the ISI, but did on the PSQI (Neukirch & Colagiuri, 2015). For ethical reasons during the planning phase of the current study, it was decided to only randomise half as many participants to the no-treatment group compared to the two placebo groups. Therefore, the desired sample size for the no-treatment group was only 23 participants. Statistically, small sample sizes like this have a higher probability that an even smaller number of subjects can disproportionally distort the group mean on a given outcome measure. It therefore might have only occurred by mere chance that the no-treatment group experienced the observed unusual improvement. Another potential reason for the unexpectedly large improvement in the no-treatment group might be related to the phenomena of small sample sizes and have to do with timing effects. Because the current study used a stratified

randomisation procedure taking participants' gender and insomnia severity into consideration when allocating participants to one of the three groups and used a prospective design, it is possible that timing effects might have biased one group such as the small no-treatment group more than the two placebo groups. A potential example could be that some no-treatment participants were allocated in such a way that they had their baseline week during a conventional work week and the second week of study participation during a public holiday. A scenario like this could possibly explain the unexpectedly large improvement in sleep in the no-treatment group.

A second limitation is that findings on the primary outcomes were not fully supported by the secondary sleep and quality-of-life outcomes. Although most of the secondary subjective outcomes showed similar trends as the primary analysis that was based on the ISI, subjective total sleep time for example showed an overall statistically significant placebo effect, but contrary to the ISI the contrast between the two placebo groups was not statistically significant. Further, the secondary objective sleep outcomes did not show any overall placebo effect, nor any differences between the active and conventional placebo group. It is interesting that the results seem to show a dichotomy between subjective and objective sleep outcomes. A potential explanation for this observation might have to do with the fact that the ISI was specifically created as a sensitive measure to detect treatment effects in prospective RCTs in sleep research (Bastien et al., 2001), while the other outcome measures used in this study were intended to generally measure symptoms. An explanation for the lack of any meaningful effects on the objective measures might have to do with the criticism that is generally aimed towards the use of actigraphy-based measures in sleep research. It is often stated that actigraphy-based measures have been insufficiently validated and therefore might not be reliable enough to detect any

significant changes in sleep outcomes (Sadeh, 2011; Sadeh & Acebo, 2002; Tryon, 2004). A systematic literature review and meta-analysis conducted by the American Academy of Sleep Medicine concluded that the quality of evidence of actigraphy for the evaluation of insomnia in adults is only moderate, mostly due to imprecision (Smith et al., 2018).

A further limitation of the study was that more than half (56%) of the participant sample consisted of students enrolled at the University of Sydney participating for course credit. This on its own would not per se classify as a limitation, but one student participant had to be excluded from the analysis due to dishonesty. During data inspection, after experimenters had been unblinded it became apparent that the participant had clearly fabricated answers on questionnaires, because the answer the most to the left had consistently been chosen across all questionnaires and study visits. During the exploratory analyses investigating the lack of an overall placebo effect that was mentioned above, it became apparent that some student participants that had been allocated to the no-treatment group indicated unusually large improvements on the ISI. During the exploratory analysis splitting the sample by students participating for course credits and paid participants, the pattern of results, although not statistically significant showed that not only had the no-treatment group of student participants the largest improvement, but student participants showed also lower improvements when taking the placebo capsules. Student participants taking the inert lactose placebo even showed no improvement at all. Using convenience or student samples for research has received criticism for multiple reasons in the past and there are valid reasons why many journals are hesitant to publish research that is based on student samples (Andrade, 2020; Peterson & Merunka, 2014). There are multiple possible explanations for the difference in result patterns between student participants and paid participants that were mostly recruited from the general population. First, as with the

single obviously dishonest participant and the four other unusual improvements, it might have been the case that some student participants were not participating out of interest in sleep research or problems with sleep, but to get course credits towards the end of the semester in order to increase their grade. An alternative explanation for student participants' sudden improvement in sleep is that they did the eligibility screening and baseline week during the time of exams and then the intervention week took place after their most stressful period of the semester had ended. While this explanation cannot be verified on the base of data collected in the current study, the participation period for student participant typically already ends before the exam weeks at the end of the semester.

Even though this study had some limitations, it successfully addressed the problems encountered in Study 1 and demonstrated as the first ever experimental study that an active placebo eliciting a target side effect can mitigate the problem of failed participant blinding. Failed participant blinding in RCTs so far had only been described using retrospective analyses in meta-analyses, and in experimental studies that relied on fake feedback or a treatment onset sensation. This study confirmed the hypothesis that the new active placebo model can be applied successfully to uphold participant blinding in RCTs. Although the observed results indicate that the active placebo caused larger improvements on the ISI as the conventional placebo, due to the lack of an overall placebo effect, it is not possible to conclusively argue that experiencing side effects can enhance the placebo effect. If at all, the findings of the current study imply that a lack of side effects in the conventional placebo group led participants to believe that they did not receive an active treatment. This might have introduced a nocebo effect in conventional placebo participants, which might be the reason why the conventional placebo group barely reported any improvement in their sleep.

Study 3 of this thesis, reported in the next chapter builds further on the current study's methodology and further looks into the effect of participants' experience of side effects, and how participants' beliefs about experiencing side effects influences the placebo response.

Chapter 5: Study 3 - Active Placebo Message Framing Study

This study builds on Study 2 that found that participants in the active placebo group were more likely to believe that they had actually received an active medication and reported larger improvements on the ISI than the conventional placebo group. The focus of the current study was to give participants receiving an active placebo different information using message framing about how experiencing the side effect of the "medications" would affect the efficacy of the treatment.

Introduction

The focus of this study was set because Study 2 could not conclusively answer if experiencing side effects actually enhanced the placebo effect, or if a lack thereof had reduced the placebo effect in the conventional placebo group. Therefore, the aim of this study was to examine whether positively, negatively, or no framed information about the active placebo's target side effect influences the observed placebo effect and how bothersome the target side effect is perceived.

Side effects are generally considered negative or unwanted, however, a WHO definition describes a side effect as any symptom that is related to the pharmacological properties of a drug, but not the active and desired drug effect. Notably, this definition does not indicate that the side effect has to be negative (Stephens, 1998). Edwards and Aronson (2000) explained that the WHO definition was formulated to include symptoms that may be beneficial to a patient as side effects, separate from the main aim of the therapy. Edwards and Aronson (2000) used the example of anticholinergic effects of a tricyclic antidepressants, that can be classified as side

effects, because they are not associated with the primary intended therapeutic effect. On the other hand, they use the example of a depressed patient who is also suffering from comorbid irritable bowel syndrome, where the anticholinergic side effect might also cause an improvement in irritable bowel symptomatology.

This means that side effects can have different valences depending on the context. They can be beneficial as in the depressed patient example, where they have other useful effects. They might reassure a suffering person's belief participating in a double-blind RCT that they have been allocated to the treatment group, hence enhancing their placebo response as Kirsch's active placebo hypothesis had suggested (Kirsch, 2014) and Study 2 demonstrated when comparing the active placebos against the conventional one for the improvement on the ISI.

As discussed in the literature review of this thesis, expectations play an important role when it comes to participants' responses to a treatment. The information given to participants play a crucial role when it comes to forming expectations. One of the earliest and best-known theories that dealt with the effects of presenting information in different manners is the prospect theory by Kahneman and Tversky (1979). Kahneman and Tversky suggested that if a person is presented with two choices, one of which is of little risk (e.g., lose $10 with 25% chance), and one is posing a higher degree of risk (e.g., lose $5 with 50% chance), the person's evaluation will depend on the manner how the choices are framed even though the mathematical expectation of the uncertain option is equal (i.e., lose $2.50). Generally, people are more willing to accept higher risks to avoid losses than to achieve gains. Gallagher et al. (2011) showed that this 'loss aversion' also applies in healthcare contexts, in that people respond more favourably in regard to clinicians' advice to loss-framed advice when risks, costs, or losses associated with *not* taking the advice (e.g., dying of cancer if not undergoing mammography screening 10%) are perceived

to be greater than the risks, costs, or losses associated with taking the advice (e.g., surviving cancer if undergoing mammography screening 90%). Tversky and Kahneman (1981) defined that gain or loss framing refers to phrasing a statement that describes a choice or outcome in terms of its positive (gain, e.g., improvement due to treatment with 60% chance) or negative (loss, e.g., *no* improvement due to treatment with 40% chance) features.

Barnes et al. (2019) reviewed the literature to find out if positive framing could be used to avoid nocebo side effects. They identified only six studies, of which four investigated attribute framing and only two used message framing strategies. According to O'Connor et al. (1996) attribute framing means that the same statistical information is presented either positively (e.g., 60% will *not* experience headaches) or negatively (e.g., 40% will experience headaches). While attribute framing is a simple strategy that has received some attention in the past, the more complex message framing has not yet received as much attention. Wicks (2005) refers to message framing as both the process of selecting and the manner in which information is presented. News messages on television for example might be framed episodically in the form of a case study, or thematically with general or abstract concepts presented.

So far, there have only been two studies conducted on message framing to either reduce the burden of side effects (Wilhelm et al., 2018) or to try to enhance the treatment response (Fernandez et al., 2019). Wilhelm et al. (2018) investigated if positive message framing could reduce the side effect burden of an antihypertensive medication (metoprolol) in a healthy participant sample that had to perform an exercise test on a bicycle ergometer before and after drug intake to provoke the drug-attributed side effects (dizziness). When the 100mg metoprolol were dispensed participants either received a standard “neutral” information that the medication might cause potentially unpleasant, but already known side effects. The participants in the

positive message framing group received the following additional statement: "… Often a feeling of dizziness also occurs. This is a sign that the drug is starting to work. If you become dizzy after taking the medication, it means that your body is responding to the beta-blocker particularly well. …" Wilhelm et al. (2018) observed that medication-specific drug-attributed side effects, although only with a small effect size, were rated statistically significantly less threatening in the positive framing group compared to the neutral standard information group. While the threat levels were differently rated, the actual rates of experiencing side effects were not statistically different.

Fernandez et al. (2019) studied the effects of positive side effect messaging, explaining to participants that the experience of side effects could be interpreted as a sign that the drug is working. The positive message framing information was compared to a control group that did not receive a suggestive information. In their experimentally-induced pain study, healthy male participants received the nonsteroidal anti-inflammatory drug diclofenac, together with atropine to induce the target side effect, i.e., dry mouth. The positive message framing group and the neutral information group did not statistically significantly differ in terms of analgesia, but they found that the total side effect load was positively correlated to analgesia in the positive framing group only. This means that the average analgesia did not differ between the two groups. But when looking at the groups individually participants in the positive message framing group who reported more side effects showed a larger treatment effect compared to participants who only reported some or no side effects and showed only a small or no treatment effect. Participants in the neutral message framing group contrary to the positive message framing group did not show this association between the rate of side effects and analgesia. Fernandez et al. (2019) concluded that a positive message framing of side effects was credible and that these findings could be used

in the clinical context to increase the treatment effect by telling patients that the experience of side effects was tied to the medication working.

Despite the interesting preliminary results of Wilhelm et al. (2018) and Fernandez et al. (2019), several limitations apply to their studies. First, both studies only focused on testing a positive against a neutral message framing. For one reason, it may be problematic that their designs did not incorporate a valid natural history control group in the form of a no-treatment group not receiving any information at all. Their neutral message framing groups still received information regarding the presence of side effects. As was discussed in the general introduction and observed in Study 1, mere verbal suggestion or written information is sufficient to introduce nocebo effects. As Colloca and Barsky (2020) have argued, it is necessary for reliable research to include no-treatment groups. Following Colloca and Barsky (2020) it may be questionable if the two control groups receiving neutral framing information can actually be called neutral when in fact they might have introduced nocebo effects. Second, a limitation that was already acknowledged by Fernandez et al. (2019) was the fact that using an actual medication to investigate framing effects might have introduced unwanted medication interactions. Both studies relied on using active medications, a blood pressure reducing beta blocker called metoprolol in the study by Wilhelm et al. (2018) and the anticholinergic medication atropine in the study by Fernandez et al. (2019). This means that a part of the observed message framing effects in their studies might have been influenced by unwanted indirect actions of the active medications.

Third, a general criticism that has been stated for placebo research in general and applies to both studies is that their studies used short experimental laboratory settings using healthy participants, meaning they have a low external validity and might not translate well into actual

clinical practice. Reviewing the placebo literature, Enck et al. (2017) argued that there is often a discrepancy between results observed in short-term experimental studies using health participants and the same studies again conducted using patients or even the effects observed in clinical practice. The discrepancy the authors described pointed towards the direction that short-term experimental studies eliciting experimental pain in healthy participants tend to overestimate the effects and demonstrated that for example the placebo effects in clinical practice might not even be of clinical significance.

In order to address these limitations, the current study was designed to implement the following methodological aspects. First, the current study was designed with four experimental groups to account for all the limitations listed above. For an extensive investigation of message framing effects, the current study included three placebo groups. Two respective placebo groups received a positive and negative message framing that was compared against a placebo group receiving no framing at all. Additionally, the current study included a no-treatment group that neither received a placebo nor any message framing information to represent a valid natural history control group. Second, all three placebo groups received the new active placebo eliciting the target side effect. This offered the advantage that unwanted indirect effects of the pharmacological agent can be ruled out, meaning that findings of the current study about message framing effects guarantee that the observations made are based on psychological processes. Third, the current study was designed to recruit a large majority of participants from the general population suffering from treatment-worthy sleep difficulties to assess the implications of message framings on the experience of side effects. Compared to the two message framing studies by Wilhelm et al. (2018) and Fernandez et al. (2019), the current study represents a setting that is more naturalistic and covers a longer time period. The findings from

the current study therefore minimise the translational gap between experimental research and real-world settings and avoid potential problems with having a convenience sample of students participating for course credits, as potentially biased Study 2. Lastly, the current study implemented changes to the randomisation methodology to omit potential problems that might have reduced the validity of Study 2. Therefore, an even randomisation ratio was chosen to guarantee the no-treatment natural history control group was more adequately sized, limiting the influence of individual participants with unusual results. Additionally, a block randomisation was chosen to guarantee that participants were allocated to the four groups during similar periods to avoid any potential bias.

Based on the earlier findings of Wilhelm et al. (2018) and Fernandez et al. (2019) regarding message framing, it was hypothesised that the active placebo group receiving the positive message framing manipulation would show the largest placebo effect for sleep and the lowest amount of bothersomeness about the target side effect. The negative message framing condition was hypothesised to show the smallest placebo effect for sleep and the highest bothersome ratings, and the no-framing condition was expected to fall between the positive and negative framing groups in terms of improvement on sleep outcomes and the side effect's bothersomeness.

Methods

The University of Sydney's Human Research Ethics Committee had reviewed and approved all ethical aspects regarding the recruitment, materials, and procedures for this study (Project Number: 2018/107). This included all necessary modifications that had to be made after

the outbreak of the SARS-CoV-2 pandemic. Detailed documentation of the ethical approvals and the necessary modifications can be found in Appendix D.

The detailed methodology of this study was registered before enrolment of the first participant on ANZCTR (Identifier ACTRN12620000232932, Australian and New Zealand Clinical Trials Registry, 2020). The current study was originally conceptualised and pre-registered as an in-lab face-to-face study similar to Study 2. Seven participants had already been enrolled and randomised when the study had to be halted because face-to-face contact was no longer allowed due to the SARS-CoV-2 pandemic. Unfortunately, none of the seven participants could finish the intervention week before the study was halted. Due to the pandemic the study methodology had to be adjusted to be delivered via an online mode using video calls and delivery of the capsules via mail. Therefore, the final study methodology deviated from the pre-registration on the ANZCTR. The methodology described in the following sections represent how the study was actually conducted. Because the methodology of the current study built on Study 2, only methodological aspects that are specific to the current study are presented to avoid unnecessary redundancies.

Participants

The participant sample for this study consisted of sleep-impaired adults recruited across all of Australia. The inclusion and exclusion criteria for the current study were equal to Study 2, with the addition of a requirement that participants had to be residing within Australia so the capsules could be delivered with Australia post. The advertisement strategy was similar to the prior experimental study in that volunteers were wanted for the evaluation of a new sleep medication. Therefore, the advertisement was asking for volunteers who regularly suffered from

sleep problems, including difficulty initiating sleep, waking up frequently during the night and having difficulties falling back asleep, or waking up earlier than desired without being able to fall back asleep. The advertisement flyer can be found in Appendix D.

Participants were also recruited via the University of Sydney's Psychology Research Participation System (SONA-PSYCH) that offers first and second year undergraduate students enrolled in psychology courses to participate in research studies to receive course credits as remuneration. Compared to Study 2, participants were no longer recruited receiving cash payment as remuneration because the online delivery mode no longer required participants to travel to the University of Sydney Camperdown campus. Student participants that had already participated in Study 1 and 2 were prohibited prom signing up to the current study to avoid any bias, because they had already been informed about the real purpose of the studies during the delayed informed consent. In addition to the University's SONA data bases and local advertisement using flyers around the University of Sydney Camperdown campus, participants were recruited online using Google Advertisement, Facebook Advertisement, and the sleep research volunteer page of the Australian Sleep Health Foundation. Because volunteers no longer had to attend face-to-face meetings on campus the online advertisement was adjusted to all of Australia.

Eligibility criteria or the current study were identical to Study 2 with the addition that participants had to be residing within Australia to be able to receive the capsules delivered with Australia post.

Student participants received two hours' worth of course credit for completing study participation. If a student participant withdrew from the study, remuneration was proportionate to the time they had invested up until their withdrawal. Volunteers recruited online from the general

population did not receive any reimbursement for study participation. The recruitment period for this study started in March 2020 and was finished in May 2021.

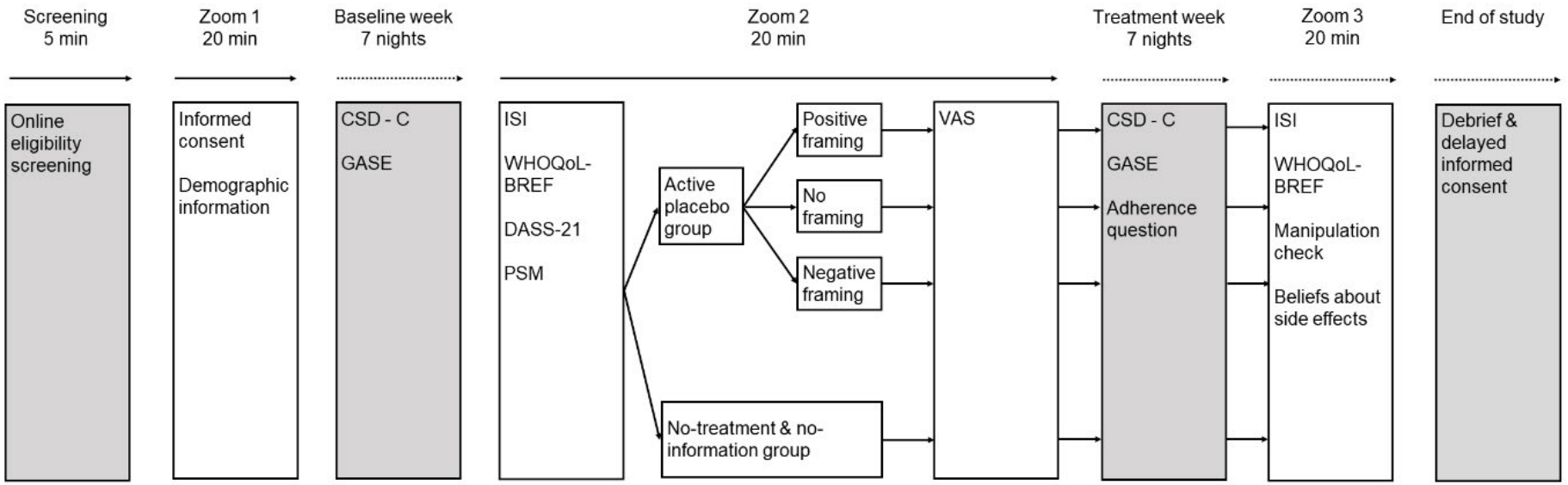
Design

The experiment used a 4 x 2-mixed design with treatment (positive message framing vs. negative message framing vs. no framing vs. no-treatment) and time (baseline vs. intervention) as factors. Figure 5.1 shows the study design with the randomisation to the four study groups and the two time periods. The factor time was divided into a baseline and a treatment period, each covering seven nights. During online appointment two, participants were instructed to open the satchel that was sent to them by another independent lab member so that their random group allocated to one of the groups was revealed. The satchels for participants in the three message framing conditions received a capsule container containing placebo treatment together with their specific framing information. Participants in the no-treatment group were sent an empty capsule container in the satchel along with information that they had been allocated to the control condition. Participants were randomly allocated to the no-treatment, positive framing group, no framing group, negative framing group, or the no-treatment group using a 1:1:1:1 ration, so that all groups would be equal in size. The experiment used a block randomisation to avoid any unwanted timing effects that might have negatively affected Study 2. All participants led to believe they would receive a novel medication in the three framing groups received an active placebo containing beetroot extract.

The independent variable was group allocation to one of the four experimental groups and the independent variables were the Insomnia Severity Index and participants' perceived bothersomeness of the target side effect.

Figure 5.1

Study Design and Procedure with all Measures



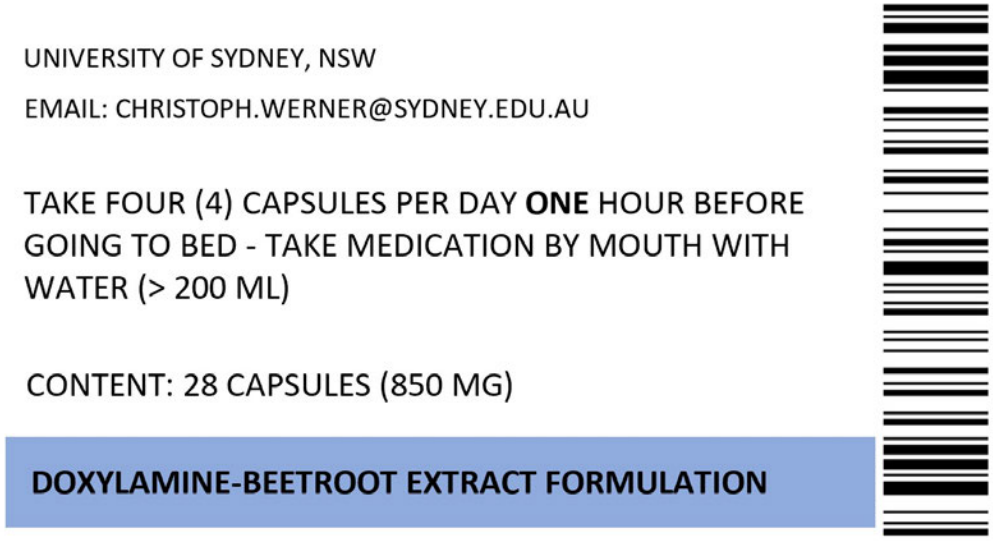
Materials

Placebo capsules.

The label that was used in this study is displayed in Figure 5.2. Capsules were identical to Study 2. The capsule containers in the current study were labelled containing a beetroot-doxylamine mix following the cover story that a randomised trial was conducted for sleep problems testing the active medication against a no-treatment group.

Figure 5.2

Capsule Container Label



Cover story.

The current study used an identical cover story to Study 1 regarding the guise of the new medication. For the full cover story regarding the allegedly new medication, please revisit the cover story for Study 1 (full cover story here). Briefly, participants were deceivingly informed that they were participating in a trial that tested the effect of a new compound consisting of an antihistamine (doxylamine) and an antioxidant (beetroot extract). It is important to note that the current study similar to Study 1 used a cover story that participants would either receive the potent medication or no-treatment, meaning the study was described as an open label medication study.

To increase standardisation of information given to participants during participants' visits, especially to make sure that the message framing in the three active placebo groups were delivered consistently, an experimenter manual was created, and the experimenter trained to deliver the information according to the manual and in a standardised fashion following the script. The full experimenter manual for the three online appointments can be found in Appendix D. All three active placebo groups received the same information about the treatment under investigation and potentially associated side effects, but the three different message framing groups received different information about how experiencing the target side effect (beeturia) would affect the efficacy of the treatment. Table 5.1 shows the three text passages the three framing conditions received both in a written form and as a verbal explanation.

Table 5.1

The Message Framing Delivered to Placebo Participants

Group	Message Framing received both verbally and in written form
Positive Message Framing	[…] I just quickly want to mention the red urine colouration. Because the drug contains beetroot extract, which is a red powder, it is very likely that your urine, especially in the morning after taking the capsules will be unusually darker or have a reddish colour. This is absolutely harmless. **In general, if you experience an unusual colouration of your urine or stool this is a sign that your body *absorbed* and metabolised the medication well. This means that the drug is *more* likely to help you improve your sleep quality, time to fall asleep and maintaining your sleep during the night especially well.** […]
No Message Framing	[…] I just quickly want to mention the red urine colouration. Because the drug contains beetroot extract, which is a red powder, it is very likely that your urine, especially in the morning after taking the capsules will be unusually darker or have a reddish colour. This is absolutely harmless. […]
Negative Message Framing	[…] I just quickly want to mention the red urine colouration. Because the drug contains beetroot extract, which is a red powder, it is very likely that your urine, especially in the morning after taking the capsules will be unusually darker or have a reddish colour. This is absolutely harmless. **In general, if you experience an unusual colouration of your urine or stool this is a sign that your body *did not absorb* and metabolise the medication well. This means that the drug is *less* likely to help you improve your sleep quality, time to fall asleep and maintaining your sleep during the night especially well.** […]

Note. The information that the positive and negative message framing groups received on top of the general information provided to the no-framing group are bolded. The information that further dictated the positive or negative message framing character between the positive and negative message framing conditions is further italicised.

The experimenter was blinded towards participants' allocation to the four experimental group until the allocation was revealed by the participant opening the satchel during the second online appointment, once all baseline measures had been collected.

Measures

The hierarchy of outcome measures and definition of assessment time points was pre-defined and registered within the ANZCTR (Identifier ACTRN12620000232932, Australian and New Zealand Clinical Trials Registry, 2020). Table 5.2 defines the primary and secondary outcome measures and when these measures were assessed over the course of the study. Because the study delivery mode had to be changed from a face-to-face in-lab visit to online appointments conducted using the online video call application Zoom (Zoom Video Communications Inc, 2020) due to the pandemic, it was no longer possible to equip participants with an actigraphy device. Therefore, the objective actigraphy-based sleep measures (sSOL and sTST) that were pre-registered could not be collected and are not included in the table.

Table 5.2

Definitions of Outcome Measures and Assessment Time Points

Outcome Hierarchy	Outcome Measure and Definition of Assessment Time Point for Statistical Analyses
Primary outcome (1)	Insomnia severity assessed using the Insomnia Severity Index (ISI)
Timepoint (1)	Baseline (assessed at a single timepoint on the day starting treatment, for the 7 nights period prior to receiving treatment) Treatment (assessed at a single timepoint on the day after finishing 7 nights of treatment, for the 7 nights period receiving treatment)
Primary outcome (2)	Reports of daily bothersome of red urine colouration assessed as part of an amended 10-item version of the General Assessment of Side Effects (GASE). We shortened the GASE to only include the most relevant items and added a question regarding urine colouration – the target side effect. We assessed how bothersome the complaints are on a visual analogue scale (VAS) ranging from 0 (= "not bothersome at all") to 100 (= "absolutely bothersome"), with 33 (= "mildly bothersome") and 66 (= "moderately bothersome")
Timepoint (2)	Baseline (assessed daily during the 7 nights prior to randomisation) Treatment (assessed daily during the 7 nights receiving treatment after randomisation)
Secondary outcome (1)	Sleep quality (self-report) using the Consensus Sleep Diary Version C (CSD-C)
Timepoint (1)	Baseline (assessed daily during the 7 nights prior to randomisation) Treatment (assessed daily during the 7 nights receiving treatment after randomisation)
Secondary outcome (2)	Total sleep time (self-report) using the Consensus Sleep Diary Version C (CSD-C)
Timepoint (2)	Baseline (assessed daily during the 7 nights prior to randomisation) Treatment (assessed daily during the 7 nights receiving treatment after randomisation)
Secondary outcome (3)	Sleep onset latency (self-report) using the Consensus Sleep Diary Version C (CSD-C)
Timepoint (3)	Baseline (assessed daily during the 7 nights prior to randomisation) Treatment (assessed daily during the 7 nights receiving treatment after randomisation)
Secondary outcome (4)	Adherence to medications (self-report), using a single numerical question about the daily capsule intake

Timepoint (4)	Treatment (assessed daily during the 7 nights receiving treatment after randomisation)
Secondary outcome (5)	Quality of life using the World Health Organisation's quality of life assessment (WHOQOL-BREF)
Timepoint (5)	Baseline (assessed at a single timepoint on the day starting treatment, for the 7 nights period prior to receiving treatment) Treatment (assessed at a single timepoint on the day after finishing 7 nights of treatment, for the 7 nights period receiving treatment)
Secondary outcome (6)	Recall and recognition of the message framing (RMF), using an open recall and a forced-choice recognition question with the options (0=no framing, 1=positive framing, 2=negative framing)
Timepoint (6)	Post-treatment (assessed at a single timepoint at the end of the seven-night period receiving treatment)

Note. The hierarchy and time points of these measures have been registered on ANZCTR. Please note that the measures in some cases were assessed more often over the course of the study as defined here for the purpose of the statistical analyses. Objective TST and SOL assessed using actigraphy were pre-registered on the ANZCTR but not listed in this table due to them not being assessed for SARS-CoV-2 related reasons.

Measures used identically as in the first two studies were omitted from being described again in this section. Full descriptions of the ISI, CSD – C, WHOQOL-BREF, DASS-21, PSM, VAS measuring expectancies, and adherence measures can be found in the subsections about measures within the chapter describing the measures used in Study 1.

General Assessment of Side Effects (GASE; Amended)

The GASE was identical as in Study 2, with the addition of an item asking participants to rate the bothersomeness of a symptom they had experienced. If participants had indicated that they had experienced a symptom they were presented with the additional question asking them "How bothersome is this complaint for you?" and they had to rate the bothersomeness on a 100-point VAS ranging from 0 (="not bothersome at all") to 100 ("absolutely bothersome").

Recall and Recognition of the Message Framing (RMF)

To assess how well participants were able to recall and recognise the message framing information they had received during the second online appointment self-designed items were created. For the recall item, participants were asked to answer the following open-ended question: "Last visit you were informed about the relationship between experiencing side effects and the efficacy of the medication. Please explain in your own words what you were informed about.". For the recognition item, participants were presented with three options they had to choose from. The positive and negative message framings were presented, together with an option that said that participants did not receive any information to cover the placebo group that did not receive any framing.

REDCap Survey System

Following Study 1 and 2, REDCap was chosen again for data collection because it was specifically designed to support data collection in longitudinal RCTs (Harris et al., 2019; Harris et al., 2009). Contrary to the first two experimental chapters REDCap was not used in this study to randomly allocate participants to their condition because this had to be done already before sending out the satchels to participants. To present participants with their allocated message framing information REDCap's branching logic was used. This allowed participants to read their specific information without realising that unbeknownst to them they had not just been selected to receive the "medication" but also were presented with different versions of medication leaflets. The information was presented digitally in REDCap, but were also in the satchels in paper format. See Appendix D for the full information provided to participants during the second appointment or Table 5.1 for the abbreviated message framing each group received.

Randomisation and Blinding

The randomisation to the four experimental groups was contained in the satchels that were mailed to participants with Australia Post. The satchels were prepared by an independent member of the Colagiuri lab and the experimenter conducting the Zoom appointments with the participant was blind to participants' allocation until they had finished all baseline measures during the second visit. To make sure participants were blind to their allocation during the baseline week and baseline assessments during the second appointment they had to show the intact and unopened satchel at the start of the second online appointment. Once participants were finished with the baseline assessments the experimenter told them to open the satchel at which point their allocation was revealed both to the participant and the experimenter. Unbeknownst to

the participants in the alleged "drug" group, participants were further allocated to one of three message framing groups. The participants were instructed to tell the experimenter what "tracking code" was written on top of the capsule containers. These codes were no "tracking codes" but informed the experimenter if the participants had to receive the positive, negative, or no message framing.

The current study used a block randomisation. A block always consisted of one participant being allocated to the four experimental groups (active placebo plus positive message framing, active placebo plus negative message framing, active placebo without framing, and no-treatment control group without receiving any information), but the order in which these four satchels were sent out was random, so that the experimenter who had contact with the participants did not know to what group they would be allocated. The randomisation ration was even to guarantee that all four groups had the same sample size.

After the first online appointment was finished the independent lab member (Biya Tang) was transferred the participants name and address and made sure to send them the satchel containing a capsule container and the appropriate information to what group the participant had been randomised.

Procedure

The total duration of the study covered a two-week period, the first week was used for baseline assessment and the second week was the treatment period. Participants attended online meetings via the application Zoom for three times, each visit lasting on average 20 minutes. The study design and procedure are depicted in Figure 5.1 in the Design section above.

The procedures of making contact and establishing eligibility were identical with Study 2, with the exception that the appointments took place online. All three visits were conducted using the secure, password protected individual meetings via Zoom that were not recorded and participants could only join upon being accepted by the experimenter to guarantee participant privacy (Zoom Video Communications Inc, 2020).

Zoom Appointment 1

During the first online appointment, participants were informed about the aims and the procedures of the study and completed an informed consent document using an online form via REDCap that allowed participants to digitally sign their approval. The aims of the study were described and explained according to the cover story to make participant believe they would participate in a RCT about a new sleep remedy. After all potential questions had been answered to participants' satisfaction, they were asked to provide demographic information. At the end of the first meeting, participants were familiarised with the daily online survey (CSD-C and GASE) that they were asked to fill out each morning as soon as possible after their final awakening.

Zoom Appointment 2

After one week of baseline measures at home, the second online appointment took place. First, participants were asked to fill out the first block of retrospective baseline questionnaires, namely the ISI, WHOQOL-BREF, DASS-21, and the PSM. Participants were then instructed to open the satchel and reveal to the experimenter if they had been allocated to the drug group (capsule container filled with 28 capsules) or if they were allocated to the no-treatment natural history control condition. Unbeknownst to the participants in the alleged "drug" group, participants were further allocated to one of three message framing groups. After this hidden randomisation the participants were instructed to read through the "medication" leaflet they had

received informing them about general considerations while taking the capsules, potential side effects, and the message framing condition. After this the experimenter made sure to go through the most important pieces of information together with the participant and told the participant again the message framing, making sure every single participant had heard the information.

After having been presented with all that information, participants in the three active placebo conditions were told to start taking the four capsules each evening, one hour before their intended sleep onset. Participants in the no-treatment group were told that they were not receiving any medical treatment and that they will act as a control group for the natural course of sleep problems, general health, and well-being.

The subsequent procedures and questionnaires during the treatment week were identical to Study 2.

Zoom Appointment 3

One week later, participants attended the third and last Zoom meeting. First, all participants were instructed to show the capsule container and asked if they had taken all capsules. Then participants started to fill out the retrospective intervention week questionnaires, namely the ISI and WHOQOL-BREF. After that participants did a manipulation check that consisted of a recall and recognition task asking them about the framing condition, they had received during the second online appointment. Participants then filled out a couple of exploratory questionnaires including a recall and recognition task asking them about the side effect information they had received during the second online appointment and an additional questionnaire that asked them about their general beliefs about the relationship between experiencing side effects and the efficacy of medications. Eventually, participants were fully debriefed on the real aims of the study.

Sample Size Calculation

For the sample size estimation regarding message framing effects, an earlier study about the influence of different message framing conditions and their effects on treatment efficacy was referred to (Wilhelm et al., 2018). Based on the results of Wilhelm et al. (2018) the required sample size would be 37 per framing condition ($f = 0.3$, $\alpha = .05$, and 1-beta = .80).

The power analysis was conducted with the software environment R version 3.6.1 (R Core Team, 2019b) and the WebPower package version 0.6 (Zhang & Yuan, 2018). Based on this power analyses participants were allocated to the four groups using a 1:1:1:1 randomisation ratio for active placebo receiving a positive framing: active placebo receiving a negative framing: active placebo receiving no framing: no-treatment control, until the desired total sample size of 148 participants was reached.

Because of the interruptions caused by the SARS-CoV-2 pandemic, it was not possible during the course of this thesis to reach the desired sample size. Due to the pandemic, the recruitment had to be halted for nearly four months to change the study methodology to an online delivery mode and get the study re-approved by the ethics committee. Additionally, once the study was recruiting participants again, it became apparent that not as many participants were interested in participating, as compared with recruitment of Study 2.

Data Analysis

Exclusion of participants from data analysis, assessment of baseline characteristics, the statistical program, and definition of statistical significance in the current study was identical to Study 1.

Outcome measures (ISI, GASE, CSD-C, WHOQOL-BREF) between the baseline and intervention week were analysed with analysis of co-variance (ANCOVAs), using orthogonal planned contrasts to analyse differences between 1) the three placebo groups vs. the no-treatment group, 2) the negative and no framing group vs. the positive framing group, 3) the negative framing group vs. the no framing group. Baseline scores for each respective outcome measure were included as covariates.

The outcome measure adherence between the three placebo groups during the intervention week were analysed with an analysis of variance (ANOVA), using orthogonal planned contrasts to analyse differences between 1) the negative and no framing group vs. the positive framing group, 2) the negative framing group vs. the no framing group.

Differences in the outcome measure RMF at the end of treatment were analysed using logistic regression to analyse differences between 1) the three placebo groups vs. the no-treatment group, 2) the negative and no framing group vs. the positive framing group, 3) the negative framing group vs. the no framing group. All analyses were carried out using the software environment R version 4.1.0 (R Core Team, 2021), with $\alpha = .05$.

Mediation analyses were conducted to examine the influence of expectations regarding the (1) treatment efficacy, (2) likelihood of experiencing side effects, and (3) concerns about side effects. Mediation analyses were also conducted on (4) PSM on all primary and secondary

outcomes, if there was a statistically significant difference between the three placebo groups in the primary and secondary analyses. The mediation analyses were calculated using the mediation package version 4.5.0 (Tingley et al., 2014) within the software environment R version 4.1.0 (R Core Team, 2021).

Results

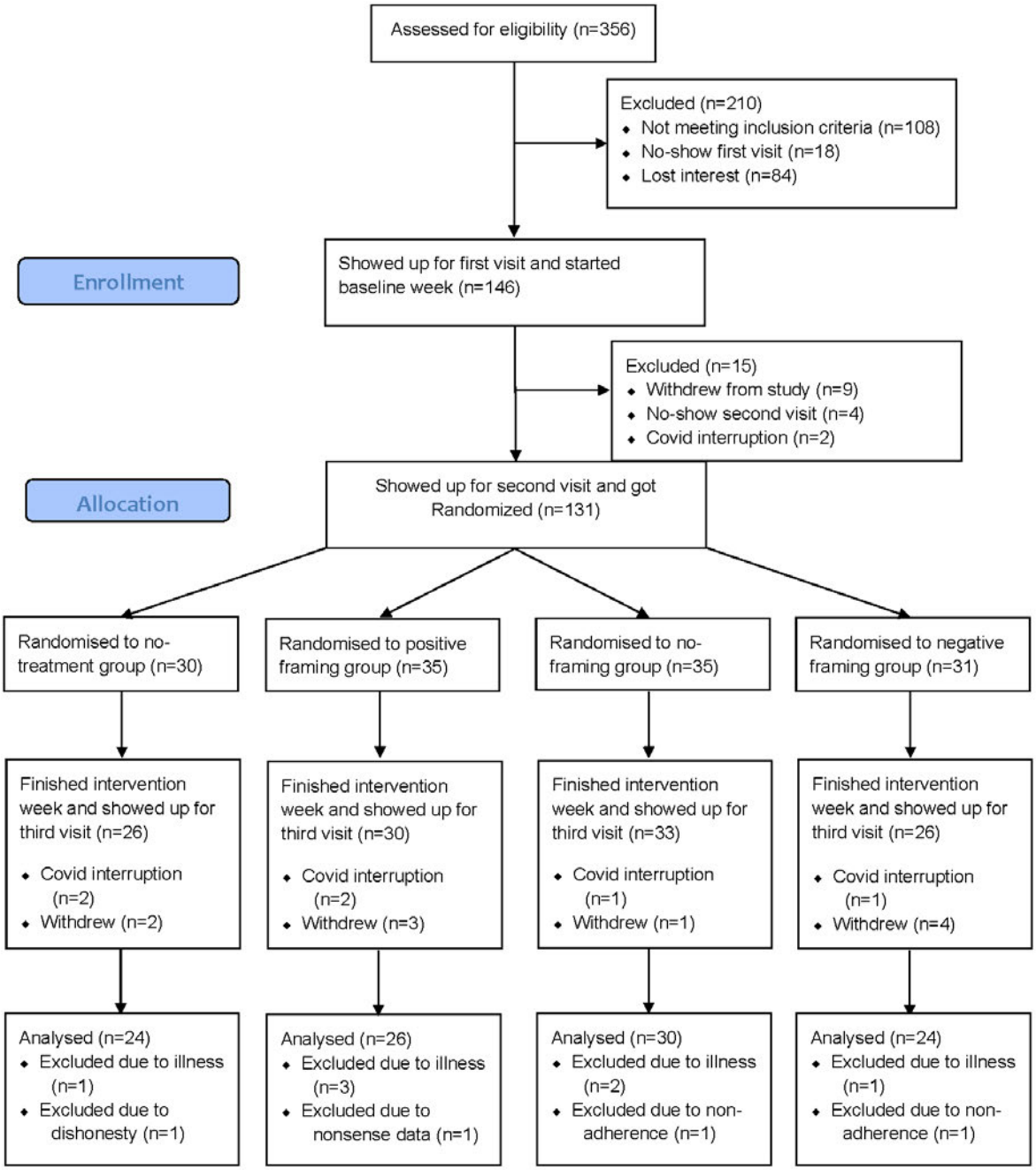
The recruitment process of this study was originally started early in the year 2020 with face-to-face appointments, with participants visiting the University of Sydney Camperdown campus for the three appointments. Because of the SARS-CoV-2 pandemic the study had to be interrupted and the methodology was changed to an online delivery mode using the Zoom application for video calls and delivery of the study capsules via Australia post. Due to delays in recruitment related to the corona pandemic it was not possible to achieve the originally calculated sample size of 148 participants.

The participant flow from the online eligibility screening to the final analysis sample is presented in Figure 5.3. A total of 356 people completed the online eligibility screening. After the screening, 108 participants had to be excluded because they did not fulfil all inclusion criteria, 84 lost interest and cancelled the first appointment, and an additional 18 participants failed to show up for their first appointment without giving any reason. From the 146 participants showing up to the first online appointment, nine declined participation during the baseline week for various reasons and four did not attend the second online meeting and never replied again, and two participants had to be excluded during the baseline week because the university decided that visitors to the University of Sydney's Camperdown campus were no longer allowed

following SARS-CoV-2 related contact restrictions. Out of the 131 participants who were randomly allocated to the four groups, 115 finished the treatment week and showed up for the last Zoom meeting. A total of ten participants withdrew from the study and six participants had to be excluded during the treatment week due to the introduced SARS-CoV-2 restrictions. As for the predefined exclusion criteria, some participants needed to be excluded from analyses to uphold the quality of data. Most participants had to be excluded because they got ill (n = 7; e.g., gastroenteritis, common cold, or influenza). Less common reasons for exclusion from data analyses included data provided by participants that did not make any sense (n = 1), non-adherence (n = 2), and dishonesty (n = 1). The attrition rate from randomisation during the second appointment and the analysis sample did not statistically significantly differ between the four groups, X^2 (3, 131) = 1.49, p = .684.

Figure 5.3

CONSORT Flow Diagram



Note. CONSORT Flow diagram amended from Moher et al. (2001).

Demographics and Descriptive Data

Specific demographic information and sample characteristics for each of the four experimental groups is presented in Table 5.3. The analysis sample for this study comprised 104 participants (69 women), between the age of 18 and 73 years (M = 37.99, SD = 14.29). Contrary to the first two studies that primarily consisted of young international students, the current study only included a small minority of four participants studying at the University of Sydney participating for course credits. The changed advertisement strategy including participants from all across Australia resulted in a large majority of participants (n = 100) enrolling via online advertisement and were volunteers from the general population. Contrary to the first two studies, Australia citizens (n = 66) made up nearly two thirds of the sample with the largest following nationality being New Zealand and England with only four participants each. As the study was advertised online across all of Australia participants' general education level was heterogeneous.

Demographics and other characteristics as assessed at the first visit or as baseline measures during the second visit were similar across groups for most characteristics. The only exception was participants' self-reported weight that statistically significantly differed between the four groups. Participants in the no-treatment group were heavier than the three framing groups receiving placebo treatment.

Table 5.3

Demographic Information and Baseline Characteristics

Variable	No treatment (n = 24)	Positive framing (n = 26)	No framing (n = 30)	Negative framing (n = 24)	Omnibus tests of statistically significant between group differences
	M (SD)	*M (SD)*	*M (SD)*	M (SD)	
Age (years)	40.67 (14.94)	36.42 (14.01)	35.83 (12.30)	39.71 (16.35)	F (3, 100) = 0.72, p = .541
Gender					X^2 (3, 104) = 0.35, p = .950
Women	16	17	19	17	
Men	8	9	11	7	
Education level					X^2 (24, 104) = 33.48, p = .094
Less than year 12	2	1	1	3	
Year 12 or equivalent	3	2	3	7	
Vocational qualification	2	1	2	3	
Associate diploma	0	2	7	2	
Undergraduate diploma	4	1	1	0	
Bachelor’s degree	5	12	9	5	
Postgraduate diploma	1	2	1	2	
Master’s degree	6	3	6	2	
Doctorate	1	2	0	0	

Relationship status					X^2 (12, 104) = 18.89, p = .091
Single	10	16	11	5	
In a relationship	11	7	16	17	
Living separated	0	2	2	1	
Divorced	1	1	1	1	
Widowed	2	0	0	0	
Height (cm)	170.75 (10.93)	170.77 (7.86)	165.90 (10.20)	168.33 (5.96)	F (3, 100) = 1.85, p = .144
Weight (kg)	**86.00 (27.26)**	**74.12 (22.71)**	**71.10 (15.88)**	**71.67 (17.05)**	***F* (3, 100) = 2.72, *p* = .048**
ISI	15.86 (3.33)	16.92 (4.21)	15.23 (3.33)	14.46 (4.14)	F (3, 100) = 1.95, p = .127
DASS-21	15.96 (12.05)	18.77 (11.21)	17.47 (11.90)	12.58 (9.91)	F (3, 100) = 1.39, p = .252
WHOQOL-BREF	3.49 (0.57)	3.47 0.65)	3.66 (0.56)	3.79 (0.52)	F (3, 100) = 1.70, p = .172
VAS Prior experience with medications	73.83 (18.35)	75.31 (20.94)	68.33 (23.30)	68.58 (28.03)	F (3, 100) = 0.64, p = .590
VAS Prior experience with side effects of medications	78.21 (23.07)	73.77 (27.21)	78.83 (20.89)	72.92 (22.43)	F (3, 100) = 0.43, p = .729

Note. The information provided in this table represents the analysis sample. Statistically significant differences between the four experimental groups as indicated by omnibus ANOVA and Chi-square tests are bolded.

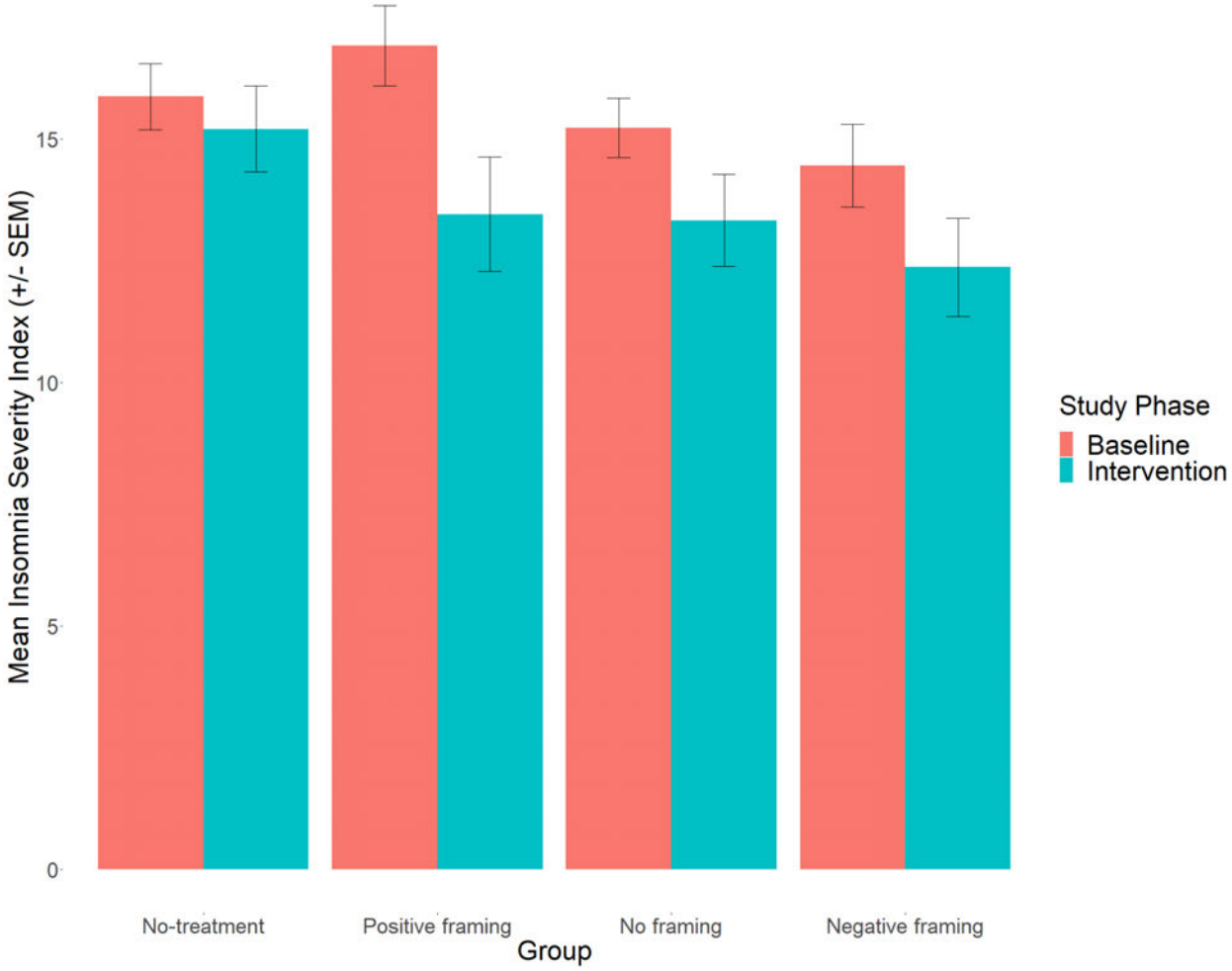
Primary Outcomes

Insomnia Severity Index

Figure 5.4 shows the four groups' ISI score for the baseline and intervention week. Adjusting for the ISI during the baseline week, the ANCOVA analysis did not show a statistically significant main effect for the factor group on the ISI during the treatment week, F (3,99) = 1.42, p = .242. The baseline-adjusted ISI scores fitted for the ANCOVA model were M = 15.027 (SE = 0.899) for the no-treatment group, M = 12.519 (SE = 0.876) for the positive framing group, M = 13.618 (SE = 0.805) for the no-framing group, and M = 13.222 (SE = 0.908) for the negative framing group. The first planned orthogonal contrast comparing the no-treatment group to the three placebo groups showed that the three placebo groups did not differ statistically significantly in terms of improvement on the ISI from baseline to intervention compared to the no-treatment group (F (1,103) = 1.86, p = .066. The second planned orthogonal contrast comparing the positive framing group to the no-framing and negative framing groups did not indicate a statistically significantly larger improvement on the ISI for the positive framing group compared to other framing groups (F (1,79) = 0.83, p = .407). The last orthogonal contrast between the no-framing and the negative framing group did not result in a statistically significant difference between the two group in terms of improvements on the ISI (F (1,53) = 0.33, p = .744).

Figure 5.4

Mean Baseline and Treatment Insomnia Severity Index across Groups



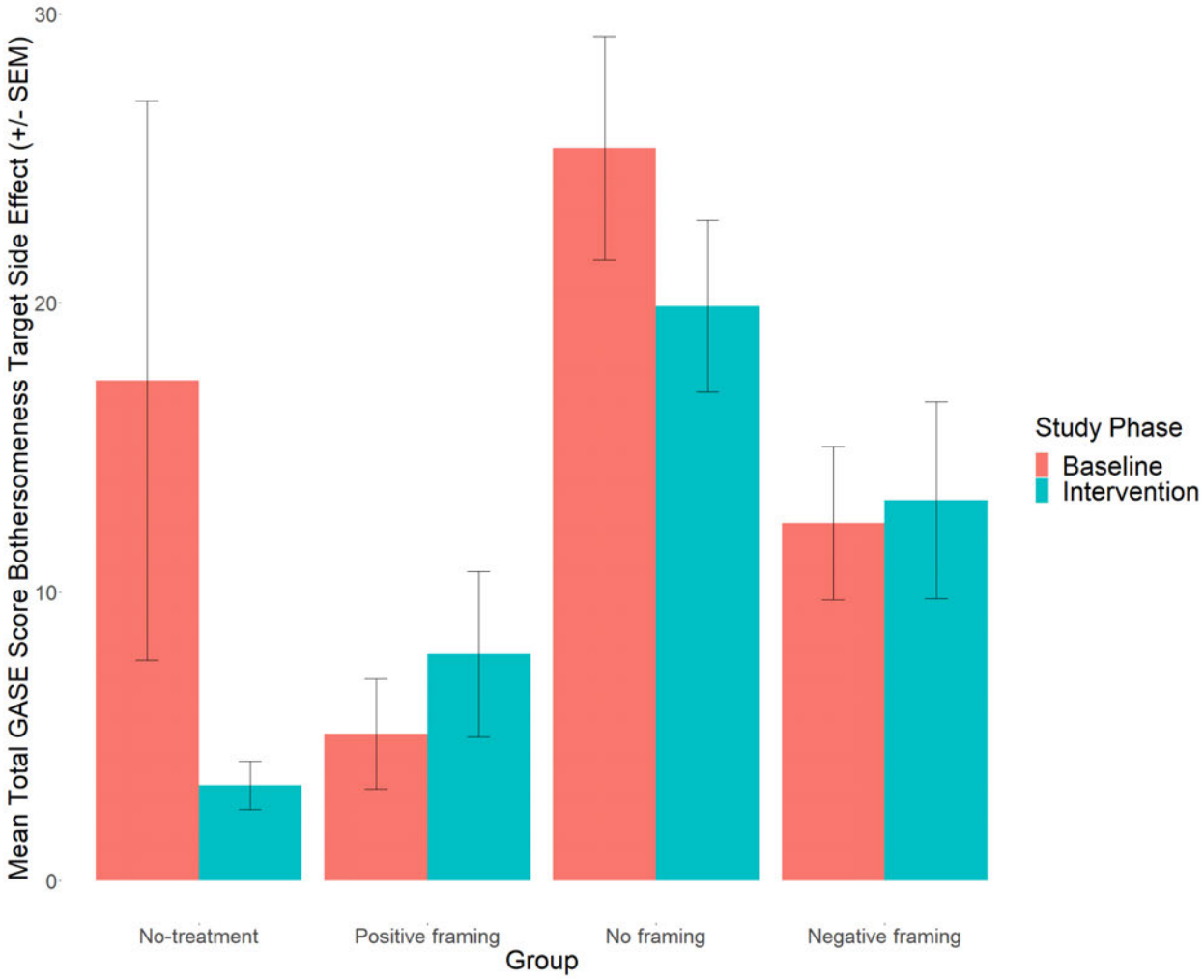
Bothersomeness Target Side Effect

Figure 5.5 shows the four groups' ratings about how bothersome they perceived the target side effect was for the baseline and intervention week. Adjusting for the bothersomeness during the baseline week, the ANCOVA analysis did not show a statistically significant main effect for the factor group on the bothersomeness during the treatment week, F (3,99) = 1.06, p = .369. The baseline-adjusted bothersomeness of the target side effect fitted for the ANCOVA model were *M*

= 15.19 (*SE* = 10.72) for the no-treatment group, *M* = 29.30 (*SE* = 10.33) for the positive framing group, *M* = 38.79 (*SE* = 9.69) for the no-framing group, and *M* = 19.71 (*SE* = 10.67) for the negative framing group. The first planned orthogonal contrast comparing the no-treatment group to the three placebo groups showed that the three placebo groups did not differ statistically significantly in terms of change in the target side effect's bothersomeness from baseline to intervention compared to the no-treatment group (F (1,103) = 1.15, p = .253). The second planned orthogonal contrast comparing the positive framing group to the no-framing and negative framing groups did not indicate a statistically significantly different change in bothersomeness for the positive framing group compared to other framing groups (F (1,79) = 0.00, p = .997). The last orthogonal contrast between the no-framing and the negative framing group did not result in a statistically significant difference between the two groups in terms of bothersomeness of the target side effect (F (1,53) = 1.33, p = .188).

Figure 5.5

Mean Baseline and Treatment Bothersomeness of the Target Side Effect across Groups



Secondary Outcomes

Subjective Sleep Quality

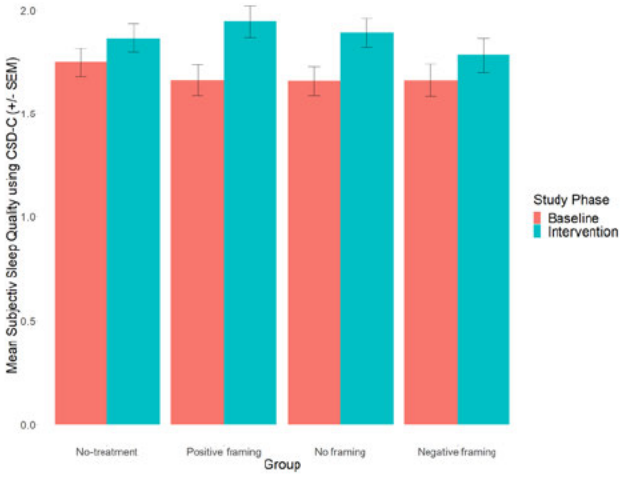
Figure 5.6a shows the four groups' sSQ for the baseline and intervention week. Adjusting for the sSQ during the baseline week, the ANCOVA analysis did not show a statistically significant main effect for the factor group on subjective sleep quality during the treatment week,

F (3,99) = 0.52, p = .667. The baseline-adjusted sSQ fitted for the ANCOVA model were M = 1.821 (SE = 0.111) for the no-treatment group, M = 1.961 (SE = 0.107) for the positive framing group, M = 1.911 (SE = 0.099) for the no-framing group, and M = 1.793 (SE = 0.111) for the negative framing group. The first planned orthogonal contrast comparing the no-treatment group to the three placebo groups showed that the three placebo groups did not differ statistically significantly in terms of improvement on sSQ from baseline to intervention compared to the no-treatment group (F (1,103) = 0.53, p = .599). The second planned orthogonal contrast comparing the positive framing group to the no-framing and negative framing groups did not indicate a statistically significantly larger improvement on sSQ for the positive framing group compared to other framing groups (F (1,79) = 0.84, p = .402). The last orthogonal contrast between the no-framing and the negative framing group did not show a statistically significant difference between the two group in terms of improvements for subjective sleep quality (F (1,53) = 0.80, p = .428).

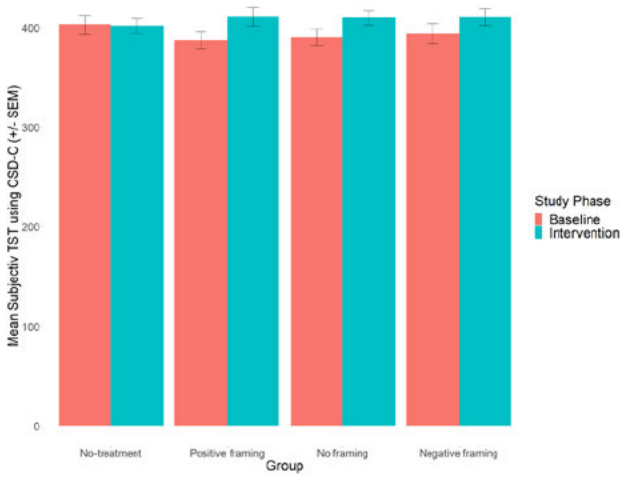
Figure 5.6

Mean Baseline and Treatment Outcomes across Groups for Secondary Outcomes

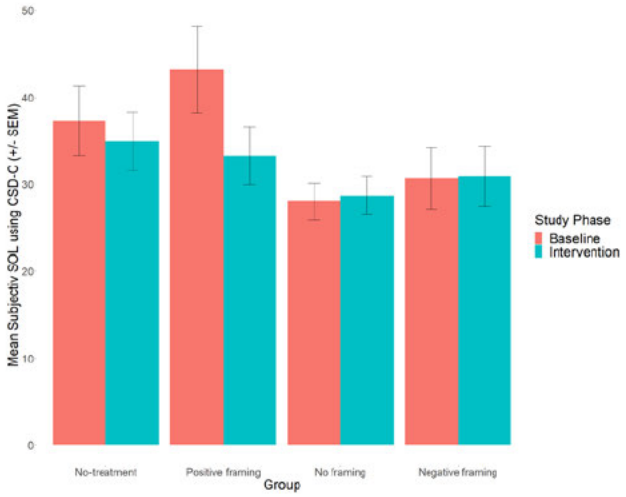
a.)



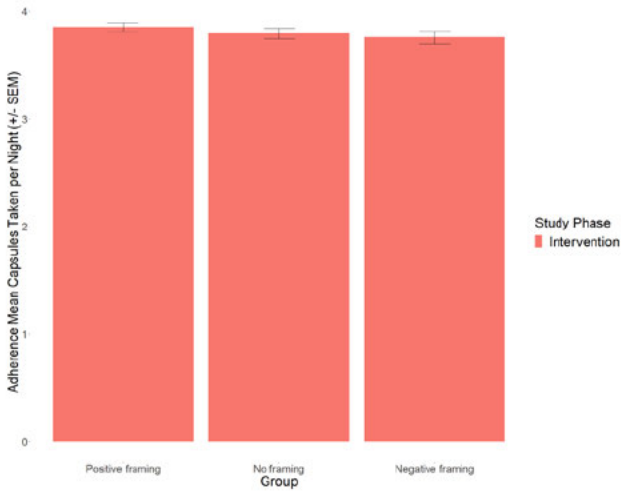
b.)



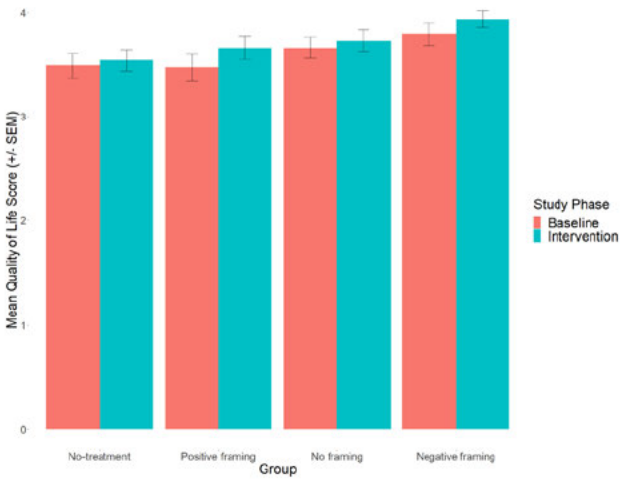
c.)



d.)



e.)



Subjective Total Sleep Time

Figure 5.6b shows the four groups' sTST for the baseline and intervention week. Adjusting for the sTST during the baseline week, the ANCOVA analysis did not show a statistically significant main effect for the factor group on subjective total sleep time during the treatment week, F (3,99) = 0.74, p = .529. The baseline-adjusted sTST fitted for the ANCOVA model were M = 395.448 (SE = 10.617) for the no-treatment group, M = 415.602 (SE = 10.189) for the positive framing group, M = 412.326 (SE = 9.480) for the no-framing group, and M = 411.656 (SE = 10.597) for the negative framing group. The first planned orthogonal contrast comparing the no-treatment group to the three placebo groups showed that the three placebo groups did not differ statistically significantly in terms of improvement on sTST from baseline to intervention compared to the no-treatment group (F (1,103) = 1.46, p = .146). The second planned orthogonal contrast comparing the positive framing group to the no-framing and negative framing groups did not indicate a statistically significantly larger improvement on sTST for the positive framing group compared to other framing groups (F (1,79) = 0.29, p = .772). The last orthogonal contrast between the no-framing and the negative framing group did not show a statistically significant difference between the two group in terms of improvements for subjective total sleep time (F (1,53) = 0.05, p = .963).

Subjective Sleep Onset Latency

Figure 5.6c shows the four groups' sSOL for the baseline and intervention week. Adjusting for the sSOL during the baseline week, the ANCOVA analysis did not show a statistically significant main effect for the factor group on subjective sleep onset latency during

the treatment week, F (3,99) = 0.73, p = .540. The baseline-adjusted sSOL for the ANCOVA model were M = 33.056 (SE = 3.377) for the no-treatment group, M = 27.448 (SE = 3.268) for the positive framing group, M = 32.938 (SE = 3.034) for the no-framing group, and M = 33.099 (SE = 3.380) for the negative framing group. The first planned orthogonal contrast comparing the no-treatment group to the three placebo groups showed that the three placebo groups did not differ statistically significantly in terms of improvement on sSOL from baseline to intervention compared to the no-treatment group (F (1,103) = 0.49, p = .624). The second planned orthogonal contrast comparing the positive framing group to the no-framing and negative framing groups did not indicate a statistically significantly larger improvement on sSOL for the positive framing group compared to other framing groups (F (1,79) = 1.39, p = .168). The last orthogonal contrast between the no-framing and the negative framing group did not show a statistically significant difference between the two group in terms of improvements for subjective sleep onset latency (F (1,53) = 0.04, p = .972).

Adherence

Figure 5.6d shows how adherently the three placebo groups took their capsules during the intervention week. The ANOVA analysis did not show a statistically significant main effect for the factor group on adherence during the treatment week, F (2,77) = 0.19, p = .829. Out of the maximum of 28 capsules each placebo participant received the means from ANOVA model were M = 26.385 (SE = 0.765) for the positive framing group, M = 25.800 (SE = 0.712) for the no-framing group, and M = 25.833 (SE = 0.796) for the negative framing group. The first planned orthogonal contrast comparing the positive framing group to the no-framing and negative

framing groups did not indicate a statistically significant difference in adherence for the positive framing group compared to other framing groups (F (1,79) = 0.70, p = .489). The second orthogonal contrast between the no-framing and the negative framing group did not show a statistically significant difference between the two group in terms of adherence (F (1,53) = 0.04, p = .972).

Quality of Life

Figure 5.6e shows the four groups' quality of life for the baseline and intervention week. Adjusting for the QOL during the baseline week, the ANCOVA analysis did not show a statistically significant main effect for the factor group on quality of life during the treatment week, F (3,99) = 1.87, p = .139. The baseline-adjusted QOL for the ANCOVA model were M = 3.627 (SE = 0.054) for the no-treatment group, M = 3.760 (SE = 0.052) for the positive framing group, M = 3.684 (SE = 0.049) for the no-framing group, and M = 3.789 (SE = 0.055) for the negative framing group. The first planned orthogonal contrast comparing the no-treatment group to the three placebo groups showed that the three placebo groups did not quite differ statistically significantly in terms of improvement for QOL from baseline to intervention compared to the no-treatment group (F (1,103) = 1.89, p = .062). The second planned orthogonal contrast comparing the positive framing group to the no-framing and negative framing groups did not indicate a statistically significantly larger improvement on QOL for the positive framing group compared to other framing groups (F (1,79) = 0.36, p = .720). The last orthogonal contrast between the no-framing and the negative framing group did not show a statistically significant difference between the two group in terms of improvements for quality of life (F (1,53) = 1.45, p = .152).

Recall and Recognition Message Framing

The results for the manipulation check, asking participants to recall and recognise the framing information they received during the second online appointment are presented in Table 5.4. For the recall participants had to describe the information they had received during the second online appointment in their own words. Results regarding the open recall indicated that the four groups differed statistically significantly in their ability to correctly indicate the information they had been exposed to (X^2 (6, 104) = 49.37, p < .001). Generally, a majority of participants in the placebo groups were not able to correctly recall the information they had received. When comparing the three placebo groups they did not differ statistically significantly in their ability to remember their allocation (X^2 (4, 80) = 3.42, p = .490). While the groups' ability to correctly recall the received framing information was limited the majority of participants managed to correctly recognise the information they had been exposed to once it was presented to them in the forced recognition item. The pattern of results for the recognition resembled that of the recall. The overall test indicated that the four groups differed statistically significantly in their ability to correctly recognise their framing information (X^2 (3, 104) = 11.41, p = .010), but the three placebo groups did not differ statistically significantly in their ability to correctly identify the framing information they had received (X^2 (2, 80) = 2.75, p = .253).

Table 5.4

Recall and Recognition about the Framing Condition

Indicated Choice \ Group	No treatment (n = 24)	Positive framing (n = 26)	No framing (n = 30)	Negative framing (n = 24)	Omnibus tests	Three placebo conditions
Open recall:					**X^2 (6, 104) = 49.37, p < .001**	X^2 (4, 80) = 3.42, p = .490
Incorrect/no recall	2	19	27	20		
Correct recall	22	6	3	3		
Partial recall	0	1	0	1		
Recognition:					**X^2 (3, 104) = 11.41, p = .010**	X^2 (2, 80) = 2.75, p = .253
Correct	22	19	16	13		
Incorrect	2	7	14	11		

Note. The recognition item presented participants with the actual two message framing information as presented in the method section, and a third option stating that no information had been presented. Statistically significant results are bolded.

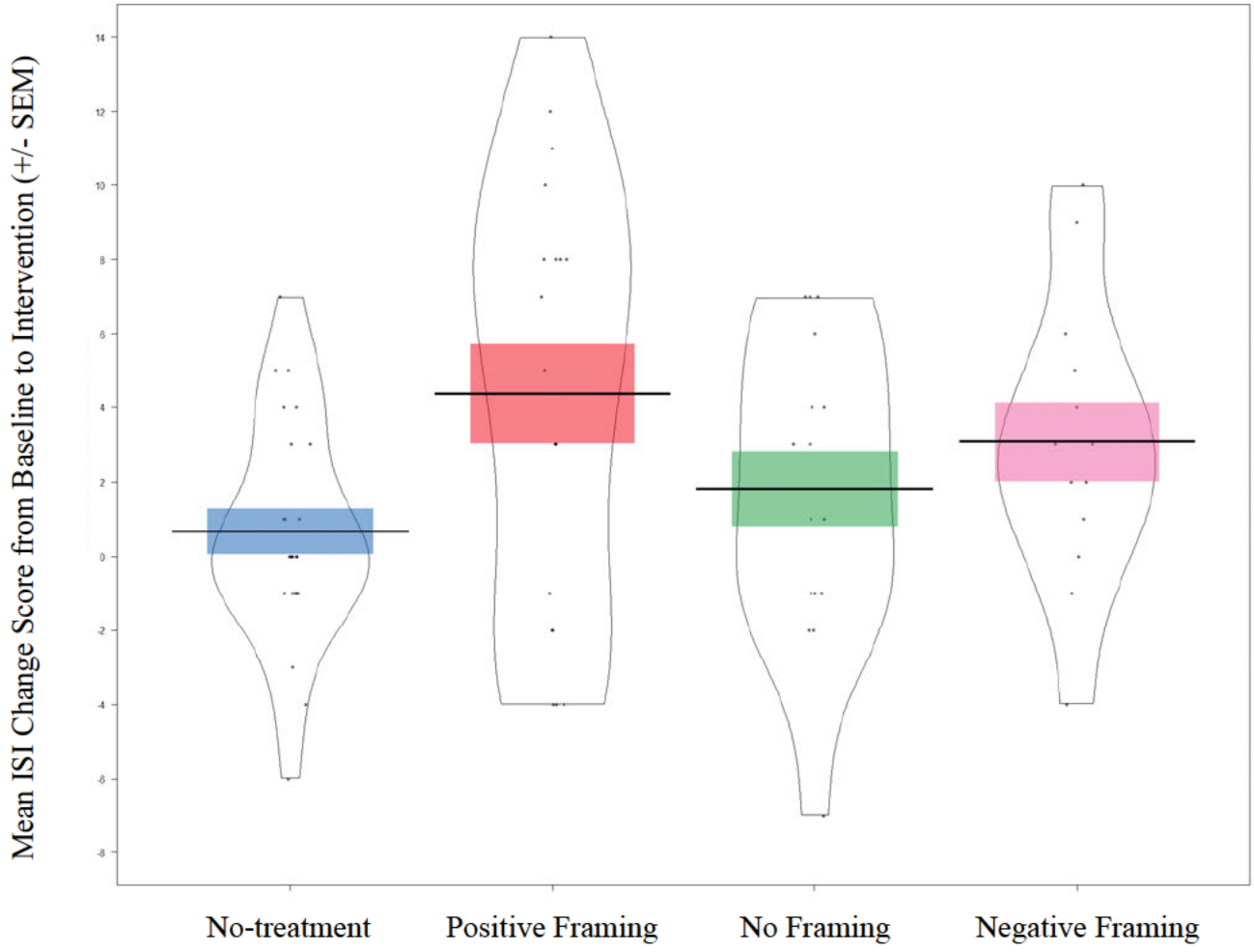
Exploratory Analyses

Exploratory analyses were calculated with the subsample of participants who correctly recognised the message framing information they had received, to investigate if this subsample showed differences as hypothesised on the two primary outcomes.

Figure 5.7 shows the four groups' ISI change score from baseline to intervention for the subsample of participants who correctly recognised the message framing information they had received during the second online appointment. Adjusting for the ISI during the baseline week, the ANCOVA analysis did not show a statistically significant main effect for the factor group on the ISI during the treatment week, F (3,67) = 2.33, p = .082. The baseline-adjusted ISI scores fitted for the ANCOVA model were M = 15.26 (SE = 0.85) for the no-treatment group, M = 12.04 (SE = 0.98) for the positive framing group, M = 13.77 (SE = 1.06) for the no-framing group, and M = 12.66 (SE = 1.16) for the negative framing group. The first planned orthogonal contrast comparing the no-treatment group to the three placebo groups for the subsample showed that the three placebo groups differed statistically significantly in terms of improvement on the ISI from baseline to intervention compared to the no-treatment group (F (1,69) = 2.32, p = .023. The second planned orthogonal contrast comparing the positive framing group to the no-framing and negative framing groups did not indicate a statistically significantly larger improvement on the ISI for the positive framing group compared to other framing groups (F (1,47) = 0.91, p = .364). The last orthogonal contrast between the no-framing and the negative framing group did not result in a statistically significant difference between the two group in terms of improvements on the ISI (F (1,28) = 0.71, p = .479).

Figure 5.7

Subsample Mean ISI Change Scores for the Subsample recognising their Message Framing Information



Note. The figure above shows participants ISI change scores from baseline to the intervention week on the y-axis. Larger scores indicate larger improvements on the ISI. The horizontal lines are the means and *SEMs* are plotted as the boxes around the mean. The points in the density distributions are participants' individual ISI change scores.

The second exploratory analysis analysing the subsamples bothersomeness of the target side effect did not result in any significant result and therefore can be found in Appendix D.

Discussion

This study aimed to evaluate the effects of positive, no, and negative message framing on the efficacy of an active placebo model and if the framing manipulation altered how bothersome placebo participants rated the target side effect they experienced.

Regarding the primary hypothesis, while the results numerically reflected the hypothesised pattern, with the no-treatment group nearly showing no improvement on the ISI, the positive message framing group having the numerically largest improvement compared to the no and negative framing condition, the variation within the groups was either too large or the effect size of the message framing was too small given that none of the differences resulted in a statistically significant result. The first contrast was aimed at observing a placebo effect, hence comparing the no-treatment group to the three active placebo groups. While this contrast came close to the pre-defined level for statistical significance it did not reach statistical significance. The more nuanced contrasts between the different framing conditions did not show any support in favour of the hypothesis. In terms of both primary outcomes, the ISI and how bothersome participants rated the target side effect of the active placebo did not show any statistically significant effect. Similar to the ISI, the daily assessed subjective sleep quality, total sleep time, sleep onset latency, quality of life, and adherence indicated a trend showing the expected results with the three placebo groups improving more than the no-treatment group, and the positive message framing group outperforming the other framing conditions. As was the case for the primary outcomes, none of the secondary outcomes resulted in statistically significant differences. Interestingly, when calculating the exploratory analysis with the subsample of participants who were able to correctly recognise the message framing information, they had

received during the second appointment, there was an overall statistically significant placebo effect on the ISI.

Several reasons might explain why this study did not find statistically significant support for a placebo effect and why the message framing did not result in different outcomes. First, because of the reduced recruitment success due to the SARS-CoV-2 pandemic it was not possible to enrol as many participants as were deemed necessary according to the power calculation. Obviously, a sufficiently sized sample is needed for the detection of statistically significant differences when comparing different groups. This is especially important if the effect size of a given intervention or manipulation is best classified as small.

A second reason that might explain the diminished effects might have to do with the settings of the two earlier studies conducted by Wilhelm et al. (2018) and Fernandez et al. (2019). Both studies tested the effects of framing manipulations in short laboratory-based sessions. Although both tried to conceal the underlying aim of the experiments, participants received the framing within minutes before the effects of the intervention could take place and were assessed. The current study, on the other hand, covered a time period across two weeks and most importantly the target side effect of the active placebo was not experienced immediately because participants took the capsules before going to bed and might have only observed beeturia the next morning. The delay between receiving the message frame information and the experience of the side effect might cause some participants to forget what they were informed about. Participants forgetting the message framing or other important aspects from the cover story might influence participants formation or retention of expectations and therefore reduce the placebo effect or any effects of the message framing.

Further, the study was presented to participants as a pharmacological RCT meaning a large amount of information was transferred. Of all this information, the actual message framing manipulation only made up a small fraction. Therefore, the effects of the framing manipulation in the current study might not have had the same extent as in single-session, in-person experimental sessions. This argumentation is supported by the recall data. At the end of the third online appointment, participants were asked to recall the framing information they had received one week ago together with the other treatment relevant information that is typically needed in actual pharmacological RCTs for ethical reasons. Across all three placebo groups, only 12 out of 80 (15%) participants were able to correctly recall the framing information received, a further two were partially able to recall the information. The large majority (66 out of 80, 82.5%) placebo participants did not recall any information at all or recalled the information incorrectly. Unfortunately, this very low number of participants correctly recalling the information did not allow for subgroup analyses to see if there were any differences in the extent of the placebo effect or the bothersomeness of the side effect for accurate recall. Interestingly, the recall rates between the three framing conditions did not differ. The same was observed for recognition rates, that did not differ between the three framing conditions. As would be expected, in comparison to the recall, the overall recognition rates were much better with 48 out of 80 placebo participants correctly identifying the correct framing information once they were re-exposed to them.

Depending on the point of view, one might argue that the poor memory retention rate of the message framing manipulation could be seen as a limitation of this study. From the perspective of primarily studying the effects of differently framed information to better understand how people react to differently framed instructions this is certainly the case. On the

other hand, it is possible to argue that the current study closely resembled the situation encountered in clinical practice where there are time delays between receiving information and the potential experience of side effect by participants, or even to a certain degree in clinical trials. Therefore the absence of strong or long-lasting framing effects might reassure trialists and clinicians that at least the way in which information is communicated to patients and clinical trial participants might not be as impactful as was originally suspected or argued (Glare et al., 2018). Although this finding is further supported by an experimental study by Faasse et al. (2019) who observed that positive framing reduced side effects if assessed just after the application of a treatment but not anymore at follow-up assessment occurring only after a 24-hour period, there might be potential alternative explanations besides delays in time between information and experience of side effects like a plethora of other potentially more important or salient information capturing participants attention as to why message framing was not as impactful in the current study as previously shown in experimental message framing studies.

Additional support for the lack of any framing effects in the current study came from the exploratory subgroup analysis of participants who were able to recognise their message framing information, but still did not show any differences in bothersomeness ratings for the target side effect. Not only were placebo participants' bothersomeness ratings for the target side effect equal, but the three placebo groups also reported similar overall symptom loads during the treatment week. The finding that overall symptom rates is not affected by message framing is consistent with Wilhelm et al. (2018), who showed that positive message framing reduced the burden of side effects, but not the actual rate.

Because the current study could not reach the desired sample size due to SARS-CoV-2 related reasons, some caution is required before drawing final conclusions for clinical practise or

clinical trial methodology based on the presented data. Nevertheless, the evidence for message framing effects pointed towards certain trends that statistically significant effects might be observed if investigated on a larger scale. While the effect size difference between the three placebo groups and the no-treatment group was around a Cohen's *d* of 0.44 on the ISI, meaning a small to medium effect size of the current study, the effect size difference between the positive framing group and the two other framing groups was small with a Cohen's *d* of around 0.29. This means that the necessary sample size to achieve a statistically significant test result between the positive framing group and the negative and no framing group would at least require around a sample size of 190 participants, assuming a power of 1-beta of 0.8, a significance level of 0.05, and a two-sample design. Based on the effect size calculation above it is important to mention that the current study would have theoretically observed an overall statistically significant placebo effect had the study reached the desired sample size.

Although the required sample size of 190 would be considered quite large for an experimental study, the effect of message framing should not be neglected when considering how many people are potentially exposed to information regarding side effects, both as participants in RCTs and as patients in clinical practice. Due to the problems with participants not being able to recall the received framing information because it might have been presented too subtly in the current study, it might be of particular interest that future research investigates in how message framing effects influence patients in clinical practise, and in larger trials that are sufficiently powered to detect the small effects of message framing on the bothersomeness or even overall rate of side effects.

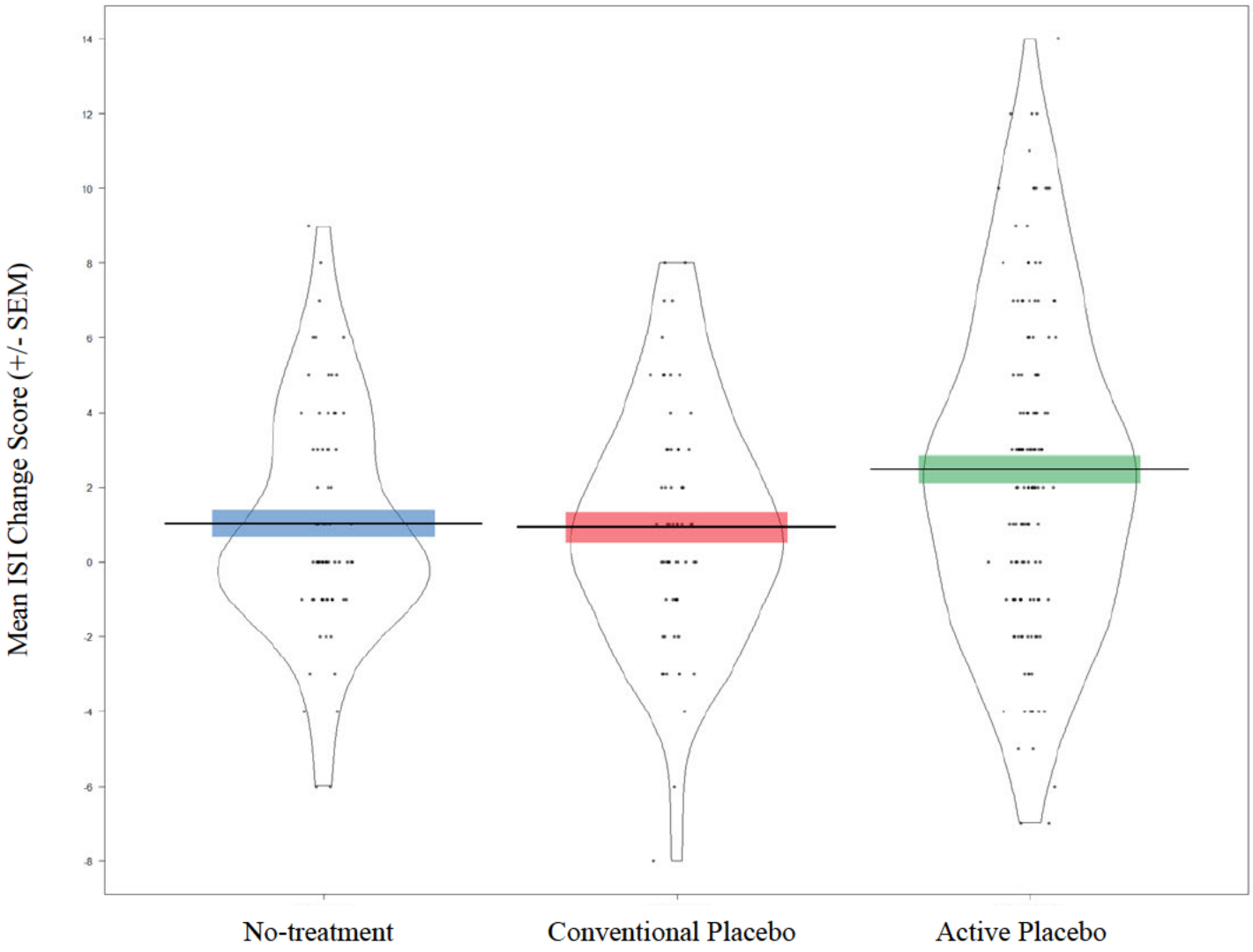
Chapter 6: Pooled Analysis

Prior experimental research and a meta-analysis comparing placebo treatment to no-treatment found statistically significant placebo effects, including within the current laboratory (Neukirch & Colagiuri, 2015; Yeung et al., 2020; Yeung et al., 2018). Because the three individual experimental studies conducted here did not find any statistically significant overall placebo effect on the ISI, a pooled analysis including all participants from the three experimental studies was carried out to investigate if there was an overall placebo effect across the three studies on the Insomnia Severity Index. The change scores of on the ISI from baseline to intervention for the no-treatment group, conventional placebo group, and the active placebo group are depicted in Figure 6.1.

Adjusting for the ISI during the baseline week, the ANCOVA analysis did not show a main effect for the factor group on the ISI during the treatment week, F (2,268) = 2.02, p = .134. The baseline-adjusted ISI scores fitted for the ANCOVA model were M = 11.53 (SE = 0.44) for the no-treatment group, M = 11.23 (SE = 0.47) for the conventional placebo group, and M = 10.49 (SE = 0.32) for the active placebo group. The first planned orthogonal contrast comparing the no-treatment group to the two placebo groups showed that the two placebo groups did not differ in terms of improvement on the ISI from baseline to intervention compared to the no-treatment group (F (1,271) = 1.29, p = .199. The second planned orthogonal contrast comparing the conventional placebo group to the active placebo group did not indicate a statistically significantly larger improvement on the ISI for the active placebo group (F (1,199) = 1.27, p = .204).

Figure 6.1

Mean ISI Change Scores from Baseline to Intervention from Participants across all three Experimental Studies



Note. The figure above shows participants ISI change scores from baseline to the intervention week on the y-axis. Larger scores indicate larger improvements on the ISI. The horizontal lines are the means and *SEMs* are plotted as the boxes around the mean. The points in the density distributions are participants' individual ISI change scores.

The pooled analysis neither found any statistically significant overall placebo effects, nor did any of the pairwise comparisons indicate that there was a statistically significant placebo effect. While pooling participants across the three studies increases statistical power to detect a statistically significant placebo effect, it cannot address limitations of the individual studies.

The pooled analysis still suffered from the problem that participants from Study 1 were healthy young adults that did not suffer from sleep problems, which in fact might only have increased the within-group variability in terms of the spread of the ISI scores, therefore making it even more difficult to find any statistically significant effects. Further, the problem with student participants from Study 2 that enrolled for course credits still remained unsolved, because Study 3 only tested active placebo group, hence not adding any new participants that received a conventional placebo.

While it was important to calculate a pooled analysis to clarify if there was a statistically significant overall placebo effect, it might be most useful for the discussion of placebo effects to focus on the findings from Study 3 that focused on investigating people suffering from sleep problems from the general population. The participant sample of Study 3 could be referred to as the least biased sample due to the recruitment strategy and the fact that they did not receive any financial or other externally motivating reimbursements like course credits that might have motivated them to participate other than their intrinsic motivation to get better sleep.

Chapter 7: General Discussion and Conclusions

The main goal of this thesis was to test a new model of an active placebo to investigate how side effects influence the placebo effect. Literature about the validity of randomised controlled trials suggested that side effects of active medications might boost the placebo effect within the drug group of RCTs compared with inert treatments. If this were the case, then it could mean that double-blind placebo-controlled randomised trials, the gold standard in evidence-based medicine to evaluate and license new pharmacological treatments, might overestimate the drug-placebo difference and incorrectly assess side effects.

The present thesis therefore had several aims. The first aim was to conduct a systematic literature review and meta-analysis to gain an overview of side effects in placebo groups of pharmacological insomnia trials, with the exploratory aim to see if adverse event rates in the placebo groups would correlate with the placebo response. The second aim was to evaluate and test a new model of an active placebo eliciting beeturia in people with sleep difficulty. Therefore, a series of three experimental studies were conducted. The goal of Study 1 was to test if the new active placebo model reliably elicited the target side effect and to evaluate the study design and methodology. Study 2 tested the active placebo against a conventional lactose placebo under (fake) double-blind conditions to see if participants in the active placebo group were more likely to believe that they had actually received a medication and if the experience of the side effect enhanced the placebo effect. Finally, Study 3 manipulated the information given to participants receiving active placebos. The information given to participants was either framed positively such that experiencing the side effect meant the treatment was working particularly well, framed negatively suggesting that experiencing the side effect meant that the treatment was not going to be efficient, or they received no

information. The current chapter summarises the findings of this thesis and the implications, discusses strengths and limitations, and suggests directions for future research.

Summary of findings

The systematic literature review and meta-analysis included 88 distinct RCTs with a total of 27,885 insomnia patients. The trials included the most typical drug classes used in the treatment of insomnia, including benzodiazepines, non-benzodiazepine hypnotics, melatonin agonists, antidepressants, orexin receptor agonists, and herbal remedies. When analysing placebo groups by the drug category the RCT investigated, the analysis showed that participants in placebo groups experienced statistically significantly different adverse event rates and specific adverse event profiles depending on the drug category investigated. These observations seemed to be robust, as they did not differ, even when controlling for the most relevant sample characteristics and methodological aspects. Analysing the specific adverse event profiles head-to-head between the placebo groups and their corresponding drug group, indicated a statistically highly significant positive correlation. The exploratory analysis did not show a statistically significant association between the overall adverse event rate in placebo groups and the placebo response, but the data for this analysis was of questionable quality, because there was no consistent assessment of adverse events across trials.

The analysis sample of the Study 1 consisted of 71 adult participants. The participants were mostly healthy students enrolled in an undergraduate course at the University of Sydney who participated for course credit. Study participation included three in-person visits on campus and consisted of seven baseline and seven intervention nights, where participants' sleep, other health-related outcomes, and daily symptoms were assessed on a daily basis. Participants were made to believe that they would either receive an actual medication or no-treatment. The active placebo successfully elicited beeturia as the target side effect in 50% of

active placebo participants, while only 23% in the conventional placebo group indicated that they had experienced an unusual urine colouration. The active beetroot placebo was considered a success because it was able to elicit the target side effect at a similar rate as did atropine. Atropine is a potent medication and was most commonly used as an active placebo in past pharmacological RCTs. Although atropine is a potent medication, Berna et al. (2017) reported that it only elicited dry mouth as a target side effect in 58% of their study. Even though Study 1 established the new active beetroot placebo there was no statistically significant placebo effect observed but this could be explained by the recruitment of healthy volunteers not necessarily experiencing sleep difficulty.

Study 2 focused on participants' perceived treatment allocation and the active placebo's potential to enhance the placebo effect. The study design was similar to Study 1, with the exception that participants were told that they would either be allocated to a medication, placebo, or no-treatment group. Ninety-seven sleep-impaired adults participated in Study 2 and reported analysable data. After one-week of placebo treatment participants' perceived treatment allocation differed statistically significantly between the two placebo groups. In the conventional placebo group merely 9 out of 38 participants believed that they had received an actual medication, whereas nearly half of participants (18 out of 38) believed that they had actually taken a medication during the last week. This result clearly indicated failed participant blinding for the conventional placebo group and effective blinding in the active placebo group (because participants' guesses were at the equivalent of chance). While the overall analysis for the ISI did not indicate a statistically significant placebo effect when comparing the two placebo groups to the no-treatment group, this was most likely due to the unusually large improvement in some of the no-treatment participants and unusually small placebo effects of student participants in the conventional placebo group. The contrast between the two placebo groups indicated that participants in the active placebo group

showed statistically significantly larger improvements on the insomnia severity index compared to the conventional placebo groups who on average showed nearly no improvement at all.

Study 3 investigated the effects of message framing on the influence of side effects on the placebo response. Therefore, participants were again recruited under the guise of testing a new sleep medication and thought they were allocated to a medication or a no-treatment group. In reality, the study comprised four different groups, a no-treatment group not receiving any information at all serving as natural history control, a positive message framing group that were told experiencing the side effect was beneficial for the efficacy of the drug, a no-framing group that did not receive any framing information but still received the general information, and a negative message framing group that was told that experiencing the side effect would mean the medication is less likely to help them with their sleep problems. The study included a total of 104 people with sleep difficulty who finished the study and provided data. The analysis revealed that there was no statistically significant overall placebo effect on the ISI between the three placebo groups and the no-treatment group. When analysing a subsample of participants that were able to recognise the message framing information they had received, the overall placebo effect on the ISI was statistically significant. Contrary to earlier studies investigating framing effects (Faasse et al., 2019; Fernandez et al., 2019; Wilhelm et al., 2018) there was no statistically significant effect of message framing on participants perceived bothersomeness or any of the secondary outcomes.

Theoretical and clinical implications

The systematic literature review and meta-analysis of this thesis demonstrated that strong nocebo responses in clinical trials of pharmacological interventions for insomnia can

be observed. It is especially important for clinicians and patients to understand that the side effect information provided in medication leaflets is not based on the same quality evidence as for the efficacy analysis. This is because side effects reported in medication leaflets typically refer to *any* side effect reported in those receiving the drug, irrespective of whether that side effect occurred at an equivalent rate in the placebo group. As long as side effect information from double-blind RCTs are assessed and reported with a methodological quality that is lagging behind the efficacy outcomes, the side effect information provided to clinicians and patients might be misleading, most likely even exaggerated. This in turn might cause nocebo effects in the general patient population causing unnecessary and avoidable harm, reduced well-being, or might even lead to non-adherence with fatal consequences (Cooper et al., 2015).

The experimental studies further emphasise the need to consider the role of side effects in RCTs and clinical practice. First and foremost, they demonstrated that active placebos are more likely to lead to effective blinding of placebo groups than conventional inert placebos. Interestingly, however, it appeared that the experience of side effects when taking an active placebo might not amplify the placebo effect, as was hypothesised by Kirsch (2014) and has been demonstrated by Rief and Glombiewski (2012), but that the lack of experiencing any distinct side effects might cause nocebo effects in conventional lactose placebos, that diminish the placebo response. This was because there was no overall placebo effect in Study 2, with the conventional placebo actually appearing to produce less improvement in ISI than no treatment. Study 3 attempted to further improve the understanding of message framing manipulations in the context of an active placebo but failed to elicit statistically significant effects of message framing on the placebo response and burden caused by side effects. Although an exploratory analysis of participants who were able to correctly recognise their message framing showed an overall statistically significant

placebo effect and a power analysis demonstrated that the overall placebo effect had been statistically significant had the desired sample size not been limited due to the SARS-CoV-2 pandemic, none of the main analysis reached the a priori level for statistical significance. Although, no final conclusion could be formulated about the effects of message framing from Study 3, the power analysis indicated that the message framing effect size was small with a Cohen's $d = 0.29$, meaning that larger scale studies with multiple hundreds of participants should be able to observe statistically significant message framing effects. This further implies that message framing effects should not be underestimated in clinical practice where potentially many people are exposed to any influential information. Because prior research had demonstrated a small to medium placebo effect for sleep, a pooled analysis across all three experimental studies of this thesis did not find a statistically significant placebo effect as assessed by the Insomnia Severity Index.

Although there appeared to be no statistically significant overall placebo effect for sleep observable in the current thesis, it did appear that side effects elicited by active placebo might have an influence on the size of the placebo effect in at least some circumstances, e.g., in the context of a double-blind RCT as in Study 2. Given that ClinicalTrials.gov currently lists more than 300,000 registered investigational studies, including millions of patients (ClinicalTrials.gov, 2021), if the side effects disproportionately affect patients' perceived treatment allocation, hence enhance the placebo effect in drug arms of even as few as 5-10% of trials, then this could affect many decisions regarding the licencing and use of pharmacological treatments. This in turn might have cost fortunes in follow up studies that failed to replicate the original findings or might have affected millions of patients exposed to insufficiently effective medications causing further harm or even causalities. Therefore, even if side effects might only have a small effect amplifying the efficacy of the active treatment

under investigation compared to the inactive placebo control, then they could still significantly impact clinical research and practice.

Due to the lack of a statistically clearly significant overall placebo effect in the experimental studies it might be important for clinical research and practice that placebo effects for sleep might not be as reliable and large as assumed in the past. While Yeung et al. (2018)'s meta-analysis found an overall placebo effect for self-reported sleep quality of moderate effect size with a Hedges' *d* of 0.58, the current study indicated only a small to moderate effect size of Cohen's $d = 0.44$. This findings might indicate that the aforementioned meta-analysis, that was based on only 10 studies for sleep quality might have overestimated placebo effects for subjective sleep quality. Especially because the other subjective sleep outcomes, sleep onset latency and total sleep time that were reported in the meta-analysis also showed placebo effects that were only of small to moderate effect size. These findings indicate that overall placebo effects might not be as robust and large for sleep as had been assumed so far. The findings from this thesis imply that a placebo effect of small to moderate effect sizes in sleep might be a more realistic estimation and therefore be of similar magnitude as had been observed in prior research in the area of depression and pain research that also showed small to moderate placebo effects (Ashar et al., 2017; Khan et al., 2005).

Limitations

This thesis does not come without limitations. The meta-analysis still showed large heterogeneity, even after trying to account for it using moderator and sensitivity analyses. This means there might be undetected factors that are driving the observed effects. A further problem with the systematic review and meta-analysis was that many of the included RCTs

had either methodological problems, or poorly assessed and reported their findings. This might have increased the risk of bias and impacted the quality of evidence presented in this thesis' meta-analysis.

Specific limitations for each of the experimental studies have been mentioned in the relevant chapters. However, there are some more general limitations worth mentioning here. First of all, there are some aspects about the generalisability of the research conducted for this thesis that should be noted. Although the studies within this thesis tried to mimic clinical practice as well as possible, there are always translational gaps between research and the reality encountered in practice. One of the potential factors contributing to problems with generalisability can often be observed in the samples that sign up to participate for studies. Especially in a country like Australia that has a universal health care system, participants suffering from health issues that decide to sign up for research studies might be classified as a subsample of all patients because they either were not able to find a solution for their health problems within the health care system, or alternatively they might be more research affine. Another more general aspect impacting on the generalisability of this thesis might be the experimenters responsible for conducting the studies. On one hand the experimenters seeing participants were all young adults under the age of 30 years, whereas most of the health-care personnel typically encountered by patients in real-world practice might be older, more experienced in handling patients, and therefore might be more credible. Adding to a potential credibility issue might have been the fact that the Colagiuri lab is increasingly well known for placebo research. Therefore, participants curious enough to research the involved experimenters and researchers might have found out that they were actually participating in placebo research. Even if they might have just had doubts about the true purpose of the study they participated in, this uncertainty might have undermined their expectations, hence impacted their placebo effect.

A further aspect that needs to be discussed in terms of limitations of the current thesis is the fact that the new active placebo model only elicited the target side effect of redder urine or stool colouration in about 50% of participants. Additionally, to only half the sample experiencing the target side effect, this symptom did not actually cause a burden to participants, because it only resulted in a noticeable, but harmless reddish colouration of urine or stool. It might therefore be possible the observed effects of the active placebo might not have been as noticeable or poignant than other side effects that actually burden patients in real clinical settings like headaches, nausea, or daytime drowsiness that are commonly experienced by insomnia patients.

When considering the validity of experimental research in general, it is important to consider if the manipulation of the independent variable was causing the changes on the dependent variables, or if the differences in between-group designs might have been biased by confounding factors. The experimental approaches in this thesis tried to eliminate as many biasing factors as possible using elaborated experimental designs as well as existing and proven experimental methods to test the hypotheses. Nevertheless, one might question whether for example the instructions used to elicit different expectations between groups were the most effective available. This applies particularly to the message framing manipulations used in Study 3. In the absence of a pilot study showing that the message framing manipulations had the desired effects and the fact that a majority of participants could not give an active recall of the message framing manipulation, findings have to be interpreted carefully. Therefore, the findings of this thesis should be interpreted carefully within the context and the aims of this thesis.

The last and potentially most important limitation of the current thesis might be that placebo effects observed in sleep might not be as robust and reliable as observed in prior research. When discussing effect sizes, it is always important to distinguish if an effect is

statistically significant, or if it is in fact as well clinically significant. While it might be possible to find a statistically significant effect with sufficiently large sample sizes this does not have to imply that it is clinically significant, meaning that a single patient is able to experience the effect.

Future directions

The findings from this thesis point towards some important directions for future research. Because of the lack of a statistically significant overall placebo effect for sleep observed in this thesis, it is suggested that a sufficiently large clinical trial is conducted testing the placebo effect in a sample of patients suffering from sleep problems. Such a trial assessing multiple sleep relevant outcomes would be able to demonstrate if the placebo effect for sleep is only present on certain outcome domains in experimental research, as was demonstrated by prior placebo research in sleep that only found statistically significant placebo effects on certain outcomes like the ISI or the PSQI (Neukirch & Colagiuri, 2015; Yeung et al., 2020) or if the improvements are generalisable across multiple domains of sleep and if the outcomes are clinically relevant, meaning the improvements caused by placebos are meaningful to a patient experiencing sleep problems. A research avenue that might be of particular interest would be a large-scale RCT comparing the new active beetroot placebo, the previously used atropine as pharmacological active placebo, an inert lactose placebo, and no treatment condition. A RCT like that could provide evidence as to how well the different placebos compare against each other regarding the placebo effect and participants perceived treatment.

A direct comparison of these different placebos would then allow to decide whether it might be worth further exploring how these findings translate to other conditions outside sleep research. If the findings showed that the active placebo indeed amplified the placebo

effect, then a future research direction might be to re-test some of the antidepressant treatments versus an active placebo that according to Kirsch (2014) only performed statistically better than a conventional lactose placebo.

As was highlighted above, even if only 5-10% of clinical trials are affected by an imbalance in side effects between the investigational drug and an inert placebo it is important from an ethical and financial point of view to further investigate the influence of side effects via active placebos. In line with the argumentation of this thesis, Jensen et al. (2017) argued that active placebos are methodological tools that merit serious consideration in clinical trials. They further argued that active placebos should be used especially if the medication to be tested elicits many noticeable side effects and for conditions where most observed effect sizes of treatments were already limited. Kirsch and Sapirstein (1999) found that the placebo response made up 75% of modern antidepressants' treatment response. Most often the antidepressants only statistically outperformed the conventional placebo but did not reach clinically significant improvement on top of the placebo response. Modern serotonergic antidepressants are highly prescribed with a point prevalence of up to 3.5% of the general population in France (Olié et al., 2002). It is therefore suggested that these modern antidepressants should be re-evaluated against an active placebo. This investigation would show if modern antidepressants actually had superior efficacy compared to a conventional placebo due to the beneficial actions of the pharmacological ingredient, or if modern antidepressants only outperformed conventional placebo because their side effects enhanced the placebo response.

Conclusions

In conclusion, the findings from this thesis indicate that double-blind placebo-controlled randomised trials, which are currently the gold standard to evaluate new

pharmacological treatments, could at least under certain circumstances suffer from methodological biases introduced by an imbalance of side effects between the inert lactose placebo and the active investigational drug. Given the large number of clinical trials conducted each year and the millions of patients relying on drugs that were licenced using these trials, it is crucial to guarantee that the clinical trial methodology is constantly tested and updated to the newest insights from research. This is especially relevant because a lack of quality in clinical trials might cause large human and financial costs it that it could result in insufficiently effective medications being licenced and prescribed to a wide patient sample. Further, inadequately assessed side effects that are then presented to patients via medication leaflets or clinician communication might further burden patients via nocebo effects potentially negatively impacting patients' expectations about the effectiveness of a treatment or introduce unnecessary side effects.

Therefore, researchers conducting RCTs should be aware that evaluating a new medication against a conventional placebo might potentially overestimate the drug-placebo difference and bias the side effect assessment. Where possible, adopting an active placebo comparator would be the best way to mitigate such effects. However, even where that is not possible, it is recommended researchers conducting clinical trials assess participants' daily symptoms more systematically across the full study duration and might additionally assess participants' perceived treatment allocation once the treatment period is finished. This would allow for correlational assessment of the relationship between side effects or perceived treatment allocation and outcomes. To better understand how patients' beliefs about side effects influence the treatment response and the experience of side effects, it is further suggested that future studies build on this thesis' message framing results to investigate if the effects of message framing are only of relevance in short-term experimental studies or if

message framing effects might be of sufficient magnitude to actually impact patients' health and well-being in clinical practice.

References

Adams, R. J., Appleton, S. L., Taylor, A. W., Gill, T. K., Lang, C., McEvoy, R. D., & Antic, N. A. (2017). Sleep health of Australian adults in 2016: results of the 2016 Sleep Health Foundation national survey. *Sleep health, 3*(1), 35-42.

Ader, R. (2001). Much ado about nothing. *Adv Mind Body Med, 17*(4), 293-295; discussion 312-298. https://www.ncbi.nlm.nih.gov/pubmed/11931053

Ader, R. (2003). Conditioned immunomodulation: research needs and directions. *Brain Behav Immun, 17 Suppl 1*, S51-57. https://doi.org/10.1016/s0889-1591(02)00067-3

Ader, R., & Cohen, N. (1982). Behaviorally conditioned immunosuppression and murine systemic lupus erythematosus. *Science, 215*(4539), 1534-1536. https://doi.org/10.1126/science.7063864

Ader, R., Mercurio, M. G., Walton, J., James, D., Davis, M., Ojha, V., Kimball, A. B., & Fiorentino, D. (2010). Conditioned pharmacotherapeutic effects: a preliminary study. *Psychosom Med, 72*(2), 192-197. https://doi.org/10.1097/PSY.0b013e3181cbd38b

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-Practice Recommendations for Defining, Identifying, and Handling Outliers. *Organizational Research Methods, 16*(2), 270-301. https://doi.org/10.1177/1094428112470848

Altman, D. G., Schulz, K. F., & Moher, D. (2004). Turning a blind eye: testing the success of blinding and the CONSORT statement. *British Medical Journal, 328*(7448), 1135.

Amanzio, M., Corazzini, L. L., Vase, L., & Benedetti, F. (2009). A systematic review of adverse events in placebo groups of anti-migraine clinical trials. *Pain, 146*(3), 261-269. https://doi.org/10.1016/j.pain.2009.07.010

Amanzio, M., Pollo, A., Maggi, G., & Benedetti, F. (2001). Response variability to analgesics: a role for non-specific activation of endogenous opioids. *Pain, 90*(3), 205-215. https://doi.org/10.1016/S0304-3959(00)00486-3

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. https://doi.org/10.1176/appi.books.9780890425596

Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcroft, W., & Pollak, C. P. (2003). The role of actigraphy in the study of sleep and circadian rhythms. *Sleep, 26*(3), 342-392. https://doi.org/10.1093/sleep/26.3.342

Andrade, C. (2020). The Inconvenient Truth About Convenience and Purposive Samples. *Indian Journal of Psychological Medicine, 43*(1), 86-88. https://doi.org/10.1177/0253717620977000

Ashar, Y. K., Chang, L. J., & Wager, T. D. (2017). Brain Mechanisms of the Placebo Effect: An Affective Appraisal Account. *Annu Rev Clin Psychol, 13*, 73-98. https://doi.org/10.1146/annurev-clinpsy-021815-093015

Atkinson, M. J., Kumar, R., Cappelleri, J. C., & Hass, S. L. (2005). Hierarchical construct validity of the treatment satisfaction questionnaire for medication (TSQM version II) among outpatient pharmacy consumers. *Value Health, 8 Suppl 1*, S9-S24. https://doi.org/10.1111/j.1524-4733.2005.00066.x

Australian and New Zealand Clinical Trials Registry. (2018a). Identifier ACTRN12618001493235. Do active placebos enhance the placebo effect for sleep compared with benign placebos? A double-blind randomised controlled trial in healthy adults. http://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=375904

Australian and New Zealand Clinical Trials Registry. (2018b). Identifier ACTRN12618002048268. Do active placebos influence perceived treatment allocation and enhance the placebo effect for sleep compared with benign placebos? A double-blind randomised controlled trial in sleep-impaired adults. http://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=376616

Australian and New Zealand Clinical Trials Registry. (2020). Identifier ACTRN12620000232932. How does positive and negative message framing affect the placebo effect in a model of active placebos for sleep difficulty? A randomised controlled study in sleep-impaired adults. http://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=379270

Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., & Andrews, T. J. (2021). Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. *Psychological Methods, 26*(3), 295-314. https://doi.org/10.1037/met0000337

Barnes, K., Faasse, K., Geers, A. L., Helfer, S. G., Sharpe, L., Colloca, L., & Colagiuri, B. (2019). Can Positive Framing Reduce Nocebo Side Effects? Current Evidence and Recommendation for Future Research. *Front Pharmacol, 10*, 167. https://doi.org/10.3389/fphar.2019.00167

Barsky, A. J., Saintfort, R., Rogers, M. P., & Borus, J. F. (2002). Nonspecific medication side effects and the nocebo phenomenon. *JAMA, 287*(5), 622-627. https://doi.org/10.1001/jama.287.5.622

Bartels, D. J., van Laarhoven, A. I., Haverkamp, E. A., Wilder-Smith, O. H., Donders, A. R., van Middendorp, H., van de Kerkhof, P. C., & Evers, A. W. (2014). Role of conditioning and verbal suggestion in placebo and nocebo effects on itch. *PLoS One, 9*(3), e91727. https://doi.org/10.1371/journal.pone.0091727

Barter, G., & Cormack, M. (1996). The long-term use of benzodiazepines: patients' views, accounts and experiences. *Fam Pract, 13*(6), 491-497. https://doi.org/10.1093/fampra/13.6.491

Bastien, C. H., Vallieres, A., & Morin, C. M. (2001). Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Med, 2*(4), 297-307. https://doi.org/10.1016/s1389-9457(00)00065-4

Beecher, H. K. (1955). The powerful placebo. *J Am Med Assoc, 159*(17), 1602-1606. https://doi.org/10.1001/jama.1955.02960340022006

Belanger, L., Vallieres, A., Ivers, H., Moreau, V., Lavigne, G., & Morin, C. M. (2007). Meta-analysis of sleep changes in control groups of insomnia treatment trials. *J Sleep Res, 16*(1), 77-84. https://doi.org/10.1111/j.1365-2869.2007.00566.x

Bell, M. L., Kenward, M. G., Fairclough, D. L., & Horton, N. J. (2013). Differential dropout and bias in randomised controlled trials: when it matters and when it may not. *BMJ : British Medical Journal, 346*, e8668. https://doi.org/10.1136/bmj.e8668

Bello, S., Moustgaard, H., & Hróbjartsson, A. (2017). Unreported formal assessment of unblinding occurred in 4 of 10 randomized clinical trials, unreported loss of blinding in 1 of 10 trials. *Journal of Clinical Epidemiology, 81*, 42-50. https://doi.org/https://doi.org/10.1016/j.jclinepi.2016.08.002

Benedetti, F., Colloca, L., Torre, E., Lanotte, M., Melcarne, A., Pesare, M., Bergamasco, B., & Lopiano, L. (2004). Placebo-responsive Parkinson patients show decreased activity in single neurons of subthalamic nucleus. *Nat Neurosci, 7*(6), 587-588. https://doi.org/10.1038/nn1250

Benedetti, F., Lanotte, M., Lopiano, L., & Colloca, L. (2007). When words are painful: unraveling the mechanisms of the nocebo effect. *Neuroscience, 147*(2), 260-271. https://doi.org/10.1016/j.neuroscience.2007.02.020

Berna, C., Kirsch, I., Zion, S. R., Lee, Y. C., Jensen, K. B., Sadler, P., Kaptchuk, T. J., & Edwards, R. R. (2017). Side effects can enhance treatment response through expectancy effects: an experimental analgesic randomized controlled trial. *Pain, 158*(6), 1014-1020. https://doi.org/10.1097/j.pain.0000000000000870

Bingel, U., Wanigasekera, V., Wiech, K., Ni Mhuircheartaigh, R., Lee, M. C., Ploner, M., & Tracey, I. (2011). The effect of treatment expectation on drug efficacy: imaging the analgesic benefit of the opioid remifentanil. *Sci Transl Med, 3*(70), 70ra14. https://doi.org/10.1126/scitranslmed.3001244

Blasini, M., Corsi, N., Klinger, R., & Colloca, L. (2017). Nocebo and pain: An overview of the psychoneurobiological mechanisms. *Pain Rep, 2*(2). https://doi.org/10.1097/PR9.0000000000000585

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to Meta-analysis*. John Wiley & Sons.

Brown, E. G., Wood, L., & Wood, S. (1999). The medical dictionary for regulatory activities (MedDRA). *Drug Safety*, *20*(2), 109-117. https://doi.org/10.2165/00002018-199920020-00002

Brzezinski, A., Vangel, M. G., Wurtman, R. J., Norrie, G., Zhdanova, I., Ben-Shushan, A., & Ford, I. (2005). Effects of exogenous melatonin on sleep: a meta-analysis. *Sleep Med Rev, 9*(1), 41-50. https://doi.org/10.1016/j.smrv.2004.06.004

Buckalew, L. W., & Coffield, K. E. (1982). An investigation of drug expectancy as a function of capsule color and size and preparation form. *J Clin Psychopharmacol, 2*(4), 245-248. https://www.ncbi.nlm.nih.gov/pubmed/7119132

Buscemi, N., Vandermeer, B., Friesen, C., Bialy, L., Tubman, M., Ospina, M., Klassen, T. P., Witmans, M., & Ospina, M. (2007). The efficacy and safety of drug treatments for chronic insomnia in adults: A meta-analysis of RCTs [Peer Reviewed]. *Journal of General Internal Medicine, .22*(9), pp. https://doi.org/10.1007/s11606-007-0251-z 17619935

Buysse, D. J., Reynolds, C. F., 3rd, Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res, 28*(2), 193-213. https://doi.org/10.1016/0165-1781(89)90047-4

Bystritsky, A., & Waikar, S. V. (1994). Inert placebo versus active medication. Patient blindability in clinical pharmacological trials. *J Nerv Ment Dis, 182*(9), 485-487. https://doi.org/10.1097/00005053-199409000-00001

Carney, C. E., Buysse, D. J., Ancoli-Israel, S., Edinger, J. D., Krystal, A. D., Lichstein, K. L., & Morin, C. M. (2012). The consensus sleep diary: standardizing prospective sleep self-monitoring. *Sleep, 35*(2), 287-302. https://doi.org/10.5665/sleep.1642

Castes, M., Palenque, M., Canelones, P., Hagel, I., & Lynch, N. (1998). Classic conditioning and placebo effects in the bronchodilator response of asthmatic children. *Neuroimmunomodulation, 5*(70), 514.

Champely, S. (2018). *pwr: Basic Functions for Power Analysis.* In https://CRAN.R-project.org/package=pwr

Cheung, J. M. Y., Bartlett, D. J., Armour, C. L., Laba, T.-L., & Saini, B. (2018). To Drug or Not to Drug: A Qualitative Study of Patients' Decision-Making Processes for Managing Insomnia. *Behavioral Sleep Medicine, 16*(1), 1-26. https://doi.org/10.1080/15402002.2016.1163702

Chong, Y., Fryer, C. D., & Gu, Q. (2013). Prescription sleep aid use among adults: United States, 2005-2010. *NCHS Data Brief*(127), 1-8. https://www.ncbi.nlm.nih.gov/pubmed/24152538

ClinicalTrials.gov. (2021). Retrieved 30.08.2021 from https://clinicaltrials.gov/ct2/resources/trends

Cohen, D., & Jacobs, D. H. (2007). Randomized controlled trials of antidepressants: clinically and scientifically irrelevant. *Debates in Neuroscience, 1*(1), 44-54. https://doi.org/10.1007/s11559-007-9002-x

Colagiuri, B. (2010). Participant expectancies in double-blind randomized placebo-controlled trials: potential limitations to trial validity. *Clin Trials, 7*(3), 246-255. https://doi.org/10.1177/1740774510367916

Colagiuri, B., & Boakes, R. A. (2010). Perceived treatment, feedback, and placebo effects in double-blind RCTs: an experimental analysis. *Psychopharmacology, 208*(3), 433-441. https://doi.org/10.1007/s00213-009-1743-9

Colagiuri, B., McGuinness, K., Boakes, R. A., & Butow, P. N. (2012). Warning about side effects can increase their occurrence: an experimental model using placebo treatment for sleep difficulty. *J Psychopharmacol, 26*(12), 1540-1547. https://doi.org/10.1177/0269881112458730

Colagiuri, B., Quinn, V. F., & Colloca, L. (2015). Nocebo Hyperalgesia, Partial Reinforcement, and Extinction. *J Pain, 16*(10), 995-1004. https://doi.org/10.1016/j.jpain.2015.06.012

Colagiuri, B., Sharpe, L., & Scott, A. (2019). The Blind Leading the Not-So-Blind: A Meta-Analysis of Blinding in Pharmacological Trials for Chronic Pain. *J Pain, 20*(5), 489-500. https://doi.org/10.1016/j.jpain.2018.09.002

Colloca, L. (2017). Treatment of Pediatric Migraine. *N Engl J Med, 376*(14), 1387-1388. https://doi.org/10.1056/NEJMc1701674

Colloca, L., & Barsky, A. J. (2020). Placebo and Nocebo Effects. *N Engl J Med, 382*(6), 554-561. https://doi.org/10.1056/NEJMra1907805

Colloca, L., & Benedetti, F. (2005). Placebos and painkillers: is mind as real as matter? *Nat Rev Neurosci, 6*(7), 545-552. https://doi.org/10.1038/nrn1705

Colloca, L., & Benedetti, F. (2006). How prior experience shapes placebo analgesia. *Pain, 124*(1-2), 126-133. https://doi.org/10.1016/j.pain.2006.04.005

Colloca, L., & Miller, F. G. (2011a). Harnessing the placebo effect: the need for translational research. *Philos Trans R Soc Lond B Biol Sci, 366*(1572), 1922-1930. https://doi.org/10.1098/rstb.2010.0399

Colloca, L., & Miller, F. G. (2011b). The nocebo effect and its relevance for clinical practice. *Psychosom Med, 73*(7), 598-603. https://doi.org/10.1097/PSY.0b013e3182294a50

Cooney, J. (1998). Active placebos in antidepressant trials. *The British Journal of Psychiatry, 172*(6), 541-542.

Cooper, V., Metcalf, L., Versnel, J., Upton, J., Walker, S., & Horne, R. (2015). Patient-reported side effects, concerns and adherence to corticosteroid treatment for asthma,

and comparison with physician estimates of side-effect prevalence: a UK-wide, cross-sectional study. *npj Primary Care Respiratory Medicine, 25*(1), 15026. https://doi.org/10.1038/npjpcrm.2015.26

Corsi, N., & Colloca, L. (2017). Placebo and Nocebo Effects: The Advantage of Measuring Expectations and Psychological Factors. *Front Psychol, 8*, 308. https://doi.org/10.3389/fpsyg.2017.00308

Crum, A., & Zuckerman, B. (2017). Changing Mindsets to Enhance Treatment Effectiveness. *JAMA, 317*(20), 2063-2064. https://doi.org/10.1001/jama.2017.4545

Crum, A. J., Corbin, W. R., Brownell, K. D., & Salovey, P. (2011). Mind over milkshakes: mindsets, not just nutrients, determine ghrelin response. *Health Psychol, 30*(4), 424-429; discussion 430-421. https://doi.org/10.1037/a0023467

Crum, A. J., & Langer, E. J. (2007). Mind-set matters: exercise and the placebo effect. *Psychol Sci, 18*(2), 165-171. https://doi.org/10.1111/j.1467-9280.2007.01867.x

Crum, A. J., Salovey, P., & Achor, S. (2013). Rethinking stress: the role of mindsets in determining the stress response. *J Pers Soc Psychol, 104*(4), 716-733. https://doi.org/10.1037/a0031201

Cuijpers, P. (2016). *Meta-analyses in mental health research. A practical guide*.

Cumpston, M., Li, T. J., Page, M. J., Chandler, J., Welch, V. A., Higgins, J. P. T., & Thomas, J. (2019). Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions. *Cochrane Database of Systematic Reviews*(10). https://doi.org/10.1002/14651858.Ed000142

Daneman, E. A. (1961). Imipramine in office management of depressive reactions. (A double blind clinical study). *Dis Nerv Syst, 22*, 213-217. https://www.ncbi.nlm.nih.gov/pubmed/13719570

de Craen, A. J., Roos, P. J., de Vries, A. L., & Kleijnen, J. (1996). Effect of colour of drugs: systematic review of perceived effect of drugs and of their effectiveness. *BMJ, 313*(7072), 1624-1626. https://doi.org/10.1136/bmj.313.7072.1624

de la Fuente-Fernandez, R., Ruth, T. J., Sossi, V., Schulzer, M., Calne, D. B., & Stoessl, A. J. (2001). Expectation and dopamine release: mechanism of the placebo effect in Parkinson's disease. *Science, 293*(5532), 1164-1166. https://doi.org/10.1126/science.1060937

de Souza, L., Benedito-Silva, A. A., Pires, M. L. N., Poyares, D., Tufik, S., & Calil, H. M. (2003). Further validation of actigraphy for sleep studies. *Sleep, 26*(1), 81-85.

Doering, B. K., & Rief, W. (2012). Utilizing placebo mechanisms for dose reduction in pharmacotherapy. *Trends Pharmacol Sci, 33*(3), 165-172. https://doi.org/10.1016/j.tips.2011.12.001

Dundar, Y., Dodd, S., Strobl, J., Boland, A., Dickson, R., & Walley, T. (2004). Comparative efficacy of newer hypnotic drugs for the short-term management of insomnia: a systematic review and meta-analysis. *Hum Psychopharmacol, 19*(5), 305-322. https://doi.org/10.1002/hup.594

Eastwood, M. A., & Nyhlin, H. (1995). Beeturia and colonic oxalic acid. *QJM, 88*(10), 711-717. https://www.ncbi.nlm.nih.gov/pubmed/7493168

Edwards, I. R., & Aronson, J. K. (2000). Adverse drug reactions: definitions, diagnosis, and management. *The Lancet, 356*(9237), 1255-1259. https://doi.org/https://doi.org/10.1016/S0140-6736(00)02799-9

EFSA Panel on Food Additives, & Nutrient Sources added to Food. (2015). Scientific Opinion on the re-evaluation of beetroot red (E 162) as a food additive. *EFSA Journal, 13*(12), 4318.

Enck, P., Klosterhalfen, S., & Weimer, K. (2017). Unsolved, forgotten, and ignored features of the placebo response in medicine. *Clinical therapeutics, 39*(3), 458-468.

Esatbeyoglu, T., Wagner, A. E., Schini-Kerth, V. B., & Rimbach, G. (2015). Betanin--a food colorant with biological activity. *Mol Nutr Food Res, 59*(1), 36-47. https://doi.org/10.1002/mnfr.201400484

Espay, A. J., Norris, M. M., Eliassen, J. C., Dwivedi, A., Smith, M. S., Banks, C., Allendorfer, J. B., Lang, A. E., Fleck, D. E., Linke, M. J., & Szaflarski, J. P. (2015). Placebo effect of medication cost in Parkinson disease: a randomized double-blind study. *Neurology, 84*(8), 794-802. https://doi.org/10.1212/WNL.0000000000001282

Faasse, K., Cundy, T., Gamble, G., & Petrie, K. J. (2013). The effect of an apparent change to a branded or generic medication on drug effectiveness and side effects. *Psychosom Med, 75*(1), 90-96. https://doi.org/10.1097/PSY.0b013e3182738826

Faasse, K., Huynh, A., Pearson, S., Geers, A. L., Helfer, S. G., & Colagiuri, B. (2019). The Influence of Side Effect Information Framing on Nocebo Effects. *Ann Behav Med, 53*(7), 621-629. https://doi.org/10.1093/abm/kay071

Fässler, M., Meissner, K., Kleijnen, J., Hrobjartsson, A., & Linde, K. (2015). A systematic review found no consistent difference in effect between more and less intensive placebo interventions. *J Clin Epidemiol, 68*(4), 442-451. https://doi.org/10.1016/j.jclinepi.2014.11.018

Fent, R., Rosemann, T., Fässler, M., Senn, O., & Huber, C. A. (2011). The use of pure and impure placebo interventions in primary care - a qualitative approach. *BMC Fam Pract, 12*, 11. https://doi.org/10.1186/1471-2296-12-11

Fergusson, D., Glass, K. C., Waring, D., & Shapiro, S. (2004). Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials. *BMJ, 328*(7437), 432. https://doi.org/10.1136/bmj.37952.631667.EE

Fernandez, A., Kirsch, I., Noël, L., Rodondi, P. Y., Kaptchuk, T. J., Suter, M. R., Décosterd, I., & Berna, C. (2019). A test of positive suggestions about side effects as a way of enhancing the analgesic response to NSAIDs. *PLoS One, 14*(1), e0209851. https://doi.org/10.1371/journal.pone.0209851

Ferracioli-Oda, E., Qawasmi, A., & Bloch, M. H. (2013). Meta-analysis: melatonin for the treatment of primary sleep disorders. *PLoS One, 8*(5), e63773. https://doi.org/10.1371/journal.pone.0063773

Figueiras, M. J., Marcelino, D., & Cortes, M. A. (2008). People's views on the level of agreement of generic medicines for different illnesses. *Pharm World Sci, 30*(5), 590-594. https://doi.org/10.1007/s11096-008-9247-y

Finegold, J. A., Manisty, C. H., Goldacre, B., Barron, A. J., & Francis, D. P. (2014). What proportion of symptomatic side effects in patients taking statins are genuinely caused by the drug? Systematic review of randomized placebo-controlled trials to aid individual patient choice. *Eur J Prev Cardiol, 21*(4), 464-474. https://doi.org/10.1177/2047487314525531

Fountoulakis, K. N., & Möller, H.-J. (2011). Efficacy of antidepressants: a re-analysis and re-interpretation of the Kirsch data. *International Journal of Neuropsychopharmacology, 14*(3), 405-412. https://doi.org/10.1017/S1461145710000957

Fu, R., Gartlehner, G., Grant, M., Shamliyan, T., Sedrakyan, A., Wilt, T. J., Griffith, L., Oremus, M., Raina, P., Ismaila, A., Santaguida, P., Lau, J., & Trikalinos, T. A. (2010). *Conducting Quantitative Synthesis When Comparing Medical Interventions: AHRQ and the Effective Health Care Program*. Agency for Healthcare Research and Quality. http://effectivehealthcare.ahrq.gov/

Gallagher, K. M., Updegraff, J. A., Rothman, A. J., & Sims, L. (2011). Perceived susceptibility to breast cancer moderates the effect of gain- and loss-framed messages on use of screening mammography. *Health Psychology, 30*(2), 145-152. https://doi.org/10.1037/a0022264

Gallardo, E. J., & Coggan, A. R. (2019). What's in Your Beet Juice? Nitrate and Nitrite Content of Beet Juice Products Marketed to Athletes. *Int J Sport Nutr Exerc Metab, 29*(4), 345-349. https://doi.org/10.1123/ijsnem.2018-0223

Geers, A. L., & Miller, F. G. (2014). Understanding and translating the knowledge about placebo effects: the contribution of psychology. *Curr Opin Psychiatry, 27*(5), 326-331. https://doi.org/10.1097/YCO.0000000000000082

Gitlin, M. (2016). Lithium side effects and toxicity: prevalence and management strategies. *Int J Bipolar Disord, 4*(1), 27. https://doi.org/10.1186/s40345-016-0068-y

Glare, P., Fridman, I., & Ashton-James, C. E. (2018). Choose Your Words Wisely: The Impact of Message Framing on Patients' Responses to Treatment Advice. *Int Rev Neurobiol, 139*, 159-190. https://doi.org/10.1016/bs.irn.2018.07.020

Glass, J., Lanctot, K. L., Herrmann, N., Sproule, B. A., & Busto, U. E. (2005). Sedative hypnotics in older people with insomnia: meta-analysis of risks and benefits. *BMJ, 331*(7526), 1169. https://doi.org/10.1136/bmj.38623.768588.47

Goebel, M. U., Trebst, A. E., Steiner, J., Xie, Y. F., Exton, M. S., Frede, S., Canbay, A. E., Michel, M. C., Heemann, U., & Schedlowski, M. (2002). Behavioral conditioning of immunosuppression is possible in humans [https://doi.org/10.1096/fj.02-0389com]. *The FASEB Journal, 16*(14), 1869-1873. https://doi.org/https://doi.org/10.1096/fj.02-0389com

Greene, P. J., Wayne, P. M., Kerr, C. E., Weiger, W. A., Jacobson, E., Goldman, P., & Kaptchuk, T. J. (2001). The powerful placebo: doubting the doubters. *Adv Mind Body Med, 17*(4), 298-307; discussion 312-298. https://www.ncbi.nlm.nih.gov/pubmed/11931055

Grünbaum, A. (1986). The placebo concept in medicine and psychiatry. *Psychol Med, 16*(1), 19-38. https://doi.org/10.1017/s0033291700002506

Gupta, A., Thompson, D., Whitehouse, A., Collier, T., Dahlof, B., Poulter, N., Collins, R., Sever, P., & Investigators, A. (2017). Adverse events associated with unblinded, but not with blinded, statin therapy in the Anglo-Scandinavian Cardiac Outcomes Trial-Lipid-Lowering Arm (ASCOT-LLA): a randomised double-blind placebo-controlled trial and its non-randomised non-blind extension phase. *Lancet, 389*(10088), 2473-2481. https://doi.org/10.1016/S0140-6736(17)31075-9

Habeck, C. W., & Schultz, A. K. (2015). Community-level impacts of white-tailed deer on understorey plants in North American forests: a meta-analysis. *Aob Plants, 7*. https://doi.org/10.1093/aobpla/plv119

Hackshaw, A. (2009). *A Concise Guide to Clinical Trials*. John Wiley & Sons Ltd.

Hariton, E., & Locascio, J. J. (2018). Randomised controlled trials - the gold standard for effectiveness research: Study design: randomised controlled trials. *BJOG, 125*(13), 1716. https://doi.org/10.1111/1471-0528.15199

Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., Duda, S. N., & Consortium, R. E. (2019). The REDCap consortium: Building an international community of software platform partners. *J Biomed Inform, 95*, 103208. https://doi.org/10.1016/j.jbi.2019.103208

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform, 42*(2), 377-381. https://doi.org/10.1016/j.jbi.2008.08.010

Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. *Br J Clin Psychol, 44*(Pt 2), 227-239. https://doi.org/10.1348/014466505X29657

Herrmann, E., Call, J., Hernandez-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: the cultural intelligence hypothesis. *Science, 317*(5843), 1360-1366. https://doi.org/10.1126/science.1146282

Herrnstein, R. J. (1962). Placebo effect in the rat. *Science, 138*(3541), 677-678. https://doi.org/10.1126/science.138.3541.677

Hess, B., Jost, C., Zipperle, L., Takkinen, R., & Jaeger, P. (1998). High-calcium intake abolishes hyperoxaluria and reduces urinary crystallization during a 20-fold normal oxalate load in humans. *Nephrol Dial Transplant, 13*(9), 2241-2247. https://doi.org/10.1093/ndt/13.9.2241

Hodgkinson, A., & Zarembski, P. M. (1968). Oxalic acid metabolism in man: a review. *Calcif Tissue Res, 2*(2), 115-132. https://doi.org/10.1007/BF02279201

Holbrook, A. M., Crowther, R., Lotter, A., Cheng, C., & King, D. (2000). Meta-analysis of benzodiazepine use in the treatment of insomnia. *CMAJ, 162*(2), 225-233. https://www.ncbi.nlm.nih.gov/pubmed/10674059

Holland, B., Welch, A., Unwin, I., Buss, D., Paul, A., & Southgate, D. (1991). *McCance and Widdowson's the composition of foods*. Royal Society of Chemistry.

Horne, R., Faasse, K., Cooper, V., Diefenbach, M. A., Leventhal, H., Leventhal, E., & Petrie, K. J. (2012). The perceived sensitivity to medicines (PSM) scale: an evaluation of validity and reliability. *Br J Health Psychol, 18*(1), 18-30. https://doi.org/10.1111/j.2044-8287.2012.02071.x

Howick, J. (2009). Questioning the methodologic superiority of 'placebo' over 'active' controlled trials. *Am J Bioeth, 9*(9), 34-48. https://doi.org/10.1080/15265160903090041

Hróbjartsson, A., Forfang, E., Haahr, M. T., Als-Nielsen, B., & Brorson, S. (2007). Blinded trials taken to the test: an analysis of randomized clinical trials that report tests for the success of blinding. *Int J Epidemiol, 36*(3), 654-663. https://doi.org/10.1093/ije/dym020

Hróbjartsson, A., & Gøtzsche, P. C. (2001). Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *N Engl J Med, 344*(21), 1594-1602. https://doi.org/10.1056/NEJM200105243442106

Hróbjartsson, A., & Gøtzsche, P. C. (2003). Unreliable analysis of placebo analgesia in trials of placebo pain mechanisms. *Pain, 104*(3), 714-715. https://doi.org/10.1016/S0304-3959(03)00164-7

Hróbjartsson, A., & Gøtzsche, P. C. (2004). Is the placebo powerless? Update of a systematic review with 52 new randomized trials comparing placebo with no treatment. *J Intern Med, 256*(2), 91-100. https://doi.org/10.1111/j.1365-2796.2004.01355.x

Hróbjartsson, A., & Gøtzsche, P. C. (2010). Placebo interventions for all clinical conditions. *Cochrane Database Syst Rev*(1), CD003974. https://doi.org/10.1002/14651858.CD003974.pub3

Iliffe, S., Curran, H., Collins, R., Yuen Kee, S., Fletcher, S., & Woods, B. (2004). Attitudes to long-term use of benzodiazepine hypnotics by older people in general practice: findings from interviews with service users and providers. *Aging & mental health, 8*(3), 242-248.

Janssens, T., Meulders, A., Cuyvers, B., Colloca, L., & Vlaeyen, J. W. S. (2019). Placebo and nocebo effects and operant pain-related avoidance learning. *Pain Rep, 4*(3), e748. https://doi.org/10.1097/PR9.0000000000000748

Jefferson, T. (1807). Thomas Jefferson to Caspar Wistar. In P. L. Ford (Ed.), *The Works of Thomas Jefferson in Twelve Volumes*.

Jensen, J. S., Bielefeldt, A. O., & Hróbjartsson, A. (2017). Active placebo control groups of pharmacological interventions were rarely used but merited serious consideration: a methodological overview. *J Clin Epidemiol, 87*, 35-46. https://doi.org/10.1016/j.jclinepi.2017.03.001

Johansen, O., Brox, J., & Flaten, M. A. (2003). Placebo and Nocebo responses, cortisol, and circulating beta-endorphin. *Psychosom Med, 65*(5), 786-790. https://doi.org/10.1097/01.psy.0000082626.56217.cf

Jonas, W. B., Crawford, C., Colloca, L., Kaptchuk, T. J., Moseley, B., Miller, F. G., Kriston, L., Linde, K., & Meissner, K. (2015). To what extent are surgery and invasive procedures effective beyond a placebo response? A systematic review with meta-analysis of randomised, sham controlled trials. *BMJ Open, 5*(12), e009655. https://doi.org/10.1136/bmjopen-2015-009655

Jütte, R. (2013). The early history of the placebo. *Complement Ther Med, 21*(2), 94-97. https://doi.org/10.1016/j.ctim.2012.06.002

Kahneman, D., & Tversky, A. (1979). On the interpretation of intuitive probability: A reply to Jonathan Cohen. *Cognition, 7*(4), 409-411. https://doi.org/https://doi.org/10.1016/0010-0277(79)90024-6

Kapadia, G. J., & Rao, G. S. (2013). Anticancer effects of red beet pigments. In *Red Beet Biotechnology* (pp. 125-154). Springer.

Kaptchuk, T. J. (2001). The double-blind, randomized, placebo-controlled trial: gold standard or golden calf? *J Clin Epidemiol, 54*(6), 541-549. https://doi.org/10.1016/s0895-4356(00)00347-4

Karanicolas, P. J., Farrokhyar, F., & Bhandari, M. (2010). Practical tips for surgical research: blinding: who, what, when, why, how? *Canadian journal of surgery. Journal canadien de chirurgie, 53*(5), 345-348. https://pubmed.ncbi.nlm.nih.gov/20858381 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2947122/

Kemmler, G., Hummer, M., Widschwendter, C., & Fleischhacker, W. W. (2005). Dropout rates in placebo-controlled and active-control clinical trials of antipsychotic drugs: a meta-analysis. *Arch Gen Psychiatry, 62*(12), 1305-1312. https://doi.org/10.1001/archpsyc.62.12.1305

Khan, A., Kolts, R. L., Rapaport, M. H., Krishnan, K. R., Brodhead, A. E., & Browns, W. A. (2005). Magnitude of placebo response and drug-placebo differences across psychiatric disorders. *Psychol Med, 35*(5), 743-749. https://doi.org/10.1017/s0033291704003873

Kienle, G. S., & Kiene, H. (1997). The powerful placebo effect: fact or fiction? *J Clin Epidemiol, 50*(12), 1311-1318. https://doi.org/10.1016/s0895-4356(97)00203-5

Kirsch, I. (1985). Response expectancy as a determinant of experience and behavior. *American Psychologist, 40*(11), 1189.

Kirsch, I. (1997). Response expectancy theory and application: A decennial review. *Applied and preventive Psychology, 6*(2), 69-79.

Kirsch, I. (2013). The placebo effect revisited: lessons learned to date. *Complement Ther Med, 21*(2), 102-104. https://doi.org/10.1016/j.ctim.2012.12.003

Kirsch, I. (2014). Antidepressants and the Placebo Effect. *Z Psychol, 222*(3), 128-134. https://doi.org/10.1027/2151-2604/a000176

Kirsch, I. (2018). Response Expectancy and the Placebo Effect. *Int Rev Neurobiol, 138*, 81-93. https://doi.org/10.1016/bs.irn.2018.01.003

Kirsch, I., Deacon, B. J., Huedo-Medina, T. B., Scoboria, A., Moore, T. J., & Johnson, B. T. (2008). Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med, 5*(2), e45. https://doi.org/10.1371/journal.pmed.0050045

Kirsch, I., & Sapirstein, G. (1999). Listening to Prozac but hearing placebo: A meta-analysis of antidepressant medications.

Kirsch, I., & Scoboria, A. (2001). Apples, oranges, and placebos: heterogeneity in a meta-analysis of placebo effects. *Adv Mind Body Med, 17*(4), 307-309; discussion 312-308. https://www.ncbi.nlm.nih.gov/pubmed/11931056

Kock, N. G., Darle, N., Hulten, L., Kewenter, J., Myrvold, H., & Philipson, B. (1977). Ileostomy. *Curr Probl Surg, 14*(8), 1-52. https://doi.org/10.1016/s0011-3840(77)80065-8

Küller, R., Ballal, S., Laike, T., Mikellides, B., & Tonello, G. (2006). The impact of light and colour on psychological mood: a cross-cultural study of indoor work environments. *Ergonomics, 49*(14), 1496-1507. https://doi.org/10.1080/00140130600858142

Laferton, J. A., Kube, T., Salzmann, S., Auer, C. J., & Shedden-Mora, M. C. (2017). Patients' Expectations Regarding Medical Treatment: A Critical Review of Concepts and Their Assessment. *Front Psychol, 8*, 233. https://doi.org/10.3389/fpsyg.2017.00233

Laferton, J. A. C., Vijapura, S., Baer, L., Clain, A. J., Cooper, A., Papakostas, G., Price, L. H., Carpenter, L. L., Tyrka, A. R., Fava, M., & Mischoulon, D. (2018). Mechanisms of Perceived Treatment Assignment and Subsequent Expectancy Effects in a Double Blind Placebo Controlled RCT of Major Depression [Original Research]. *Frontiers in Psychiatry, 9*(424). https://doi.org/10.3389/fpsyt.2018.00424

Langer, G. (1987). [The placebo: beyond pretense and the nuisance variable. Arguments in favor of re-evaluating a significant protherapeutic concept ("aura curae")]. *Wiener klinische Wochenschrift. Supplementum, 175*, 1-20. http://europepmc.org/abstract/MED/3318149

Lengen, C. (2015). The effects of colours, shapes and boundaries of landscapes on perception, emotion and mentalising processes promoting health and well-being. *Health Place, 35*, 166-177. https://doi.org/10.1016/j.healthplace.2015.05.016

Letemendia, F. J., & Harris, A. D. (1959). The influence of side-effects on the reporting of symptoms. *Psychopharmacologia, 1*, 39-47. https://doi.org/10.1007/BF00408110

Lie, J. D., Tu, K. N., Shen, D. D., & Wong, B. M. (2015). Pharmacological Treatment of Insomnia. *P T, 40*(11), 759-771. https://www.ncbi.nlm.nih.gov/pubmed/26609210

Lind, J. (1753). A treatise of the scurvy. *Three Parts. Containing an Inquiry into the Nature, Causes and Cure, of that Disease. Together with a Critical and Chronological View of what has been Published on the Subject*. https://www.jameslindlibrary.org/lind-j-1753/

Linde, K., Fässler, M., & Meissner, K. (2011). Placebo interventions, placebo effects and clinical practice. *Philos Trans R Soc Lond B Biol Sci, 366*(1572), 1905-1912. https://doi.org/10.1098/rstb.2010.0383

Lindström, D., Sundberg-Petersson, I., Adami, J., & Tönnesen, H. (2010). Disappointment and drop-out rate after being allocated to control group in a smoking cessation trial. *Contemporary Clinical Trials, 31*(1), 22-26. https://doi.org/https://doi.org/10.1016/j.cct.2009.09.003

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behav Res Ther, 33*(3), 335-343. https://doi.org/10.1016/0005-7967(94)00075-u

Maich, K. H. G., Lachowski, A. M., & Carney, C. E. (2018). Psychometric Properties of the Consensus Sleep Diary in Those With Insomnia Disorder. *Behav Sleep Med, 16*(2), 117-134. https://doi.org/10.1080/15402002.2016.1173556

McCall, W. V., D'Agostino, R., Jr., & Dunn, A. (2003). A meta-analysis of sleep changes associated with placebo in hypnotic clinical trials. *Sleep Med, 4*(1), 57-62. https://doi.org/10.1016/s1389-9457(02)00242-3

Meissner, K. (2011). The placebo effect and the autonomic nervous system: evidence for an intimate relationship. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366*(1572), 1808-1817.

Meissner, K., & Linde, K. (2018). Are Blue Pills Better Than Green? How Treatment Features Modulate Placebo Effects. *Int Rev Neurobiol, 139*, 357-378. https://doi.org/10.1016/bs.irn.2018.07.014

Merck & Co., Inc. (2021). *Zocor (simvastatin) [package insert]*. U.S. Food and Drug Administration website. Retrieved 18.04.2021 from https://www.accessdata.fda.gov/drugsatfda_docs/label/2012/019766s085lbl.pdf

Miller, F. G., & Brody, H. (2011). Understanding and harnessing placebo effects: clearing away the underbrush. *J Med Philos, 36*(1), 69-78. https://doi.org/10.1093/jmp/jhq061

Miller, F. G., & Rosenstein, D. L. (2006). The nature and power of the placebo effect. *J Clin Epidemiol, 59*(4), 331-335. https://doi.org/10.1016/j.jclinepi.2005.12.001

Moher, D., Liberati, A., Tetzlaff, J., Altmann, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med, 6*(7), e1000097. https://doi.org/doi:10.1371/journal.pmed.1000097.

Moher, D., Schulz, K. F., Altman, D. G., & Consort, G. (2001). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med, 134*(8), 657-662. https://doi.org/10.7326/0003-4819-134-8-200104170-00011

Moncrieff, J. (2001). Are antidepressants overrated? A review of methodological problems in antidepressant trials. *The Journal of nervous and mental disease, 189*(5), 288-295.

Mondaini, N., Gontero, P., Giubilei, G., Lombardi, G., Cai, T., Gavazzi, A., & Bartoletti, R. (2007). Finasteride 5 mg and sexual side effects: how many of these are related to a nocebo phenomenon? *J Sex Med, 4*(6), 1708-1712. https://doi.org/10.1111/j.1743-6109.2007.00563.x

Moseley, J. B., O'Malley, K., Petersen, N. J., Menke, T. J., Brody, B. A., Kuykendall, D. H., Hollingsworth, J. C., Ashton, C. M., & Wray, N. P. (2002). A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med, 347*(2), 81-88. https://doi.org/10.1056/NEJMoa013259

Murad, M. H., Asi, N., Alsawas, M., & Alahdab, F. (2016). New evidence pyramid. *Evid Based Med, 21*(4), 125-127. https://doi.org/10.1136/ebmed-2016-110401

Nakagawa, S., & Santos, E. S. A. (2012). Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology, 26*(5), 1253-1274. https://doi.org/10.1007/s10682-012-9555-5

Nakamura, Y., Donaldson, G. W., Kuhn, R., Bradshaw, D. H., Jacobson, R. C., & Chapman, R. C. (2012). Investigating dose-dependent effects of placebo analgesia: a psychophysiological approach. *Pain, 153*(1), 227-237. https://doi.org/10.1016/j.pain.2011.10.024

Nash, H. (1962). The double-blind procedure: rationale and empirical evaluation. *J Nerv Ment Dis, 134*, 34-47. https://www.ncbi.nlm.nih.gov/pubmed/14478428

Neukirch, N., & Colagiuri, B. (2015). The placebo effect, sleep difficulty, and side effects: a balanced placebo model. *J Behav Med, 38*(2), 273-283. https://doi.org/10.1007/s10865-014-9590-5

O'Connor, A. M., Pennie, R. A., & Dales, R. E. (1996). Framing effects on expectations, decisions, and side effects experienced: The case of influenza immunization. *Journal of Clinical Epidemiology, 49*(11), 1271-1276. https://doi.org/10.1016/S0895-4356(96)00177-1

Olié, J. P., Elomari, F., Spadone, C., & Lépine, J. P. (2002). Antidepressants consumption in the global population in France. *L'encephale, 28*(5 Pt 1), 411-417.

Oxford Centre for Evidence-Based Medicine. (2021). *Levels of Evidence*. Retrieved 12.04.2021 from https://www.cebm.ox.ac.uk/resources/levels-of-evidence/oxford-centre-for-evidence-based-medicine-levels-of-evidence-march-2009

Pavlov, I. P. (1927). *Conditioned reflexes; an investigation of the physiological activity of the cerebral cortex*. Oxford University Press: Humphrey Milford.

Peterson, R. A., & Merunka, D. R. (2014). Convenience samples of college students and research reproducibility. *Journal of Business Research, 67*(5), 1035-1041. https://doi.org/https://doi.org/10.1016/j.jbusres.2013.08.010

Petrie, K. J., Faasse, K., Crichton, F., & Grey, A. (2014). How common are symptoms? Evidence from a New Zealand national telephone survey. *BMJ Open, 4*(6), e005374. https://doi.org/10.1136/bmjopen-2014-005374

Petrie, K. J., & Rief, W. (2019). Psychobiological Mechanisms of Placebo and Nocebo Effects: Pathways to Improve Treatments and Reduce Side Effects. *Annu Rev Psychol, 70*, 599-625. https://doi.org/10.1146/annurev-psych-010418-102907

Pollo, A., Amanzio, M., Arslanian, A., Casadio, C., Maggi, G., & Benedetti, F. (2001). Response expectancies in placebo analgesia and their clinical relevance. *Pain, 93*(1), 77-84. https://doi.org/10.1016/S0304-3959(01)00296-2

Price, D. D., Riley III, L. J., & Vase, L. (2003). Reliable differences in placebo effects between clinical analgesic trials and studies of placebo analgesia mechanisms. *Pain, 104*(3), 715-716.

Quinn, V. F., & Colagiuri, B. (2016). Sources of Placebo-Induced Relief From Nausea: The Role of Instruction and Conditioning. *Psychosom Med, 78*(3), 365-372. https://doi.org/10.1097/PSY.0000000000000265

R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. In R Foundation for Statistical Computing. https://www.R-project.org/

R Core Team. (2019a). *A Language and Environment for Statistical Computing*. In http://www.R-project.org/

R Core Team. (2019b). *R: A Language and Environment for Statistical Computing*. In R Foundation for Statistical Computing. https://www.R-project.org/

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. In R Foundation for Statistical Computing. https://www.R-project.org/

Rescorla, R. A. (1988). Pavlovian conditioning. It's not what you think it is. *Am Psychol, 43*(3), 151-160. https://doi.org/10.1037//0003-066x.43.3.151

Rescorla, R. A. (1991). Associative relations in instrumental learning: The eighteenth Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology, 43*(1), 1-23.

Rief, W., Avorn, J., & Barsky, A. J. (2006). Medication-attributed adverse effects in placebo groups: implications for assessment of adverse effects. *Arch Intern Med, 166*(2), 155-160. https://doi.org/10.1001/archinte.166.2.155

Rief, W., Barsky, A. J., Glombiewski, J. A., Nestoriuc, Y., Glaesmer, H., & Braehler, E. (2011). Assessing general side effects in clinical trials: reference data from the general population. *Pharmacoepidemiol Drug Saf, 20*(4), 405-415. https://doi.org/10.1002/pds.2067

Rief, W., & Glombiewski, J. A. (2012). The hidden effects of blinded, placebo-controlled randomized trials: an experimental investigation. *Pain, 153*(12), 2473-2477. https://doi.org/10.1016/j.pain.2012.09.007

Rief, W., Nestoriuc, Y., von Lilienfeld-Toal, A., Dogan, I., Schreiber, F., Hofmann, S. G., Barsky, A. J., & Avorn, J. (2009). Differences in adverse effect reporting in placebo groups in SSRI and tricyclic antidepressant trials: a systematic review and meta-analysis. *Drug Saf, 32*(11), 1041-1056. https://doi.org/10.2165/11316580-000000000-00000

Riemenschneider, W., & Tanifuji, M. (2011). Oxalic Acid. Ullmann's Encyclopedia of Industrial Chemistry. Wiley, Germany.

Sackett, D. L. (2004). Turning a blind eye: why we don't test for blindness at the end of our trials. *British Medical Journal, 328*(7448), 1136.

Sadeh, A. (2011). The role and validity of actigraphy in sleep medicine: An update. *Sleep Medicine Reviews, 15*(4), 259-267. https://doi.org/https://doi.org/10.1016/j.smrv.2010.10.001

Sadeh, A., & Acebo, C. (2002). The role of actigraphy in sleep medicine. *Sleep Medicine Reviews, 6*(2), 113-124. https://doi.org/https://doi.org/10.1053/smrv.2001.0182

Salamone, J. D. (2000). A critique of recent studies on placebo effects of antidepressants: importance of research on active placebos. *Psychopharmacology (Berl), 152*(1), 1-6. https://doi.org/10.1007/s002130000509

Sandler, A. D., Glesne, C. E., & Bodfish, J. W. (2010). Conditioned placebo dose reduction: a new treatment in attention-deficit hyperactivity disorder? *J Dev Behav Pediatr, 31*(5), 369-375. https://doi.org/10.1097/DBP.0b013e3181e121ed

Schlarb, A. A., Classen, M., Grunwald, J., & Vogele, C. (2017). Sleep disturbances and mental strain in university students: results from an online survey in Luxembourg and Germany. *Int J Ment Health Syst, 11*, 24. https://doi.org/10.1186/s13033-017-0131-9

Schroll, J. B., Penninga, E. I., & Gøtzsche, P. C. (2016). Assessment of Adverse Events in Protocols, Clinical Study Reports, and Published Papers of Trials of Orlistat: A Document Analysis. *PLoS medicine, 13*(8), e1002101-e1002101. https://doi.org/10.1371/journal.pmed.1002101

Schünemann, H., Brożek, J., Guyatt, G., & Oxman, A. (2013). *GRADE handbook for grading quality of evidence and strength of recommendations*. https://gdt.gradepro.org/app/handbook/handbook.html

Shamloo, S., Irandoust, K., & Afif, A. H. (2019). The effect of beetroot juice supplementation on physiological fatigue and quality of sleep in male athletes. *Sleep and Hypnosis (Online), 21*(2), 97-100.

Shapiro, A. K. (1964). A Historic and Heuristic Definition of the Placebo. *Psychiatry, 27*, 52-58. https://doi.org/10.1080/00332747.1964.11023375

Siervo, M., Lara, J., Ogbonmwan, I., & Mathers, J. C. (2013). Inorganic nitrate and beetroot juice supplementation reduces blood pressure in adults: a systematic review and meta-analysis. *J Nutr, 143*(6), 818-826. https://doi.org/10.3945/jn.112.170233

Siriwardena, A. N., Qureshi, M. Z., Dyas, J. V., Middleton, H., & Orner, R. (2008). Magic bullets for insomnia? Patients' use and experiences of newer (Z drugs) versus older (benzodiazepine) hypnotics for sleep problems in primary care. *Br J Gen Pract, 58*(551), 417-422. https://doi.org/10.3399/bjgp08X299290

Skevington, S. M., Lotfy, M., O'Connell, K. A., & Group, W. (2004). The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties

and results of the international field trial. A report from the WHOQOL group. *Qual Life Res, 13*(2), 299-310. https://doi.org/10.1023/B:QURE.0000018486.91360.00

Smith, M. T., McCrae, C. S., Cheung, J., Martin, J. L., Harrod, C. G., Heald, J. L., & Carden, K. A. (2018). Use of Actigraphy for the Evaluation of Sleep Disorders and Circadian Rhythm Sleep-Wake Disorders: An American Academy of Sleep Medicine Systematic Review, Meta-Analysis, and GRADE Assessment. *Journal of Clinical Sleep Medicine, 14*(07), 1209-1230. https://doi.org/10.5664/jcsm.7228

Smith, M. T., Perlis, M. L., Park, A., Smith, M. S., Pennington, J., Giles, D. E., & Buysse, D. J. (2002). Comparative meta-analysis of pharmacotherapy and behavior therapy for persistent insomnia. *Am J Psychiatry, 159*(1), 5-11. https://doi.org/10.1176/appi.ajp.159.1.5

Spies, K., Hesse, F., & Loesch, K. (1997). Store atmosphere, mood and purchasing behavior. *International Journal of Research in Marketing, 14*(1), 1-17. https://doi.org/https://doi.org/10.1016/S0167-8116(96)00015-8

Stallone, F., Mendlewicz, J., & Fieve, R. R. (1975). Double-blind procedure: an assessment in a study of lithium prophylaxis. *Psychol Med, 5*(1), 78-82. https://doi.org/10.1017/s0033291700007248

Stephens, M. D. B. (1998). Definitions and classifications of adverse reaction terms. *The detection of new adverse reactions, 4th edn. London: Macmillan Reference*, 32-44.

Sterne, J. A. C., & Egger, M. (2005). Regression Methods to Detect Publication and Other Bias in Meta-Analysis. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, 99-110. https://doi.org/10.1002/0470870168

Stevens, J. P. (1984). Outliers and Influential Data Points in Regression-Analysis. *Psychological Bulletin, 95*(2), 334-344. https://doi.org/10.1037/0033-2909.95.2.334

Stinson, K., Tang, N. K., & Harvey, A. G. (2006). Barriers to treatment seeking in primary insomnia in the United Kingdom: a cross-sectional perspective. *Sleep, 29*(12), 1643-1646. https://doi.org/10.1093/sleep/29.12.1643

Straus, J. L., & von Ammon Cavanaugh, S. (1996). Placebo effects. Issues for clinical practice in psychiatry and medicine. *Psychosomatics, 37*(4), 315-326. https://doi.org/10.1016/S0033-3182(96)71544-X

Szabo, A., Jobbagy, L., & Koteles, F. (2018). Super Pill is Less Effective than an Ordinary Mint in Altering Subjective Psychological Feeling States within a Few Minutes. *J Gen Psychol, 145*(2), 208-222. https://doi.org/10.1080/00221309.2018.1459454

Taber, C. W. (1937). Taber's Digest of Medical Terms.

Tan, K., Petrie, K. J., Faasse, K., Bolland, M. J., & Grey, A. (2014). Unhelpful information about adverse drug reactions. *BMJ : British Medical Journal, 349*, g5019. https://doi.org/10.1136/bmj.g5019

Techniker Krankenkasse. (2017). *Schlaf gut, Deutschland*. T. Krankenkasse. https://www.tk.de/presse/themen/praevention/gesundheitsstudien/tk-schlafstudie-2017-2036638

The New Sydenham Society's Lexicon. (1899). The New Sydenham Society's lexicon of medicine and the allied sciences.

Therapeutic Goods Administration. (2011). Colourings Permitted in Medicines for Oral Use. https://www.tga.gov.au/publication/colourings-permitted-medicines-oral-use

Thomson, R. (1982). Side effects and placebo amplification. *Br J Psychiatry, 140*, 64-68. https://doi.org/10.1192/bjp.140.1.64

Timberlake, C. F., & Henry, B. S. (1988). Anthocyanins as natural food colorants. *Prog Clin Biol Res, 280*, 107-121. https://www.ncbi.nlm.nih.gov/pubmed/3051023

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: R Package for Causal Mediation Analysis. *Journal of Statistical Software, 1*(5). https://doi.org/10.18637/jss.v059.i05

Tough, E. A., White, A. R., Richards, S. H., Lord, B., & Campbell, J. L. (2009). Developing and validating a sham acupuncture needle. *Acupunct Med, 27*(3), 118-122. https://doi.org/10.1136/aim.2009.000737

Tryon, W. W. (2004). Issues of validity in actigraphic sleep assessment. *Sleep, 27*(1), 158-165.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453-458.

Twisk, J., Bosman, L., Hoekstra, T., Rijnhart, J., Welten, M., & Heymans, M. (2018). Different ways to estimate treatment effects in randomised controlled trials. *Contemporary Clinical Trials Communications, 10*, 80-85. https://doi.org/https://doi.org/10.1016/j.conctc.2018.03.008

U.S. Food and Drug Administration. (2020). *New Drug Application (NDA)*. Retrieved 28.12.2020 from https://www.fda.gov/drugs/types-applications/new-drug-application-nda

van Hees, V. T., Fang, Z., Langford, J., Assah, F., Mohammad, A., da Silva, I. C. M., Trenell, M. I., White, T., Wareham, N. J., & Brage, S. (2014). Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. *Journal of Applied Physiology, 117*(7), 738-744. https://doi.org/10.1152/japplphysiol.00421.2014

van Hees, V. T., Fang, Z., Zhao, J. H., Heywood, J., Mirkes, E., Sabia, S., & Migueles, J. H. (2021). *GGIR: Raw Accelerometer Data Analysis.* In https://CRAN.R-project.org/package=GGIR

van Hees, V. T., Sabia, S., Anderson, K. N., Denton, S. J., Oliver, J., Catt, M., Abell, J. G., Kivimäki, M., Trenell, M. I., & Singh-Manoux, A. (2015). A Novel, Open Access Method to Assess Sleep Duration Using a Wrist-Worn Accelerometer. *PLoS One, 10*(11), e0142533. https://doi.org/10.1371/journal.pone.0142533

Vase, L., Riley, J. L., 3rd, & Price, D. D. (2002). A comparison of placebo effects in clinical analgesic trials versus studies of placebo analgesia. *Pain, 99*(3), 443-452. https://doi.org/10.1016/S0304-3959(02)00205-1

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software, 36*(3), 1-48. <Go to ISI>://WOS:000281593200001

Voudouris, N. J., Peck, C. L., & Coleman, G. (1990). The role of conditioning and verbal expectancy in the placebo response. *Pain, 43*(1), 121-128. https://doi.org/10.1016/0304-3959(90)90057-K

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

Watson, W. C., Luke, R. G., & Inall, J. A. (1963). Beeturia: Its Incidence and a Clue to Its Mechanism. *Br Med J, 2*(5363), 971-973. https://doi.org/10.1136/bmj.2.5363.971

Weir, C. J., Butcher, I., Assi, V., Lewis, S. C., Murray, G. D., Langhorne, P., & Brady, M. C. (2018). Dealing with missing standard deviation and mean values in meta-analysis of continuous outcomes: a systematic review. *Bmc Medical Research Methodology, 18*. https://doi.org/ARTN 2510.1186/s12874-018-0483-0

Werner, C. P., Faasse, K., Sharpe, L., Holsteg, S., Hackett, M., Glozier, N., & Colagiuri, B. (2018). Do the adverse event profiles in placebo participants mimic the adverse event profiles of the drugs under investigation? A systematic review and meta-analysis of double-blind randomised placebo-controlled trials of pharmacological treatments in adults with insomnia. *PROSPERO: International prospective register of systematic reviews*. https://doi.org/10.15124/CRD42018097395

Wickramasekera, I. (1980). A conditioned response model of the placebo effect predictions from the model. *Biofeedback Self Regul, 5*(1), 5-18. https://doi.org/10.1007/BF00999060

Wicks, R. H. (2005). Message Framing and Constructing Meaning: An Emerging Paradigm in Mass Communication Research. *Annals of the International Communication Association, 29*(1), 335-362. https://doi.org/10.1080/23808985.2005.11679052

Wilhelm, M., Rief, W., & Doering, B. K. (2018). Decreasing the Burden of Side Effects Through Positive Message Framing: an Experimental Proof-of-Concept Study. *Int J Behav Med, 25*(4), 381-389. https://doi.org/10.1007/s12529-018-9726-z

Williams, V. S., Morlock, R. J., & Feltner, D. (2010). Psychometric evaluation of a visual analog scale for the assessment of anxiety. *Health Qual Life Outcomes, 8*, 57. https://doi.org/10.1186/1477-7525-8-57

Yang, M., Morin, C. M., Schaefer, K., & Wallenstein, G. V. (2009). Interpreting score differences in the Insomnia Severity Index: using health-related outcomes to define the minimally important difference. *Current medical research and opinion, 25*(10), 2487-2494.

Yeager, D. S., & Dweck, C. S. (2020). What can be learned from growth mindset controversies? *American Psychologist, 75*(9), 1269-1284. https://doi.org/10.1037/amp0000794

Yeung, V., Sharpe, L., Geers, A., & Colagiuri, B. (2020). Choice, Expectations, and the Placebo Effect for Sleep Difficulty. *Ann Behav Med, 54*(2), 94-107. https://doi.org/10.1093/abm/kaz030

Yeung, V., Sharpe, L., Glozier, N., Hackett, M. L., & Colagiuri, B. (2018). A systematic review and meta-analysis of placebo versus no treatment for insomnia symptoms. *Sleep Med Rev, 38*, 17-27. https://doi.org/10.1016/j.smrv.2017.03.006

Zhang, Z., & Yuan, K.-H. (2018). *Practical statistical power analysis using Webpower and R*. ISDSA Press.

Zion, S. R., & Crum, A. J. (2018). Mindsets Matter: A New Framework for Harnessing the Placebo Effect in Modern Medicine. *Int Rev Neurobiol, 138*, 137-160. https://doi.org/10.1016/bs.irn.2018.02.002

Zoom Video Communications Inc. (2020). *Zoom Cloud Meetings.* https://zoom.us/

Appendices

Appendix A – Additional Materials Literature Review and Meta-analysis

eTable 1

Detailed Literature Search Strategy

Database & Time Window	Search domain	Detailed Search Terms
MEDLINE (Ovid MEDLINE(R)) 1946 to 6th March 2020		
	Diagnosis	exp "Sleep Initiation and Maintenance Disorders"/
	Design	(((randomized controlled trial or controlled clinical trial).pt. or randomized.ab. or placebo.ab. or drug therapy fs. or randomly.ab. or trial.ab. or groups.ab.) not (exp animals/ not humans.sh.)) not (children or teenager* or adolescent* or animal* or mice or mouse or rat* or case report*) mp. not (review* or meta* or protocol* or cross*).ti.
	Intervention	exp Sleep Aids, Pharmaceutical/ or exp benzodiazepines/ or quazepam mp. or exp "Hypnotics and Sedatives"/ or zolpidem.mp. or zaleplon mp. or exp Eszopiclone/ or exp Melatonin/ or Ramelteon.mp. or exp Orexins/ or Suvorexant mp. or exp Antidepressive Agents/ or exp Barbiturates/ or Butabarbital.mp. or diphenhydramine/ or doxylamine/ or exp Trazodone/ or Mirtazapine.mp. or exp Antipsychotic Agents/ or Olanzapine.mp. or exp Valerian/ or exp Kava/
PsycINFO (Ovid PsycINFO) 1806 to 6th March 2020		
	Diagnosis	exp Insomnia/
	Design	(((randomized controlled trial or controlled clinical trial).pt. or randomized.ab. or placebo.ab. or drug therapy.ab. or randomly.ab. or trial.ab. or groups.ab.) not (exp animals/ not humans.sh.)) not (children or teenager* or adolescent* or animal* or mice or mouse or rat* or case report*) mp. not (review* or meta* or protocol* or cross*).ti.
	Intervention	exp benzodiazepines/ or quazepam mp. or exp HYPNOTIC DRUGS/ or exp SEDATIVES/ or zolpidem mp. or zaleplon mp. or eszopiclone.mp. or exp Melatonin/ or Ramelteon mp. or exp Orexin/ or Suvorexant mp. or exp Antidepressant Drugs/ or exp Barbiturates/ or Butabarbital.mp. or diphenhydramine/ or doxylamine/ or Doxylamine mp. or exp Trazodone/ or Mirtazapine mp. or exp Neuroleptic Drugs/ or Valerian mp. or Kava mp.
EMBASE (Ovid Embase) 1974 to 6th March 2020		
	Diagnosis	exp insomnia/
	Design	(crossover-procedure/ or double-blind procedure/ or randomized controlled trial/ or single-blind procedure/ or (random* or factorial*

		or crossover* or cross over* or placebo* or (doubl* adj blind*) or (singl* adj blind*) or assign* or allocat* or volunteer*).tw.) not (children or teenager* or adolescent* or animal* or mice or mouse or rat* or case report*) mp. not (review* or meta* or protocol* or cross*).ti.
	Intervention	exp Sleep Aids, Pharmaceutical/ or exp benzodiazepines/ or quazepam mp. or exp "Hypnotics and Sedatives"/ or zolpidem.mp. or zaleplon mp. or exp Eszopiclone/ or exp Melatonin/ or Ramelteon.mp. or exp Orexins/ or Suvorexant mp. or exp Antidepressive Agents/ or exp Barbiturates/ or Butabarbital.mp. or diphenhydramine/ or doxylamine/ or exp Trazodone/ or Mirtazapine.mp. or exp Antipsychotic Agents/ or Olanzapine.mp. or exp Valerian/ or exp Kava/
CENTRAL inception to 6th March 2020		
	Diagnosis	MeSH descriptor: [Sleep Initiation and Maintenance Disorders] explode all trees ***Attention: MeSH terms must be selected via MeSH Terms in Cochrane Library!!***
	Design	-
	Intervention	MeSH descriptor: [Sleep Aids, Pharmaceutical] this term only MeSH descriptor: [Hypnotics and Sedatives] this term only MeSH descriptor: [Melatonin] this term only MeSH descriptor: [Orexins] this term only MeSH descriptor: [Antidepressive Agents] this term only MeSH descriptor: [Barbiturates] this term only MeSH descriptor: [Histamine Antagonists] this term only MeSH descriptor: [Antipsychotic Agents] this term only MeSH descriptor: [Valerian] this term only MeSH descriptor: [Kava] this term only MeSH descriptor: [Benzodiazepines] this term only quazepam or zolpidem or zaleplon or Eszopiclone or Ramelteon or Suvorexant or Butabarbital or Trazodone or Mirtazapine or Olanzapine
	Additional Filter	children or teenager* or adolescent* or animal* or mice or mouse or rat* or case report* or review* or meta* or protocol* or cross* ***Attention: connect with NOT!!!*** Trials ***Attention: select on lower left after search is done!!!***
Web of Science all years to 6th March 2020		
	Diagnosis	TS=(Insomnia)
	Design	TS=((randomized controlled trial or controlled clinical trial) or randomized or placebo or (drug therapy) or randomly or trial or groups)
	Intervention	TS=(Sleep Aids or benzodiazepines or quazepam or "Hypnotics and Sedatives" or zolpidem or zaleplon or Eszopiclone or Melatonin or Ramelteon or Orexin* or Suvorexant or (Antidepressive Agents) or Barbiturates or Butabarbital or "Histamine Antagonist*" or Antihistamine* or Trazodone or Mirtazapine or "Antipsychotic Agent*" or Antipsychotic* or Olanzapine or Valerian or Kava)
	Additional Filter	NOT TS=(children or teenager* or adolescent* or animal* or mice or mouse or rat* or case report* or review* or meta* or protocol* or cross*)

eTable 2

Characteristics of all 88 RCTs included in this Systematic Review and Meta-analysis

First author (publication year)	Interventions (dosage each night)	Insomnia severity	Study Population	Reported sleep measures	Total N at randomisation (women)	Mean age in years (*SD*)	Treatment duration in nights	Region / country	Funding source
Aden (1983)	Quazepam 30mg, Placebo	Full insomnia diagnosis	Adults	sSQ, sSOL, sTST, sAwake	57 (33)	47.0 (na)	5	USA	Industry
Allain (2001)	Zolpidem 10mg, Placebo	Full insomnia diagnosis	Adults	sTST, sSQ, sSOL, sAwake	245 (188)	46.2 (10.6)	28	France	Industry
Ananth (1973)	Gaboxadol 650mg, Diazepam 10mg, Secobarbital 100mg, Placebo	Subclinical insomnia / elevated insomnia symptoms	Adults	oSOL, oTST, sSOL, sTST, sSQ	60 (33)	37.4 (na)	1	Canada	Mixed
Ancoli-Israel (2010)	Eszopiclone 2mg, Placebo	Full insomnia diagnosis	Elderly (> 65 years old)	sTST, sSOL	388 (243)	72.0 (5.1)	84	USA	Industry
Ansoms (1976)	Flunitrazepam 3mg, Fenobarbital 200mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sTST, sAwake	49 (37)	na	10	Belgium	na
Cuanang (1982)	Temazepam 10mg, Temazepam 20mg,	Full insomnia diagnosis	Adults	sSOL, sTST	60 (20)	na	5	Philippines	University / Government

	Placebo								
Davari-Tanha (2016)	Venlafaxine 75mg, Citalopram 20mg, Placebo	Self-reported sleeping problems	Menopausal women	sSQ	60 (60)	51.0 (3.5)	56	Iran	na
Dominguez (1985)	Brotizolam 50mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sAwake, sTST, SQ	67 (41)	49.0 (na)	21	USA	na
Dominguez (1986)	Estazolam 2mg, Flurazepam 30mg, Placebo	Full insomnia diagnosis	Adults	sSQ, sTST, sAwake	74 (na)	na	7	USA	na
Dorsey (2004)	Zolpidem 10mg, Placebo	Full insomnia diagnosis	Menopausal women	sSOL, sAwake, sTST	141 (141)	50.8 (4.5)	28	USA	Industry
Elie (1990)	Zopiclone 7.5mg, Flurazepam 30mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sTST, sAwake	36 (24)	37.6 (1.8)	28	Canada	na
Elie (1999)	Zaleplon 5mg, Zaleplon 10mg, Zaleplon 20mg, Zolpidem 10mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sTST, sAwake, sSQ	615 (397)	42.8 (12.4)	28	Europe & Canada	Industry
Fabre (1978)	Triazolam 0.5mg,	Full insomnia	Adults	sSQ, sAwake,	277 (144)	45.2 (na)	14	USA	na

	Placebo	diagnosis		sTST, sSOL				
Fan (2017)	Suvorexant 40mg, Placebo	Full insomnia diagnosis	Adults	sTST, sSOL, sSQ	120 (48)	51.0 (12.1)	182	China	na
Ferretti (1980)	Triazolam 0.25mg, Placebo	Full insomnia diagnosis	Elderly	sSQ, sTST, sSOL, sAwake	28 (20)	73.0 (7.3)	14	Italy	Industry
Fillingim (1982)	Temazepam 30mg, Flurazepam 30mg, Placebo	Full insomnia diagnosis	Elderly (> 60 years old)	sSQ, sSOL, sAwake, sTST	75 (67)	81.0 ()	4	USA	na
Fleming (1995)	Zolpidem 10mg, Zolpidem 20mg, Flurazepam 30mg, Placebo	Full insomnia diagnosis	Adults	oSOL, sSQ, sSOL, oSQ	144 (69)	35.0 (9.8)	3	USA & Canada	na
Fry (2000)	Zaleplon 5mg, Zaleplon 10mg, Zaleplon 20mg, Zolpidem 10mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sTST, sAwake, sSQ	595 (347)	42.0 (12.0)	28	USA	Industry
Goldenberg (1994)	Zopiclone 7.5mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sAwake, sTST, sSQ	524 (333)	42.9 (8.9)	14	Europe	na
Merck (2014)	Esmirtazapine 4.5mg,	Full	Adults	sTST, sSOL,	460 (282)	47.8 (11.3)	182	na	Industry

	Placebo	insomnia diagnosis		sAwake, sSQ					
Neurim (2018)	Piromelatine 20mg, Piromelatine 50mg, Placebo	Full insomnia diagnosis	Adults	oSOL, oAwake	137 (97)	49.6 (14.8)	28	USA	Industry
Black (2017)	Zolpidem 5mg, Sodium Oxybate 2.25mg, Placebo	Full insomnia diagnosis	Adults	sSQ	48 (32)	53.2 (12.3)	84	USA	University / Government
Zick (2013)	Chamomile 540mg, Placebo	Full insomnia diagnosis	Adults	sSQ, sSOL, sAwake, sTST	34 (25)	41.4 (14.2)	28	USA	University / Government
Sanofi-Aventis (2010)	Eplivanserin 5mg, Placebo	Full insomnia diagnosis	Adults	sAwake, sTST, sSOL, sSQ	1155 (690)	51.9 (na)	84	International	Industry
Acelion (2018)	ACT 5mg, ACT 10mg, ACT 25mg, ACT 50mg, Placebo	Full insomnia diagnosis	Adults	oSOL, sSOL	299 (190)	45.0 (11.5)	28	International	Industry
Hedner (2007)	Gaboxadol 5mg, Gaboxadol 10mg, Placebo	Full insomnia diagnosis	Elderly (> 65 years old)	sTST, sSOL, sAwake, sSQ	541 (357)	71.0 (na)	28	Europe	Industry

Hajak (2009)	Gaboxadol 5mg, Gaboxadol 10mg, Gaboxadol15mg, Zolpidem 10mg, Placebo	Full insomnia diagnosis	Adults	sTST, sAwake, sSQ, sSOL	742 (471)	48.2 (11.2)	14	International	Industry
Hedner (2000)	Zaleplon 5mg, Zaleplon 10mg, Placebo	Full insomnia diagnosis	Elderly (> 65 years old)	sSOL, sTST, sAwake, sSQ	437 (295)	72.5 (6.3)	14	Europe	Industry
Heffron (1979)	Temazepam 30mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sAwake, sTST, sSQ	55 (35)	na	4	USA	na
Heidrich (1981)	Lormetazepam 2mg, Placebo	Full insomnia diagnosis	Adults	sSQ, sSOL, sTST, sAwake	62 (42)	45.3 (2.0)	14	Germany	Industry
Herberg (2002)	Zopiclone 7.5mg, Placebo	Full insomnia diagnosis	Adults	sSQ, sTST, sSOL, sAwake	48 (24)	51.8 (8.2)	14	Germany	Industry
Herring (2016)	Suvorexant 40/30mg, Suvorexant 20/15mg, Placebo	Full insomnia diagnosis	Adults	sSQ, sTST, sSOL, sAwake oSOL	1021 (637)	56.0 (15.0)	91	International	Industry
Herring (2016)	Suvorexant 40/30mg, Suvorexant 20/15mg, Placebo	Full insomnia diagnosis	Adults	sSQ, sTST, sSOL, sAwake oSOL	1009 (671)	57.0 (15.0)	91	International	Industry

Ivgy-May (2015)	Esmirtazapine 1.5mg, Esmirtazapine 3.0mg, Esmirtazapine 4.5mg, Placebo	Full insomnia diagnosis	Adults	sTST, sSOL, sAwake, sSQ	526 (339)	45.3 (12.0)	14	USA & Canada	Industry
Ivgy-May (2015)	Esmirtazapine 3.0mg, Esmirtazapine 4.5mg, Placebo	Full insomnia diagnosis	Adults	oSOL, oTST, oAwake, sTST, sSOL, sAwake, sSQ	419 (279)	45.0 (11.0)	42	USA & Canada	Industry
Jacobson (1986)	Brotizolam 0.125mg, Placebo	Full insomnia diagnosis	Elderly (> 60 years old)	sTST, sSOL, sAwake, sSQ	57 (na)	70.5 (5.5)	4	USA	na
Kesson (1984)	Loprazolam 1mg, Placebo	Self-reported sleeping problems	Adults	sSOL, sTST, sSQ	71 (52)	52.9 (14.3)	42	UK	na
Kramer (1985)	Midazolam 7.5mg, Placebo	Full insomnia diagnosis	Elderly (> 60 years old)	sSOL, sAwake, sTST	88 (28)	68.2 (6.3)	4	USA	Industry
Krystal (2003)	Eszopiclone 3mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sTST, sAwake, sSQ	791 (498)	44.0 (11.0)	182	USA	na
Krystal (2008)	Zolpidem 12.5mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sTST, sAwake, sSQ	1025 (627)	45.7 (11.0)	168	USA	Industry
Lankford (2008)	Gaboxadol 10mg, Gaboxadol 15mg,	Full insomnia	Adults	sSQ, sTST, sSOL, sAwake, oTST,	458 (302)	44.0 (11.0)	30	USA	Industry

	Placebo	diagnosis		sSOL, sAwake					
Lankford (2008)	Gaboxadol 5mg, Gaboxadol 10mg, Placebo	Full insomnia diagnosis	Elderly (> 65 years old)	sSQ, sTST, sSOL, sAwake, oTST, sSOL, sAwake	486 (296)	71.0 (5.0)	30	USA & Canada	Industry
Lemoine (2007)	Melatonin 2mg, Placebo	Full insomnia diagnosis	Elderly (> 65 years old)	sSQ	170 (112)	68.5 (8.3)	21	France & Israel	Industry
Lingjaerde (1983)	Midazolam 15mg, Flunitrazepam 1mg, Placebo	Full insomnia diagnosis	Midwinter insomnia	sSOL, sAwake, sTST, sSQ	43 (25)	38.0 (na)	5	Norway	Industry
Luthringer (2009)	Melatonin 2mg, Placebo	Full insomnia diagnosis	Elderly (> 55 years old)	sSQ, oSQ, oSOL, oAwake, oSOL	40 (16)	60.5 (3.3)	21	France	Industry
Mayer (2009)	Ramelteon 8mg, Placebo	Full insomnia diagnosis	Adults	oSOL, oTST, sSOL, sTST, sAwake, sSQ	451 (285)	46.2 (14.8)	182	International	Industry
McCall (2006)	Eszopiclone 2mg, Placebo	Full insomnia diagnosis	Elderly (> 64 years old)	oSOL, oSQ, oAwake, sSOL, sTST, sAwake, sSQ	264 (178)	71.1 (5.1)	14	USA	Industry
Melo de Paula (1984)	Lorazepam 1mg, Lorazepam 2mg, Flurazepam 30mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sTST, sAwake	60 (44)	29.6 (9.0)	14	Brazil	na

Mendels (1983)	Quazepam 15mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sSQ, sTST, sAwake	60 (19)	46.0 (10.0)	5	USA	Industry
Minnekeer (1988)	Quazepam 15mg, Flunitrazepam 2mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sTST, sAwake	205 (131)	54.5 (13.6)	28	Belgium	University / Government
Monchesky (1989)	Zopiclone 7.5mg, Placebo	Full insomnia diagnosis	Shift work insomnia	sSOL, sTST, sSQ	50 (3)	34.9 (1.2)	13	Canada	Mixed
Monti (1987)	Midazolam 15mg, Methaqualone 300mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sTST, sAwake, sSQ	30 (19)	47.5 (11.6)	5	Uruguay	na
Monti (1994)	Triazolam 0.5mg, Zolpidem 10mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sTST, sAwake, oTST, oAwake	24 (21)	47.3 (13.0)	27	Uruguay	na
Murphy (2017)	Lemborexant 1-25mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sSQ, oSQ, oSOL	291 (182)	48.3 (14.4)	15	USA	Industry
O'Haire (1981)	Quazepam 30mg, Placebo	Full insomnia diagnosis	Adults	sSQ, sTST	60 (26)	43.0 (10.0)	5	USA	Industry
Oxman (2007)	Valerian 600mg, Placebo	Full insomnia diagnosis	Adults	sSQ, sSOL, sAwake, sTST	405 (247)	43.8 (13.4)	14	Norway	University / Government

Rickels (1986)	Brotizolam 0.25/0.5mg, Placebo	Full insomnia diagnosis	Adults	sTST, sSOL, sAwake, sSQ	63 (40)	46.0 (12.0)	21	USA	Mixed
Riemann (2002)	Lormetazepam 1mg, Trimipramine 100mg, Placebo	Full insomnia diagnosis	Adults	sSQ, oTST, oSQ, oSOL	65 (23)	47.0 (10.9)	28	Germany	Industry
Rondanelli (2011)	Mezinat, Placebo	Full insomnia diagnosis	Elderly (> 70 years old)	sSQ, oTST	43 (27)	78.3 (3.9)	56	Italy	University / Government
Roth (2010)	Gaboxadol 10mg, Gaboxadol 15mg, Placebo	Full insomnia diagnosis	Adults	sTST, sSOL, sAwake	928 (598)	43.0 (12.0)	91	USA	Industry
Roth (2010)	Gaboxadol 15mg, Placebo	Full insomnia diagnosis	Adults	sTST, sSOL, sAwake	605 (353)	45.0 (12.0)	365	USA	Industry
Roth (2006)	Ramelteon 4mg, Ramelteon 8mg, Placebo	Full insomnia diagnosis	Elderly (> 64 years old)	sSOL, sTST, sAwake, sSQ	829 (488)	72.4 (6.0)	35	USA	Industry
Roth (2006)	Zolpidem 2.5mg, Placebo	Full insomnia diagnosis	Adults	oSOL, oSQ, oAwake, sSOL, sTST, sAwake	212 (123)	44.0 (13.0)	21	USA, Canada, & Australia	Industry
Roth (1997)	Quazepam 7.5mg, Quazepam 15mg, Placebo	Full insomnia diagnosis	Elderly (> 60 years old)	oTST, oSOL	30 (15)	65.9 (4.6)	7	USA	Mixed

Schadeck (1996)	Doxylamine 15mg, Zolpidem 10mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sSQ, sTST, sAwake	338 (250)	45.9 (14.1)	14	France	na
Scharf (2007)	Indiplon 10mg, Indiplon 20mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sAwake, sTST, sSQ	702 (428)	45.6 (11.2)	91	USA, Canada, & UK	Industry
Scharf (1990)	Estazolam 2mg, Flurazepam 30mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sTST, sSQ, sAwake	244 (na)	41.5 (13.5)	7	USA	na
Scholey (2017)	Lzcomplex3, Placebo	Subclinical insomnia / elevated insomnia symptoms	Adults	sSQ	171 (96)	30.3 (9.8)	14	Australia	Industry
Sivertsen (2006)	Zopiclone 7.5mg, Placebo	Full insomnia diagnosis	Elderly (> 55 years old)	sTST, sSQ, oTST, oSQ	46 (22)	60.8 (5.4)	42	Norway	University / Government
Takeda Pharmaceuticals (2008)	Ramelteon 8mg, Ramelteon 16mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sTST, sAwake, sSQ	848 (499)	43.8 (12.2)	35	USA	Industry
Uchimura (2011)	Ramelteon 4mg, Ramelteon 8mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sTST, sAwake, sSQ	1145 (723)	48.8 (17.2)	7	Japan	Industry

Uchiyama (2011)	Ramelteon 8mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sTST, sAwake, sSQ	987 (621)	38.8 (13.8)	14	Japan	Industry
Vorbach (1996)	Valerian 600mg, Placebo	Full insomnia diagnosis	Adults	sSQ	121 (71)	47.4 (11.1)	28	Germany	na
Wade (2010)	Circadin 2mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sTST, sSQ	791 (497)	62.0 (na)	21	UK	Industry
Walsh (1998)	Trazodone 50mg, Zolpidem 10mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sTST, sAwake, sSQ	306 (193)	42.0 (na)	14	USA	Industry
Walsh (1984)	Estazolam 1mg, Estazolam 2mg, Placebo	Full insomnia diagnosis	Adults	sSQ, sSOL, sTST, sAwake	367 (190)	41.1 (12.7)	7	USA	Industry
Walsh (1998)	Zaleplon 5mg, Zaleplon 10mg, Triazolam 0.25mg, Placebo	Full insomnia diagnosis	Adults	oSOL, oTST, sSOL, sTST, sAwake, sSQ	132 (77)	40.2 (10.3)	14	USA	Industry
Walsh (2007)	Eszopiclone 3mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sTST, sAwake, sSQ	830 (506)	45.4 (11.8)	182	USA	Industry
Walsh (2007)	Indiplon 5mg,	Full insomnia	Elderly (> 65 years old)	sSOL, sTST, sAwake, sSQ	358 (196)	70.9 (0.4)	14	USA & Canada	Industry

	Indiplon 10mg, Placebo	diagnosis							
Walsh (2010)	EVT 201 1.5mg, EVT 201 2.5mg, Placebo	Full insomnia diagnosis	Elderly (> 65 years old)	sSOL, sTST, sAwake, sSQ, oSOL, oTST, oAwake	149 (96)	71.3 (4.9)	7	USA	Industry
Walsh (2008)	Zolpidem 6.25mg, Placebo	Full insomnia diagnosis	Elderly (> 65 years old)	sSOL, sTST, sAwake, sSQ, oTST, oSOL, oAwake, oSQ	205 (117)	70.2 (4.5)	21	International	Industry
Walsh (2000)	Zaleplon 10mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sTST, sAwake, oSOL, oTST, oAwake	113 (84)	42.1 (11.0)	35	USA	Industry
Wang-Weigand (2011)	Ramelteon 8mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sTST, sAwake, sSQ, oSOL	556 (359)	43.2 (12.5)	21	USA	Industry
Wheatley (1989)	Zolpidem 10mg, Zolpidem 20mg, Placebo	Subclinical insomnia / elevated insomnia symptoms	Adults	sSOL, sAwake, sTST, sSQ	88 (62)	49.0 (12.0)	21	UK	na
Winsauer (1984)	Quazepam 15mg, Placebo	Full insomnia diagnosis	Elderly (> 60 years old)	sSOL, sTST, sAwake, sSQ	60 (39)	na	5	USA	Mixed
Xu (2011)	Propofol 3000mg/l, Placebo	Full insomnia diagnosis	Refractory insomnia	sSOL, sSQ, sAwake, oTST, oSOL, oAwake	103 (59)	46.0 (15.6)	5	China	University / Government

Zammit (2007)	Ramelteon 8mg, Ramelteon 16mg, Placebo	Full insomnia diagnosis	Adults	sSOL, sTST, sSQ, oSOL, oTST, oSQ, oAwake	405 (272)	39.3 (12.0)	35	USA	Industry
Zammit (2004)	Eszopiclone 2mg, Eszopiclone 3mg, Placebo	Full insomnia diagnosis	Adults	sSQ, sSOL, sTST, sAwake, oAwake, oSOL, oSQ	308 (199)	39.8 (11.7)	44	USA	Industry

Note. na, not reported in original article; sSOL, subjective sleep onset latency; sSQ, subjective sleep quality; sTST, subjective total sleep time; sAwake, subjective number of awakenings; oSOL, objective sleep onset latency; oSQ, objective sleep quality; oTST, objective total sleep time; oAwake, objective number of awakenings

References included in the Analysis of the Meta-analysis

Actelion Pharmaceuticals Ltd. (2018). *Multi-center, double-blind, randomized, placebo-controlled, active reference, parallel-group, polysomnography dose response study to assess the efficacy and safety of ACT-541468 in adult subjects with insomnia disorder*.

Aden, G. C., & Thatcher, C. (1983). Quazepam in the short-term treatment of insomnia in outpatients. *The Journal of Clinical Psychiatry, 44*(12), 454-456. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc2&NEWS=N&AN=1984-26711-001

Allain, H., Arbus, L., & Schuck, S. (2001). Efficacy and safety of zolpidem administered 'as needed' in primary insomnia: Results of a double-blind, placebo-controlled study. *Clinical Drug Investigation, 21*(6), 391-400. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed8&AN=32553601

Ananth, J. V., Bronheim, L. A., Klingner, A., & Ban, T. A. (1973, Apr). Diazepam in the treatment of insomnia in psychiatric patients. *Current Therapeutic Research, Clinical & Experimental, 15*(4), 217-222. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med1&AN=4633283

Ancoli-lsrael, S., Krystal, A. D., McCall, W. V., Schaefer, K., Wilson, A., Claus, R., Rubens, R., & Roth, T. (2010). A 12-week, randomized, double-blind, placebo-controlled study evaluating the effect of eszopiclone 2 mg on sleep/wake function in older adults with primary and comorbid insomnia. *Sleep: Journal of Sleep and Sleep Disorders Research, 33*(2), 225-234. https://doi.org/http://dx.doi.org/10.1093/sleep/33.2.225

Ansoms, C., & Marchal, J. P. (1976). Clinical double blind trial of flunitrazepam versus phenobarbital in patients with intractable sleeplessness (Dutch). [Dutch]. *Tijdschrift voor Geneeskunde, 32*(21), 1259-1262. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emcl2&AN=7185781

Black, J. E. (2017). *Xyrem(Sodium Oxybate) and Ambien(Zolpidem Tartrate) in the Treatment of Chronic Insomnia.*

Cuanang, J. R., & Limos, L. (1982). Treatment of insomnia with temazepam: double-blind, placebo-controlled evaluation. *Clinical Therapeutics, 4*(5), 402-412. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med2&AN=6121626

Davari-Tanha, F., Soleymani-Farsani, M., Asadi, M., Shariat, M., Shirazi, M., & Hadizadeh, H. (2016, May). Comparison of citalopram and venlafaxine's role in treating sleep disturbances in menopausal women, a randomized, double-blind, placebo-controlled trial. *Archives of Gynecology & Obstetrics, 293*(5), 1007-1013. https://doi.org/https://dx.doi.org/10.1007/s00404-015-3900-1

Dominguez, R. A., Goldstein, B. J., Jacobson, A. F., & Steinbook, R. M. (1986). Estazolam in the treatment of insomnia. *Psychopharmacology Bulletin, 22*(1), 278-280. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed3&AN=16101371

Dorsey, C. M., Lee, K. A., & Scharf, M. B. (2004, Oct). Effect of zolpidem on sleep in women with perimenopausal and postmenopausal insomnia: a 4-week, randomized, multicenter, double-blind, placebo-controlled study. *Clinical Therapeutics, 26*(10), 1578-1586. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med5&AN=15598474

Elie, R., Lavoie, G., Bourgouin, J., & Le Morvan, P. (1990, Oct). Zopiclone versus flurazepam in insomnia: prolonged administration and withdrawal. *International Clinical Psychopharmacology, 5*(4), 279-286. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med3&AN=2081899

Elie, R., Ruther, E., Farr, I., Emilien, G., & Salinas, E. (1999). Sleep latency is shortened during 4 weeks of treatment with zaleplon, a novel nonbenzodiazepine hypnotic. *The Journal of Clinical Psychiatry, 60*(8), 536-544. https://doi.org/http://dx.doi.org/10.4088/JCP.v60n0806

Fabre, L. F., Brachfeld, J., Meyer, L. R., Slowe, I. A., Calvo, R., & Metzler, C. (1978). Multi-clinic double-blind comparison of triazolam (Halcion) and placebo administered for 14 consecutive nights in outpatients with insomnia. *The Journal of Clinical Psychiatry, 39*(8), 679-682. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc2&NEWS=N&AN=1980-12469-001

Fan, B., Kang, J., He, Y., Hao, M., Du, W., & Ma, S. (2017). Efficacy and safety of suvorexant for the treatment of primary insomnia among Chinese: A 6-month randomized double-blind controlled study. *Neurology Asia, 22*(1), 41-47. http://www.neurology-asia.org/articles/neuroasia-2017-22(1)-041.pdf

Ferretti, G., Ronchini, P., Maini, C., & Cucinotta, D. (1980). Effectiveness and tolerability of Triazolam vs placebo in the treatment of geriatric patients with insomnia. [Italian] [Impiego e tollerabilita del triazolam nel trattamento di pazienti ospedalizzati in confronto a placebo.]. *Giornale di Gerontologia, 28*(12), 808-815. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed2&AN=11117983

Fillingim, J. M. (1982). Double-blind evaluation of temazepam, flurazepam, and placebo in geriatric insomniacs. *Clinical Therapeutics, 4*(5), 369-380. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med2&AN=6121625

Fleming, J., Moldofsky, H., Walsh, J. K., Scharf, M., Ninomurcia, G., & Radonjic, D. (1995, Jun). Comparison of the Residual Effects and Efficacy of Short-Term Zolpidem, Flurazepam and Placebo in Patients with Chronic Insomnia. *Clinical Drug Investigation, 9*(6), 303-313. https://doi.org/10.2165/00044011-199509060-00001

Fry, J., Scharf, M., Mangano, R., & Fujimori, M. (2000). Zaleplon improves sleep without producing rebound effects in outpatients with insomnia. *International Clinical Psychopharmacology, 15*(3), 141-152. https://doi.org/http://dx.doi.org/10.1097/00004850-200015030-00003

Goldenberg, F., Hindmarch, I., Joyce, C. R. B., Le Gal, M., Partinen, M., & Pilate, C. (1994). Zopiclone, sleep and health-related quality of life. *Human Psychopharmacology: Clinical and Experimental, 9*(4), 245-251. https://doi.org/http://dx.doi.org/10.1002/hup.470090403

Hajak, G., Hedner, J., Eglin, M., Loft, H., Storustovu, S. i., Lutolf, S., & Lundahl, J. (2009, August). A 2-week efficacy and safety study of gaboxadol and zolpidem using electronic diaries in primary insomnia outpatients. *Sleep Medicine, 10*(7), 705-712. https://doi.org/http://dx.doi.org/10.1016/j.sleep.2008.09.010

Hedner, J. (2007). *A double-blind, randomised, placebo-controlled, parallel-group, phase III study evaluating the efficacy and safety of gaboxadol 5mg and 10mg daily in elderly outpatients with primary insomnia*.

Hedner, J., Yaeche, R., Emilien, G., Farr, I., & Salinas, E. (2000). Zaleplon shortens subjective sleep latency and improves subjective sleep quality in elderly patients with insomnia. *International Journal of Geriatric Psychiatry, 15*(8), 704-712. https://doi.org/http://dx.doi.org/10.1002/1099-1166%28200008%2915:8%3C704::AID-GPS183%3E3.0.CO;2-S

Heffron, W. A., & Roth, P. (1979). Double-blind evaluation of the safety and hypnotic efficacy of temazepam in insomniac outpatients. *British Journal of Clinical Pharmacology, 8*(1), 69S-72S. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med1&AN=41544

Heidrich, H., Ott, H., & Beach, R. C. (1981, Jan). Lormetazepam - a benzodiazepine derivative without hangover effect? A double-blind study with chronic insomniacs in a general practice setting. *International Journal of Clinical Pharmacology, Therapy, & Toxicology, 19*(1), 11-17. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med2&AN=6110634

Herberg, K. W., Laux, G., & Fischer, W. (2002). Analysis of the effects of a 14 days treatment with zopiclone 7.5mg/d on performance capability, actual well-being, and quality of sleep of patients with primary insomnia. [German] [Leistungsfahigkeit, befindlichkeit und schlafqualitat von patienten mit primarer insomnie unter zopiclon: Einflusse einer 14-tagigen behandlung mit 7,5 mg/Tag Zopiclon.]. *Psychopharmakotherapie, 9*(1), 25-34. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed8&AN=34205231

Herring, W. J., Connor, K. M., Ivgy-May, N., Snyder, E., Liu, K., Snavely, D. B., Krystal, A. D., Walsh, J. K., Benca, R. M., Rosenberg, R., Sangal, R. B., Budd, K., Hutzelmann, J., Leibensperger, H., Froman, S., Lines, C., Roth, T., & Michelson, D. (2016, Jan 15). Suvorexant in Patients With Insomnia: Results From Two 3-Month Randomized Controlled Clinical Trials. *Biological Psychiatry, 79*(2), 136-148. https://doi.org/https://dx.doi.org/10.1016/j.biopsych.2014.10.003

Ivgy-May, N., Roth, T., Ruwe, F., & Walsh, J. (2015, Jul). Esmirtazapine in non-elderly adult patients with primary insomnia: efficacy and safety from a 2-week randomized outpatient trial. *Sleep Medicine, 16*(7), 831-837. https://doi.org/https://dx.doi.org/10.1016/j.sleep.2015.03.005

Ivgy-May, N., Ruwe, F., Krystal, A., & Roth, T. (2015, Jul). Esmirtazapine in non-elderly adult patients with primary insomnia: efficacy and safety from a randomized, 6-week sleep laboratory trial. *Sleep Medicine, 16*(7), 838-844. https://doi.org/https://dx.doi.org/10.1016/j.sleep.2015.04.001

Jacobson, A. F., Dominguez, R. A., Goldstein, B. J., & Steinbook, R. M. (1986). Efficacy of brotizolam in geriatric patients with insomnia. *Current Therapeutic Research - Clinical and Experimental, 39*(4), 528-536. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed3&AN=17027294

Kesson, C. M., Gray, J. M. B., Lawson, D. H., & Ankier, S. I. (1984). Long-term efficacy and tolerability of a new hypnotic-loprazolam. *British Journal of Clinical Practice, 38*(9), 306-312. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed3&AN=14031057

Kramer, M., Bonnet, M., Schoen, L. S., & Dexter, J. R. (1985). Efficacy and safety of midazolam 7.5 mg. in geriatric insomniacs. *Current Therapeutic Research, 38*(3), 414-422. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc2&NEWS=N&AN=1986-22550-001

Kramer, M., Schoen, L. S., & Scharf, M. (1984). Effects of alprazolam on sleep and performance in geriatric insomniacs. *Current Therapeutic Research, 36*(1), 67-76. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc2&NEWS=N&AN=1985-15303-001

Krystal, A., Walsh, J., Roth, T., Amato, D. A., & Wessel, T. (2003, May). The sustained efficacy and safety of eszopiclone over six months of nightly treatment: A placebo-controlled study in patients with chronic insomnia. *Sleep, 26*, A310-A310.

Krystal, A. D., Erman, M., Zammit, G. K., Soubrane, C., & Roth, T. (2008). Long-term efficacy and safety of zolpidem extended-release 12.5 mg, administered 3 to 7 nights per week for 24 weeks, in patients with chronic primary insomnia: A 6-month, randomized, double-blind, placebo-controlled, parallel-group, multicenter study. *Sleep: Journal of Sleep and Sleep Disorders Research, 31*(1), 79-90. https://doi.org/http://dx.doi.org/10.1093/sleep/31.1.79

Lankford, D. A., Corser, B. C., Zheng, Y. P., Li, Z. R., Snavely, D. B., Lines, C. R., & Deacon, S. (2008, Oct). Effect of gaboxadol on sleep in adult and elderly patients with primary insomnia: Results from two randomized, placebo-controlled, 30-night polysomnography studies. *Sleep, 31*(10), 1359-1370. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2572741/pdf/aasm.31.10.1359.pdf

Lemoine, P., Nir, T., Laudon, M., & Zisapel, N. (2007). Prolonged-release melatonin improves sleep quality and morning alertness in insomnia patients aged 55 years and older and has no withdrawal effects. *Journal of Sleep Research, 16*(4), 372-380. https://doi.org/http://dx.doi.org/10.1111/j.1365-2869.2007.00613.x

Lingjaerde, O., Bratlid, T., Westby, O. C., & Gordeladze, J. O. (1983). Effect of midazolam, flunitrazepam, and placebo against midwinter insomnia in northern Norway. *Acta Psychiatrica Scandinavica, 67*(2), 118-129. https://doi.org/http://dx.doi.org/10.1111/j.1600-0447.1983.tb06731.x

Luthringer, R., Muzet, M., Zisapel, N., & Staner, L. (2009). The effect of prolonged-release melatonin on sleep measures and psychomotor performance in elderly patients with insomnia. *International Clinical Psychopharmacology, 24*(5), 239-249. https://doi.org/http://dx.doi.org/10.1097/YIC.0b013e32832e9b08

Mayer, G., Wang-Weigand, S., Roth-Schechter, B., Lehmann, R., Staner, C., & Partinen, M. (2009). Efficacy and safety of 6-month nightly ramelteon administration in adults with chronic primary insomnia. *Sleep: Journal of Sleep and Sleep Disorders Research, 32*(3), 351-360. https://doi.org/http://dx.doi.org/10.1093/sleep/32.3.351

McCall, W. V., Erman, M., Krystal, A. D., Rosenberg, R., Scharf, M., Zammit, G. K., & Wessel, T. (2006, Sep). A polysomnography study of eszopiclone in elderly patients with insomnia. *Current Medical Research & Opinion, 22*(9), 1633-1642. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med5&AN=16968566

Melo de Paula, A. J. (1984). Comparative study of lormetazepam and flurazepam in the treatment of insomnia. *Clinical Therapeutics, 6*(4), 500-508. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med2&AN=6380724

Mendels, J., & Stern, S. (1983). Evaluation of the short-term treatment of insomnia in out-patients with 15 milligrams of quazepam. *Journal of International Medical Research, 11*(3), 155-161. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med2&AN=6347747

Merck Sharp & Dohme Corp. (2014). *A 6- Month Efficacy and Safety Study of Org 50081 in Adult Patients With Chronic Primary Insomnia (21106/P05701/MK-8265-002)*.

Minnekeer, R. J., Marchal, J. L., Van De Velde, L., & Herman, P. (1988). A double-blind study comparing the efficacy and safety of quazepam with flunitrazepam and placebo in patients with chronic insomnia. *Acta Therapeutica, 14*(2), 159-170. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed4&AN=18147695

Monchesky, T. C., Billings, B. J., Phillips, R., & Bourgouin, J. (1989). Zopiclone in insomniac shiftworkers. Evaluation of its hypnotic properties and its effects on mood and work performance. *International Archives of Occupational & Environmental Health, 61*(4), 255-259. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med3&AN=2656527

Monti, J. M., Alterwain, P., & Debellis, J. (1987). Comparative double-blind clinical study of midazolam, methaqualone, and placebo in outpatients with slight-to-moderate insomnia. *Current Therapeutic Research - Clinical and Experimental, 41*(4), 437-443.

http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed3&AN=17065324

Monti, J. M., Attali, P., Monti, D., Zipfel, A., de la Giclais, B., & Morselli, P. (1994). Zolpidem and rebound insomnia: A double-blind, controlled polysomnographic study in chronic insomniac patients. *Pharmacopsychiatry, 27*(4), 166-175. https://doi.org/http://dx.doi.org/10.1055/s-2007-1014298

Murphy, P., Moline, M., Mayleben, D., Rosenberg, R., Zammit, G., Pinner, K., Dhadda, S., Hong, Q., Giorgi, L., & Satlin, A. (2017). Lemborexant, a dual orexin receptor antagonist (DORA) for the treatment of insomnia disorder: Results from a Bayesian, adaptive, randomized, double-blind, placebo-controlled study. *Journal of Clinical Sleep Medicine, 13*(11), 1289-1299. https://doi.org/http://dx.doi.org/10.5664/jcsm.6800

Neurim Pharmaceuticals Ltd. (2018). *Sleep Laboratory Study to Investigate the Safety and Efficacy of Neu-P11 in Primary Insomnia Patients*.

O'Hair, D. E., & Winsauer, H. J. (1981). Evaluation of short-term treatment with 30 mg of quazepam in insomniac outpatients. *Clinical Therapeutics, 4*(4), 291-301. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med2&AN=7037183

Oxman, A. D., Flottorp, S., Havelsrud, K., Fretheim, A., Odgaard-Jensen, J., Austvoll-Dahlgren, A., Carling, C., Pallesen, S., & Bjorvatn, B. (2007, 17 Oct). A televised, web-based randomised trial of an herbal remedy (Valerian) for insomnia. *PLoS ONE, 2 (10) (no pagination)*(e1040). https://doi.org/http://dx.doi.org/10.1371/journal.pone.0001040

Rickels, K., Morris, R. J., & Mauriello, R. (1986). Brotizolam, a triazolothienodiazepine, in insomnia. *Clinical Pharmacology and Therapeutics, 40*(3), 293-299. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed3&AN=16017856

Riemann, D., Voderholzer, U., Cohrs, S., Rodenbeck, A., Hajak, G., Ruther, E., Wiegand, M. H., Laakmann, G., Baghai, T., Fischer, W., Hoffmann, M., Hohagen, F., Mayer, G., & Berger, M. (2002). Trimipramine in primary insomnia: Results of a polysomnographic double-blind controlled study. *Pharmacopsychiatry, 35*(5), 165-174. https://doi.org/http://dx.doi.org/10.1055/s-2002-34119

Rondanelli, M., Opizzi, A., Monteferrario, F., Antoniello, N., Manni, R., & Klersy, C. (2011). The effect of melatonin, magnesium, and zinc on primary insomnia in long-term care facility residents in Italy: A double-blind, placebo-controlled clinical trial. *Journal of the American Geriatrics Society, 59*(1), 82-90. https://doi.org/http://dx.doi.org/10.1111/j.1532-5415.2010.03232.x

Roth, T., Lines, C., Vandormael, K., Ceesay, P., Anderson, D., & Snavely, D. (2010, Feb). Effect of Gaboxadol on Patient-reported Measures of Sleep and Waking Function in Patients with Primary Insomnia: Results from Two Randomized, Controlled, 3-month Studies. *Journal of Clinical Sleep Medicine, 6*(1), 30-39. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2823273/pdf/jcsm.6.1.30.pdf

Roth, T., Seiden, D., Sainati, S., Wang-Weigand, S., Zhang, J., & Zee, P. (2006, Jun). Effects of ramelteon on patient-reported sleep latency in older adults with chronic insomnia. *Sleep Medicine, 7*(4), 312-318.

http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med5&AN=16709464

Roth, T., Soubrane, C., Titeux, L., & Walsh, J. K. (2006). Efficacy and safety of zolpidem-MR: A double-blind, placebo-controlled study in adults with primary insomnia. *Sleep Medicine, 7*(5), 397-406. https://doi.org/http://dx.doi.org/10.1016/j.sleep.2006.04.008

Roth, T. G., Roehrs, T. A., Koshorek, G. L., Greenblatt, D. J., & Rosenthal, L. D. (1997). Hypnotic effects of low doses of quazepam in older insomniacs. *Journal of Clinical Psychopharmacology, 17*(5), 401-406. https://doi.org/http://dx.doi.org/10.1097/00004714-199710000-00009

Sanofi-Aventis SA. (2010). *Efficacy and safety of eplivanserin 5 mg/day on sleep maintenance insomnia: a 12-week multicenter, randomized, double-blind, placebo-controlled study followed by an open label extension phase with eplivanserin for 40-week period.*

Schadeck, B., Chelly, M., Amsellem, D., Cohen, A., Peraudeau, P., & Scheck, F. (1996, Jun). Comparative efficacy of doxylamine (15 mg) and zolpidem (10 mg) for the treatment of common insomnia. A placebo-controlled study (vol 72, pg 428, 1996). *Semaine des Hopitaux, 72*(21-22), 649-649.

Scharf, M. B., Black, J., Hull, S., Landin, R., & Farber, R. (2007, Jun). Long-term nightly treatment with indiplon in adults with primary insomnia: results of a double-blind, placebo-controlled, 3-month study. *Sleep, 30*(6), 743-752. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med5&AN=17580596

Scharf, M. B., Roth, P. B., Dominguez, R. A., & Ware, J. C. (1990, May). Estazolam and flurazepam: a multicenter, placebo-controlled comparative study in outpatients with insomnia. *Journal of Clinical Pharmacology, 30*(5), 461-467. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med3&AN=1971831

Scholey, A., Benson, S., Gibbs, A., Perry, N., Sarris, J., & Murray, G. (2017). Exploring the Effect of Lactium™ and Zizyphus Complex on Sleep Quality: a Double-Blind, Randomized Placebo-Controlled Trial. *Nutrients, 9*(2). http://cochranelibrary-wiley.com/o/cochrane/clcentral/articles/258/CN-01401258/frame.html

Sivertsen, B., Omvik, S., Pallesen, S., Bjorvatn, B., Havik, O. E., Kvale, G., Nielsen, G. H., & Nordhus, I. H. (2006, Jun 28). Cognitive behavioral therapy vs zopiclone for treatment of chronic primary insomnia in older adults: a randomized controlled trial. *JAMA, 295*(24), 2851-2858. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med5&AN=16804151

Takeda Pharmaceuticals North America Inc. (2008). *A Phase III, Randomized, Double-Blind, Placebo-Controlled, Outpatient, Safety and Efficacy Study of TAK-375 in Adults With Chronic Insomnia*.

Uchimura, N., Ogawat, A., Hamamura, M., Hashimoto, T., Nagata, H., & Uchiyama, M. (2011). Efficacy and safety of ramelteon in Japanese adults with chronic insomnia: A randomized, double-blind, placebo-controlled study. *Expert Review of Neurotherapeutics, 11*(2), 215-224. https://doi.org/http://dx.doi.org/10.1586/ern.10.197

Uchiyama, M., Hamamura, M., Kuwano, T., Nishiyama, H., Nagata, H., & Uchimura, N. (2011, Feb). Evaluation of subjective efficacy and safety of ramelteon in Japanese subjects with chronic insomnia. *Sleep Medicine, 12*(2), 119-126. https://doi.org/https://dx.doi.org/10.1016/j.sleep.2010.08.010

Vorbach, E. U., Gortelmeyer, R., & Bruning, J. (1996, August). Therapy of insomnia. The efficacy and tolerability of valerian. [German] [Therapie von Insomnien. Wirksamkeit und Vertraglichkeit eines Baldrianpraparats.]. *Psychopharmakotherapie, 3*(3), 109-115. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed7&AN=26278238

Wade, A. G., Ford, I., Crawford, G., McConnachie, A., Nir, T., Laudon, M., & Zisapel, N. (2010, 16 Aug). Nightly treatment of primary insomnia with prolonged release melatonin for 6 months: A randomized placebo controlled trial on age and endogenous melatonin as predictors of efficacy and safety. *BMC Medicine, 8 (no pagination)*(51). https://doi.org/http://dx.doi.org/10.1186/1741-7015-8-51

Walsh, J. K., Erman, M., Erwin, C. W., Jamieson, A., Mahowald, M., Regestein, Q., Scharf, M., Tigel, P., Vogel, G., & Ware, J. C. (1998). Subjective hypnotic efficacy of trazodone and zolpidem in DSMIII-R primary insomnia. *Human Psychopharmacology: Clinical and Experimental, 13*(3), 191-198. https://doi.org/http://dx.doi.org/10.1002/%28SICI%291099-1077%28199804%2913:3%3C191::AID-HUP972%3E3.0.CO;2-X

Walsh, J. K., & et al. (1984). A multi-center clinical investigation of estazolam: Short-term efficacy. *Current Therapeutic Research, 36*(5, Sect 2), 866-874. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc2&NEWS=N&AN=1985-26213-001

Walsh, J. K., Fry, J., Erwin, C. W., Scharf, M., Roth, T., & Vogel, G. W. (1998). Efficacy and tolerability of 14-day administration of zaleplon 5 mg and 10 mg for the treatment of primary insomnia. *Clinical Drug Investigation, 16*(5), 347-354. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed7&AN=28526380

Walsh, J. K., Krystal, A. D., Amato, D. A., Rubens, R., Caron, J., Wessel, T. C., Schaefer, K., Roach, J., Wallenstein, G., & Roth, T. (2007). Nightly treatment of primary insomnia with eszopiclone for six months: Effect on sleep, quality of life, and work limitations. *Sleep: Journal of Sleep and Sleep Disorders Research, 30*(8), 959-968. https://doi.org/http://dx.doi.org/10.1093/sleep/30.8.959

Walsh, J. K., Moscovitch, A., Burke, J., Farber, R., & Roth, T. (2007, Nov). Efficacy and tolerability of indiplon in older adults with primary insomnia. *Sleep Medicine, 8*(7-8), 753-759. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med5&AN=17825616

Walsh, J. K., Salkeld, L., Knowles, L. J., Tasker, T., & Hunneyball, I. M. (2010, Jan). Treatment of elderly primary insomnia patients with EVT 201 improves sleep initiation, sleep maintenance, and daytime sleepiness. *Sleep Medicine, 11*(1), 23-30. https://doi.org/https://dx.doi.org/10.1016/j.sleep.2009.07.012

Walsh, J. K., Soubrane, C., & Roth, T. (2008). Efficacy and safety of zolpidem extended release in elderly primary insomnia patients. *Special Issue: Clinical studies for late-life disorders, 16*(1), 44-57. https://doi.org/http://dx.doi.org/10.1097/JGP.0b013e3181256b01

Walsh, J. K., Vogel, G. W., Scharf, M., Erman, M., Erwin, C. W., Schweitzer, P. K., Mangano, R. M., & Roth, T. (2000, Feb). A five week, polysomnographic assessment of zaleplon 10 mg for the treatment of primary insomnia. *Sleep Medicine, 1*(1), 41-49. https://doi.org/10.1016/S1389-9457(99)00006-4

Wang-Weigand, S., Watissee, M., & Roth, T. (2011). Use of a post-sleep questionnaire-interactive voice response system (PSQ-IVRS) to evaluate the subjective sleep effects of ramelteon in adults with chronic insomnia. *Sleep Medicine, 12*(9), 920-923. https://doi.org/http://dx.doi.org/10.1016/j.sleep.2011.06.008

Wheatley, D. (1989). Zolpidem: a new imidazopyridine hypnotic. *Psychopharmacology Bulletin, 25*(1), 124-127. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med3&AN=2672067

Winsauer, H. J., O'Hair, D. E., & Valero, R. (1984). Quazepam: Short-term treatment of insomnia in geriatric outpatients. *Current Therapeutic Research, 35*(2), 228-234. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc2&NEWS=N&AN=1985-02016-001

Xu, Z., Jiang, X., Li, W., Gao, D., Li, X., & Liu, J. (2011, Jul). Propofol-induced sleep: efficacy and safety in patients with refractory chronic primary insomnia. *Cell Biochemistry & Biophysics, 60*(3), 161-166. https://doi.org/https://dx.doi.org/10.1007/s12013-010-9135-7

Zammit, G., Erman, M., Wang-Weigand, S., Sainati, S., Zhang, J., & Roth, T. (2007, Aug 15). Evaluation of the efficacy and safety of ramelteon in subjects with chronic insomnia.[Erratum appears in J Clin Sleep Med. 2008 Oct 15;4(5):table of contents], [Erratum appears in J Clin Sleep Med. 2007 Oct 15;3(6):table of contents]. *Journal of Clinical Sleep Medicine, 3*(5), 495-504. http://ezproxy.library.usyd.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med5&AN=17803013

Zammit, G. K., McNabb, L. J., Caron, J., Amato, D. A., & Roth, T. (2004, December). Efficacy and safety of eszopiclone across 6-weeks of treatment for primary insomnia. *Current Medical Research and Opinion, 20*(12), 1979-1991. https://doi.org/http://dx.doi.org/10.1185/174234304X15174

Zick, S. (2013). *Chamomile for Chronic Primary Insomnia*. https://clinicaltrials.gov/ct2/show/NCT01286324?term=drug&recrs=eh&type=Intr&cond=Insomnia&intr=Drug&age=12&draw=3

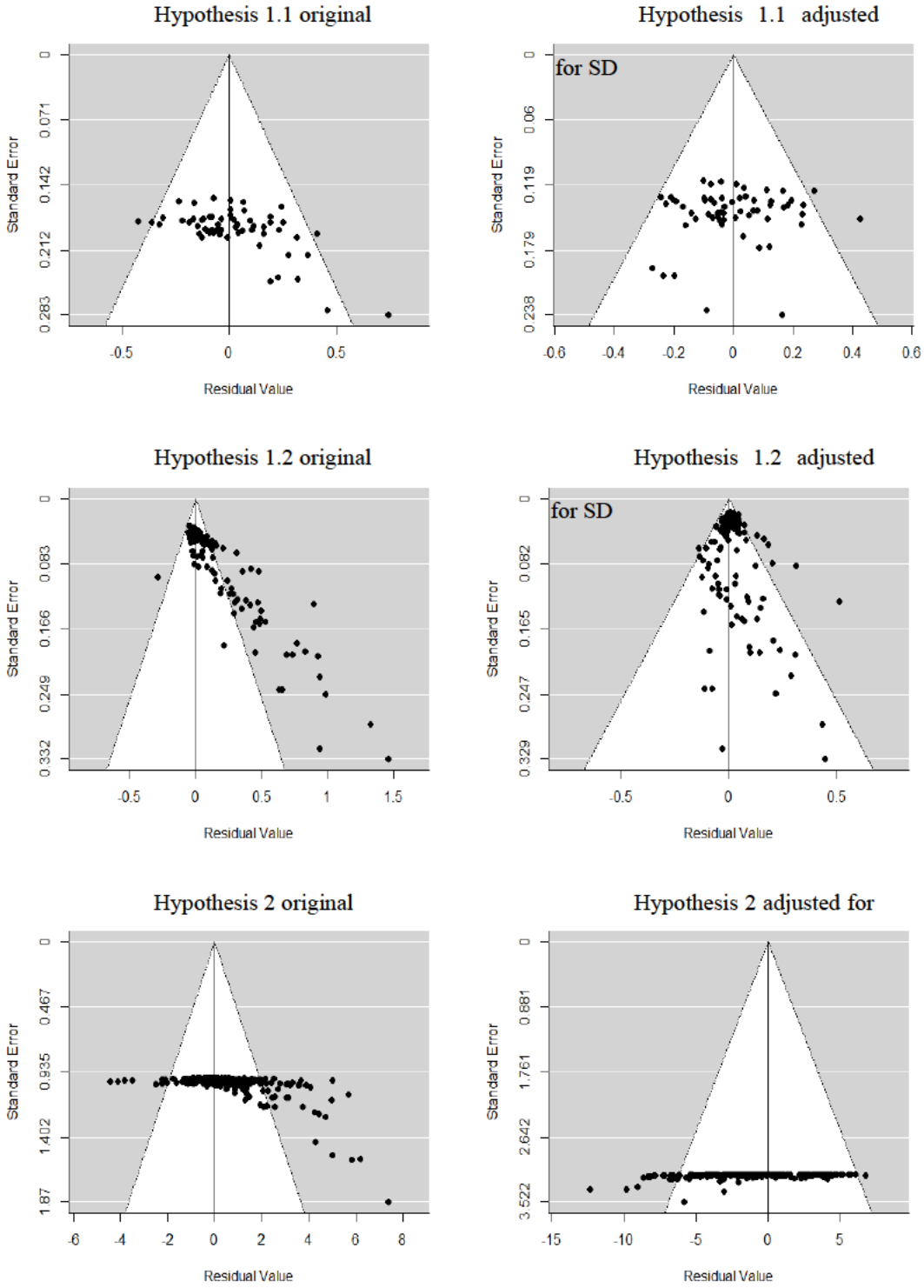
Publication Bias

In accordance with the three funnel plots in the right column in eFigure1 the following three R outputs represent the three analyses calculated to account for the publication bias. We ran the main analyses with the standard deviation of the effect size as an additional predictor to account for publication bias. Adding the standard deviation as an additional predictor follows the logic of Egger's regression test (Sterne & Egger, 2005). When the intercept of these regression analyses deviated statistically significantly ($p \leq .05$) from zero, the overall association between the precision and the size of studies included in the analysis is considered asymmetrical and therefore biased.

As can be seen in the three R outputs below adding the additional predictor to adjust for the study imprecision resolved all issues with asymmetry in the first analysis. There was still some statistically significant asymmetry present, as can be seen by the statistically significant intercept term in the model. Although the asymmetry in the latter two analyses is still statistically significant the extent was reduced to a large extent compared to the original analysis, as can be observed when comparing the funnel plots in the right column that represent the adjusted analyses compared with the funnel plots on the left representing the original analyses. For the interpretation of the results, it is essential to see that the factors of interest remained statistically significant. This means that it can be ruled out that the effects observed in the main analyses in the Result section were not caused by the presence of a publication bias.

eFigure 1

Publication Bias: Visual Analysis using Funnel Plots



Note. This illustration shows funnel plots for the three main analyses. In conventional meta-analytical investigations funnel plots are a form of scatter plots that show the effect sizes against a measure of study precision and are primarily a visual aid to assess publication bias. Ideally, one would expect to see symmetrically distributed dots in an inverted funnel shape, whereas an asymmetric distribution on the x-axis might indicate the presence of a publication bias. Here we have plotted funnel plots for multilevel / multivariate (mixed-effects) linear models with the standard error as the measure of precision on the y-axis and the residuals of the regression, instead of the effect sizes on the x-axis. Therefore, they might look slightly unusual. In the left column are the funnel plots for the original analyses as reported in the Result section and the right column shows these models adjusted with the standard deviation of the predicted outcome measures as an additional predictor to adjust for study precision. The adjusted models visibly improved the symmetry of the funnel plots.

Hypothesis 1.1 analysis adjusted with the standard deviation of the outcome measure to account for a potential publication bias.

```
> summary(rma.uni(IR_AEn_es~as.factor(drug_category_drug1)+sqrt(IR_AEn_var), IR_AEn_var, data =
subset(d_h2h, drug_category_drug1 %in% c(1,2,3,4,6,10,11)), method = "REML"))

Mixed-Effects Model (k = 61; tau^2 estimator: REML)

  logLik  deviance       AIC       BIC      AICc
 24.3669  -48.7337  -30.7337  -13.0011  -26.5477

tau^2 (estimated amount of residual heterogeneity):     0.0190 (SE = 0.0044)
tau (square root of estimated tau^2 value):             0.1379
I^2 (residual heterogeneity / unaccounted variability): 91.50%
H^2 (unaccounted variability / sampling variability):   11.76
R^2 (amount of heterogeneity accounted for):            56.79%

Test for Residual Heterogeneity:
QE(df = 53) = 570.3475, p-val < .0001

Test of Moderators (coefficients 2:8):
QM(df = 7) = 73.4120, p-val < .0001

Model Results:

                            estimate      se     zval    pval    ci.lb   ci.ub
intrcpt                      -0.0882  0.0585  -1.5090  0.1313  -0.2028  0.0264
as.factor(drug_category_drug1)2     0.3237  0.0550   5.8826  <.0001   0.2159  0.4316  ***
as.factor(drug_category_drug1)3     0.2461  0.0635   3.8745  0.0001   0.1216  0.3706  ***
as.factor(drug_category_drug1)4     0.1805  0.0938   1.9235  0.0544  -0.0034  0.3644    .
as.factor(drug_category_drug1)6     0.2759  0.0787   3.5051  0.0005   0.1216  0.4302  ***
as.factor(drug_category_drug1)10    0.1282  0.0876   1.4638  0.1433  -0.0435  0.2999
as.factor(drug_category_drug1)11    0.3408  0.0726   4.6959  <.0001   0.1986  0.4830  ***
sqrt(IR_AEn_var)                    4.0089  0.6159   6.5088  <.0001   2.8018  5.2161  ***
```

Hypothesis 1.2 analysis adjusted with the standard deviation of the outcome measure to account for a potential publication bias.

```
> summary(rma.mv(IR_AEspec_es~as.factor(drug_category)+sqrt(IR_AEspec_var), IR_AEspec_var,
random = ~ factor(group_AE_spec_all) | factor(code), method = "REML", data = subset(d_h2h_h1.1,
drug_category %in% c(1,2,3,4,6,9,10,11))))

Multivariate Meta-Analysis Model (k = 697; method: REML)

   logLik    Deviance         AIC         BIC        AICc
 2000.3027  -4000.6055  -3978.6055  -3928.7338  -3978.2149

Variance Components:

outer factor: factor(code)             (nlvls = 70)
inner factor: factor(group_AE_spec_all) (nlvls = 33)
          estim    sqrt  fixed
```

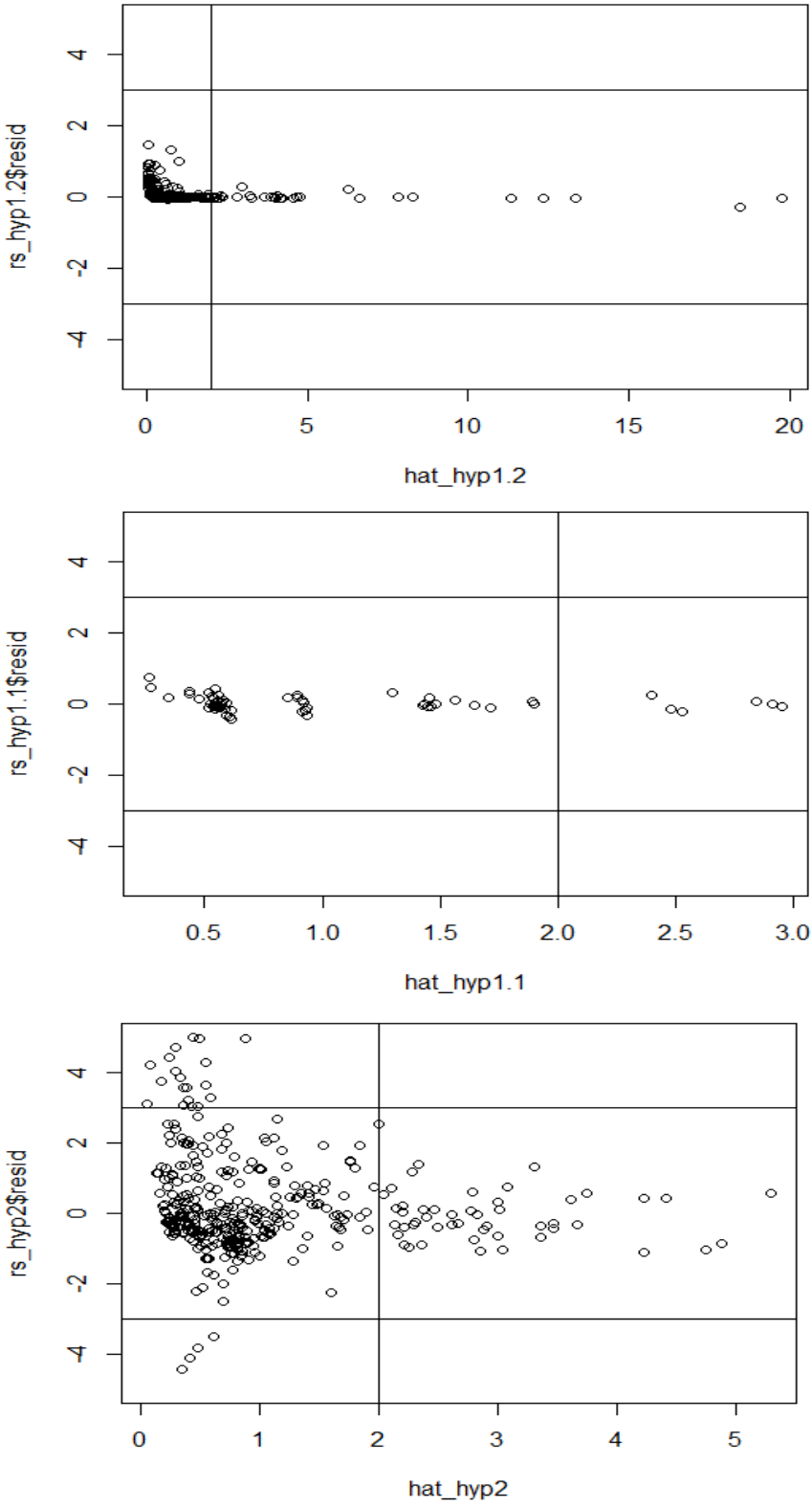
```
tau^2      0.0004  0.0196     no
rho        1.0000             no

Test for Residual Heterogeneity:
QE(df = 688) = 698.3526, p-val = 0.3837

Test of Moderators (coefficients 2:9):
QM(df = 8) = 1554.5142, p-val < .0001

Model Results:

                          estimate      se     zval    pval    ci.lb    ci.ub
intrcpt                    -0.0573  0.0064  -8.9467  <.0001  -0.0699  -0.0448  ***
as.factor(drug_category)2   0.0363  0.0074   4.9073  <.0001   0.0218   0.0508  ***
as.factor(drug_category)3   0.0431  0.0088   4.8966  <.0001   0.0259   0.0604  ***
as.factor(drug_category)4   0.0399  0.0130   3.0639  0.0022   0.0144   0.0654   **
as.factor(drug_category)6   0.0343  0.0109   3.1476  0.0016   0.0129   0.0556   **
as.factor(drug_category)9   0.0480  0.0569   0.8440  0.3987  -0.0635   0.1594
as.factor(drug_category)10  0.0279  0.0156   1.7882  0.0737  -0.0027   0.0585    .
as.factor(drug_category)11  0.0380  0.0079   4.7995  <.0001   0.0225   0.0536  ***
sqrt(IR_AEspec_var)         3.4373  0.0902  38.1015  <.0001   3.2605   3.6142  ***
```

Hypothesis 2 analysis adjusted with the standard deviation of the outcome measure to account for a potential publication bias.

```
> summary(rma.mv(ES_overall, VAR_overall,  mods=AE_freq_overall+sqrt(VAR_overall), random=~
ES|code, data = d_h2_pla_all))

Multivariate Meta-Analysis Model (k = 382; method: REML)

  logLik   Deviance        AIC        BIC       AICc
-862.4723  1724.9446  1732.9446  1748.7052  1733.0512

Variance Components:

outer factor: code (nlvls = 49)
inner factor: ES   (nlvls = 8)

           estim    sqrt  fixed
tau^2     10.1441  3.1850     no
rho        0.9647             no

Test for Residual Heterogeneity:
QE(df = 380) = 15817.6525, p-val < .0001

Test of Moderators (coefficient 2):
QM(df = 1) = 1995.3528, p-val < .0001

Model Results:

        estimate      se      zval    pval    ci.lb    ci.ub
intrcpt   -5.3975  0.4743  -11.3800  <.0001  -6.3271  -4.4679  ***
mods      11.6531  0.2609   44.6694  <.0001  11.1418  12.1644  ***
```

eFigure 2

Scatterplots showing Standardized Residuals and Hat Values



Note. The three scatterplots above show the statistical model's standardized residuals on the y-axis and the hat value on the x-axis. Any point having a standardized residual above three or below minus three and a hat value higher than two is considered as an outlier / influential data.

Appendix B – Additional Materials Study 1

Ethics Approval Study 1

Research Integrity & Ethics Administration
Human Research Ethics Committee

Friday, 2 March 2018

Dr Ben Colagiuri
Psychology; Faculty of Science
Email: ben.colagiuri@sydney.edu.au

Dear Ben

The University of Sydney Human Research Ethics Committee (HREC) has considered your application.

I am pleased to inform you that your project has been approved.

Approval is granted for a period of four years from **02 March 2018** to **02 March 2022**

Project title: **The development and evaluation of a new model of perceptible placebos**

Project no.: **2018/107**

First Annual Report due: **02 March 2019**

Authorised Personnel: **Colagiuri Ben; Werner Christoph;**

Documents Approved:

Date Uploaded	Version number	Document Name
29/01/2018	Version 1	Advertisement GENERAL study1&2- perceptible placebo
29/01/2018	Version 1	Advertisement GENERAL study3- perceptible placebo
29/01/2018	Version 1	CSD
29/01/2018	Version 1	DASS-21
29/01/2018	Version 1	DASS-21 publication
29/01/2018	Version 1	Debrief & delayed consent - perceptible placebo
29/01/2018	Version 1	Figure 1 Study Design
29/01/2018	Version 1	GASE
29/01/2018	Version 1	GASE_additional items
29/01/2018	Version 1	human-ethics-pcf - perceptible placebo
29/01/2018	Version 1	human-ethics-pis visit1 study1&2- perceptible placebo
29/01/2018	Version 1	human-ethics-pis visit1 study3- perceptible placebo
29/01/2018	Version 1	human-ethics-pis visit2 drug information placebo&perceptible
29/01/2018	Version 1	human-ethics-pis visit2 information no-treatment group
29/01/2018	Version 1	human-ethics-pis visit2 information study 3
29/01/2018	Version 1	ISI
29/01/2018	Version 1	PSM
29/01/2018	Version 1	PSQI
29/01/2018	Version 1	TSQM-II
29/01/2018	Version 1	WHOQoL-BREF

Research Integrity & Ethics Administration
Level 2, Margaret Telfer Building (K07)
The University of Sydney
NSW 2006 Australia

T +61 2 9036 9161
E human.ethics@sydney.edu.au
W sydney.edu.au/ethics

ABN 15 211 513 464
CRICOS 00026A

THE UNIVERSITY OF SYDNEY

Special Condition/s of Approval

1. As this is a PhD project, please include the Supervisor's name on the study advertisement.
2. Please provide the Compounding Pharmacy certification for Good Manufacturing Practice (GMP). The Clinical Trials Governance Office can provide guidance if required.

It is noted that in the response to Q351/352 that data may be used to inform future studies. Please be reminded that use of the data outside the aims described in this application, or sharing of this data with other researchers will require separate ethics approval.

Condition/s of Approval

- Research must be conducted according to the approved proposal.
- An annual progress report must be submitted to the Ethics Office on or before the anniversary of approval and on completion of the project.
- You must report as soon as practicable anything that might warrant review of ethical approval of the project including:
  - Serious or unexpected adverse events (which should be reported within 72 hours).
  - Unforeseen events that might affect continued ethical acceptability of the project.
- Any changes to the proposal must be approved prior to their implementation (except where an amendment is undertaken to eliminate *immediate* risk to participants).
- Personnel working on this project must be sufficiently qualified by education, training and experience for their role, or adequately supervised. Changes to personnel must be reported and approved.
- Personnel must disclose any actual or potential conflicts of interest, including any financial or other interest or affiliation, as relevant to this project.
- Data and primary materials must be retained and stored in accordance with the relevant legislation and University guidelines.
- Ethics approval is dependent upon ongoing compliance of the research with the *National Statement on Ethical Conduct in Human Research*, the *Australian Code for the Responsible Conduct of Research*, applicable legal requirements, and with University policies, procedures and governance requirements.
- The Ethics Office may conduct audits on approved projects.
- The Chief Investigator has ultimate responsibility for the conduct of the research and is responsible for ensuring all others involved will conduct the research in accordance with the above.

This letter constitutes ethical approval only.

Please contact the Ethics Office should you require further information or clarification.

Sincerely

[redacted]

THE UNIVERSITY OF SYDNEY

Professor Glen Davis
Chair
Human Research Ethics Committee (HREC 2)

The University of Sydney HRECs are constituted and operate in accordance with the National Health and Medical Research Council's (NHMRC) National Statement on Ethical Conduct in Human Research (2007) and the NHMRC's Australian Code for the Responsible Conduct of Research (2007).

Thursday, May 31, 2018 at 9:51:25 AM Australian Eastern Standard Time

Subject: [2018/107] Human Ethics: Compliance with special conditions outcome
Date: Monday, 7 May 2018 at 4:45:03 pm Australian Eastern Standard Time
From: Human Ethics
To: Ben Colagiuri
CC: Christoph Werner

Dear Dr Colagiuri

Project Title: The development and evaluation of a new model of perceptible placebos
Project number: 2018/107

Compliance with Special Conditions Outcome

Thank you for providing clarification addressing the following special condition(s) of approval:

Please provide the Compounding Pharmacy certification for Good Manufacturing Practice (GMP).

The Committee noted your response with thanks and are pleased to advise that you will not be required to fulfil the condition stated as above.

You should retain a copy of this email with your study records.

Please contact us if you have any queries.

Regards,
The Ethics Office

Research Integrity & Ethics Administration | Research Portfolio
THE UNIVERSITY OF SYDNEY
Level 2 Margaret Telfer Building (K07) | The University of Sydney | NSW | 2006
T +61 2 9036 9161 | **E** human.ethics@sydney.edu.au | **W** http://sydney.edu.au/ethics

THE UNIVERSITY OF SYDNEY

Research Integrity & Ethics Administration
HUMAN RESEARCH ETHICS COMMITTEE

Thursday, 30 August 2018

Dr Ben Colagiuri
Psychology; Faculty of Science
Email: ben.colagiuri@sydney.edu.au

Dear Ben,

Your request to modify this project, which was submitted on 16/07/2018, has been considered.

After consideration of your response to the comments raised, this project has been approved to proceed with the proposed amendments.

Details of the approval are as follows:

Project No.: **2018/107**
Project Title: **The development and evaluation of a new model of perceptible placebos**

Next Annual Report due: 02/03/2019

New Approved Documents:

Date Uploaded	Type	Document Name
22/08/2018	Advertisements/Flyer	02a_Advertisement SONA_PAID study1&2-perceptible placebo
22/08/2018	Advertisements/Flyer	02b_Advertisement SONA_PAID study3-perceptible placebo
22/08/2018	Participant Info Statement	final_04a_human-ethics-pis visit1 study1&2
22/08/2018	Participant Info Statement	final_04b_human-ethics-pis visit1 study3

Please contact the ethics office should you require further information.

Sincerely,

[redacted]

Dr Clifton Chan
Chair, Modification Review Committee (MRC 2)

The University of Sydney of Sydney HRECs are constituted and operate in accordance with the National Health and Medical Research Council's (NHMRC) National Statement on Ethical Conduct in Human Research (2007) and the NHMRC's Australian Code for the Responsible Conduct of Research (2007)

THE UNIVERSITY OF SYDNEY

Research Integrity & Ethics Administration
Human Research Ethics Committee

Wednesday, 8 August 2018

Dr Ben Colagiuri
Psychology; Faculty of Science
Email: ben.colagiuri@sydney.edu.au

Dear Ben

Your request to modify this project, which was submitted on 25/06/2018, has been considered.

After consideration of your response to the comments raised the project has been approved to proceed with the proposed amendments.

Details of the approval are as follows:

Project Title: **The development and evaluation of a new model of perceptible placebos**

Project No.: **2018/107**

New Approved Documents:

Date	Type	Document
25/06/2018	Advertisements/Flyer	final_02a_Advertisement GENERAL study1&2_v1.2
25/06/2018	Advertisements/Flyer	final_02b_Advertisement GENERAL study3_v1.2
25/06/2018	Participant Info Statement	final_04a_human-ethics-pis visit1 study1&2_v1.1
25/06/2018	Participant Info Statement	final_04b_human-ethics-pis visit1 study3_v1.1

Please contact the Ethics Office should you require further information or clarification.

Sincerely

[redacted]

Associate Professor Stephen Fuller
Chair
Modification Review Committee Chair (MRC 3)

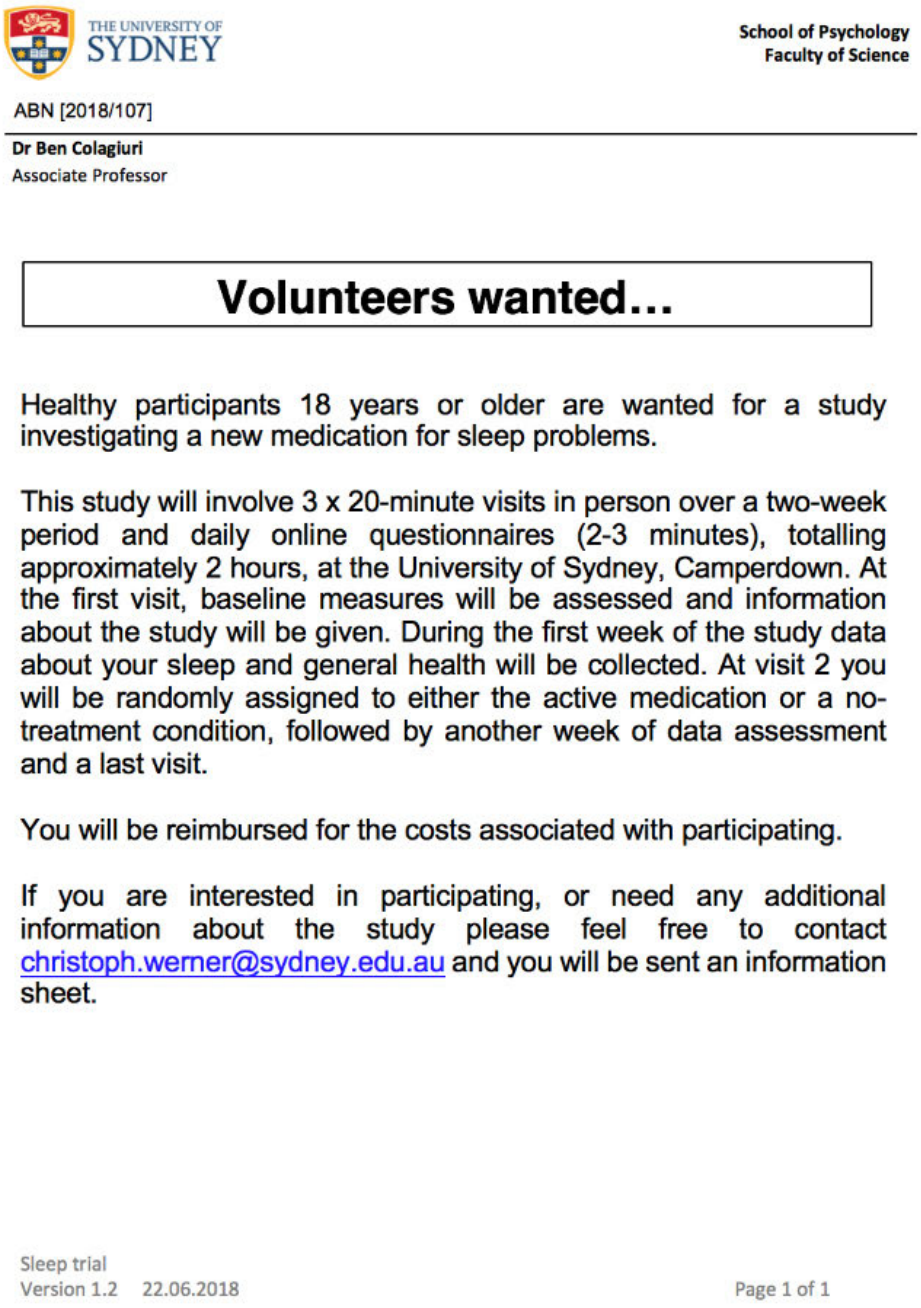
The University of Sydney HRECs are constituted and operate in accordance with the National Health and Medical Research Council's (NHMRC) National Statement on Ethical Conduct in Human Research (2007) and the NHMRC's Australian Code for the Responsible Conduct of Research (2007).

Research Integrity & Ethics Administration
Research Portfolio
Level 2, Margaret Telfer Building (K07)
The University of Sydney
NSW 2006 Australia

T +61 2 9036 9161
E human.ethics@sydney.edu.au
W sydney.edu.au/ethics

ABN 15 211 513 464
CRICOS 00026A

Advertisement Flyer Study 1



THE UNIVERSITY OF SYDNEY

School of Psychology
Faculty of Science

ABN [2018/107]

Dr Ben Colagiuri
Associate Professor

Volunteers wanted…

Healthy participants 18 years or older are wanted for a study investigating a new medication for sleep problems.

This study will involve 3 x 20-minute visits in person over a two-week period and daily online questionnaires (2-3 minutes), totalling approximately 2 hours, at the University of Sydney, Camperdown. At the first visit, baseline measures will be assessed and information about the study will be given. During the first week of the study data about your sleep and general health will be collected. At visit 2 you will be randomly assigned to either the active medication or a no-treatment condition, followed by another week of data assessment and a last visit.

You will be reimbursed for the costs associated with participating.

If you are interested in participating, or need any additional information about the study please feel free to contact christoph.werner@sydney.edu.au and you will be sent an information sheet.

Sleep trial
Version 1.2 22.06.2018

Page 1 of 1

Experimenter Manual Study 1

THE UNIVERSITY OF SYDNEY

School of Psychology
Faculty of Science

ABN [2018/107]

Experimenter Manual Sleep Study 1

First visit:

✓ Preparation finished according to check list

Hello, my name is XXX. I'm part of the sleep study and today I will introduce you to the study. Please follow me to the consulting room. Please take a seat. Can I offer you a glass of water?

Guide participant to room.

Before we start with the study I want to explain the overall procedure and the goals of the study and want to emphasise that it is important for us that you understand everything. So, if you have any questions please feel free to interrupt.
As you've hopefully already read the goal of our study is to test a new compound of an antihistamine (doxylamine) and an antioxidant (beetroot extract). Effective antihistamines are already available over-the-counter to treat sleep difficulties. Previous research demonstrated that our digestive system is not able to absorb the total amount of the pharmacological agent included in pills and capsules due to the "unfriendly" environment in our stomach and colon. Microbiota and enzymes already start to degrade the agents before they are absorbed into our blood stream, therefore hindering the full potential of the drug.
To address this problem we added an antioxidant, which will protect the antihistamines against the "unfriendly" environment and guarantee full absorption which will result in better sleep.
Do you have any questions so far?
Alright, if everything is clear let's move on to the procedure. As you've already read this study covers a two-week period and includes three short consultations. In the first week we will assess your sleep quality and other aspects as a baseline, so we can compare it to the second week, in which you will either take four capsules of our new compound each evening or just continue as in the first week, if you have been randomised to the no-treatment group. The randomisation will happen at the next consultation. During the two weeks you will be required to answer short questionnaires as well as wearing the Actiwatch on a daily basis.
Do you have any questions about the procedure?
If not, we would proceed to the informed consent form.

Hand participant information sheet and informed consent form to participant and go through them.

Please read these documents thoroughly if you haven't done so already and please feel free to ask for clarifications if needed. Please fill out these details and sign the document if all your questions have been answered to your satisfaction.

Explain if something is unclear. Ask them to fill out and sign the consent form.

Alright, thank you very much for participating in our research study. I now would like to ask you to fill in the following questionnaires until it states that you are finished with this block.

Let participant fill out questionnaires.

Now I want to explain the Actiwatch and what you have to do over the next two weeks.

Hand over Actiwatch

This is the Actiwatch, it includes a sensor that measures movements, which we will use as an objective measure besides your sleep ratings. Please put it on now. It doesn't matter on which wrist you are wearing it. Please push the button when you go to bed and get up again. It is important, that you always wear the watch, except when you take a bath, shower or go for a swim. If you should experience any problems with the device, please contact Mr Christoph Werner via the known email address.
Besides wearing the Actiwatch we would like you to fill out short surveys twice a day. In the morning when you get up we would like you to fill out the CSD-C, which is asking questions about your sleep. In the evening we want you to answer questions about your daily symptoms and some general questions about your daily intake of liquids, cigarettes etc. These data are important since these substances can influence your sleep.
You will receive an email in the morning and evening containing the links to the online surveys. The one in the morning will be sent to you at 5am with a reminder email at 10 am and the email in the evening will be sent at 8pm with a reminder at 11 pm.
Please keep up with the scheduled online surveys, since this is very important for the quality of the data and could jeopardise the quality of our study to the point that we are not able to use your inputs. We are able to see if and when you did the online surveys and if you fail to fill them out properly we have to exclude you from the study.

Mention that they should use the 24-hour format (e.g. 1115pm = 2315).

Alright, if you do not have any questions we are done for today and I can show you the way out.
Thank you very much for participating in our study. Have a nice day and see you next week.

Guide/show participant the way out.

Second visit:

- ✓ Preparation finished according to check list

Hello and welcome back. Please follow me to the consulting room. Please take a seat. Can I offer you a glass of water?

Guide participant to room.

I hope you had a good week and everything worked well. Did you experience any problems with the Actiwatch or the surveys? If not, I would like you to fill in the same questionnaires like last week. Please progress until you are told to inform the experimenter that you finished the first block.

Since we like to do the randomisation process the old-fashioned way I would like you to draw a number from this envelope. You will have to enter this number in the next survey.

Let participant draw a number.

Drug Group [2,4,5,6,7,9]

I would like to inform you, that you have been randomised to the drug group of our study. This means that you have to take 4 capsules of our newly developed formulation of antihistamine and beetroot approximately one hour before going to bed each evening for the next seven days, starting this evening. Since you will get a drug we have to inform you about potential side effects. Therefore, please read this information thoroughly. You can keep it, if you want.

Hand out drug information sheet and capsules to participant.

Do you have any questions about your assignment to the drug group, the drug itself or possible side effects?
I just quickly want to focus on the red colouration of urine. It is very likely that your urine, especially in the morning after taking the capsules will be a little darker or have a slight reddish colour. This is due to the beetroot that is contained in the capsules and is absolutely harmless. So, you don't have to worry about it. Please do not forget to mention this in the evening survey about your daily symptoms, if you experienced it.
Just in case you would experience fever, or pain in the kidney area together with red urine, you have to let us know, since this would not be caused by our medication. In this extremely unlikely case of these two symptoms it's likely that you by chance got a kidney infection, which as well as beetroot can cause red urine. I just tell you about this as a precaution, as said this is nothing that can be caused by our medication. But we have to inform you about the possibility since a kidney infection too causes red urine.
If you should experience side effects that are too unpleasant you are allowed to reduce the dose for the next evening from 4 to e.g. 2 capsules. If 2 capsules are no problem then please start to increase the dose again. If 2 capsules are still to unpleasant you can further reduce the dose. But the goal is still to take 4 capsules.
If you do not have any questions, I would like you to answer the last questionnaires for today's visit.

Highlight (1)alcohol, (2) water, (3) empty stomach, (4) 1hour. Let participant fill out questionnaires. Store number in drawer.

Alright that was it for today. Here are your capsules and a document stating that you take part in a pharmacological study just in case you need to explain why you are carrying these capsules around. This is just a precaution, since approximately 3 or 4 years ago had a sports coach suspecting his student to take doping medications.

Hand out participation statement.

No-treatment Group [1,3,8]

I would like to inform you, that you have been randomised to the no-treatment group of our study. This essentially means that you do not have to take any medications and act as a control group so we can control for the natural course of sleep. Please read this information. You can keep it, if you want.

Hand out no-treatment information sheet to participant.

Do you have any questions about your assignment to the no-treatment control?
If not, I would like you to answer some last questions for today's visit.

Let participant fill out questionnaire. Store number in drawer.

I would like to remind you again, please keep up with the scheduled online surveys, since this is very important for the quality of the data.
Alright, if you do not have any last questions we are done for today and I can show you the way out.
Thank you very much for participating in our study. Have a nice day and see you next week.

Guide/show participant the way out.

Third visit:

- ✓ Preparation finished according to check list

Hello. Please follow me to the consulting room. Please take a seat. Can I offer you a glass of water?

Guide participant to room.

I hope you had a good week. Could I please have the Actiwatch and the capsule bottle back, if you got any of them? Thank you very much.
Did you experience any problems with the Actiwatch or the surveys? If not, I would like you to fill in some questionnaires. Please progress until you reach a section, that the study team needs to fill out.

Collect Actiwatch plus bottle. Let participant fill out questionnaires, retrieve data from Actiwatch and label the capsule bottle with the ID and group number.

Thank you very much for filling out the last questionnaires of this study. Now I want you to please read this debrief information.

Give participant debrief info and let them read.

SONA participant

Do you still have questions after reading this information?
If not, we are at the end of the study. You will get your credits for participating in this study within the next days through SONA. Thank you very much for participating in our study. Please keep in mind to not talk about the actual topic of this study with potential participants, because it could compromise the validity of our study. Have a nice day.

Finish completion section on REDCap and give credits on SONA.

Paid participant

Do you still have questions after reading this information?
If not, we are at the end of the study. Here are your 50$, please fill out this list and sign that you've received the reimbursement. Thank you very much for participating in our study. Please keep in mind to not talk about the actual topic of this study with potential participants, because it could compromise the validity of our study. Have a nice day.

Hand out Reimbursement to participant, let him sign the receipt and finish completion section on REDCap.

Appendix C – Additional Materials Study 2

Ethics Approval Study 2

THE UNIVERSITY OF SYDNEY

Research Integrity & Ethics Administration
HUMAN RESEARCH ETHICS COMMITTEE

Monday, 26 November 2018

Dr Ben Colagiuri
Psychology, Faculty of Science
Email: ben.colagiuri@sydney.edu.au

Dear Ben,

Your request to modify this project, which was submitted on 29/10/2018, has been considered.

After consideration of your response to the comments raised, this project has been approved to proceed with the proposed amendments.

Protocol Number: **2018/107**
Protocol Title: **The development and evaluation of a new model of perceptible placebos**
Annual Report Due: **02/03/2019**

Documents Approved:

Date Uploaded	Version Number	Document Name
22/11/2018	Version 1.1	final_03_human-ethics-pcf-perceptibleplacebo_CW22.11.18_v1.1
29/10/2018	Version 1.3	final_04b_pis visit1 perceptible placebo_CW29.10.18_v1.3
29/10/2018	Version 1.4	final_02b_Advertisement_perceptible placebo_CW29.10.18_v1.4
29/10/2018	Version 1.1	final_05c_pis visit2-perceptible placebo_CW29.10.18_v1.1

Special Condition/s of Approval

- It will be a condition of approval that the additional tick box has the '?' removed at the end of the sentence prior to distribution ('I agree to perform a mouth swab (saliva test) to assess if you have been adherently taking the medication during the intervention week?')

Please contact the ethics office should you require further information.

Sincerely,

[redacted]

Dr Helen Mitchell
Chair, Modification Review Committee (MRC 1)

The University of Sydney of Sydney HRECs are constituted and operate in accordance with the National Health and Medical Research Council's (NHMRC) National Statement on Ethical Conduct in Human Research (2007) and the NHMRC's Australian Code for the Responsible Conduct of Research (2007)

Research Integrity & Ethics Administration
Research Portfolio
Level 3, F23 Administration Building
The University of Sydney
NSW 2006 Australia

T +61 2 9036 9161
E human.ethics@sydney.edu.au
W sydney.edu.au/ethics

ABN 15211 513 464
CRICOS 00026A

Advertisement Flyer Study 2

THE UNIVERSITY OF SYDNEY

School of Psychology
Faculty of Science

ABN [2018/107]

Dr Ben Colagiuri
Associate Professor

Volunteers wanted…

Participants regularly suffering from sleep problem, including difficulty initiating sleep, waking up frequently during the night and having difficulties falling back asleep, or waking up earlier than desired without being able to fall back asleep, 18 years or older are wanted for a study investigating a new medication for sleep problems.

This study will involve 3 x 20-minute visits in person over a two-week period and daily online questionnaires (2-3 minutes), totalling approximately 2 hours, at the University of Sydney, Camperdown. At the first visit, baseline measures will be assessed and information about the study will be given. During the first week of the study data about your sleep and general health will be collected. At visit 2 you will be randomly assigned to either the active medication, placebo or a no-treatment condition, followed by another week of data assessment and a last visit.

You will be reimbursed for the costs associated with participating.

If you are interested in participating, or need any additional information about the study please feel free to contact christoph.werner@sydney.edu.au and you will be sent an information sheet.

Experimenter Manual Study 2

THE UNIVERSITY OF SYDNEY

School of Psychology
Faculty of Science

ABN [2018/107]

Experimenter Manual Sleep Study 2

First visit:

✓ Preparation finished according to check list

Hello, my name is XXX. I'm part of the sleep study and today I will introduce you to the study. Please follow me to the consulting room. Please take a seat. Can I offer you a glass of water?

Guide participant to room.

Before we start with the study I want to explain the goal and procedure of the study. Should you have any questions please feel free to interrupt.
The goal of our study is to test a new compound of an antihistamine (doxylamine) and an antioxidant (beetroot extract). Effective antihistamines are already available over-the-counter to treat sleep difficulties. But we realised that they are not as effective as they could be, because our digestive system is not able to fully absorb the antihistamines due to the "unfriendly" environment in our stomach and colon.
To address this problem, we added the beetroot extract which will protect the antihistamines and guarantee better absorption, which will result in better sleep. These are the capsules I was talking about, just so you know what to expect next week.
Do you have any questions so far?
Alright, if everything is clear I'd like to explain the procedure. This double-blind study covers a two-week period and includes three short visits. Double-blind means that you and the experimenters in this study don't know if the capsules you will get contain the active drug or is a fake medication. This is important that nobody can influence or bias the results of the study.
In the first week we will assess your sleep quality and other aspects as a baseline, so we can compare it to the second week, in which you will either take four capsules of our new compound each evening, a placebo, or nothing. Could you please tell me what your understanding of a placebo is?
The random allocation to the three groups - no-treatment, placebo or drug - takes place during the next visit. This means that an independent computer program will put you in any of the three groups by chance.
At the end of the study you will be informed to which group you have been randomised.
During the last visit we will do a mouth swab to objectively test if you have been adherently taking the medication.
During the two weeks of the study you will be required to answer a short online questionnaire about your sleep and daily symptoms each morning, as well as wearing the GENEActive on a daily basis.
Now I want to explain the GENEActive and what you have to do over the next two weeks.

Set up GENEActive with participant and hand it over

This is the GENEActive, it includes sensors that measure movements, body temperature, and light intensity. We will use this information as an objective measure besides your sleep ratings. Please put it on now. Please push this button when you go to bed and wake up again. It is important, that you always wear the watch. You can keep it on for all activities, except scuba diving.
Besides wearing the GENEActive we would like you to fill out a short online survey after you wake up. The survey will ask you questions about your sleep last night and daily symptoms you experienced during the last 24 hours.
You will always receive an email with the link for the online survey in the morning at 5am and a reminder email every five hours later, if you haven't filled out the survey in the meantime.
Please keep up with the scheduled online surveys, since this is very important for the quality of the data. We are able to see if and when you did the online surveys and if you fail to fill them out properly we have to exclude you from the study. Please use the 24-hour format (e.g. 1115pm = 2315) to answer the questions. Please be honest in all your answers, it is very important for our research, that you tell us the truth.
Do you have any questions about the procedure? If not, we would proceed to the informed consent form.

Hand information sheet and informed consent form to participant and go through them.

Please read these documents thoroughly if you haven't done so already and please ask for clarifications if needed. Please fill out these details and sign the document if all your questions have been answered to your satisfaction.

Explain if something is unclear. Ask them to fill out and sign the consent form. Check if everything is filled out and readable.

Alright, thank you very much for participating in our study. I now would like to ask you to fill in the demographic questionnaire.

Let participant fill out questionnaire.

If you do not have any last questions we are done for today and I can show you the way out.
Have a nice day and see you next week.

Guide/show participant the way out.

Second visit:

- ✓ Preparation finished according to check list

Hello and welcome back. Please take a seat. Can I offer you a glass of water?

Guide participant to room.

I would like you to answer the baseline questionnaires asking you about the last week. Please progress until you are told to inform the experimenter that you finished the first block.

RANDOMISE participant!!!

Since this is a double-blind study you and me do not know if you will get the drug or the placebo. Therefore, the randomisation is done automatically by our survey program. Now please click submit.

Intervention Groups

You have been randomised to one of the intervention groups. This means that you have to take 4 capsules one hour before going to bed each evening for the next seven days, starting this evening. The capsules will either contain the active drug, meaning our newly developed formulation of doxylamine plus beetroot extract or they will just contain sugar in the form of lactose fibres in case you are in the placebo group. Since you might get a drug we have to inform you about potential side effects. Please read this information thoroughly.

Hand out drug information sheet and capsules to participant.

Do you have any questions about your assignment, the capsules or the side effects?
I just quickly want to mention the red urine colouration. If you receive the actual drug it is very likely that your urine, especially in the morning after taking the capsules will be a little darker or have a slight reddish colour. This is due to the beetroot and is absolutely harmless. Please do not forget to mention this in the survey about your daily symptoms, if you experienced it.
In case you experience fever, or pain in the kidney area together with red urine, you have to let us know, since this cannot be caused by our drug. In the very unlikely case of red urine and fever or pain in the kidney area it's likely that you got a kidney infection. Kidney infections are not related to our formulation, but I have to inform you about this as a precaution that you do not falsely attribute this to our drug.
So, in summary, the most likely thing you will experience if you receive the drug is a harmless change in the colour of your urine and you should not be worried if you observe this. It's just the colour of the beetroot extract.
Should you experience side effects that are really unpleasant you are allowed to reduce the dose for the next evening from 4 to 2 capsules. If 2 capsules are no problem then please start to increase the dose again. If 2 capsules are still too unpleasant you can further reduce the dose. But the goal is still to take 4 capsules if possible.
Do you have any questions? If not, please answer the last questionnaires for today's visit.

Highlight (1) alcohol, (2) water, (3) empty stomach, (4) 1hour. Let participant fill out questionnaires.

Alright that was it for today. Here is a document stating that you take part in our study just in case you need to explain why you are carrying these capsules around. Please bring back the container with any left-over capsules next week. From now on the online surveys you get will include questions asking how many capsules you took the evening before, please answer honestly. I'd like to remind you that we'll do a mouth swab next visit to determine if you took the capsules. So I can reliably perform the mouth swab next week you are not allowed to consume anything in the 10 minutes prior to testing next week.

Hand out participation statement.

No-treatment Group

I would like to inform you, that you have been randomised to the no-treatment group of our study. This means that you do not have to take any capsules so we can control for the natural course of sleep. Please read this information. You can keep it, if you want.

Hand out no-treatment information sheet to participant.

Do you have any questions about your assignment to the no-treatment control?
If not, I would like you to answer the last questions for today's visit.

Let participant fill out questionnaire.

I would like to remind you to keep up with the scheduled online surveys.
Alright, if you do not have any last questions we are done for.
Have a nice day and see you next week.

Guide/show participant the way out.

Third visit:

- ✓ Preparation finished according to check list

Hello. Please follow me to the consulting room. Please take a seat.

Guide participant to room.

Intervention Groups

Did you consume anything in the last 10 minutes?

- NO: Good this means I can reliably perform the mouth swab to examine if you've been taking the medication adherently. I will swipe the left and right side of the back of your mouth and under your tongue. Please open your mouth.
- Yes: This means that we will delay the mouth swab until the end of today's visit.

Put on gloves and perform mouth swab.

Could I please have the GENEActive and the capsule container back? Thank you very much.
I would like you to fill in some questionnaires. Please continue until you reach the section that the study team needs to fill out.

Collect GENEActive plus container. Let participant fill out questionnaires, retrieve data from GENEActive and label the container with the ID.

Thank you very much for filling out the last questionnaires of this study. Could you please tell me if you are currently enrolled as a student? If so are you an undergraduate or postgraduate student?
You will be sent the debrief information at the end of the study.

Ask about enrolment as student.

SONA participant

We are now at the end of the study. You will get your credits for participating in this study within the next days through SONA. Thank you very much for participating in our study. Have a nice day.

Finish completion section on REDCap and give credits on SONA.

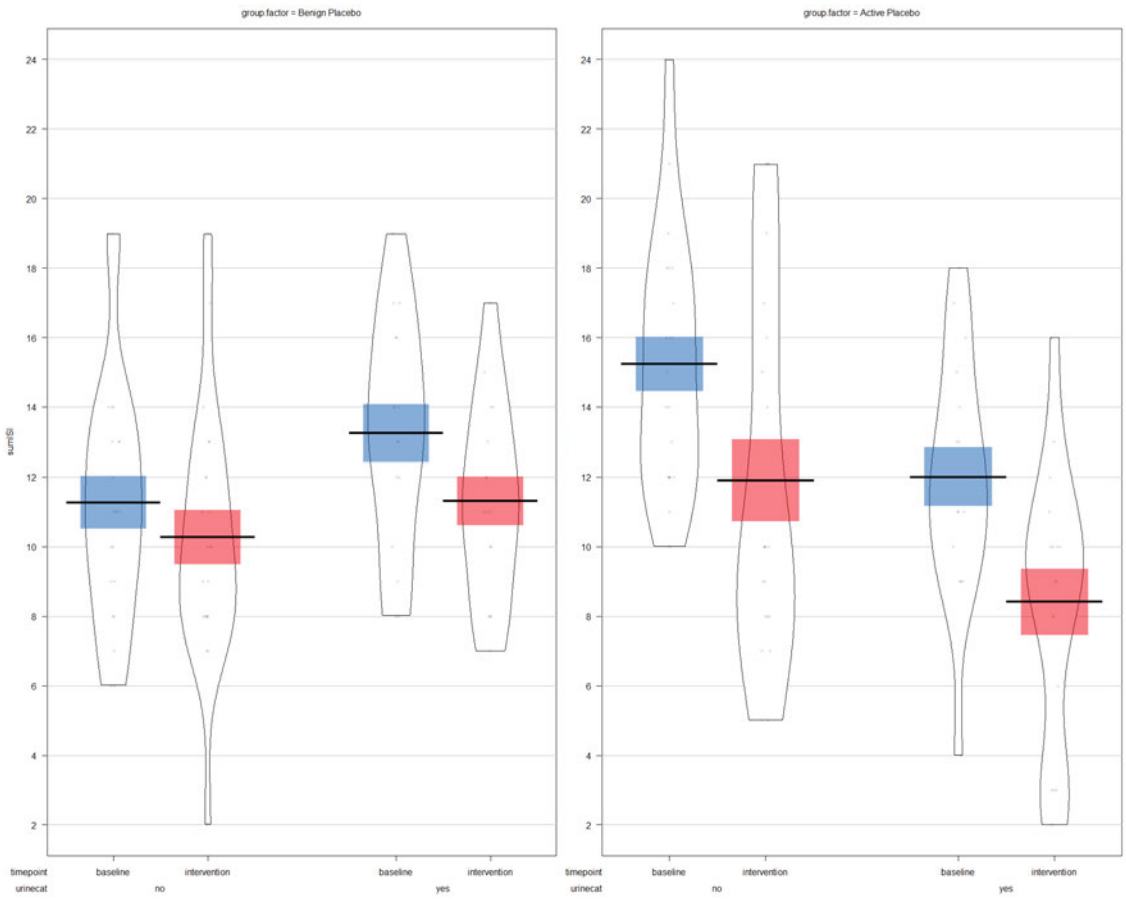
Paid participant

We are now at the end of the study. Here are your 50$, please fill out this list and sign that you've received the money. Thank you very much for participating in our study. Have a nice day.

Hand out money, let participant sign the receipt and finish completion section on REDCap.

Additional Subgroup Analyses

The figure below plots the baseline and intervention ISI scores of the active and conventional placebo groups.

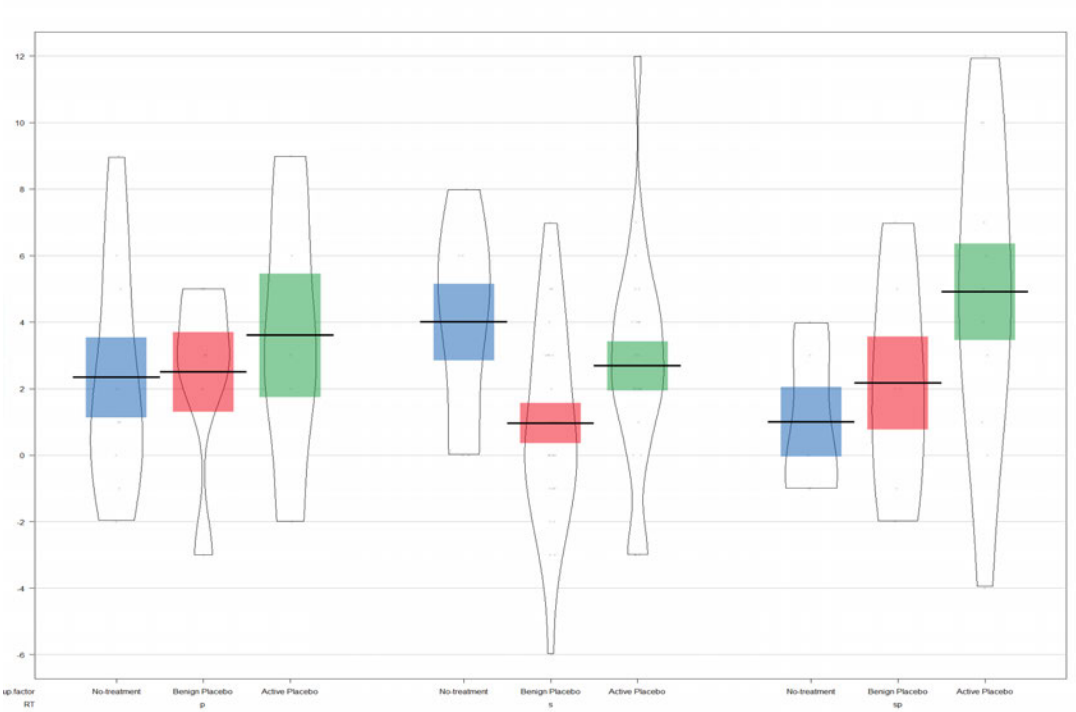


The left panel shows the ISI scores for the conventional placebo group and the right panel depicts ISI scores for the active placebo group. The x-axis shows the mean ISI scores with the boxes around the means plotting SEMs. On the y-axis groups are divided for the factor time (baseline in blue, and intervention in red) and the dichotomous covariate "urinecat" with the levels no (= participants did *not* report a unusually redder urine colouration during intervention) and yes (=participants reported at least once a unusually redder urine colouration during intervention).

The R output below shows the ANCOVA model calculated with the additional variable "urinecat" as described above. As can be seen the inclusion of the covariate urinecat was not statistically significant, as was the group factor.

```
> suboutISI=aov(sumISI.intervention ~ sumISI.baseline+urinecat*group.factor, data = dataANCOVA1) #interaction mo
del; test for main effect (group and time) and their interactions (*); note: (+) would only test main effects
> Anova(suboutISI, type = 3, singular.ok=TRUE) #Type III ANOVA, because data not balanced (remember: ratio 1:2:2
 (NT:BP:AP))
Anova Table (Type III tests)

Response: sumISI.intervention
                    Sum Sq Df F value    Pr(>F)
(Intercept)          19.36  1  1.7623   0.18770
sumISI.baseline     611.13  1 55.6397 5.231e-11 ***
urinecat              6.98  1  0.6356   0.42741
group.factor         61.56  2  2.8023   0.06597 .
urinecat:group.factor 58.09  2  2.6443   0.07657 .
Residuals            988.53 90
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The bean plot below depicts the ISI change scores from baseline to intervention (larger scores indicating more improvement, i.e., better sleep) for the three experimental groups split by enrolment type. The factor enrolment type had three factors (p="paid participants from the general population, reimbursed in cash"; sp="students participating for cash reimbursement"; and s="students participating for course credits"). As can be seen in the bean plot below students participating for course credits seem to differ from the other two categories who participated for money.



Although paid participants showed the expected pattern of results, excluding students participating for course credits, that made up the majority (n=55) of all participants did not result in any statistically significant between group effects when repeating the main analysis with the subsample of paid participants. “group factor1” below stands for the contrast for the overall placebo effect comparing no-treatment to the two placebo groups and “group factor2” stands for the contrast comparing the two placebo groups.

```
> #Inspect planned orthogonal contrasts
> summary.lm(suboutISI) #dummy variable group.factor1 gives sign for
ves second contrast (BPvsAP)

Call:
aov(formula = sumISI.intervention ~ sumISI.baseline + group.factor,
    data = subset(dataANCOVA1, RT %in% c("p", "sp")))

Residuals:
   Min     1Q Median     3Q    Max
-7.292 -2.371 -0.048  2.369  8.036

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)       1.6523     1.9663   0.840    0.406
sumISI.baseline   0.6679     0.1379   4.842 2.17e-05 ***
group.factor1    -0.5228     0.3846  -1.359    0.182
group.factor2    -0.8550     0.6773  -1.262    0.215
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.512 on 38 degrees of freedom
Multiple R-squared:  0.4048,    Adjusted R-squared:  0.3578
F-statistic: 8.613 on 3 and 38 DF,  p-value: 0.0001731
```

Appendix D – Additional Materials Study 3

Ethics Approval Study 3

THE UNIVERSITY OF SYDNEY

Research Integrity & Ethics Administration
HUMAN RESEARCH ETHICS COMMITTEE

Friday, 13 December 2019

Dr Ben Colagiuri
Psychology; Faculty of Science
Email: ben.colagiuri@sydney.edu.au

Dear Ben,

Your request to modify this project, which was submitted on 19 November 2019 has been considered.

After consideration of your response to the comments raised, this project has been approved to proceed with the proposed amendments.

Protocol Number: **2018/107**

Protocol Title: **The development and evaluation of a new model of perceptible placebos**

Documents Approved:

Date Uploaded	Version Number	Document Name
19/11/2019	Version 1.1	original_pis visit2 information study 3- perceptible placebo
19/11/2019	Version 1.2	final_pis visit2 information study 3-noframing_CW10.10.19_v1
19/11/2019	Version 1.2	final_pis visit2 information study 3-positiveframing_CW10.10
19/11/2019	Version 1.2	final_pis visit2 information study 3-negativeframing_CW10.10
19/11/2019	Other Type	study_design_messageframing

Please contact the ethics office should you require further information.

Sincerely,

[redacted]

Dr Clifton Chan
Acting Chair, Modification Review Committee (MRC 1)

The University of Sydney of Sydney HRECs are constituted and operate in accordance with the National Health and Medical Research Council's (NHMRC) National Statement on Ethical Conduct in Human Research (2007) and the NHMRC's Australian Code for the Responsible Conduct of Research (2007)

Research Integrity & Ethics Administration
Research Portfolio
Level 3, F23 Administration Building
The University of Sydney
NSW 2006 Australia

T +61 2 9036 9161
E human.ethics@sydney.edu.au
W sydney.edu.au/ethics

ABN 15 211 513 464
CRICOS 00026A

THE UNIVERSITY OF SYDNEY

Research Integrity & Ethics Administration
HUMAN RESEARCH ETHICS COMMITTEE

Monday, 18 May 2020

Dr Ben Colagiuri
Psychology; Faculty of Science
Email: ben.colagiuri@sydney.edu.au

Dear Ben,

Your request to modify this project, which was submitted on 8 May 2020, has been considered.

This project has been approved to proceed with the proposed amendments.

Protocol Number: **2018/107**

Protocol Title: **The development and evaluation of a new model of perceptible placebos**

Documents Approved:

Date Uploaded	Version Number	Document Name
08/05/2020	Version 1.5	final_Advertisement General study3b_CW20200414_v1.5
08/05/2020	Version 1.2	final_pcf visit1 study 3b_CW20200414_v1.2
08/05/2020	Version 1.4	final_pis visit1 study 3b_CW20200414_v1.4
08/05/2020	Other Type	study3b_design_messageframing

Please contact the ethics office should you require further information.

Sincerely,

[redacted]

Dr Clifton Chan
Chair
Modification Review Committee (MRC 1)

The University of Sydney of Sydney HRECs are constituted and operate in accordance with the National Health and Medical Research Council's (NHMRC) National Statement on Ethical Conduct in Human Research (2007) and the NHMRC's Australian Code for the Responsible Conduct of Research (2007)

Research Integrity & Ethics Administration
Research Portfolio
Level 3, F23 Administration Building
The University of Sydney
NSW 2006 Australia

T +61 2 9036 9161
E human.ethics@sydney.edu.au
W sydney.edu.au/ethics

ABN 15 211 513 464
CRICOS 00026A

Advertisement Flyer Study 3

THE UNIVERSITY OF SYDNEY

School of Psychology
Faculty of Science

ABN [2018/107]

Dr Ben Colagiuri
Associate Professor

Volunteers wanted…

Participants regularly suffering from sleep problem, including difficulty initiating sleep, waking up frequently during the night and having difficulties falling back asleep, or waking up earlier than desired without being able to fall back asleep, 18 years or older are wanted for a study investigating a new medication for sleep problems.

This study will involve 3 x 20-minute online visits (via Zoom) over a two-week period and daily online questionnaires (2-3 minutes), totalling approximately 2 hours. At the first visit, baseline measures will be assessed and information about the study will be given. During the first week of the study data about your sleep and general health will be collected. At visit 2 you will be randomly assigned to either the active medication or a no-treatment condition, followed by another week of data assessment and a last visit.

If you are interested in participating, or need any additional information about the study please feel free to contact christoph.werner@sydney.edu.au and you will be sent an information sheet.

Experimenter Manual Study 3

THE UNIVERSITY OF SYDNEY

School of Psychology
Faculty of Science

ABN [2018/107]

Experimenter Manual Sleep Study 3

First meeting:

- ✓ Preparation finished according to check list

Hello, my name is XXX. I'm part of the sleep study and today I will introduce you to the study. Can you hear and see me?

Before we start with the study, I want to explain the goal and procedure of the study. Should you have any questions please feel free to interrupt.
The goal of our study is to test a new compound of an antihistamine (doxylamine) and an antioxidant (beetroot extract). Effective antihistamines are already available over-the-counter to treat sleep difficulties. But we realised that they are not as effective as they could be, because our digestive system is not able to fully absorb the antihistamines due to the "unfriendly" environment in our stomach and colon.
To address this problem, we added the beetroot extract which we think will protect the antihistamines, guarantee better absorption and therefore better sleep.
Do you have any questions so far?
Alright, if everything is clear I'd like to explain the procedure. This study covers a two-week period and includes three short meetings.
In the first week we will assess your sleep quality and other aspects as a baseline, so we can compare it to the second week, in which you will either take four capsules of our new medicine each evening or nothing.
During the two weeks of the study you will be required to answer a short online questionnaire about your sleep and daily symptoms each morning.
Do you have any questions about the procedure? If not, we would proceed to the informed consent form.

Send link to PIS, PCF, and demographics, go through them if necessary.

Please read the information thoroughly if you haven't done so already and please ask for clarifications if needed. If you don't have any questions or if all your questions have been answered to your satisfaction, please fill out the details and sign the online form.

Explain if something is unclear. Ask them to fill out and sign the consent form. Check if everything is filled out and readable.

Now I want to explain what you have to do over the next two weeks in a little more detail.
The first part of the daily survey will ask you questions about your sleep last night, when you went to bed, what time you fell asleep, how many times you woke up etc. and a comment field to indicate if this night was not representative for your typical sleep. For example, because you were sick, had a late night out with friends, or other reasons.
As soon as you click the submit button, you'll see the second part of the survey asking about the daily symptoms you experienced during the last 24 hours. Each symptom (e.g. headache, stomach pain etc.) is rated from 0 to 100. Everything above zero indicates that you experienced this symptom somehow and a second question will pop up asking you if you think this symptom is related to the study medication, please answer this question with no during the baseline week as you didn't get any medications from us yet.
The invitation emails with the link for the online survey are always sent at 5am in the morning with a reminder email every five hours later, if you haven't filled out the survey in the meantime. It's no problem if you receive the reminders, but you are required to keep up with the scheduled online surveys. The survey system notifies us if you've failed to do the surveys within 24 hours. In this case, I'll send you a warning email and if you then still fail to do the survey, I have to exclude you from the study.
Please keep in mind to use the 24-hour format (Military time; e.g. 1115pm is 2315).

> *Please be honest in all your answers, it is very important for our research, that you tell us the truth. Should we realise that you have been lying to us or that you have given us misleading information we will put this forward to the Academic Misconduct Committee of the University of Sydney as you are required to be truthful when participating in research according to the Code of Conduct.*

The random allocation to the two groups - no-treatment or drug – will be included in the satchel you will receive before next week. It is important that you do NOT open the satchel as we don't want this information to influence the baseline measures. Therefore, you have to open the satchel only after I have instructed to do so during the next week's zoom meeting. Could you please send me a quick email notification once you received the satchel so that I know we can continue with the second meeting as scheduled?

If you do not have any last questions we are done for today and I'll see you again next week.
Have a good week.

Guide/show participant the way out if required.

Second meeting:

- ✓ Preparation finished according to check list

Hello and welcome back. Can you hear and see me?

I would like you to fill out the baseline questionnaires asking you about the last week. Please only continue until you are presented with the information that you finished the first block and should inform the experimenter about it. It will say "STOP, do not press the submit button" on this specific instruction page.

Now we'll figure out to what group you've been randomly allocated. First of all, could you please tell me if the satchel is still intact, or has been opened?
Okay, now I'd like you to open the satchel and tell me if you received any capsules or not.
Additionally, I'd like to know the number on the capsule container seal and if the seal is still intact.
These checks are not to check your behaviour, but to make sure that everything is still clean and safe for you to use.

Enter Randomisation Code!

XX0=No-treatment group, XX1=Positive Framing (F+), XX2=No Framing (F0), XX3=Negative Framing (F-)

No-treatment Group (XX0)

You have been randomised to the no-treatment group. This means that you don't get any capsules. Please read this information.

From now on the daily surveys will contain an adherence question for the drug group.
Unfortunately, it can't be hidden completely for the no-treatment group. You will just see an empty survey form, please still just press the submit button and continue until it says that you have finished the survey and can close the tab.
Do you have any questions about your assignment to the no-treatment control?
If not, I would like you to answer the last questions for today's meeting.

Let participant fill out questionnaire.

I would like to remind you to keep up with the scheduled online surveys.
Alright, if you do not have any last questions we are done for.
Have a nice day and see you next week.

Drug Group F+ (XX1=Positive Framing)

You have been randomised to the drug group. Starting today, you have to take 4 capsules approximately 30min to 1 hour before you want to fall asleep for each of the next 7 evenings. Because you get the doxylamine-beetroot formulation we have to inform you about potential side effects. The satchel should contain two documents, a drug information sheet and a research participation statement. Please read the drug information sheet thoroughly.

Instruct to read info.

Do you have any questions about your assignment, the capsules or the side effects?
I just quickly want to mention the red urine colouration. Because the drug contains beetroot extract, which is a red powder, it is very likely that your urine, especially in the morning after taking the capsules will be unusually darker or have a reddish colour. This is absolutely harmless.

In general, if you experience an unusual colouration of your urine or stool this is a sign that your body absorbed and metabolised the medication well. This means that the drug is more likely to help you improve your sleep quality, time to fall asleep and maintaining your sleep during the night especially well.

In case you experience fever, or pain in the kidney area together with red urine, you have to let us know, since this cannot be caused by our drug. In the very unlikely case of red urine and fever or pain in the kidney area it's likely that you got a kidney infection. Kidney infections are not related to our formulation, but I have to inform you about this as a precaution that you do not falsely attribute this to our drug.
So, in summary, the most likely side effect you will experience is a harmless change in the colour of your urine and you should not be worried if you observe this. It's just the colour of the beetroot extract. Should you experience this, please don't forget to mention this in the daily survey because we need to know how many people experience this symptom.

Should you experience side effects that are too unpleasant you are allowed to reduce the dose by 1 capsule per day. If you've reduced the capsules and you are no longer experiencing anything unpleasant then please start to increase the dose again by 1 capsule per day, as the goal is to take 4 capsules which equates to the standard dose of 25mg doxylamine.
Do you have any questions? If not, please answer the last questionnaires for today's meeting.

Highlight (1) 30min-1h, (2) water & 1by1, (3) alcohol & acid, (4) empty stomach, (5) frame.
Let participant fill out questionnaires.

The second document you got is simply stating that you take part in our study just in case you need to explain why you are carrying these capsules around. Feel free to insert your name and the time span from today until next week. From now on the online surveys you get will include questions asking how many capsules you took the evening before, please answer honestly.

I would like to remind you to keep up with the scheduled online surveys.
Alright, if you do not have any last questions we are done for.
Have a nice day and see you next week.

Drug Group F0 (XX2=No Framing),

You have been randomised to the drug group. Starting today, you have to take 4 capsules approximately 30min to 1 hour before you want to fall asleep for each of the next 7 evenings. Because you get the doxylamine-beetroot formulation we have to inform you about potential side effects. The satchel should contain two documents, a drug information sheet and a research participation statement. Please read the drug information sheet thoroughly.

Instruct to read info.

Do you have any questions about your assignment, the capsules or the side effects?
I just quickly want to mention the red urine colouration. Because the drug contains beetroot extract, which is a red powder, it is very likely that your urine, especially in the morning after taking the capsules will be unusually darker or have a reddish colour. This is absolutely harmless.

In case you experience fever, or pain in the kidney area together with red urine, you have to let us know, since this cannot be caused by our drug. In the very unlikely case of red urine and fever or pain in the kidney area it's likely that you got a kidney infection. Kidney infections are not related to our formulation, but I have to inform you about this as a precaution that you do not falsely attribute this to our drug.
So, in summary, the most likely side effect you will experience is a harmless change in the colour of your urine and you should not be worried if you observe this. It's just the colour of the beetroot extract. Should you experience this, please don't forget to mention this in the daily survey because we need to know how many people experience this symptom.

Should you experience side effects that are too unpleasant you are allowed to reduce the dose by 1 capsule per day. If you've reduced the capsules and you are no longer experiencing anything unpleasant then please start to increase the dose again by 1 capsule per day, as the goal is to take 4 capsules which equates to the standard dose of 25mg doxylamine.
Do you have any questions? If not, please answer the last questionnaires for today's meeting.

Highlight (1) 30min-1h, (2) water & 1by1, (3) alcohol & acid, (4) empty stomach.
Let participant fill out questionnaires.

The second document you got is simply stating that you take part in our study just in case you need to explain why you are carrying these capsules around. Feel free to insert your name and the time span from today until next week. From now on the online surveys you get will include questions asking how many capsules you took the evening before, please answer honestly.

I would like to remind you to keep up with the scheduled online surveys.
Alright, if you do not have any last questions we are done for.
Have a nice day and see you next week.

Drug Group F- (XX3=Negative Framing)

You have been randomised to the drug group. Starting today, you have to take 4 capsules approximately 30min to 1 hour before you want to fall asleep for each of the next 7 evenings. Because you get the doxylamine-beetroot formulation we have to inform you about potential side effects. The satchel should contain two documents, a drug information sheet and a research participation statement. Please read the drug information sheet thoroughly.

Instruct to read info.

Do you have any questions about your assignment, the capsules or the side effects?
I just quickly want to mention the red urine colouration. Because the drug contains beetroot extract, which is a red powder, it is very likely that your urine, especially in the morning after taking the capsules will be unusually darker or have a reddish colour. This is absolutely harmless.

In general, if you experience an unusual colouration of your urine or stool this is a sign that your body did not absorb and metabolise the medication well. This means that the drug is less likely to help you improve your sleep quality, time to fall asleep and maintaining your sleep during the night especially well.

In case you experience fever, or pain in the kidney area together with red urine, you have to let us know, since this cannot be caused by our drug. In the very unlikely case of red urine and fever or pain in the kidney area it's likely that you got a kidney infection. Kidney infections are not related to our formulation, but I have to inform you about this as a precaution that you do not falsely attribute this to our drug.
So, in summary, the most likely side effect you will experience is a harmless change in the colour of your urine and you should not be worried if you observe this. It's just the colour of the beetroot extract. Should you experience this, please don't forget to mention this in the daily survey because we need to know how many people experience this symptom.

Should you experience side effects that are too unpleasant you are allowed to reduce the dose by 1 capsule per day. If you've reduced the capsules and you are no longer experiencing anything unpleasant then please start to increase the dose again by 1 capsule per day, as the goal is to take 4 capsules which equates to the standard dose of 25mg doxylamine.
Do you have any questions? If not, please answer the last questionnaires for today's meeting.

Highlight (1) 30min-1h, (2) water & 1by1, (3) alcohol & acid, (4) empty stomach, (5) frame.
Let participant fill out questionnaires.

The second document you got is simply stating that you take part in our study just in case you need to explain why you are carrying these capsules around. Feel free to insert your name and the time span from today until next week. From now on the online surveys you get will include questions asking how many capsules you took the evening before, please answer honestly.

I would like to remind you to keep up with the scheduled online surveys.
Alright, if you do not have any last questions we are done for.
Have a nice day and see you next week.

Third meeting:

- ✓ Preparation finished according to check list

Hello and welcome back. Can you hear and see me?

Drug Group

Could you please show me the capsule container and let me know if you took all capsules, or how many are left?

Please continue until you reach end of the surveys. The first will again be the ISI, then the QoL, and a couple of exploratory questionnaires.

Let participant fill out questionnaires.

You will be sent the debrief information and a lay summary of the results at the end of the study.

SONA participant

We are now at the end of the study. You will get your credits for participating in this study within the next days through SONA. Thank you very much for participating in our study. Have a nice day.

Finish completion section on REDCap and give credits on SONA.

Volunteering participant

We are now at the end of the study. Thank you very much for participating in our study. Have a nice day.

Give debrief information about whole research project (influence of side effects on efficacy) and specific purpose of this study (framing information)

Fill out "Experimenter" part in REDCap (returned capsules, which experimenter conducted what meeting, and AE_recall metrics etc.)

Positive Message Framing Information

THE UNIVERSITY OF SYDNEY

School of Psychology
Faculty of Science

ABN [2018/107]

Dr Ben Colagiuri
Associate Professor

Room 486
Griffith Taylor, A19
The University of Sydney
NSW 2006 AUSTRALIA
Telephone: +61 2 9351 4589
Facsimile: +61 2 9351 5223
Email: ben.colagiuri@sydney.edu.au
Web: http://www.sydney.edu.au/

INFORMATION FOR DRUG GROUP – SLEEP TRIAL

Overview:
This new drug formulation including doxylamine and beetroot extract has been newly developed to treat sleep problems. Beetroot extract mainly consists of betanin, a potent antioxidant, which interacts with doxylamine to boost its efficacy for better sleep.

For all participants taking the new drug formulation (sleep aid) please tell the examiner:

- If you have an allergy to antihistamine or any other part of this medicine.
- If you have a sensitive stomach or gastric problems (especially acid reflux)
- If you are allergic to any drugs like this one, any other drugs, foods, or other substances. Tell your examiner about the allergy and what signs you had, like rash; hives; itching; shortness of breath; wheezing; cough; swelling of face, lips, tongue, or throat; or any other signs.
- This medicine may interact with other drugs or health problems. Therefore, it is important that you inform your examiner about all of your drugs (prescription or over-the-counter, natural products, vitamins) and health problems. You must check to make sure that it is safe for you to take this medicine. Do not start, stop, or change the dose of any drug without checking with the research team.

What are some things you need to know or do while you take this new drug formulation?

- Avoid driving and doing other tasks or actions that call for you to be alert after you take this medicine. You may still feel sleepy the day after you take this new drug. Avoid these tasks or actions until you feel fully awake.
- Avoid drinking alcohol or other acidic foods, especially in the evening while taking this medicine.
- Do not use other drugs and natural products during the two weeks of this study, unless discussed with the research team.
- Be careful in hot weather or while being active. Drink lots of fluids to stop fluid loss.

Side Effects:
General:

- Most side effects are mild and short-lived; serious side effects are rare (1 out of 1000).

Musculoskeletal:

- Muscle weakness (frequency not known)

Nervous system:
- Sedation, dizziness, impaired coordination, drowsiness, headache, anticholinergic effects (e.g., dry mouth, nose, and/or throat) (frequency not known)

Respiratory:
- Thicker nasal discharge (frequency not known)

Renal:
- Yellow to red excretion, urine or stool (very common, 4 out of 5)

Dosage:

Usual adult dose for sleeping difficulties:
- Four (4) capsules orally once a day 1 hour before bedtime.
- Duration of therapy no more than 2 weeks consecutively.

Administration Advice:
- Swallow the capsules with a glass of water (≥200 ml).

General:
- Helps to reduce difficulty in falling asleep, increases quality of sleep and sleeping through the night.
- There is no evidence that this drug is addictive.
- This drug has calmative and anti-nausea properties which can be useful in relieving tension and nausea associated with pain.
- Reactions associated with over dosage may vary from central nervous depression to stimulation and gastro-intestinal problems (e.g. upset stomach).

Participant Advice:
- Avoid alcohol and acidic foods (e.g. citrus fruits) while taking this drug.
- Avoid taking this drug on an empty stomach.
- This drug may cause drowsiness that can affect your ability to perform activities which require complete alertness such as driving and operating machinery; avoid such activities until you know how this drug affects you.
- This medication may cause yellow to red urine (or stool), which is a normal physiological process caused by the betanin contained in beetroot and is completely harmless. But, this symptom may as well be caused by serious kidney problems. If you experience this symptom together with fever or pain in the kidney region or lower back, immediately inform the research team for further medical evaluations.
- In general, if you experience an unusual colouration of your urine or stool this is a sign that your body absorbed and metabolised the medication well. This means that the drug is more likely to help you improve your sleep quality, time to fall asleep and maintaining your sleep during the night especially well.

No Message Framing Information

THE UNIVERSITY OF SYDNEY

School of Psychology
Faculty of Science

ABN [2018/107]

Dr Ben Colagiuri
Associate Professor

Room 486
Griffith Taylor, A19
The University of Sydney
NSW 2006 AUSTRALIA
Telephone: +61 2 9351 4589
Facsimile: +61 2 9351 5223
Email: ben.colagiuri@sydney.edu.au
Web: http://www.sydney.edu.au/

INFORMATION FOR DRUG GROUP – SLEEP TRIAL

Overview:
This new drug formulation including doxylamine and beetroot extract has been newly developed to treat sleep problems. Beetroot extract mainly consists of betanin, a potent antioxidant, which interacts with doxylamine to boost its efficacy for better sleep.

For all participants taking the new drug formulation (sleep aid) please tell the examiner:

- If you have an allergy to antihistamine or any other part of this medicine.
- If you have a sensitive stomach or gastric problems (especially acid reflux)
- If you are allergic to any drugs like this one, any other drugs, foods, or other substances. Tell your examiner about the allergy and what signs you had, like rash; hives; itching; shortness of breath; wheezing; cough; swelling of face, lips, tongue, or throat; or any other signs.
- This medicine may interact with other drugs or health problems. Therefore, it is important that you inform your examiner about all of your drugs (prescription or over-the-counter, natural products, vitamins) and health problems. You must check to make sure that it is safe for you to take this medicine. Do not start, stop, or change the dose of any drug without checking with the research team.

What are some things you need to know or do while you take this new drug formulation?

- Avoid driving and doing other tasks or actions that call for you to be alert after you take this medicine. You may still feel sleepy the day after you take this new drug. Avoid these tasks or actions until you feel fully awake.
- Avoid drinking alcohol or other acidic foods, especially in the evening while taking this medicine.
- Do not use other drugs and natural products during the two weeks of this study, unless discussed with the research team.
- Be careful in hot weather or while being active. Drink lots of fluids to stop fluid loss.

Side Effects:
General:

- Most side effects are mild and short-lived; serious side effects are rare (1 out of 1000).

Musculoskeletal:

- Muscle weakness (frequency not known)

Nervous system:
- Sedation, dizziness, impaired coordination, drowsiness, headache, anticholinergic effects (e.g., dry mouth, nose, and/or throat) (frequency not known)

Respiratory:
- Thicker nasal discharge (frequency not known)

Renal:
- Yellow to red excretion, urine or stool (very common, 4 out of 5)

Dosage:

Usual adult dose for sleeping difficulties:
- Four (4) capsules orally once a day 1 hour before bedtime.
- Duration of therapy no more than 2 weeks consecutively.

Administration Advice:
- Swallow the capsules with a glass of water (≥200 ml).

General:
- Helps to reduce difficulty in falling asleep, increases quality of sleep and sleeping through the night.
- There is no evidence that this drug is addictive.
- This drug has calmative and anti-nausea properties which can be useful in relieving tension and nausea associated with pain.
- Reactions associated with over dosage may vary from central nervous depression to stimulation and gastro-intestinal problems (e.g. upset stomach).

Participant Advice:
- Avoid alcohol and acidic foods (e.g. citrus fruits) while taking this drug.
- Avoid taking this drug on an empty stomach.
- This drug may cause drowsiness that can affect your ability to perform activities which require complete alertness such as driving and operating machinery; avoid such activities until you know how this drug affects you.
- This medication may cause yellow to red urine (or stool), which is a normal physiological process caused by the betanin contained in beetroot and is completely harmless. But, this symptom may as well be caused by serious kidney problems. If you experience this symptom together with fever or pain in the kidney region or lower back, immediately inform the research team for further medical evaluations.

Negative Message Framing Information

THE UNIVERSITY OF SYDNEY

School of Psychology
Faculty of Science

ABN [2018/107]

Dr Ben Colagiuri
Associate Professor

Room 486
Griffith Taylor, A19
The University of Sydney
NSW 2006 AUSTRALIA
Telephone: +61 2 9351 4589
Facsimile: +61 2 9351 5223
Email: ben.colagiuri@sydney.edu.au
Web: http://www.sydney.edu.au/

INFORMATION FOR DRUG GROUP – SLEEP TRIAL

Overview:
This new drug formulation including doxylamine and beetroot extract has been newly developed to treat sleep problems. Beetroot extract mainly consists of betanin, a potent antioxidant, which interacts with doxylamine to boost its efficacy for better sleep.

For all participants taking the new drug formulation (sleep aid) please tell the examiner:

- If you have an allergy to antihistamine or any other part of this medicine.
- If you have a sensitive stomach or gastric problems (especially acid reflux)
- If you are allergic to any drugs like this one, any other drugs, foods, or other substances. Tell your examiner about the allergy and what signs you had, like rash; hives; itching; shortness of breath; wheezing; cough; swelling of face, lips, tongue, or throat; or any other signs.
- This medicine may interact with other drugs or health problems. Therefore, it is important that you inform your examiner about all of your drugs (prescription or over-the-counter, natural products, vitamins) and health problems. You must check to make sure that it is safe for you to take this medicine. Do not start, stop, or change the dose of any drug without checking with the research team.

What are some things you need to know or do while you take this new drug formulation?

- Avoid driving and doing other tasks or actions that call for you to be alert after you take this medicine. You may still feel sleepy the day after you take this new drug. Avoid these tasks or actions until you feel fully awake.
- Avoid drinking alcohol or other acidic foods, especially in the evening while taking this medicine.
- Do not use other drugs and natural products during the two weeks of this study, unless discussed with the research team.
- Be careful in hot weather or while being active. Drink lots of fluids to stop fluid loss.

Side Effects:
General:

- Most side effects are mild and short-lived; serious side effects are rare (1 out of 1000).

Musculoskeletal:

- Muscle weakness (frequency not known)

Nervous system:
- Sedation, dizziness, impaired coordination, drowsiness, headache, anticholinergic effects (e.g., dry mouth, nose, and/or throat) (frequency not known)

Respiratory:
- Thicker nasal discharge (frequency not known)

Renal:
- Yellow to red excretion, urine or stool (very common, 4 out of 5)

Dosage:

Usual adult dose for sleeping difficulties:
- Four (4) capsules orally once a day 1 hour before bedtime.
- Duration of therapy no more than 2 weeks consecutively.

Administration Advice:
- Swallow the capsules with a glass of water (≥200 ml).

General:
- Helps to reduce difficulty in falling asleep, increases quality of sleep and sleeping through the night.
- There is no evidence that this drug is addictive.
- This drug has calmative and anti-nausea properties which can be useful in relieving tension and nausea associated with pain.
- Reactions associated with over dosage may vary from central nervous depression to stimulation and gastro-intestinal problems (e.g. upset stomach).

Participant Advice:
- Avoid alcohol and acidic foods (e.g. citrus fruits) while taking this drug.
- Avoid taking this drug on an empty stomach.
- This drug may cause drowsiness that can affect your ability to perform activities which require complete alertness such as driving and operating machinery; avoid such activities until you know how this drug affects you.
- This medication may cause yellow to red urine (or stool), which is a normal physiological process caused by the betanin contained in beetroot and is completely harmless. But, this symptom may as well be caused by serious kidney problems. If you experience this symptom together with fever or pain in the kidney region or lower back, immediately inform the research team for further medical evaluations.
- In general, if you experience an unusual colouration of your urine or stool this is a sign that your body did not absorb and metabolise the medication well. This means that the drug is less likely to help you improve your sleep quality, time to fall asleep and maintaining your sleep during the night especially well.

No-treatment Group Information

THE UNIVERSITY OF SYDNEY

School of Psychology
Faculty of Science

ABN [2018/107]

Dr Ben Colagiuri
Associate Professor

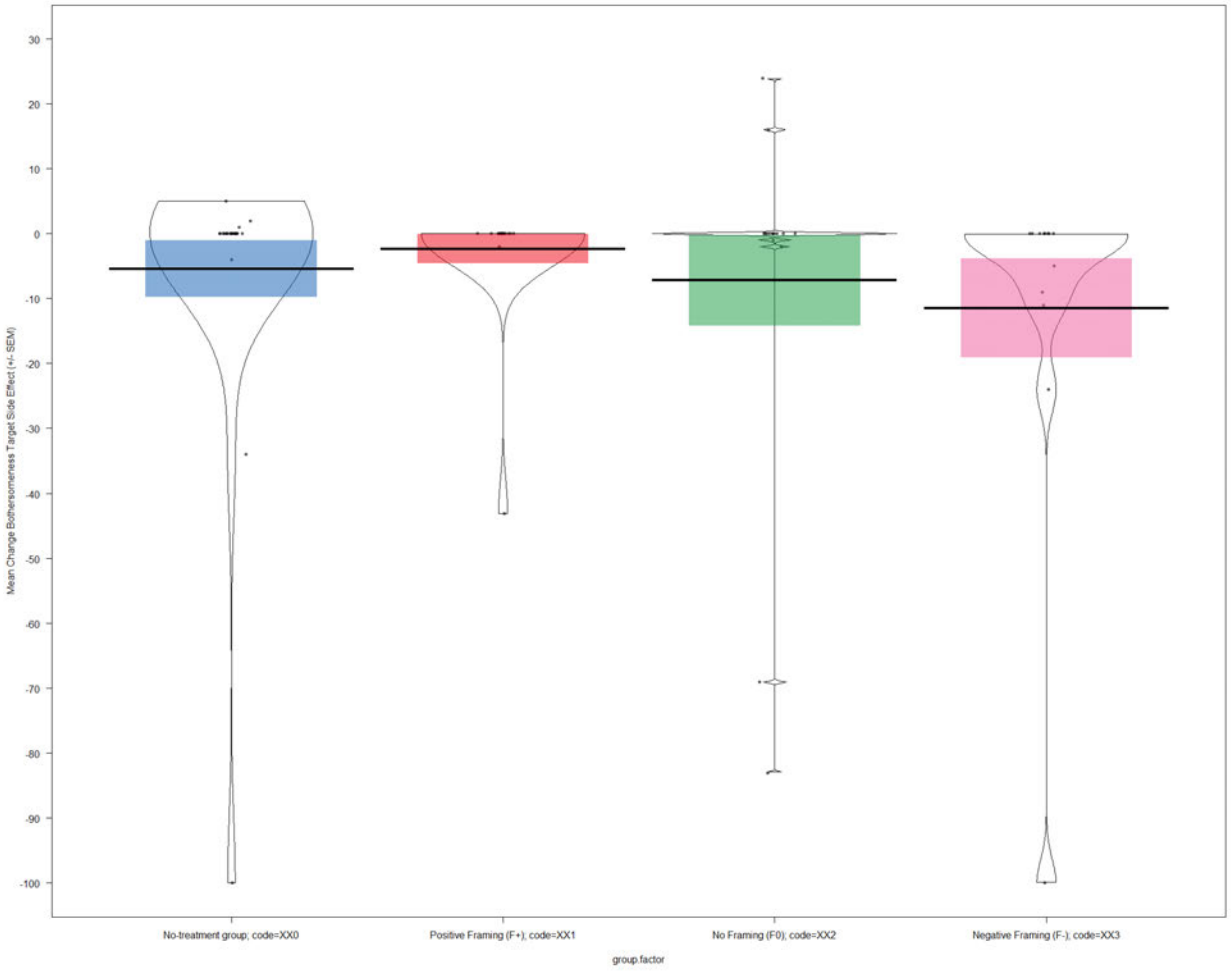
Room 486
Griffith Taylor, A19
The University of Sydney
NSW 2006 AUSTRALIA
Telephone: +61 2 9351 4589
Facsimile: +61 2 9351 5223
Email: ben.colagiuri@sydney.edu.au
Web: http://www.sydney.edu.au/

INFORMATION FOR NO-TREATMENT GROUP – SLEEP TRIAL

General information:
You have been randomised to the no-treatment group. Therefore, your will not receive any medication during the next week. To accurately assess the effects of a medication it is crucial to include a natural history for the condition to treat in a study, due to the fact, that a medical condition changes over time for natural occurring reasons.

Additional Subgroup Analyses

The figure below shows the mean change scores from baseline to intervention for the mean bothersomeness (+/- SEM) of the target side effect, i.e., red urine colouration for the four experimental groups.



As can be seen from the graph above and the R output below, none of the orthogonal contrasts indicated any statistically significant differences. The factor "group.factor1" shows the result for the first orthogonal contrast comparing no-treatment to the three message framing groups receiving the active placebo, "group.factor2" gives second orthogonal contrast comparing the

positive message framing group against the no message framing and the neutral message framing group, and "group.factor3" gives the third orthogonal contrast comparing the no message framing group to the neutral message framing group.

```
> #Inspect planned orthogonal contrasts
> summary.lm(primouturinecolourbothersub) #dummy variable group.facto
0 F-), group.factor2 gives second contrast (F+ vs F0 F-), group.facto

Call:
aov(formula = urinecolourbothertotal.post ~ urinecolourbothertotal.p
    group.factor, data = dataANCOVA3sub)

Residuals:
    Min      1Q  Median      3Q     Max
-35.091  -0.570  -0.027   0.795  28.642

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                0.64249    1.02910   0.624    0.535
urinecolourbothertotal.pre 0.35886    0.03038  11.811   <2e-16 ***
group.factor1              0.47921    0.50536   0.948    0.346
group.factor2              0.54714    0.80289   0.681    0.498
group.factor3              1.12261    1.50637   0.745    0.459
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.026 on 67 degrees of freedom
Multiple R-squared:  0.6952,    Adjusted R-squared:  0.677
F-statistic:  38.2 on 4 and 67 DF,  p-value: < 2.2e-16
```