Placebos and their Consequences in Clinical and Basic Research: Effects and Possible Ways to Harness Them

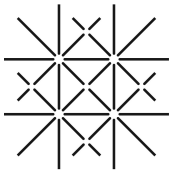
Inauguraldissertation
zur
Erlangung der Würde
eines Doktors der Philosophie
vorgelegt der
Fakultät für Psychologie
der Universität Basel

von

M Sc Cosima Locher

aus Basel

Basel, 2017



Genehmigt von der Fakultät für Psychologie

auf Antrag von

Prof. Dr. Jens Gaab
Prof. Dr. Rainer Greifeneder

Basel, den _____

_____
Prof. Dr. Roselind Lieb

UNI
BASEL

Acknowledgements

An dieser Stelle möchte ich mich bei all denjenigen bedanken, die mich während der Dissertationsjahre unterstützt haben und die dadurch massgeblich zu der vorliegenden Arbeit beigetragen haben.

Mein ausserordentlicher Dank gilt Jens Gaab, der mich während des gesamten Prozesses betreut gefördert, und inspiriert hat. Neben fachlicher Unterstützung habe ich durch Jens auch andere Dinge erleben dürfen: in einem Team eingebettet zu sein, sich ernst genommen zu fühlen, sowie bei einem Anliegen auf ein offenes Ohr und Herz zu treffen. Darüber hinaus hat er mich stetig dazu ermuntert, neue Wege und Gedankengänge zu wagen, wodurch ich mit reichen Erfahrungen und Erlebnissen beschenkt worden bin.

Ich möchte mich bei der gesamten Abteilung bedanken. Mein besonderes Dankeschön gilt folgenden Personen: Heike Gerger, mit welcher ich mich sehr verbunden fühle. Den inhaltlichen Austausch, sowie das gemeinsame Arbeiten und Entwickeln von Ideen und Konzepten schätze ich aufs Äusserste. Mein herzlicher Dank gilt auch Joe Kossowsky, der mich von Anfang an für das Thema begeistern konnte und der für mich ein guter Ratgeber sowie zuverlässiger Freund geworden ist. Ebenso möchte ich mich bei Sibylle Meier bedanken: über die Zeit ist eine wertvolle Freundschaft entstanden, welche Raum für Persönliches ermöglicht hat. Weiter hat mir die intensive und gelungene Zusammenarbeit mit Antje Frey Nascimento und Helen Koechlin viel Freude bereitet, was ich sehr schätze. Danke auch Boby Xavier Thaiparambil für die anregenden Gespräche. Sebastian Hasler möchte ich an dieser Stelle ebenso danken, auch für sein pflichtbewusstes Korrekturlesen.

Bedeutsam waren für mich all diejenigen, welche für mich als Forscherin zum Vorbild und zur Quelle der Inspiration geworden sind. So hat das gesamte Werk von Irving Kirsch grossen Eindruck auf mich gemacht und meine eigene Arbeit geprägt.

Ein herzliches Dankeschön gilt meiner Familie und Freunden. Besonders bedanken möchte ich mich bei meinen Eltern – ihre Ermutigung, ihr Zuhören, ihr Humor, sowie ihr Vertrauen in mich haben mir viel Kraft und Energie gegeben.

Tief verbunden und dankbar bin ich Martin Müller, für all die unzähligen kleinen Dinge, die mein Leben und Schaffen bunter und leichter machen und für seine treue Begleitung auf dem gesamten Weg.

Declaration of Independence

The submitted studies in partial fulfilment of the requirements for the degree of Doctor of Philosophy were written in collaboration with the mentioned co-authors. Neither the author, co-authors nor any other persons published the studies elsewhere. All citations are indicated and only the tools cited were used.

For the purpose of the cumulative dissertation, the following studies have been submitted for publication in various journals. Copies of the studies can be found in the appendix:

Study I:

Locher, C., Koechlin, H., Zion, S. R., Werner, C., Pine, S. D., Kirsch, I., Kessler, R.C., & Kossowsky, J. (2017). *Efficacy and safety of SSRIs, SNRIs, and placebo in common psychiatric disorders: A comprehensive meta-analysis in children and adolescents*. Manuscript submitted for publication.

Study II:

Locher, C., Kossowsky, J., Gaab, J., Kirsch, I., Bain, P., & Krummenacher, P. (2015). Moderation of antidepressant and placebo outcomes by baseline severity in late-life depression: A systematic review and meta-analysis. *Journal of Affective Disorders, 181*, 50–60. doi:10.1016/j.jad.2015.03.062

Study III:

Locher, C., Frey Nascimento, A., Kossowsky, J., Meyer, A., & Gaab, J. (2017). *Is the rationale more important than deception? A randomized controlled trial of open-label placebo analgesia*. Manuscript submitted for publication.

With my signature, I testify that all statements are true and complete.

Basel, 24. Mai 2017

Table of Contents

APPENDICES

A. STUDY I

B. STUDY II

C. STUDY III

D. CURRICULUM VITAE

Abstract

While for adults, a plurality of studies examines placebo responses and potential moderators in antidepressant trials, comparable investigations in pediatric and geriatric patients are scarce. This is especially problematic since the efficacy and safety of antidepressants is controversial in these sensitive populations: effect sizes are small and severe side effects such as an increased risk of suicidal thoughts and behavior have been reported. Here, it has been hypothesized that the lack of a consistent significant benefit of antidepressants over placebo could be associated with an increased response to placebo. Therefore, it is worth considering whether the potential of placebos can be harnessed without undermining patients' autonomy through deception.

The emphasis of the current dissertation was twofold: first, to investigate the efficacy of placebos and potential moderators in pediatric and geriatric patients. Second, to experimentally test the necessity of deception. For this reason, two different statistical approaches were indicated. For the first aim, a meta-analytic approach was applied in order to assess differences between antidepressant and placebo interventions in pediatric major depressive disorder (MDD), anxiety disorders, obsessive-compulsive disorder, and posttraumatic stress disorder (Locher, et al., 2017; Study I), as well as in geriatric MDD, dysthymia and minor depression (Locher et al., 2015; Study II) along with variation in placebo responses and moderators. For the second goal, a basic research approach was chosen in order to compare the effects of openly prescribed placebos with a deceptive placebo administration in a standardized heat pain experiment with healthy participants (Locher, Frey Nascimento, Kossowsky, Meyer, & Gaab, 2017; Study III). Here, basic research represents an excellent way to experimentally compare these treatments in accordance with ethical principles.

The meta-analyses revealed that antidepressants are more effective than placebo at treating MDD in children and adolescents (Hedges' $g = 0.20$; Study I), as well as in elderly people (Hedges' $g = 0.37$; Study II). However, the effects were only small and did not reach the proposed cut-off for clinical significance. Also, placebo responses in depressed youth (Hedges' $g = 1.57$; Study I), as well as in depressed elderly (Hedges' $g = 0.96$; Study II) were significant and substantial. Findings of the heat pain experiment with healthy participants revealed that open-label placebos do not differ in their effects from deceptive placebos in subjective outcomes (i.e., heat pain intensity ratings: $p = .136$ and heat pain unpleasantness ratings: $p = .481$; Study III).

Placebo responses are large and meaningful in children, adolescents and elderly people with depression. Pediatric as well as geriatric patients seem to respond well to clinician contact

that promotes the therapeutic alliance and other common factors such as the patients' expectations and hopes of improvement. Furthermore, the ubiquitously assumed necessity of concealment in placebo administration is questioned and new ways in order to harness the potential of placebos should be considered.

1. Theoretical Background

1.1. The Placebo Effect

Time has passed since Beecher (1955) claimed that placebos are powerful—a meanwhile numerously confirmed finding in both healthy individuals and patients with various clinical conditions (Forsberg, Martinussen, & Flaten, 2016). In particular, relevant placebo effects have been reported in medical conditions which are amenable to psychological factors (Wampold, Minami, Tierney, Baskin, & Bhati, 2005), such as pain (Tuttle et al., 2015), Parkinson's disease (Schmidt, Braun, Wager, & Shohamy, 2014), asthma (Wechsler et al., 2011), and nausea (Quinn & Colagiuri, 2016); as well as in mental diseases such as depression (Furukawa et al., 2016; Rutherford et al., 2017) and anxiety (Sugarman, Loree, Baltes, Grekin, & Kirsch, 2014). The original conceptualization of a placebo as an inert agent or procedure was unavoidably linked with a paradox: By definition, something that is inert can't cause an effect (Price, Finniss, & Benedetti, 2008). Inevitably, the focus has been shifted to the concept of placebo effects as genuine psychobiological events (Finniss, Kaptchuk, Miller, & Benedetti, 2010) within a psychosocial context (Moerman & Jonas, 2002). In this understanding, the doctor-patient relationship, consisting of both, emotional (e.g., trust, empathy, respect, acceptance and warmth), as well as informational (e.g., patient education, treatment information, and expectation management) components (Di Blasi, Harkness, Ernst, Georgiou, & Kleijnen, 2001; Lucassen & Olesen, 2016), is indispensable for the success of any treatment (Kelley, Kraft-Todd, Schapira, Kossowsky, & Riess, 2014). Accordingly, it has been shown that an enhanced relationship with a practitioner, combined with the therapeutic ritual of a placebo administration, promotes the most robust benefit when compared to placebo administration with only limited social support and to a waitlist control group (Kaptchuk et al., 2008).

Environmental and psychosocial mechanisms that contribute to placebo effects are numerous (Benedetti, 2008); however, two factors are most established: expectations, which are reinforced through verbal suggestions (Benedetti, 2002; Jepma & Wager, 2015; Kirsch & Weixel, 1988; Pollo et al., 2001), as well as classical conditioning, a learning of relations among events (Rescorla, 1988) through direct experience (Benedetti et al., 2016; Schedlowski & Pacheco-López, 2010; Voudouris, Peck, & Coleman, 1985) and through social observation (Colloca & Benedetti, 2009). Whereas expectations affect conscious physiological functions such as pain and motor performance, conditioning has an additional impact on unconscious physiological functions such as hormonal levels and immune responses (Benedetti et al., 2003). While the relation of these two psychological mechanisms is an ongoing subject of discussion

(Kirsch, 2004; Kirsch et al., 2014; Stewart-Williams & Podd, 2004), new and multimodal conceptual frameworks for placebo effects are likewise proposed. For example, the "somatic focus" model is based on the assumption that sensations, somatic attention, construal, and bodily states are connected (Alfano, 2015; Lundh, 1987). This model proposes that individuals with positive expectations selectively attend to signs of somatic improvement and interpret them as evidence that the placebo intervention has been successful even if their physical health is unchanged and the perceived signs are only part of a natural variability (Geers, Helfer, Weiland, & Kosbab, 2006; Walker et al., 2006). Similarly, the awareness of being treated promotes placebo effects and improves clinical outcomes (Colloca & Benedetti, 2016), which stands in contrast to hidden applications in which patients are not informed about being treated—what substantially minimizes the effect of placebos (Colloca, Lopiano, Lanotte, & Benedetti, 2004). Further, the Bayesian models of perceptual decision is becoming established as an innovative framework in placebo research (Geuter, Koban, & Wager, 2017). The core assumption is that sensory bottom-up signals trigger top-down predictions or expectations which result from prior knowledge or experience (Friston, 2003). If the incoming signals are in line with the prior knowledge, the prediction is confirmed. However, if they are not congruent, a prediction error signal is generated and the expectation may be adapted through a learning rule (Wiech, 2016). Importantly, the content of sensory signals is represented by means of probability (Büchel, Geuter, Sprenger, & Eippert, 2014) and we are "forever trying to stay one step ahead of the incoming waves of sensory stimulation" (Clark, 2015, p. 21). Also, a prediction error might not automatically lead to an adaption of the expectation (Wiech, 2016) —an additional argument why expectations are not necessarily linked to self-fulling prophecies (Crombez & Wiech, 2011).

Further, and in order to understand the complex psychological processes involved in placebo effects, a much broader scope of psychological variables is warranted (Geers & Miller, 2014). Placebo effects are dynamically influenced by idiosyncratic and top-down constructs, such as expectations and learning experiences, yet also by other psychological variables such as meaning, mindsets, hope, and beliefs (Crum, Phillips, Scott, Kosslyn, & Pinkerton, 2015). Therefore, the placebo effect has been defined as a meaning response (Barrett et al., 2006; Moerman & Jonas, 2002), a process which is essentially evoked by narrative language (Brody, 1994; Bruner, 1990). Along similar lines, there is the proposition that individual mindsets—the frame of mind which orients an individual to a particular set of associations and expectations—shape how individuals respond to placebos (Crum, Akinola, Martin, & Fath, 2017; Crum & Langer, 2007). Further, patients themselves usually do not mention specific expectations, yet

rather spontaneously express hope (Di Blasi, Crawford, Bradley, & Kleijnen, 2005; Kaptchuk et al., 2009; D. A. Stone, Kerr, Jacobson, Conboy, & Kaptchuk, 2005). However, they are at the same time realistic and do not hope so much as to risk despair when the treatment doesn't show immediate effects (Kaptchuk, 2011). This conceptualization of hope turns out to be a kind of tragic optimism and is seen as more "visceral" than the cognitive approach that expectancy entails (Eaves, Nichter, & Ritenbaugh, 2016; Eaves, Ritenbaugh, Nichter, Hopkins, & Sherman, 2014).

Taken together and without the claim of completeness, the placebo effect is closely related to other constructs, most prominently conditioning and expectancy, but also "somatic focus", the awareness of being treated, predictions, various psychological variables and the patient-physician relationship. The question of relatedness, that is to what degree these constructs overlap or diverge, requires further investigation and would contribute to an enhanced theoretical understanding of these constructs and the placebo effect itself.

1.2. Placebo Responses in Antidepressant Trials

Almost twenty years ago, researchers had the idea to look at placebo responses in antidepressant trials for depression (Kirsch & Sapirstein, 1998). They assumed that major depressive disorder (MDD) would be good source for the investigation of placebo effects since an effective antidepressant treatment enables patients to elevate the suffering by replacing the sense of hopelessness with hopefulness (Kirsch, 2016), a mechanism which is also called remoralization (Frank, 1973, 1974) and which has been shown to be related to placebo effects (e.g., Kaptchuk et al., 2009; see above). Kirsch and Sapirstein (1998) found that improvements in the placebo groups correspond to 75% of the improvements in the antidepressant groups. The finding that only a small amount of the antidepressant response is due to the administration of an active medication, was replicated since then with correspondence rates up to 82% (Kirsch et al., 2008). Accordingly, standardized effect sizes for the antidepressant-placebo difference range from 0.15 for unpublished studies up to 0.37 for published studies (Turner, Matthews, Linardatos, Tell, & Rosenthal, 2008) and from 0.11 for mild to moderate depression up to 0.47 for very severe depression (Fournier et al., 2010). In other words, 35% to 40% of patients respond to placebo (Enck, 2016; Furukawa et al., 2016) compared with a mean antidepressant response rate of around 50% (Rutherford & Roose, 2013; Walsh, Seidman, Sysko, & Gould, 2002). Beyond statistical significance, the National Institute for Health and Clinical Excellence (NICE) proposed a definition for clinical significance in their depression guidelines from 2004 (which they replaced by the term "clinical importance" in 2010), defined as a 3-point difference between antidepressant and placebo on the Hamilton Depression Rating Scale (HDRS;

Hamilton, 1967) or a mean drug-placebo standardized mean difference of ≥ 0.50. In response to criticism that argued that the NICE criterion is arbitrary (e.g., Turner & Rosenthal, 2008), Moncrieff and Kirsch (2015) empirically demonstrated clinical significance: They compared the HDRS and the clinician-rated Clinical Global-Impression-Improvement (CGI-I) scale, showing that a 3-point difference is undetectable by clinicians using the CGI-I scale. Remarkably, the clinical significance of antidepressants is failed to be reported in meta-analyses including published and unpublished trials (Kirsch, 2016) as the range of the antidepressant-placebo difference lays between 1.80 to 2.51 points on the HDRS (Khin, Chen, Yang, Yang, & Laughren, 2011; Kirsch et al., 2008; Sugarman et al., 2014). In conclusion, there are relatively small but statistically significant differences between antidepressants and inert placebos—while, however, clinical meaningful and visible significance is not given.

Nevertheless, prescriptions of the most common antidepressants, selective serotonin reuptake inhibitors (SSRIs) and serotonin-noradrenaline reuptake inhibitors (SNRIs), have doubled in the last decade according to the NHS research (2016). In 2001, for example, an influencing trial concluded that the "treatment with paroxetine results in clinically relevant improvement in depression scores" and that their findings "provide evidence of the […] safety of the SSRI" (Keller et al., 2001, p. 770). However, general concerns were raised about relying on published research to reflect the efficacy of antidepressants. First, there is the issue of misreported trials, meaning that the study has been erroneously reported (Doshi, Dickersin, Healy, Vedula, & Jefferson, 2013). Indeed, this was the case in the Keller et al. (2001) study: A re-evaluation of the data concluded that the efficacy of the SSRI (i.e., paroxetine) is not different from placebo, whether statistically nor clinically. Moreover, "there were clinically significant increases in harms, including suicidal ideation and behavior" (Le Noury et al., 2015, p. 1) and the side effect profiles differed between paroxetine and placebo (Le Noury et al., 2015). Although placebos evoke side effects through negative expectations (Rief, Bingel, Schedlowski, & Enck, 2011), antidepressants produce significantly more adverse effects than inert placebos (Sharma, Guski, Freund, & Gøtzsche, 2016). Further, there is another basic problem in the field of antidepressant research: Many trials remain unpublished (Doshi et al., 2013). The publication bias is particularly caused by multiple publication, selective publication, and selective reporting in trials sponsored by the pharmaceutical industry (Melander, Ahlqvist-Rastad, Meijer, & Beermann, 2003). Hence, looking at the published literature, around 94% of the antidepressant studies are associated with positive outcomes. In contrast, the register of the Food and Drug Administration (FDA) indicates that only 51% of the analyses show an advantage of antidepressants over placebo (Turner et al., 2008). Notable, the criteria for

antidepressant drug approval require two "adequate and well-controlled" clinical trials indicating that the antidepressant is better than a placebo—however, there is no limit to the amount of studies and negative studies do not count (Kirsch, 2009).

Furthermore, practice guidelines identify antidepressants, including SSRIs and SNRIs, not only for MDD, yet also for anxiety disorders (AD), obsessive-compulsive disorder (OCD), and posttraumatic stress disorder (PTSD) as first line pharmaceutical treatments (American Psychiatric Association, 2010; Bandelow et al., 2012). However, few analyses have focused on these other conditions and, initially, anxiety symptom relief due to antidepressants was only investigated for panic disorder (Mitte, 2005; Otto, Tuby, Gould, McLean, & Pollack, 2001). A recent meta-analysis targeting on the efficacy of SSRIs in generalized anxiety disorder and panic disorder, as well as in depression using published and unpublished trials (Sugarman et al., 2014) provided important insights: The SSRI-placebo effect size was modest and significant, yet not clinically significant, for both, anxiety disorders (d = 0.27) and depression (d = 0.32) while the two effect sizes did not differ from each other. Also, the trend of larger placebo pre-post effect sizes in depression (d = 1.03) than in anxiety disorders (d = 0.96), was not significant, indicating that placebos are equally effective and meaningful in both disorders.

1.3. Possible Moderators

To clarify the differentiation between antidepressant and placebo groups further, the source of symptom change in antidepressant trials has been grouped into factors influencing natural history (e.g., improvement, worsening), measurement factors (e.g., regression to the mean, rater bias and response bias), treatment factors (e.g., therapeutic setting and expectancy-based placebo effects), as well as disorder characteristics (e.g., severity and duration of the disorder) (Rutherford & Roose, 2013). Hence, placebo effects are one component of the placebo response observed in clinical trials, while other components influence the symptom changes in patients randomized to the placebo arm (Rutherford & Roose, 2013). General treatment factors that contribute to the placebo response can be found in the first Section of this thesis. However, some specific findings regarding placebo effects in antidepressant trials should be mentioned. Relating to the therapeutic setting in antidepressant trials, it has been shown that the clinician administering the treatment explains more of the variability of outcomes than the psychopharmacological treatment itself (McKay, Imel, & Wampold, 2006). Also, a good way to indirectly measure expectancy-based placebo effects in antidepressant trials is to compare antidepressant response between active comparator trials (i.e., one or more antidepressant with no placebo group) and placebo-controlled trials (i.e., one or more antidepressant compared with placebo). Patients in comparator trials are aware that they have a 100% chance of receiving an

antidepressant, which, in turn, increases their expectancy of a treatment benefit, leading to enhanced placebo effects and greater antidepressant responses, which is opposed to patients in placebo-controlled trials who are aware that they may receive a placebo and show lower antidepressant response rates (Rutherford, Sneed, Devanand, Eisenstadt, & Roose, 2010; Rutherford, Sneed, & Roose, 2009). Similarly, the number of treatment arms is negatively correlated with a significant benefit of the antidepressant over placebo—a greater number of treatment arms enhances the probability of receiving a verum which may increase patients' expectations and, accordingly, placebo responses (Khan, Kolts, Thase, Krishnan, & Brown, 2004).

Regarding disorder characteristics, one of the most reported findings for depression is that the mean differences between antidepressant and placebo groups become larger as baseline severity increases. However, it is controversial whether the increasing benefits, as severity increases, of antidepressants over placebo are due to a decrease in the responsiveness to placebo treatment (Kirsch et al., 2008), or an increase in responsiveness to pharmacological intervention (Fournier et al., 2010; Khan, Leventhal, Khan, & Brown, 2002). Furthermore, a re-analysis of the Kirsch et al. (2008) data set did not find that initial severity determined antidepressant-placebo differences (Fountoulakis, Veroniki, Siamouli, & Möller, 2013), a result which has been also reported in a recent patient-level analysis from 34 randomized-controlled trials (RCTs) of antidepressants (Rabinowitz et al., 2016). Interestingly, in the latter study, the trial-level meta-analysis—which was calculated simultaneously to the patient-level analysis—supported previous findings of an association between baseline depression severity and drug-placebo differences (Rabinowitz et al., 2016)[1]. However, one should note that of four additional patient-level analyses, three reported that baseline severity of depression was associated with antidepressant efficacy (Fournier et al., 2010; Khan, Bhat, Faucett, Kolts, & Brown, 2011; Khan, Brodhead, Kolts, & Brown, 2005), whereas another large investigation did not find a significant correlation (Gibbons, Hur, Brown, Davis, & Mann, 2012).

1.4. Placebo Responses in Antidepressant Trials over the Lifespan

While for adults, a plurality of studies examines placebo responses in antidepressant trials, comparable investigations in children and elderly people are scarce. For pediatrics, the lack of research in this field is partially due to the fact that placebo research in children faces even bigger ethical considerations than placebo research in adults, since it is acknowledged that children may not fully understand the associated risks or benefits of a clinical trial and may be

[1] It is well known that trial-level meta-analyses are linked to ecologically fallacious findings, where the cumulative association fails to reproduce associations at the individual level (Spoerri et al., 2010).

more suggestible than adults (Parellada et al., 2012; Simmons et al., 2014). Further, methodological challenges in pediatric research include difficulties enrolling children, small market share for active substances aimed at children, and low prevalence of many pediatric diseases (Martinez-Castaldi, Silverstein, & Bauchner, 2008). These ethical and methodological barriers in pediatric placebo research have led to a scarcity of high-quality RCTs in children as compared to adults (Klassen, Hartling, Craig, & Offringa, 2008). Currently, more than half of the pharmacological treatments used for hospitalized children are off-label or unlicensed drugs (Conroy et al., 2000; 't Jong et al., 2000) and child health care providers must often rely on evidence that has been generated on adult populations (Cramer et al., 2005). This is especially problematic since both the safety and efficacy profiles of medications may be significantly different for children than for adults due to divergences in developmental physiology, disease pathophysiology, or developmental pharmacokinetics and pharmacodynamics (Caldwell, Murphy, Butow, & Craig, 2004; Janiaud et al., 2017). Also, when examining a complex phenomenon such as the placebo effect, it is substantial to consider the fact that age and neural development remarkably affect overall cognition (Lau et al., 2011; Simmons et al., 2014). To investigate the clinical significance of antidepressants and the proportion of the drug response which is duplicated by placebo is therefore most warranted in patients of different ages (Alamo, López-Muñoz, García-García, & García-Ramos, 2014; Henry, Kisicki, & Varley, 2012).

For youth taking antidepressants, severe side effects such as an increased risk of suicidal thoughts and behavior have been reported (Mann et al., 2006), leading to the "black box" warning on the labels of all antidepressants for pediatric use by the FDA in 2004. However, this remains controversial due to contradictory findings of re-analyses of the data (M. B. Stone, 2014). The limited meta-analytic evaluations in the field of pediatric depression reveal that antidepressants have only small to moderate effect sizes (Garland, Kutcher, Virani, & Elbe, 2016). A recent network meta-analysis indicated that for the primary outcome, only fluoxetine was significantly more effective than placebo with a moderate standardized mean difference of 0.51 on various standardized rating scales; however, all other antidepressants—including second-generation SSRIs, SNRIs, yet also first-generation tricyclic antidepressants—did not significantly differ from placebo treatments (Cipriani et al., 2016). Similarly, another meta-analysis on pediatric depression found that while SSRIs differed significantly from placebo, SNRIs and tricyclic antidepressants did not (Rojas-Mirquez et al., 2014). Even more, despite some correlation between response to antidepressant and to placebo in a meta-analysis (r = 0.47), the placebo response explained more of the variance in the efficacy than the antidepressant response (Bridge, Birmaher, Iyengar, Barbe, & Brent, 2009). The lack of a

consistent significant benefit of antidepressants over placebo in pediatric depressive disorders (Emslie, 2009; Parellada et al., 2012) has been associated with an increased response to placebo rather than a decreased antidepressant response (D. Cohen et al., 2008). This contrast is not unique: Children tend to improve more with placebo across a wide variety of diseases (Janiaud et al., 2017) and depression (Rutherford & Roose, 2013) when compared to adults.

For elderly patients with depression, the safety and efficacy of antidepressants is likewise controversial: Geriatric patients are more likely to have serious medical conditions and thus receiving polypharmacotherapy, which often causes adverse side effects and interactions between medications (Coupland et al., 2011). Age-related declines in the drug metabolism are also associated with increased rates of antidepressant side effects (Taylor, 2014). Methodologically, there is an underrepresentation of elderly people in RCTs, which can be related to study protocol restrictions with exclusion criteria on age, polypharmacotherapy, and comorbidities (Konrat et al., 2012; Zimmerman et al., 2016). Further, there is a lack of RCTs exclusively designed for geriatric patients and findings from mixed-age studies are often extrapolated from adults to elderly people (Broekhuizen, Pothof, de Craen, & Mooijaart, 2015). Again, this is troubling since pharmacokinetic and pharmacodynamics changes in the geriatric population, such as declines in the drug metabolism and extended elimination half-life, result in altered drug responses (DeVane & Pollock, 1999).

While antidepressants have shown to be more effective than placebo in depressive patients aged 55–65 years regarding remission and response rates (e.g., Kok, Nolen, & Heeren, 2012; Nelson, Delucchi, & Schneider, 2008), recent studies indicate that SSRI treatment for depression in people 65 years of age and older do not offer any benefits over a placebo (Tedeschini et al., 2011; Tham et al., 2016). Accordingly, in a RCT with depressed participants, authors reported a decrease of the average HDRS difference score between the antidepressant and placebo group with increasing age: Patients aged 45 showed a HDRS difference score of 5.6 points, while patients aged 65 reported a negligible HDRS difference score of –0.04 (Rutherford et al., 2017). However, the reason for the reduced effect of antidepressants in elderly depressed patients when compared to middle-aged patients is not yet completely understood. One possible explanation is that expectancies are minimized by multiplied negative treatment experiences throughout the lifespan (Bingel, Colloca, & Vase, 2011). This is supported by a recent investigation reporting that depressed patients aged 55 and older show a diminished expectancy effect when compared to younger participants (Rutherford et al., 2017). Also, elderly patients with early-onset depression and corresponding diminished motivation and reward-seeking behavior (Shankman, Klein, Tenke, & Bruder, 2007) report higher

posttreatment depression scores in the placebo arm than elderly patients with late-onset depression (Sneed et al., 2007). It is worth adding that in these analyses (Rutherford et al., 2017; Sneed et al., 2007), patients with cerebrovascular diseases and cognitive impairments such as dementia were not excluded; however, executive dysfunction and related learning impairments in elderly depressed patients have been associated with a loss of expectancy-related mechanisms (Benedetti et al., 2006) and a lower probability of placebo responses (Alexopoulos et al., 2005; Sneed et al., 2010). Further, there is the assumption that the reduced effect of antidepressants over placebo in late-life depression is evoked by a limited antidepressant response rather than a decreased placebo response when compared to middle-aged patients (Alexopoulos et al., 2007; Walsh & Sysko, 2005). Along similar lines, it has been shown that the placebo response remains powerful, also in late-life depression (Rutherford, Tandler, Brown, Sneed, & Roose, 2014; Sneed et al., 2008). Nevertheless, studies focusing on the placebo response and its mechanism in late-life depression are even rarer than those in pediatric depression.

1.5. Harnessing Placebo Effects

Given that patients of different ages demonstrate robust placebo responses in antidepressant trials, it is worth considering whether this can be harnessed in clinical practice (Simmons et al., 2014). Also, a genuine placebo effect in depression is extremely plausible, as the combination of placebo and supportive care has been shown to be more effective than supportive care alone in depressed patients (Leuchter, Hunter, Tartter, & Cook, 2014). However, one should bear in mind that the clinical use of placebos is not warranted since placebo administration involves deception and the violation of patients' autonomy (Gold & Lichtenberg, 2014; Trachsel & Gaab, 2016) and the ethical maxim states that "withholding medical information from patients without their knowledge or consent is ethically unacceptable" (American Medical Association, 2016: Opinion 2.1.4). Also, patients may judge the use of placebos unacceptable (Bishop, Aizlewood, & Adams, 2014). Interestingly, though, some recent research questions whether deception is indeed an unalienable characteristic of the placebo effect and insinuates the possibility of openly prescribed placebos with full transparency (Carvalho et al., 2016; Kam-Hansen et al., 2014; Kaptchuk et al., 2010; Kelley, Kaptchuk, Cusin, Lipkin, & Fava, 2012; Sandler & Bodfish, 2008). In open-label placebo studies, the placebo is provided with a scientific rationale (Kaptchuk et al., 2010). A positive albeit realistic expectancy is verbally fostered by conveying four scientific findings: (a) placebos are effective (b) classical conditioning as a possible mechanism (c) compliance is important for outcome (d) positive expectations increase placebo effects, but are not necessary.

Remarkably, patients in the open-label placebo group experienced a significantly higher symptom reduction of irritable bowel syndrome (Kaptchuk et al., 2010) and juvenile ADHD (Sandler & Bodfish, 2008; Sandler, Glesne, & Bodfish, 2010) when compared to patients in a waitlist control group, and two further studies underpinned the effectiveness of open-label placebos also compared to treatment as usual conditions in patients suffering from rhinitis (Schaefer, Harke, & Denke, 2016) and chronic back pain (Carvalho et al., 2016). In contrast, a pilot open-label study with patients suffering from major depression did not find significant improvements compared to a waitlist control group; yet a medium effect size for open-label placebos was found, exceeding standardized drug-placebo differences reported in antidepressant RCTs (Kelley et al., 2012).

Despite these promising results, open-label placebos with full disclosure have not been directly compared to deceptive placebo administration yet. Still, given the long hold belief of an inextricable interconnection between deceit and placebo usage (Foddy, 2009; Miller, Wendler, & Swartzman, 2005), an empirical investigation testing the necessity of deception in placebo application is most warranted. Such a study design requires a real life "deceptive administration" where participants in the deceptive placebo group are fully deceived at the beginning of the pharmacological treatment and have a 100% expectation of receiving an active medication, while, in fact, get a placebo. This stands in contrast to the "double-blind administration" where participants know that they have a 50% chance of receiving a placebo (Kirsch & Weixel, 1988; Koshi & Short, 2007). In the case of depressed patients, however, the comparison of an open-label placebo administration and a real life "deceptive administration" is ethically not appropriate: For instance, according to the Declaration of Helsinki in its most recent formulation, the use of placebo is only justifiable under the condition that no proven intervention exists but also where "for compelling scientifically sound methodological reasons, [it] is necessary to determine the efficacy or safety of an intervention", provided that "the patients who receive placebo . . . will not be subject to additional risks of serious or irreversible harm" (World Medical Association, 2013). Antidepressants are considered as a proven intervention and to discontinue or delay a pharmacological treatment in order to receive a deceptive placebo administration could result in a potential risk. Under these circumstances, basic research with healthy participants represent a viable alternative. In that case, also the breach of trust in physicians and the medical profession can be spared (Wendler & Miller, 2004). Here, one of the best examined conditions in placebo research with healthy volunteers is experimental pain (Benedetti, 2014) where it has been shown that the placebo treatment reveals a large effect size (Vase, Riley, & Price, 2002). Finally, at least to a certain extent, the

generalizability of experimental pain to clinical conditions is emphasized (Forsberg et al., 2016). Therefore, it can be assumed that experimental pain offers an empirical starting point for a conceptual rethinking of the necessity of deception in placebo application.

2. Aims of the Thesis

The aims of this thesis were twofold: One was to examine the efficacy of placebo and second-generation antidepressants in a pediatric and geriatric population, focusing on potential moderators of placebo outcomes. The identification of clinically significant placebo responses at certain stages over the lifespan where the efficacy and safety of antidepressants are of special concern would offer adapted treatment approaches. Here, it is worth considering whether powerful placebo responses can be harnessed in clinical practice—without violating ethical key principles of openness and transparency. Therefore, the second aim was to experimentally test the necessity of deception, by comparing ethically feasible open-label placebos with deceptive placebos.

For this reason, two different statistical approaches were indicated. For the first aim, a meta-analytic approach was applied. Meta-analyses have the potential to reduce the complexity and scope of research findings as they statistically combine the evidence of individual studies with regard to the particular research question (Guyatt et al., 1995). By calculating a pooled estimate of the intervention effect by means of the intervention effects extracted from each of the included studies, a more reliable result can be obtained (Glass, 1976).

For the second goal, a basic research approach with healthy participants was chosen. Experimental studies have the potential to provide greater insight into sources of variability of placebo effects since they allow to control for potential confounders (Price et al., 2008).

The three investigations described in this thesis were therefore developed to add insight to the following research questions:

(1) What potential do placebos have in pediatric and geriatric patients with depression when compared to antidepressants?

Study I. The use of SSRIs and SNRIs in children and adolescents is still debated. Nevertheless, SSRIs and SNRIs are first line pharmaceutical treatments not only for MDD, yet also for AD, OCD and PTSD. At present, there is only one study comparing antidepressants in children and adolescents across these disorders. About ten years ago, Bridge et al. (2007) reported that the between antidepressant-placebo effect was strongest for anxiety disorders (Hedges' g = 0.69), intermediate for OCD (Hedges' g = 0.41), and only modest in MDD

(Hedges' $g = 0.20$). This could be due to disorder-specific differences in placebo responses, which were not reported in the meta-analyses. Accordingly, D. Cohen et al. (2008) stated in a systematic review that the placebo response rate was significantly higher in pediatric studies on depression than in those AD and OCD. This is opposed to adult studies that found no significant differences in placebo effect size between depression and anxiety (Sugarman et al., 2014). Further and in contrast to adult studies, the influence of baseline severity on antidepressant-placebo differences has rarely been studied in pediatric meta-analyses and did not emerge as a significant moderator in some studies (Bridge et al., 2009; Gibbons et al., 2012), whereas another meta-analysis reported the moderating effect of initial severity (Tsapakis, Soldani, Tondo, & Baldessarini, 2008). Thus, the goal of Study I was to conduct an updated and extended review to assess continuous mean differences between antidepressant and placebo interventions in pediatric MDD, AD, OCD, and PTSD along with between-disorder variation in placebo responses and moderators.

Study II. Regarding late-life depression, the safety and efficacy of antidepressants is likewise controversial. Meta-analyses reveal that overall antidepressant effects are only modest for response and for remission when compared to a placebo administration (Kok et al., 2012; Nelson et al., 2008; Tedeschini et al., 2011; Tham et al., 2016). However, for all of these studies, statistical differences were only calculated for response and remission rates, yet not for continuous outcome data which would enable to evaluate clinical significance according to the NICE criterion. Also, dichotomizing continuous scores into categorical outcome data leads to a loss of information, reduces power and creates an artificial boundary (Altman & Royston, 2006; Moncrieff & Kirsch, 2005). In contrast to mixed-aged findings, meta-analyses reported no association between initial severity and antidepressant over placebo efficacy in elderly people (Gibbons et al., 2012; Nelson, Delucchi, & Schneider, 2013) with the exception of patients suffering from depression for longer than 10 years (Nelson et al., 2013). However, these investigations rely on a limited number of studies and focused on MDD, only. It should be noted that a minority of depressed elderly patients fulfill the diagnostic criteria for MDD, yet elderly patients tend to underreport their symptoms, often relate it to physical burden and ageing itself (Giron, Fastbom, & Winblad, 2005), thus the rate of sub-threshold depression rises with age (Pinquart, Duberstein, & Lyness, 2006). Hence, in Study II, the goal was to conduct a meta-analysis in order to evaluate continuous mean differences between antidepressant and placebo interventions and to test whether baseline severity has an influence on outcome in a geriatric population with MDD, dysthymia and minor depression.

(2) How important is a deceptive placebo administration when compared with an ethically feasible open-label placebo administration in healthy participants?

Study III. To date, open-label placebos with full disclosure have not been directly compared to deceptive placebo administration. Here, basic research with healthy participants represents an excellent way to experimentally compare these treatments in accordance with ethical principles. Therefore, the aim of study III was to compare the effects of open-label placebos with a deceptive placebo administration in a standardized heat pain experiment (Gaab et al., 2017; Krummenacher et al., 2014; Maeoka, Hiyamizu, Matsuo, & Morioka, 2015) in healthy participants. Additional control groups were an open-label placebo group without a rationale and a no treatment group.

3. Methods

3.1. Efficacy and Safety of SSRIs, SNRIs, and Placebo in Common Psychiatric Disorders: A Comprehensive Meta-Analysis in Children and Adolescents (Study I)

Search strategy and study selection. We searched PubMed, Embase, PsycINFO, Cochrane, Web of Science, Clinicaltrials.gov and fda.gov and checked references of originally identified papers and reviews. Randomized, double blind, placebo-controlled trials of SSRIs and SNRIs in children and adolescents < 18 years of age published through August 2016 were included. Subjects were required to have a diagnosis of MDD, AD, OCD, or PTSD based on DSM-III, DSM-III-R, or DSM-IV-TR criteria.

Outcome measures and data extraction. The primary outcome as defined by authors was chosen as the sole outcome measure for each individual study. Pre- and post-intervention data or mean change data had to be available. Outcomes had to be reported on a well-validated disorder specific scale (e.g., CDRS-R, MASC, CY-BOCS) or on a general severity scale (i.e., CGI-S). Only continuous outcome data were included.

Data analysis. Comprehensive Meta-Analysis V3 (Biostat, Englewood, NJ) and R 3.2.1 (R Foundation; Vienna, Austria) were used for calculations and analyses. Effect sizes were calculated as Hedges' *g* (Hedges & Olkin, 1985). Random-effects models rather than fixed-effects models were used. Fixed-effects model assumes that there is one true effect size for all studies and any variations are due to sampling error, whereas a random-effects model assumes that variations in effect sizes for the samples are a combination of sampling error and true variance in effect sizes (Borenstein, Hedges, Higgins, & Rothstein, 2010). Random effect sizes

were preferable for this meta-analysis as the studies we included were heterogeneous and the number of studies for the sub-analyses was relatively small. Three effect sizes were calculated for each included study. First, differences in mean change scores between groups were evaluated. Then, within-group pre-post effect sizes for both antidepressant and placebo were also calculated. They inform about whether a small difference between groups is explained by a small change in either group or a meaningful change in both groups over treatment time. Effect size differences between subgroups were analyzed using a mixed-effects model (Borenstein et al., 2010). Heterogeneity was assessed by calculating the Q statistic (Cochran, 1950), the τ^2, and the I^2 (Higgins, Thompson, Deeks, & Altman, 2003). A statistically significant Q indicates a heterogeneous distribution, meaning that systematic differences between studies are present and it rejects the null-hypothesis that all the variation in effects is due to random error. Similarly, the higher the Q value, the more variation in the studies can be explained by a true variance of effects between studies (Cochran, 1954). In addition, the I^2 statistic was used to quantify inconsistency. It measures the proportion of observed variance across studies, that is a result of real heterogeneity rather than chance. An I^2 value of 0% indicates no heterogeneity, a value of 25% is classified as low, 50% as moderate, and 75% as high (Higgins et al., 2003). The τ^2 offers an estimate of the variance among true effect sizes (Higgins, 2008). Treatment emergent adverse events (TEAEs) and serious adverse events (SAEs) were examined across diagnosis categories to determine if emergent events differ between antidepressant and placebo administration. We also examined whether baseline severity of the studies is related to the between-group effect sizes. As various scales were used to assess baseline severity, we standardized the baseline and outcome values by dividing the mean values by the *SD*. The Z statistic was used to test the significance of the slope. In the case that data conforms to the null hypothesis, Z has a normal distribution. A significant Z would indicate that the slope is probably not zero, and hence that baseline depression severity moderates the effect size (Borenstein, Hedges, Higgins, & Rothstein, 2011).

3.2. Moderation of Antidepressant and Placebo Outcomes by Baseline Severity in Late-Life Depression: A Systematic Review and Meta-Analysis (Study II)

Search strategy and study selection. We performed searches in Cochrane, Embase, PsycINFO, PubMed, and Web of Science on studies published through September 30, 2014. We included peer-reviewed randomized, double-blind, placebo-controlled clinical trials of depressed elderly individuals in a placebo group with depressed elderly individuals in an intervention group receiving second-generation antidepressants. The minimum age criterion was set at a mean or median age of 55 years, or described as elderly, geriatric or older adults.

We included studies investigating patients with MDD or subclinical depressive symptoms (i.e., minor depressive disorder or dysthymia) based on DSM-III, DSM-III-R, DSM-IV or DSM-IV-TR. Studies in which patients had depression following cerebrovascular disease, a cognitive impairment, or Parkinson's disease were excluded.

Outcome measures and data extraction. Outcomes had to be reported as mean change in depressive symptoms on a continuous mood scale, such as the HDRS (Hamilton, 1967) or Montgomery-Asberg Depression Rating Scale (MADRS; Montgomery & Asberg, 1979).

Data analysis. Comprehensive Meta-Analysis V2 (Biostat, Englewood, NJ) was used for calculations and analyses. Effect sizes were calculated as Hedges' *g* (Hedges & Olkin, 1985). For the analyses we chose to use random-effects models rather than fixed-effects models (Borenstein et al., 2010; see Study I for details). Differences in mean change scores between groups were evaluated. Moreover, within-group pre-post effect sizes for both antidepressant and placebo were calculated. Heterogeneity was assessed by calculating the Q statistic (Cochran, 1950) and the I^2 (Higgins et al., 2003; see Study I for details). To assess the moderating effect of baseline depression severity on outcome measure (i.e., mean change in depressive symptoms), we conducted within-group and between-group comparisons. We first converted the different mood scales to a standardized scale (range 0–100), using the largest point of each mood scale as 100%. The Z statistic was used to test the significance of the slope (see Study I for details). To test the moderating effect of baseline depression severity on outcome measures within each group, we performed meta-regression using method-of-moments analyses (random-effects model). Further, we analyzed the mean difference scores to test the hypothesis that between-groups mean differences increase as a function of baseline depression severity. We calculated the overall baseline severity for each study (i.e., mean of antidepressant and placebo baseline severity, weighted by number of participants) and the antidepressant-placebo difference in improvement. Pearson's correlation between those two variables was calculated.

3.3. Is the Rationale More Important than Deception? A Randomized Controlled Trial of Open-Label Placebo Analgesia (Study III)

Between January 2016 and July 2016, healthy adults from the general Swiss population were recruited via advertisements for "a novel mind-body management study of individual pain perception". They had to be healthy by self-report, right-handed, aged between 18 and 65 years and have sufficient German language skills.

Data analysis. Participants were randomly assigned to no treatment (NT; $N = 40$), open-label placebo without rationale (OPR^-; $N = 40$), open-label placebo with rationale (OPR^+; $N =$

40), and deceptive placebo (DP; $N = 40$). Upon arrival, all participants performed an objective baseline assessment of heat pain, as well as subjective heat pain ratings. After baseline measurements, the treatment phase was conducted. Participants in the NT group did not receive any treatment and were told that they are in the "no treatment group". All participants in the three other groups (OPR^-, OPR^+, and DP) received an inert white placebo cream. In the OPR^- group, participants were informed that they are receiving an inert placebo cream and no additional information regarding placebo mechanisms was provided. In the OPR^+ group, participants were informed that they are receiving an inert placebo cream and obtained scientific explanations of the effects and mechanisms of placebos which were in accordance with the study of Kaptchuk et al. (2010). In the DP group, participants were told that they are receiving an analgesic cream—named "Antidolor", containing the active substance Lidocaine—in fact they received an inert placebo cream. After the treatment phase, heat pain measurement procedures were performed again (posttreatment), regardless of group allocation.

Primary outcomes. Pain sensation was assessed using the suprathreshold method of the Thermo Sensory Analyser (TSA-II). Objective heat pain tolerance was determined using the method of limits: Participants were asked to stop the increasing heat stimulus at the moment they could not stand the heat any longer. Three measurements were taken, each starting at 32 °C, with a rise of 0.5 °C every second (Granot, Sprecher, & Yarnitsky, 2003). Following each heat pain tolerance stimulation, participants were asked to rate pain intensity and unpleasantness on Visual Analogue Scales (VAS) (Price et al., 1999). The intensity VAS with a range from 0 up to 100 was titled at the left by the descriptors "no pain sensation" and at the right by "the most intense pain sensation imaginable". Similarly, the unpleasantness VAS (ranged from 0–100) was anchored by the descriptors "not at all unpleasant" and "the most unpleasant imaginable".

Statistical analyses. Primary outcomes were objective heat pain tolerance and the corresponding subjective intensity and unpleasantness ratings. For posttreatment heat pain tolerance, we calculated one one-way analysis of covariance (ANCOVA), using the treatment group (NT, OPR^-, OPR^+, and DP) as independent between-subject factor and baseline heat pain tolerance as covariate. Regarding intensity and unpleasantness ratings for heat pain tolerance at posttreatment, we conducted two separate one-way ANCOVAs with treatment group as between-subject factor and the corresponding outcome variable measured at baseline as covariate. For all three primary outcome ANCOVAs, we tested three orthogonal planned contrasts (one-tailed): (c1) NT group vs groups with a cream application (OPR^-, OPR^+, and DP), (c2) OPR^- group vs groups with a rationale (OPR^+ and DP) in order to test the significance

of the rationale; and (c3) OPR^+ group vs DP group in order to evaluate the significance of deception. We calculated Cohen's *d* in order to describe the standardized mean difference of an effect. We interpreted effect sizes as small ($d = 0.2$), medium ($d = 0.5$), and large ($d = 0.8$) based on benchmarks suggested by Cohen (J. Cohen, 1988). R 3.3.2 (R Foundation; Vienna, Austria) was used for calculations and analyses.

4. Summary of the Results

4.1. Efficacy and Safety of SSRIs, SNRIs, and Placebo in Common Psychiatric Disorders: A Comprehensive Meta-Analysis in Children and Adolescents (Study I)

Our search identified thirty-six randomized, double blind trials including 6778 participants that compared an SSRI or an SNRI against placebo in patients < 18 years with a diagnosis of MDD ($N = 17$), AD ($N = 10$), OCD ($N = 8$), or PTSD ($N = 1$). No disorder-specific subgroup analyses were calculated for PTSD, since only one study fit our inclusion criteria.

In the between group analysis stratified by disorder, AD ($g = 0.56$, 95% CI [0.40, 0.72], $p < .001$) and OCD ($g = 0.39$, 95% CI [0.25, 0.54], $p < .001$) did not differ from each other ($p = .14$) but both were significantly higher (AD vs MDD: $p < .001$ and OCD vs MDD: $p = .02$) than the MDD group ($g = 0.20$, 95% CI [0.13, 0.27], $p < .001$). The within antidepressant group analysis stratified by disorder yielded no significant difference ($p = .06$) between studies of AD ($g = 1.58$, 95% CI [1.35, 1.81], $p < .001$) and MDD ($g = 1.85$, 95% CI [1.7, 2.0], $p < .001$), yet both yielded significantly larger effect sizes (ps < .001) than studies of OCD ($g = 1.01$, 95% CI [0.88, 1.14], $p < .001$). The within placebo group analysis stratified by disorder yielded a large placebo response for studies of MDD ($g = 1.57$, 95% CI [1.36, 1.78], $p < .001$), which was significantly larger ($p < .001$) than the effect size for studies of AD ($g = 1.02$, 95% CI [0.85, 1.20], $p < .001$). The moderate placebo response in the OCD group ($g = 0.63$, 95% CI [0.47, 0.79], $p < .001$) was significantly lower than both the MDD ($p < .001$) and AD ($p = .001$) groups.

Side effect analysis. Compared to placebo, patients taking both SSRIs and SNRIs reported significantly more SAEs (SSRI: 6.80%, SNRI: 7.59%) compared to placebo (3.32%; $p = .01$), but reported no significant difference in TEAEs.

Moderator analysis. The relationship between effect size and baseline severity was not significant, whether in the MDD ($Z = 1.21$, $p = .23$), nor in the AD ($Z = -1.18$, $p = .24$), or OCD ($Z = -0.33$, $p = .74$) subgroup analyses.

4.2. Moderation of Antidepressant and Placebo Outcomes by Baseline Severity in Late-Life Depression: A Systematic Review and Meta-Analysis (Study II)

In total, 19 studies met inclusion criteria, including a total of 5737 elderly depressed patients, of whom 3226 received active drug and 2511 received placebo. Based on HDRS17, classification of baseline depression severity ranged from mild to very severe.

Combined over all mood scales, patients in the treatment groups showed a significantly higher mean change in depressive symptoms than patients in the placebo groups ($g = 0.37$, 95% CI [0.27, 0.46], $p < .001$). Pre-post effect sizes revealed that there is a significant treatment improvement in antidepressant groups ($g = 1.35$, 95% CI [1.14, 1.57], $p < .000$), as well as in placebo groups ($g = 0.96$, 95% CI [0.79, 1.13], $p < .000$). To clarify further the clinical significance of the differences between the two treatments, we referred to the NICE guidelines (2004). For elderly patients with mild to moderate depression (HDRS17 score of ≤ 18), Cohen's d was .24 (95% CI [0.02, 0.47], $p = .031$), for patients with an HDRS17 score in the severe range (19–22), Cohen's d was .39 (95% CI [0.25, 0.52], $p < .001$), and for patients with HDRS17 score in the very severe range (≥ 23), we found an effect size of $d = .37$ (95% CI [0.09, 0.66], $p = .011$). In summary, none of the values reached the proposed cut-off of $d = .5$ for clinical significance. However, the criterion of a difference of ≥ 3 points on the HDRS was met for baseline HDRS17 scores of ≥ 21.

Moderator analysis. The slope representing the overall relationship between baseline severity and change in symptoms was not significant in either antidepressant groups ($Z = 1.47$, $p = .142$) or placebo groups ($Z = 1.38$, $p = .168$), nor was baseline severity significantly correlated with drug-placebo differences ($r = .17$, $p = .392$). In contrast, subgroup analyses of studies using the clinician-rated HDRS mood scale indicated that mean change in depressive symptoms increased significantly in antidepressant trials ($Z = 2.67$, $p = .008$, $R^2 = .40$) and placebo trials ($Z = 4.46$, $p < .000$, $R^2 = .50$) as a function of HDRS baseline severity. However, this would be expected as a result of regression toward the mean, and mean differences between groups did not increase ($r = .19$, $p = .469$) as a function of baseline severity.

4.3. Is the Rationale More Important than Deception? A Randomized Controlled Trial of Open-Label Placebo Analgesia (Study III)

Participants had a mean age of 27.15 ($SD = 9.51$) years and 68% of the participants were female.

Objective heat pain tolerance. Planned contrasts indicated that the groups did not differ regarding their objective heat pain tolerance at posttreatment (NT vs. OPR^-, OPR^+, and

DP: $t(146) = 0.35$, $p = .724$; OPR^- vs. OPR^+ and DP: $t(146) = 1.15$, $p = .254$; OPR^+ vs. DP: $t(146) = 0.37$, $p = .711$).

Subjective heat pain ratings. Planned contrasts indicated that the NT group and the other three groups did not differ in heat pain intensity ratings at posttreatment (c1: $t(146) = -0.44$, $p = .658$). However, the two groups with a rationale (OPR^+ and DP group) showed significantly lower ratings at posttreatment when compared to the OPR^- group (c2: $t(146) = -2.15$, $p = .033$, $d = 0.43$). Further, the OPR^+ and DP group did not differ from each other (c3: $t(146) = -1.10$, $p = .272$). Results for heat pain unpleasantness ratings at posttreatment were similar. No difference was found between the NT group and the three other groups (c1: $t(146) = -1.38$, $p = .169$). Participants in the two groups with a rationale (OPR^+ group and DP) reported lower ratings at posttreatment compared to participants from the OPR^- group (c2: $t(146) = -2.43$, $p = .016$, $d = 0.49$), and the OPR^+ and DP group did not differ from each other (c3: $t(146) = -0.05$, $p = .961$).

5. General Discussion

One aim of this thesis was to assess differences between antidepressant and placebo interventions in pediatric (Study I) and geriatric (Study II) patients in order to evaluate the potential of placebos in age categories where the efficacy and safety of antidepressants are of special concern. Given that patients of different ages demonstrate robust placebo responses in antidepressant trials, it is worth considering whether this can be harnessed in clinical practice. Despite promising results of an ethically feasible placebo administration with full disclosure, open-label placebos have not been directly compared to deceptive placebo administration yet. Therefore, a further goal of this thesis (Study III) was to compare the effects of open-label placebos with a deceptive placebo administration in a standardized heat pain experiment.

Results of study I and II showed that second-generation antidepressants are more effective than placebo at treating depression in children and adolescents (Hedges' $g = 0.20$, Study I), as well as in elderly people (Hedges' $g = 0.37$, Study II). However, in both cases, the effect is only small and did not reach the proposed cut-off of a standardized effect size of 0.5 for clinical significance according to the NICE criterion. The small effect size between antidepressants and placebo in pediatric and geriatric MDD might be due to lack of a clear depression phenotype. This was apparent in DSM-5 Field Trials, which found a low and questionable test-retest reliability (kappa: 0.28) for MDD (Regier et al., 2013). Especially for children, it may be more challenging to diagnose depression, as the symptoms are assumed to be more nonspecific than in adult depression (Parellada et al., 2012). Similarly, there is substantial heterogeneity in late-life depression and sub-threshold or minor depressive disorders

among elderly people are common (Taylor & Doraiswamy, 2004). Also, geriatric patients showed a smaller within-antidepressant group effect size (Hedges' g = 1.35, Study II) when compared to pediatric patients with depression (Hedges' g = 1.85, Study I). These preliminary findings are consistent with the assumption that geriatric patients have a limited response to antidepressants or require more time to respond to medication treatment (Alexopoulos et al., 2007; Walsh & Sysko, 2005).

Also, findings of our meta-analyses revealed that the within-placebo group effect size in depressed youth (Hedges' g = 1.57, Study I), as well as in depressed elderly (Hedges' g = 0.96, Study II), is substantial. On the one hand, it is possible that placebo effects might be even more substantial than reported. Thus, placebo effects are larger in experimental studies which explicitly investigate mechanisms of placebos than in RCTs where placebos are only used as a control condition (Vase et al., 2002). Further, the test of the blind in double-blind designs usually reveals that group allocation is penetrated and thus susceptible to the researcher's assumption that the active drug will be more effective than the placebo (Fisher & Greenberg, 1993). On the other hand, the observed placebo response can be attributed to a variety of factors such as natural history factors (e.g., improvement, worsening), and measurement factors (e.g., regression to the mean, rater bias and response bias [Rutherford & Roose, 2013]), as well as confounding variables such as unreported co-interventions (Benedetti, 2008). Moreover, the additivity assumption, that the difference between antidepressant response and placebo response is attributable to the pharmacological effect of the antidepressant, has rarely been tested in scientific research (Wager & Roy, 2010) and may be incorrect (Kirsch, 2000), thus leading to underestimated antidepressant effects in RCTs (Lund, Vase, Petersen, Jensen, & Finnerup, 2014). Generally, it is noteworthy that the placebo efficacy in late-life depression is about the same when compared to mixed-aged studies (e.g., Sugarman et al., 2014), while, however, children and adolescents show a substantially higher placebo response. These comparisons—which should be directly compared in a future meta-analytical or experimental investigation in order to give conclusive answers—underline the finding that children and adolescents tend to improve more with placebos when compared to adults (Janiaud et al., 2017; Rutherford & Roose, 2013).

Study I and II did not confirm the severity hypothesis proposed in mixed-aged studies, in which an increasing advantage of antidepressants over placebo was reported with increasing baseline severity (Fournier et al., 2010; Khan et al., 2002; Kirsch et al., 2008). Our failure to find an association between initial severity and antidepressant over placebo efficacy is similar to previous pediatric and geriatric studies (Bridge et al., 2009; Gibbons et al., 2012; Nelson et

al., 2013). Here, it should be questioned whether the placebo moderators that have been explored so far are uniquely important for adults as opposed to children and elderly people.

Overall, the significant response to placebo in pediatric and geriatric MDD indicates that children and elderly people with depression might benefit from innovative treatment modalities that attempt to harness the power of the placebo effect in an ethical fashion. Here, findings of study III revealed that open-label placebos do not differ in their effects from deceptive placebos in subjective outcomes. Therefore, the ubiquitously assumed necessity of concealment in placebo administration is questioned.

5.1 Efficacy and Safety of SSRIs, SNRIs, and Placebo in Common Psychiatric Disorders: A Comprehensive Meta-Analysis in Children and Adolescents (Study I)

The robust response to placebo in pediatric MDD is associated with the finding that especially depressed youths have a maximum benefit from contact with research staff who invests in the therapeutic alliance and promotes confidence and self-esteem (D. Cohen et al., 2008; Rutherford et al., 2011). Interestingly, patient expectancy in clinical trials seems to play a minor role in the treatment of depressed children and adolescents when compared to adults since for pediatric depression, antidepressant response in comparator and placebo-controlled trials do not differ significantly (Rutherford et al., 2011). Here, it has been proposed that children are less cognitively primed to understand the rationale of the study in which they are participating, and that they receive less information at the study enrollment since the parents provide informed consent (Rutherford & Roose, 2013). However, especially in depressed pediatric patients, family members have an emotional response to the benefit of the treatment and hope that the intervention will work (Grelotti & Kaptchuk, 2011; Simmons et al., 2014) and parental expectancies, in turn, have a great impact on the treatment outcome of the child (Whalley & Hyland, 2013).

While it appears that the response to placebo is robust in pediatric MDD, children and adolescents with anxiety disorders, who respond to antidepressants to the same degree as those with MDD, do not appear to exhibit such a robust placebo response. While in line with older reviews in children (D. Cohen et al., 2008) this is in contrast to adult studies that found no significant differences in placebo effect size between depression and anxiety (Sugarman et al., 2014). One explanation might be that children and adolescents with MDD may be more demoralized than patients with anxiety disorders and are therefore more sensitive to changes in hope and favorable meanings (D. Cohen et al., 2008). Further, patients with OCD exhibited a significantly smaller response to both antidepressant and placebo treatment compared to AD and DD. Here, another feasible explanation is possible: Treatment expectations could vary

between disorders since antidepressants have been widely promoted for depression (Lacasse & Leo, 2005), yet considerably received less focus in the general population for the treatment of anxiety disorders and OCD. However, as no pediatric trial included a no-treatment group that could serve as a control for the natural course of the disorders, the difference in placebo response may also reflect differences in the probability of spontaneous improvement between the pediatric disorders rather than differences in the placebo effect.

With regard to side effects, our meta-analysis found that pediatric patients taking antidepressants do not report more treatment emergent adverse events than those assigned to placebo. Here, it has been shown that negative expectations from investigators and patients can influence adverse effect reporting (Rief et al., 2011) and that depressed patients may attribute pre-existing symptoms, which are common in the general population (e.g., headache, abdominal discomfort) to the intake of antidepressants (Rief et al., 2009). However, these explanatory approaches may not apply to serious adverse events. In our study, the serious side effects profiles for antidepressants significantly exceeded that of the placebo group. Here, our results regarding serious side effect profiles support concerns about the safety of antidepressants in children and adolescents (Bridge et al., 2007; Sharma et al., 2016).

5.2. Moderation of Antidepressant and Placebo Outcomes by Baseline Severity in Late-Life Depression: A Systematic Review and Meta-Analysis (Study II)

The finding of a meaningful placebo response in late-life depression is in line with present knowledge. Hence, augmented clinical visits and supportive care generate greater placebo, yet not antidepressant response in elderly depressed patients (Rutherford et al., 2014). Also, in clinical trials with depressed elderly patients, perceived social support is associated with subsequent decrease in depressive symptoms (Oxman & Hull, 2001) and increases probability of recovery (Bosworth, Hays, George, & Steffens, 2002). It has been argued that the impact of social support on treatment outcome is especially relevant for elderly patients, as they often live alone and may have little social contact (Bingel et al., 2011). Likewise, it has been reported that self-rated reclusiveness predicts response in a supportive patient-practitioner relationship (Conboy et al., 2010). Also, as in mixed-aged studies, antidepressant response rates are significantly higher (60%) in comparator trials than those in placebo-controlled trials (46%) in late-life depression, indirectly underlining the impact of expectancy on treatment outcome (Sneed et al., 2008).

The finding of a non-association of baseline severity and antidepressant-placebo differences may be influenced by the initial severity grades of the trials under investigation. We included only one study of very severely depressed patients reporting a HDRS baseline severity

of 26.9 (Heun et al., 2013). The fact that only a few investigations have focused on severely depressed elderly patients (Kok et al., 2012) may be closely related to restrictive inclusion criteria in clinical trials for late-life depression (Konrat et al., 2012; Zimmerman et al., 2016).

5.3. Is the Rationale More Important than Deception? A Randomized Controlled Trial of Open-Label Placebo Analgesia (Study III)

We found that healthy participants given open-label placebos with a persuasive rationale showed a decrease in subjective heat pain intensity and unpleasantness ratings of pain tolerance which, surprisingly, did not differ significantly from deceptive placebo administration. Hence, the necessity of deception in placebo application—at least in healthy participants—needs to be reconsidered. This indicates that the ethically troubling component of placebos—a counterfeit rationale (Blease, Colloca, & Kaptchuk, 2016)—may under certain circumstances be comparable to a transparent and scientific rationale.

In line with our hypothesis, the provision of a convincing rationale, either open or deceptive (OPR^+ and DP), did outperform the placebo cream application without any rationale (OPR^- group) with regard to subjective ratings for heat pain tolerance. This confirms the finding of an older study, showing that for a verum (i.e., naproxen) as well as for a placebo, the analgesic effect is significantly better in the informed-consent group when compared to the uninformed group (Bergmann et al., 1994). The impact of the comprehensible theoretical embedding, which was offered to the participants in both groups, underlines the importance of plausibility and conviction of treatment rationales (Borkovec & Nau, 1972). Our finding is also interesting when focusing on the one pilot study with only 20 participants investigating open-label placebo administration in patients suffering from depression (Kelley et al., 2012) which reported ambiguous findings: There were no significant improvements in comparison to a waitlist control group, though a medium effect size for open-label placebos was found (d = 0.54), which is larger than typical antidepressant-placebo differences in clinical trials and above the criterion for clinically significant improvement defined by NICE. It should be noted, however, that the rationale of the open-label placebo administration was in accordance with the Kaptchuk et al. (2010) study, which was originally developed for patients suffering from pain conditions (see Section 1.5 for details). Here, it is most likely that additional openly communicated and disorder-specific explanatory mechanisms besides classical conditioning, such as the impact of the patient-physician relationship and psychological variables (e.g., the transformation from hopelessness into hope, called remoralization), would make the rationale more plausible—an aspect which is utmost important for any treatment to be effective (Frank, 1986; Wampold & Imel, 2015).

5.4. Limitations

Our non-significant findings regarding the severity hypothesis in study I and II must be considered with caution. First, for both studies, heterogeneity was small to moderate between trials, making it very unlikely to find moderators. Also, multivariate analyses in study I may have lacked the power to reveal statistically significant interaction effects (Gerger, Hlavica, Gaab, Munder, & Barth, 2015). A randomized trial in which participants are stratified by the predictor under investigation (Cuijpers, Van Straten, Warmerdam, & Smits, 2008) or a patient-level meta-analysis (Fournier et al., 2010) might be better suited to investigate the association between baseline severity and effect sizes. Regarding our meta-analyses with depressed children and adolescents (Study I), none of the included trials directly compared the effectiveness across disorders and, therefore, only indirect conclusion with regard to disorder specificity can be made. Also, it is well known that if studies are genuinely diverse, the pooled effect-size may be invalid (the so-called apple and oranges problem) (Gerger & Gaab, 2016; Sharpe, 1997). However, we mainly focused on effect sizes stratified by disorder, rather than the combined effect sizes across all disorders. The meta-analysis with geriatric patients (Study II) was limited to published data, which may have resulted in a considerable bias towards studies reporting a benefit of antidepressants over placebo (Turner et al., 2008).

Regarding study III, subjective intensity and unpleasantness ratings are not independent of the corresponding objective heat pain tolerance. Thus, differences in subjective heat pain ratings could be suppressed since assessing heat pain tolerance allows participants to stop the pain stimulus at different points in time. However, in our study, objective heat pain tolerance and the corresponding subjective ratings were not correlated. Further, we did not find the hypothesized group differences between the no treatment group and the combined effect of the three other groups. This may be due to the conceptual heterogeneity between the three groups receiving the placebo cream. Also, we only found significant group differences in subjective heat pain ratings and not in objective heat pain tolerance. Nevertheless, our findings are in line with the view that placebos primarily affect subjective self-report (Kaptchuk & Miller, 2015)

5.5. Conclusions and Implications for Future Research

Despite these limitations, the result of this thesis—that there is a substantial placebo response in depressed children and elderly people—shows that those patients seem to respond well to clinician contact that promotes the therapeutic alliance and other common factors such as the patients' expectations and hopes of improvement. This is especially relevant since both the safety and efficacy of antidepressants is controversial in these sensitive populations. Here, a stepwise approach is proposed, i.e., to initially offer depressed children and elderly people

psychosocial interventions—a safer alternative—and only consider antidepressants if patients do not response. Further, the results of this thesis point to the need for additional research into the factors that moderate the efficacy of antidepressants and placebos in children and elderly people; similarly, additional research in understanding the developmental ontogeny of the placebo response is highly warranted.

Besides the use of psychosocial interventions, another way to harness the powerful effect of placebos in clinical practice without the violation of patients' autonomy are open-label placebos. Here, survey findings suggest that adult patients accept the idea of using placebos within the clinic, but their attitudes depend on several factors, such as transparency (Fässler, Gnädinger, Rosemann, & Biller-Andorno, 2011). Similarly, a recent review of parental attitudes about placebo use in children revealed that parents would acknowledge the use of open-label placebos (Faria et al., 2017). Authors of earlier studies already argued that open-label placebos could be prescribed with a "wait and watch" strategy before the administration of drugs (Kaptchuk et al., 2010) or to be administered after repeated intake of active drugs to achieve drug-like effects (Colloca, Enck, & DeGrazia, 2016). Moreover, open-label placebos may have the potential to work in treatment resistant patients, assumedly due to a form of empowerment (Carvalho et al., 2016). The question arises whether open-label placebos could be used as vehicles to boost placebo effects in depression. Experimental open-label placebo studies investigating the role of a plausible rationale in the field of depression are surely warranted (see Section 5.3. for details). Moreover, one should bear in mind that antidepressants are viewed as a long-term treatment including an acute, continuation and maintenance phase (Cartwright, Gibson, Read, Cowan, & Dehar, 2016). Withdrawal reactions when antidepressants are discontinued include a wide range of physical and psychological symptoms and occur with any type of antidepressants (Fava, Gatti, Belaise, Guidi, & Offidani, 2015). When patients taper down their dosage, withdrawal symptoms are usually attributed to pathophysiological mechanisms. However, it is well known that an open (i.e., expected) interruption of drugs is accompanied by fear and negative expectations of symptom relapse, mechanism which are characteristic for the nocebo effect (Benedetti, Lanotte, Lopiano, & Colloca, 2007). Here, open-label placebos could bear the potential to be deployed as a replacement therapy. With regard to clinical care, findings of study III, indicate that an open application of a placebo with a convincing rationale is more effective than an open application of a placebo without a rationale or theoretical embedding and call attention to whether physicians may best benefit from placebo effects by enhancing patients' expectation through communication and a convincing story behind any intervention.

6. References

Alamo, C., López-Muñoz, F., García-García, P., & García-Ramos, S. (2014). Risk-benefit analysis of antidepressant drug treatment in the elderly. *Psychogeriatrics, 14*, 261–268. doi:10.1111/psyg.12057

Alexopoulos, G. S., Kanellopoulos, D., Murphy, C., Gunning-Dixon, F., Katz, R., & Heo, M. (2007). Placebo response and antidepressant response. *American Journal of Geriatric Psychiatry, 15*, 149–158. doi:10.1097/01.JGP.0000232206.91841.d9

Alexopoulos, G. S., Kiosses, D. N., Heo, M., Murphy, C. F., Shanmugham, B., & Gunning-Dixon, F. (2005). Executive dysfunction and the course of geriatric depression. *Biological Psychiatry, 58*, 204–210. doi:10.1016/j.biopsych.2005.04.024

Alfano, M. (2015). Placebo effects and informed consent. *American Journal of Bioethics, 15*(10), 3–12. doi:10.1080/15265161.2015.1074302

Altman, D. G., & Royston, P. (2006). The cost of dichotomising continuous variables. *BMJ, 332*, 1080. doi:10.1136/bmj.332.7549.1080

American Medical Association. (2016). *Code of Medical Ethics: Opinion 2.1.4 – Use of Placebo in Clinical Practice*. Council on Ethical and Judicial Affairs. Retrieved from http://www.ama-assn.org/delivering-care/ama-code-medical-ethics

Bandelow, B., Sher, L., Bunevicius, R., Hollander, E., Kasper, S., Zohar, J., . . . WFSBP Task Force on Anxiety Disorders, OCD and PTSD. (2012). Guidelines for the pharmacological treatment of anxiety disorders, obsessive-compulsive disorder and posttraumatic stress disorder in primary care. *International Journal of Psychiatry in Clinical Practice, 16*, 77–84. doi:10.3109/13651501.2012.667114

Barrett, B., Muller, D., Rakel, D., Rabago, D., Marchand, L., & Scheder, J. C. (2006). Placebo, meaning, and health. *Perspectives in Biology and Medicine, 49*, 178–198. doi:10.1353/pbm.2006.0019

Beecher, H. K. (1955). The powerful placebo. *Journal of the American Medical Association, 159*, 1602–1606. doi:10.1001/jama.1955.02960340022006

Benedetti, F. (2002). How the doctor's words affect the patient's brain. *Evaluation & the Health Professions, 25*, 369–386. doi:10.1177/0163278702238051

Benedetti, F. (2008). Mechanisms of placebo and placebo-related effects across diseases and treatments. *European Neuropsychopharmacology, 20*(Suppl. 1), S60–S61. doi:10.1016/S0924-977X(10)70072-9

Benedetti, F. (2014). *Placebo effects* (2nd ed.). New York, NY: Oxford University Press.

Benedetti, F., Arduino, C., Costa, S., Vighetti, S., Tarenzi, L., Rainero, I., & Asteggiano, G. (2006). Loss of expectation-related mechanisms in Alzheimer's disease makes analgesic therapies less effective. *Pain, 121*, 133–144. doi:10.1016/j.pain.2005.12.016

Benedetti, F., Frisaldi, E., Carlino, E., Giudetti, L., Pampallona, A., Zibetti, M., . . . Lopiano, L. (2016). Teaching neurons to respond to placebos. *Journal of Physiology, 594*, 5647–5660. doi:10.1113/JP271322

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

Benedetti, F., Pollo, A., Lopiano, L., Lanotte, M., Vighetti, S., & Rainero, I. (2003). Conscious expectation and unconscious conditioning in analgesic, motor, and hormonal placebo/nocebo responses. *Journal of Neuroscience, 23*, 4315–4323.

Bergmann, J. F., Chassany, O., Gandiol, J., Deblois, P., Kanis, J. A., Segrestaa, J. M., . . . Dahan, R. (1994). A randomised clinical trial of the effect of informed consent on the analgesic activity of placebo and naproxen in cancer pain. *Clinical Trials and Meta-Analysis, 29*(1), 41–47.

Bingel, U., Colloca, L., & Vase, L. (2011). Mechanisms and clinical implications of the placebo effect: Is there a potential for the elderly? A mini-review. *Gerontology, 57*, 354–363. doi:10.1159/000322090

Bishop, F. L., Aizlewood, L., & Adams, A. E. M. (2014). When and why placebo-prescribing is acceptable and unacceptable: A focus group study of patients' views. *PLoS ONE, 9*, e101822. doi:10.1371/journal.pone.0101822

Blease, C., Colloca, L., & Kaptchuk, T. J. (2016). Are open-label placebos ethical? Informed consent and ethical equivocations. *Bioethics, 30*, 407–414. doi:10.1111/bioe.12245

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. Chichester, England: Wiley.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*, 97–111. doi:10.1002/jrsm.12

Borkovec, T. D., & Nau, S. D. (1972). Credibility of analogue therapy rationales. *Journal of Behavior Therapy and Experimental Psychiatry, 3*, 257–260. doi:10.1016/0005-7916(72)90045-6

Bosworth, H. B., Hays, J. C., George, L. K., & Steffens, D. C. (2002). Psychosocial and clinical predictors of unipolar depression outcome in older adults. *International Journal of Geriatric Psychiatry, 17*, 238–246. doi:10.1002/gps.590

Bridge, J. A., Birmaher, B., Iyengar, S., Barbe, R. P., & Brent, D. A. (2009). Placebo response in randomized controlled trials of antidepressants for pediatric major depressive disorder. *American Journal of Psychiatry, 166*, 42–49. doi:10.1176/appi.ajp.2008.08020247

Bridge, J. A., Iyengar, S., Salary, C. B., Barbe, R. P., Birmaher, B., Pincus, H. A., . . . Brent, D. A. (2007). Clinical response and risk for reported suicidal ideation and suicide attempts in pediatric antidepressant treatment: A meta-analysis of randomized controlled trials. *JAMA, 297*, 1683–1696. doi:10.1001/jama.297.15.1683

Brody, H. (1994). "My story is broken; can you help me fix it?" Medical ethics and the joint construction of narrative. *Literature and Medicine, 13*, 79–92. doi:10.1353/lm.2011.0169

Broekhuizen, K., Pothof, A., de Craen, A. J., & Mooijaart, S. P. (2015). Characteristics of randomized controlled trials designed for elderly: A systematic review. *PLoS ONE, 10*, e0126709. doi:10.1371/journal.pone.0126709

Bruner, J. S. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.

Büchel, C., Geuter, S., Sprenger, C., & Eippert, F. (2014). Placebo analgesia: A predictive coding perspective. *Neuron, 81*, 1223–1239. doi:10.1016/j.neuron.2014.02.042

Caldwell, P. H. Y., Murphy, S. B., Butow, P. N., & Craig, J. C. (2004). Clinical trials in children. *Lancet, 364*, 803–811. doi:10.1016/S0140-6736(04)16942-0

Cartwright, C., Gibson, K., Read, J., Cowan, O., & Dehar, T. (2016). Long-term antidepressant use: Patient perspectives of benefits and adverse effects. *Patient Preference and Adherence, 10*, 1401–1407. doi:10.2147/Ppa.S110632

Carvalho, C., Caetano, J. M., Cunha, L., Rebouta, P., Kaptchuk, T. J., & Kirsch, I. (2016). Open-label placebo treatment in chronic low back pain: A randomized controlled trial. *Pain, 157*, 2766–2772. doi:10.1097/j.pain.0000000000000700

Cipriani, A., Zhou, X., Del Giovane, C., Hetrick, S. E., Qin, B., Whittington, C., . . . Xie, P. (2016). Comparative efficacy and tolerability of antidepressants for major depressive disorder in children and adolescents: A network meta-analysis. *Lancet, 388*, 881–890. doi:10.1016/S0140-6736(16)30385-3

Clark, A. (2015). Embodied prediction. In T. K. Metzinger & J. M. Windt (Eds.), *Open MIND*. Frankfurt am Main, Germany: MIND Group. doi:10.15502/9783958570115

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika, 37*, 256–266. doi:10.1093/biomet/37.3-4.256

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics, 10*, 101–129. doi:10.2307/3001666

Cohen, D., Deniau, E., Maturana, A., Tanguy, M. L., Bodeau, N., Labelle, R., . . . Guile, J. M. (2008). Are child and adolescent responses to placebo higher in major depression than in anxiety disorders? A systematic review of placebo-controlled trials. *PLoS ONE, 3*, e2632. doi:10.1371/journal.pone.0002632

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Colloca, L., & Benedetti, F. (2009). Placebo analgesia induced by social observational learning. *Pain, 144*, 28–34. doi:10.1016/j.pain.2009.01.033

Colloca, L., & Benedetti, F. (2016). Placebo and nocebo effects: Unfolding the complex interplay between distinct phenotypes and physiological mechanisms. *Psychology of Consciousness (Washington, D.C.), 3*, 162–174. doi:10.1037/cns0000054

Colloca, L., Enck, P., & DeGrazia, D. (2016). Relieving pain using dose-extending placebos: A scoping review. *Pain, 157*, 1590–1598. doi:10.1097/j.pain.0000000000000566

Colloca, L., Lopiano, L., Lanotte, M., & Benedetti, F. (2004). Overt versus covert treatment for pain, anxiety, and Parkinson's disease. *Lancet Neurology, 3*, 679–684. doi:10.1016/S1474-4422(04)00908-1

Conboy, L. A., Macklin, E., Kelley, J., Kokkotou, E., Lembo, A., Davis, R. & Kaptchuk, T. J. (2010). Which patients improve: Characteristics increasing sensitivity to a supportive patient–practitioner relationship. *Social Science & Medicine, 70*, 479–484. doi:10.1016/j.socscimed.2009.10.024

Conroy, S., Choonara, I., Impicciatore, P., Mohn, A., Arnell, H., Rane, A., . . . van den Anker, J. (2000). Survey of unlicensed and off label drug use in paediatric wards in European countries. European Network for Drug Investigation in Children. *BMJ, 320*, 79. doi:10.1136/bmj.320.7227.79

Coupland, C. A., Dhiman, P., Barton, G., Morriss, R., Arthur, A., Sach, T., & Hippisley-Cox, J. (2011). A study of the safety and harms of antidepressant drugs for older people: A cohort study using a large primary care database. *Health Technology Assessment, 15*(28). doi:10.3310/hta15280

Cramer, K., Wiebe, N., Moyer, V., Hartling, L., Williams, K., Swingler, G., & Klassen, T. P. (2005). Children in reviews: Methodological issues in child-relevant evidence syntheses. *BMC Pediatrics, 5*, 38. doi:10.1186/1471-2431-5-38

Crombez, G., & Wiech, K. (2011). You may (not always) experience what you expect: In search for the limits of the placebo and nocebo effect. *Pain, 152*, 1449–1450. doi:10.1016/j.pain.2011.02.028

Crum, A. J., Akinola, M., Martin, A., & Fath, S. (2017). The role of stress mindset in shaping cognitive, emotional, and physiological responses to challenging and threatening stress. *Anxiety, Stress, and Coping*. Advance online publication. doi:10.1080/10615806.2016.1275585

Crum, A. J., & Langer, E. J. (2007). Mind-set matters: Exercise and the placebo effect. *Psychological Science, 18*, 165–171. doi:10.1111/j.1467-9280.2007.01867.x

Crum, A. J., & Phillips, D. (2015). Self-fulfilling prophesies, placebo effects, and the social–psychological creation of reality. In R. Scott & S. Kosslyn (Eds.), *Emerging Trends in the Social and Behavioral Sciences*. Wiley Publishers. doi:10.1002/9781118900772.etrds0296

Cuijpers, P., Van Straten, A., Warmerdam, L., & Smits, N. (2008). Characteristics of effective psychological treatments of depression: A metaregression analysis. *Psychotherapy Research, 18*, 225–236. doi:10.1080/10503300701442027

DeVane, C. L., & Pollock, B. G. (1999). Pharmacokinetic considerations of antidepressant use in the elderly. *Journal of Clinical Psychiatry, 60*(Suppl. 20), 38–44.

Di Blasi, Z., Crawford, F., Bradley, C., & Kleijnen, J. (2005). Reactions to treatment debriefing among the participants of a placebo controlled trial. *BMC Health Services Research, 5*, 30. doi:10.1186/1472-6963-5-30

Di Blasi, Z., Harkness, E., Ernst, E., Georgiou, A., & Kleijnen, J. (2001). Influence of context effects on health outcomes: A systematic review. *Lancet, 357*, 757–762. doi:10.1016/S0140-6736(00)04169-6

Doshi, P., Dickersin, K., Healy, D., Vedula, S. S., & Jefferson, T. (2013). Restoring invisible and abandoned trials: A call for people to publish the findings. *BMJ, 346*, f2865. doi:10.1136/bmj.f2865

Eaves, E. R., Nichter, M., & Ritenbaugh, C. (2016). Ways of hoping: Navigating the paradox of hope and despair in chronic pain. *Culture, Medicine and Psychiatry, 40*, 35–58. doi:10.1007/s11013-015-9465-4

Eaves, E. R., Ritenbaugh, C., Nichter, M., Hopkins, A. L., & Sherman, K. J. (2014). Modes of hoping: Understanding hope and expectation in the context of a clinical trial of complementary and alternative medicine for chronic pain. *Explore (N.Y.), 10*, 225–232. doi:10.1016/j.explore.2014.04.004

Emslie, G. J. (2009). Understanding placebo response in pediatric depression trials. *American Journal of Psychiatry, 166*, 1–3. doi:10.1176/appi.ajp.2008.08101541

Enck, P. (2016). Placebo response in depression: Is it rising? *Lancet Psychiatry, 3*, 1005–1006. doi:10.1016/S2215-0366(16)30308-X

National Institute for Clinical Excellence. (2004). *Depression: Management of depression in primary and secondary care*. National Clinical Practice Guideline, No. 23. London, England: National Institute for Clinical Excellence. Retrieved from http://www.nice.org.uk/guidance/cg23

National Institute for Health & Clinical Excellence. (2010). *Depression: The treatment and management of depression in adults (Updated Edition)*. National Clinical Practice Guideline, No. 90. Leicester/London, England: The British Psychological Society & the Royal College of Psychiatrists. Retrieved from http://www.nice.org.uk/guidance/cg90/

Faria, V., Kossowsky, J., Petkov, M. P., Kaptchuk, T. J., Kirsch, I., Lebel, A., & Borsook, D. (2017). Parental attitudes about placebo use in children. *Journal of Pediatrics, 181*, 272–278.e210. doi:10.1016/j.jpeds.2016.10.018

Fässler, M., Gnädinger, M., Rosemann, T., & Biller-Andorno, N. (2011). Placebo interventions in practice: A questionnaire survey on the attitudes of patients and physicians. *British Journal of General Practice, 61*, 101–107. doi:10.3399/bjgp11X556209

Fava, G. A., Gatti, A., Belaise, C., Guidi, J., & Offidani, E. (2015). Withdrawal symptoms after selective serotonin reuptake inhibitor discontinuation: A systematic review. *Psychotherapy and Psychosomatics, 84*, 72–81. doi:10.1159/000370338

Finniss, D. G., Kaptchuk, T. J., Miller, F., & Benedetti, F. (2010). Biological, clinical, and ethical advances of placebo effects. *Lancet, 375*, 686–695. doi:10.1016/S0140-6736(09)61706-2

Fisher, S., & Greenberg, R. P. (1993). How sound is the double-blind design for evaluating psychotropic drugs? *Journal of Nervous and Mental Disease, 181*, 345–350. doi:10.1097/00005053-199306000-00002

Foddy, B. (2009). A duty to deceive: Placebos in clinical practice. *American Journal of Bioethics, 9*(12), 4–12. doi:10.1080/15265160903318350

Forsberg, J. T., Martinussen, M., & Flaten, M. A. (2016). The placebo analgesic effect in healthy individuals and patients: A meta-analysis. *Psychosomatic Medicine, 79*, 388–394. doi:10.1097/PSY.0000000000000432

Fountoulakis, K. N., Veroniki, A. A., Siamouli, M., & Möller, H. J. (2013). No role for initial severity on the efficacy of antidepressants: Results of a multi-meta-analysis. *Annals of General Psychiatry, 12*, 26. doi:10.1186/1744-859x-12-26

Fournier, J. C., DeRubeis, R. J., Hollon, S. D., Dimidjian, S., Amsterdam, J. D., Shelton, R. C., & Fawcett, J. (2010). Antidepressant drug effects and depression severity: A patient-level meta-analysis. *JAMA, 303*, 47–53. doi:10.1001/jama.2009.1943

Frank, J. D. (1973). *Persuasion and healing: Comparative study of psychotherapy* (Rev. ed.). Baltimore, MD: Johns Hopkins University Press.

Frank, J. D. (1974). Psychotherapy: The restoration of morale. *American Journal of Psychiatry, 131*, 271–274. doi:10.1176/ajp.131.3.271

Frank, J. D. (1986). Psychotherapy—The transformation of meanings: Discussion paper. *Journal of the Royal Society of Medicine, 79*, 341–346. doi:10.1176/ajp.131.3.271

Friston, K. (2003). Learning and inference in the brain. *Neural Networks, 16*, 1325–1352. doi:10.1016/j.neunet.2003.06.005

Furukawa, T. A., Cipriani, A., Atkinson, L. Z., Leucht, S., Ogawa, Y., Takeshima, N., . . . Salanti, G. (2016). Placebo response rates in antidepressant trials: A systematic review of published and unpublished double-blind randomised controlled studies. *Lancet Psychiatry, 3*, 1059–1066. doi:10.1016/S2215-0366(16)30307-8

Gaab, J., Jiménez, J., Voneschen, L., Oschwald, D., Meyer, A. H., Nater, U. M., & Krummenacher, P. (2017). Psychosocial stress-induced analgesia: An examination of effects on heat pain threshold and tolerance and of neuroendocrine mediation. *Neuropsychobiology, 74*, 87–95. doi:10.1159/000454986

Garland, E. J., Kutcher, S., Virani, A., & Elbe, D. (2016). Update on the use of SSRIs and SNRIs with children and adolescents in clinical practice. *Journal of the Canadian Academy of Child and Adolescent Psychiatry, 25*, 4–10.

Geers, A. L., Helfer, S. G., Weiland, P. E., & Kosbab, K. (2006). Expectations and placebo response: A laboratory investigation into the role of somatic focus. *Journal of Behavioral Medicine, 29*, 171–178. doi:10.1007/s10865-005-9040-5

Geers, A. L., & Miller, F. G. (2014). Understanding and translating the knowledge about placebo effects: The contribution of psychology. *Current Opinion in Psychiatry, 27*, 326–331. doi:10.1097/Yco.0000000000000082

American Psychiatric Association. (2010). *Practice guideline for the treatment of patients with major depressive disorder* (3rd ed.). Washington, D.C.: American Psychiatric Publishing. Retrieved from http://psychiatryonline.org/pb/assets/raw/sitewide/practice_guidelines/guidelines/mdd.pdf

Gerger, H., & Gaab, J. (2016). Post-traumatic stress disorder outcome research: Why moderators should not be neglected. In G. El-Baalbaki (Ed.), *A multidimensional approach to post-traumatic stress disorder - From theory to practice*. Rijeka, Croatia: InTech. doi:10.5772/64266

Gerger, H., Hlavica, M., Gaab, J., Munder, T., & Barth, J. (2015). Does it matter who provides psychological interventions for medically unexplained symptoms? A meta-analysis. *Psychotherapy and Psychosomatics, 84*, 217–226. doi:10.1159/000380914

Geuter, S., Koban, L., & Wager, T. D. (2017). The cognitive neuroscience of placebo effects: Concepts, predictions, and physiology. *Annual Review of Neuroscience, 40*. Advance online publication. doi:10.1146/annurev-neuro-072116-031132

Gibbons, R. D., Hur, K., Brown, C. H., Davis, J. M., & Mann, J. J. (2012). Benefits from antidepressants synthesis of 6-week patient-level outcomes from double-blind placebo-controlled randomized trials of fluoxetine and venlafaxine. *Archives of General Psychiatry, 69*, 572–579. doi:10.1001/archgenpsychiatry.2011.2044

Giron, M. S., Fastbom, J., & Winblad, B. (2005). Clinical trials of potential antidepressants: To what extent are the elderly represented: A review. *International Journal of Geriatric Psychiatry, 20*, 201–217. doi:10.1002/gps.1273

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*, 3–8. doi:10.3102/0013189X005010003

Gold, A., & Lichtenberg, P. (2014). The moral case for the clinical placebo. *Journal of Medical Ethics, 40*, 219–224. doi:10.1136/medethics-2012-101314

Granot, M., Sprecher, E., & Yarnitsky, D. (2003). Psychophysics of phasic and tonic heat pain stimuli by quantitative sensory testing in healthy subjects. *European Journal of Pain, 7*, 139–143. doi:10.1016/S1090-3801(02)00087-3

Grelotti, D. J., & Kaptchuk, T. J. (2011). Placebo by proxy. *BMJ, 343*, d4345. doi:10.1136/bmj.d4345

Guyatt, G. H., Sackett, D. L., Sinclair, J. C., Hayward, R., Cook, D. J., & Cook, R. J. (1995). Users' guides to the medical literature. IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group. *JAMA, 274*, 1800–1804. doi:10.1001/jama.1995.03530220066035

Hamilton, M. (1967). Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology, 6*, 278–296. doi:10.1111/j.2044-8260.1967.tb00530.x

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York, NY: Academic Press.

Henry, A., Kisicki, M. D., & Varley, C. (2012). Efficacy and safety of antidepressant drug treatment in children and adolescents. *Molecular Psychiatry, 17*, 1186–1193. doi:10.1038/mp.2011.150

Heun, R., Ahokas, A., Boyer, P., Giménez-Montesinos, N., Pontes-Soares, F., & Olivier, V. (2013). The efficacy of agomelatine in elderly patients with recurrent major depressive disorder: A placebo-controlled study. *Journal of Clinical Psychiatry, 74*, 587–594. doi:10.4088/JCP.12m08250

Higgins, J. P. T. (2008). Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. *International Journal of Epidemiology, 37*, 1158–1160. doi:10.1093/ije/dyn204

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ, 327*, 557–560. doi:10.1136/bmj.327.7414.557

Janiaud, P., Cornu, C., Lajoinie, A., Djemli, A., Cucherat, M., & Kassai, B. (2017). Is the perceived placebo effect comparable between adults and children? A meta-regression analysis. *Pediatric Research, 81*, 11–17. doi:10.1038/pr.2016.181

Jepma, M., & Wager, T. D. (2015). Conceptual conditioning: Mechanisms mediating conditioning effects on pain. *Psychological Science, 26*, 1728–1739. doi:10.1177/0956797615597658

Kam-Hansen, S., Jakubowski, M., Kelley, J. M., Kirsch, I., Hoaglin, D. C., Kaptchuk, T. J., & Burstein, R. (2014). Altered placebo and drug labeling changes the outcome of episodic migraine attacks. *Science Translational Medicine, 6*, 218ra5. doi:10.1126/scitranslmed.3006175

Kaptchuk, T. J. (2011). Placebo studies and ritual theory: A comparative analysis of Navajo, acupuncture and biomedical healing. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 366*, 1849–1858. doi:10.1098/rstb.2010.0385

Kaptchuk, T. J., Friedlander, E., Kelley, J. M., Sanchez, M. N., Kokkotou, E., Singer, J. P., . . . Lembo, A. J. (2010). Placebos without deception: A randomized controlled trial in irritable bowel syndrome. *PLoS ONE, 5*, e15591. doi:10.1371/journal.pone.0015591

Kaptchuk, T. J., Kelley, J. M., Conboy, L. A., Davis, R. B., Kerr, C. E., Jacobson, E. E., . . . Lembo, A. J. (2008). Components of placebo effect: Randomised controlled trial in patients with irritable bowel syndrome. *BMJ, 336*, 999–1003. doi:10.1136/bmj.39524.439618.25

Kaptchuk, T. J., & Miller, F. G. (2015). Placebo effects in medicine. *New England Journal of Medicine, 373*, 8–9. doi:10.1056/NEJMp1504023

Kaptchuk, T. J., Shaw, J., Kerr, C. E., Conboy, L. A., Kelley, J. M., Csordas, T. J., . . . Jacobson, E. E. (2009). "Maybe I made up the whole thing": Placebos and patients' experiences in a randomized controlled trial. *Culture, Medicine and Psychiatry, 33*, 382–411. doi:10.1007/S11013-009-9141-7

Keller, M. B., Ryan, N. D., Strober, M., Klein, R. G., Kutcher, S. P., Birmaher, B., . . . McCafferty, J. P. (2001). Efficacy of paroxetine in the treatment of adolescent major depression: A randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 762–772. doi:10.1097/00004583-200107000-00010

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

Kelley, J. M., Kraft-Todd, G., Schapira, L., Kossowsky, J., & Riess, H. (2014). The influence of the patient-clinician relationship on healthcare outcomes: A systematic review and meta-analysis of randomized controlled trials. *PLoS ONE, 9*, e94207. doi:10.1371/journal.pone.0094207

Khan, A., Bhat, A., Faucett, J., Kolts, R., & Brown, W. A. (2011). Antidepressant-placebo differences in 16 clinical trials over 10 years at a single site: Role of baseline severity. *Psychopharmacology, 214*, 961–965. doi:10.1007/s00213-010-2107-1

Khan, A., Brodhead, A. E., Kolts, R. L., & Brown, W. A. (2005). Severity of depressive symptoms and response to antidepressants and placebo in antidepressant trials. *Journal of Psychiatric Research, 39*, 145–150. doi:10.1016/j.jpsychires.2004.06.005

Khan, A., Kolts, R. L., Thase, M. E., Krishnan, K. R., & Brown, W. (2004). Research design features and patient characteristics associated with the outcome of antidepressant clinical trials. *American Journal of Psychiatry, 161*, 2045–2049. doi:10.1176/appi.ajp.161.11.2045

Khan, A., Leventhal, R. M., Khan, S. R., & Brown, W. A. (2002). Severity of depression and response to antidepressants and placebo: An analysis of the Food and Drug Administration database. *Journal of Clinical Psychopharmacology, 22*, 40–45. doi:10.1097/00004714-200202000-00007

Khin, N. A., Chen, Y. F., Yang, Y., Yang, P., & Laughren, T. P. (2011). Exploratory analyses of efficacy data from major depressive disorder trials submitted to the US Food and Drug Administration in support of new drug applications. *Journal of Clinical Psychiatry, 72*, 464–472. doi:10.4088/JCP.10m06191

Kirsch, I. (2000). Are drug and placebo effects in depression additive? *Biological Psychiatry, 47*, 733–735. doi:10.1016/S0006-3223(00)00832-5

Kirsch, I. (2004). Conditioning, expectancy, and the placebo effect: Comment on Stewart-Williams and Podd (2004). *Psychological Bulletin, 130*, 341–343. doi:10.1037/0033-2909.130.2.341

Kirsch, I. (2009). *The emperor's new drugs: Exploding the antidepressant myth*. London, England: Bodley Head.

Kirsch, I. (2016). The placebo effect in the treatment of depression. *Verhaltenstherapie, 26*, 55–61. doi:10.1159/000443542

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 Medicine, 5*, e45. doi:10.1371/journal.pmed.0050045

Kirsch, I., Kong, J., Sadler, P., Spaeth, R., Cook, A., Kaptchuk, T. J., & Gollub, R. (2014). Expectancy and conditioning in placebo analgesia: Separate or connected processes? *Psychology of Consciousness (Washington, D.C.), 1*, 51–59. doi:10.1037/cns0000007

Kirsch, I., & Sapirstein, G. (1998). Listening to Prozac but hearing placebo: A meta-analysis of antidepressant medication. *Prevention & Treatment, 1*(2), 2a. doi:10.1037/1522-3736.1.1.12a

Kirsch, I., & Weixel, L. J. (1988). Double-blind versus deceptive administration of a placebo. *Behavioral Neuroscience, 102*, 319–323. doi:10.1037/0735-7044.102.2.319

Klassen, T. P., Hartling, L., Craig, J. C., & Offringa, M. (2008). Children are not just small adults: The urgent need for high-quality trial evidence in children. *PLoS Medicine, 5*, e172. doi:10.1371/journal.pmed.0050172

Kok, R. M., Nolen, W. A., & Heeren, T. J. (2012). Efficacy of treatment in older depressed patients: A systematic review and meta-analysis of double-blind randomized controlled trials with antidepressants. *Journal of Affective Disorders, 141*, 103–115. doi:10.1016/j.jad.2012.02.036

Konrat, C., Boutron, I., Trinquart, L., Auleley, G. R., Ricordeau, P., & Ravaud, P. (2012). Underrepresentation of elderly people in randomised controlled trials. The example of trials of 4 widely prescribed drugs. *PLoS ONE, 7*, e33559. doi:10.1371/journal.pone.0033559

Koshi, E. B., & Short, C. A. (2007). Placebo theory and its implications for research and clinical practice: A review of the recent literature. *Pain Practice, 7*, 4–20. doi:10.1111/j.1533-2500.2007.00104.x

Krummenacher, P., Kossowsky, J., Schwarz, C., Brugger, P., Kelley, J. M., Meyer, A., & Gaab, J. (2014). Expectancy-induced placebo analgesia in children and the role of magical thinking. *Journal of Pain, 15*, 1282–1293. doi:10.1016/j.jpain.2014.09.005

Lacasse, J. R., & Leo, J. (2005). Serotonin and depression: A disconnect between the advertisements and the scientific literature. *PLoS Medicine, 2*, e392. doi:10.1371/journal.pmed.0020392

Lau, J. Y., Britton, J. C., Nelson, E. E., Angold, A., Ernst, M., Goldwin, M., . . . Pine, D. S. (2011). Distinct neural signatures of threat learning in adolescents and adults. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 4500–4505. doi:10.1073/pnas.1005494108

Le Noury, J., Nardo, J. M., Healy, D., Jureidini, J., Raven, M., Tufanaru, C., & Abi-Jaoude, E. (2015). Restoring Study 329: Efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. *BMJ, 351*, h4320. doi:10.1136/bmj.h4320

Leuchter, A. F., Hunter, A. M., Tartter, M., & Cook, I. A. (2014). Role of pill-taking, expectation and therapeutic alliance in the placebo response in clinical trials for major depression. *British Journal of Psychiatry, 205*, 443–449. doi:10.1192/bjp.bp.113.140343

Locher, C., Frey Nascimento, A., Kossowsky, J., Meyer, A., & Gaab, J. (2017). *Is the rationale more important than deception? A randomized controlled trial of open-label placebo analgesia*. Manuscript submitted for publication.

Locher, C., Koechlin, H., Zion, S. R., Werner, C., Pine, S. D., Kirsch, I., . . . Kossowsky, J. (2017). *Efficacy and safety of SSRIs, SNRIs, and placebo in common psychiatric disorders: A comprehensive meta-analysis in children and adolescents*. Manuscript submitted for publication.

Locher, C., Kossowsky, J., Gaab, J., Kirsch, I., Bain, P., & Krummenacher, P. (2015). Moderation of antidepressant and placebo outcomes by baseline severity in late-life depression: A systematic review and meta-analysis. *Journal of Affective Disorders, 181*, 50–60. doi:10.1016/j.jad.2015.03.062

Lucassen, P., & Olesen, F. (2016). Context as a drug: Some consequences of placebo research for primary care. *Scandinavian Journal of Primary Health Care, 34*, 428–433. doi:10.1080/02813432.2016.1249065

Lund, K., Vase, L., Petersen, G. L., Jensen, T. S., & Finnerup, N. B. (2014). Randomised controlled trials may underestimate drug effects: Balanced placebo trial design. *PLoS ONE, 9*, e84104. doi:10.1371/journal.pone.0084104

Lundh, L. G. (1987). Placebo, belief, and health. A cognitive-emotional model. *Scandinavian Journal of Psychology, 28*, 128–143. doi:10.1111/j.1467-9450.1987.tb00747.x

Maeoka, H., Hiyamizu, M., Matsuo, A., & Morioka, S. (2015). The influence of repeated pain stimulation on the emotional aspect of pain: A preliminary study in healthy volunteers. *Journal of Pain Research, 8*, 431–436. doi:10.2147/JPR.S86732

Mann, J. J., Emslie, G., Baldessarini, R. J., Beardslee, W., Fawcett, J. A., Goodwin, F. K., . . . Wagner, K. D. (2006). ACNP Task Force report on SSRIs and suicidal behavior in youth. *Neuropsychopharmacology, 31*, 473–492. doi:10.1038/sj.npp.1300958

Martinez-Castaldi, C., Silverstein, M., & Bauchner, H. (2008). Child versus adult research: The gap in high-quality study design. *Pediatrics, 122*, 52–57. doi:10.1542/peds.2007-2849

McKay, K. M., Imel, Z. E., & Wampold, B. E. (2006). Psychiatrist effects in the psychopharmacological treatment of depression. *Journal of Affective Disorders, 92*, 287–290. doi:10.1016/j.jad.2006.01.020

Melander, H., Ahlqvist-Rastad, J., Meijer, G., & Beermann, B. (2003). Evidence b(i)ased medicine--Selective reporting from studies sponsored by pharmaceutical industry: Review of studies in new drug applications. *BMJ, 326*, 1171–1173. doi:10.1136/bmj.326.7400.1171

Miller, F. G., Wendler, D., & Swartzman, L. C. (2005). Deception in research on the placebo effect. *PLoS Medicine, 2*, e262. doi:10.1371/journal.pmed.0020262

Mitte, K. (2005). A meta-analysis of the efficacy of psycho- and pharmacotherapy in panic disorder with and without agoraphobia. *Journal of Affective Disorders, 88*, 27–45. doi:10.1016/j.jad.2005.05.003

Moerman, D. E., & Jonas, W. B. (2002). Deconstructing the placebo effect and finding the meaning response. *Annals of Internal Medicine, 136*, 471–476. doi:10.7326/0003-4819-136-6-200203190-00011

Moncrieff, J., & Kirsch, I. (2005). Efficacy of antidepressants in adults. *BMJ, 331*, 155–157. doi:10.1136/bmj.331.7509.155

Moncrieff, J., & Kirsch, I. (2015). Empirically derived criteria cast doubt on the clinical significance of antidepressant-placebo differences. *Contemporary Clinical Trials, 43*, 60–62. doi:10.1016/j.cct.2015.05.005

Montgomery, S. A., & Asberg, M. (1979). A new depression scale designed to be sensitive to change. *British Journal of Psychiatry, 134*, 382–389. doi:10.1192/bjp.134.4.382

Nelson, J. C., Delucchi, K., & Schneider, L. S. (2008). Efficacy of second generation antidepressants in late-life depression: A meta-analysis of the evidence. *American Journal of Geriatric Psychiatry, 16*, 558–567. doi:10.1097/JGP.0b013e3181693288

Nelson, J. C., Delucchi, K. L., & Schneider, L. S. (2013). Moderators of outcome in late-life depression: A patient-level meta-analysis. *American Journal of Psychiatry, 170*, 651–659. doi:10.1176/appi.ajp.2012.12070927

NHS Digital. (2016, July 5). *Antidepressants show greatest increase in number of prescription items dispensed*. Retrieved from http://content.digital.nhs.uk/article/7159/Antidepressants-show-greatest-increase-in-number-of-prescription-items-dispensed

Otto, M. W., Tuby, K. S., Gould, R. A., McLean, R. Y., & Pollack, M. H. (2001). An effect-size analysis of the relative efficacy and tolerability of serotonin selective reuptake inhibitors for panic disorder. *American Journal of Psychiatry, 158*, 1989–1992. doi:10.1176/appi.ajp.158.12.1989

Oxman, T. E., & Hull, J. G. (2001). Social support and treatment response in older depressed primary care patients. *Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 56*, P35–P45. doi:10.1093/geronb/56.1.P35

Parellada, M., Moreno, C., Moreno, M., Espliego, A., de Portugal, E., & Arango, C. (2012). Placebo effect in child and adolescent psychiatric trials. *European Neuropsychopharmacology, 22*, 787–799. doi:10.1016/j.euroneuro.2011.09.007

Pinquart, M., Duberstein, P. R., & Lyness, J. M. (2006). Treatments for later-life depressive conditions: A meta-analytic comparison of pharmacotherapy and psychotherapy. *American Journal of Psychiatry, 163*, 1493–1501. doi:10.1176/appi.ajp.163.9.1493

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

Price, D. D., Finniss, D. G., & Benedetti, F. (2008). A comprehensive review of the placebo effect: Recent advances and current thought. *Annual Review of Psychology, 59*, 565–590. doi:10.1146/annurev.psych.59.113006.095941

Price, D. D., Milling, L. S., Kirsch, I., Duff, A., Montgomery, G. H., & Nicholls, S. S. (1999). An analysis of factors that contribute to the magnitude of placebo analgesia in an experimental paradigm. *Pain, 83*, 147–156. doi:10.1016/S0304-3959(99)00081-0

Quinn, V. F., & Colagiuri, B. (2016). Sources of placebo-induced relief from nausea: The role of instruction and conditioning. *Psychosomatic Medicine, 78*, 365–372. doi:10.1097/PSY.0000000000000265

Rabinowitz, J., Werbeloff, N., Mandel, F. S., Menard, F., Marangell, L., & Kapur, S. (2016). Initial depression severity and response to antidepressants v. placebo: Patient-level data analysis from 34 randomised controlled trials. *British Journal of Psychiatry, 209*, 427–428. doi:10.1192/bjp.bp.115.173906

Regier, D. A., Narrow, W. E., Clarke, D. E., Kraemer, H. C., Kuramoto, S. J., Kuhl, E. A., & Kupfer, D. J. (2013). DSM-5 field trials in the United States and Canada, Part II: Test-retest reliability of selected categorical diagnoses. *American Journal of Psychiatry, 170*, 59–70. doi:10.1176/appi.ajp.2012.12070999

Rescorla, R. A. (1988). Pavlovian conditioning. It's not what you think it is. *American Psychologist, 43*, 151–160. doi:10.1037/0003-066X.43.3.151

Rief, W., Bingel, U., Schedlowski, M., & Enck, P. (2011). Mechanisms involved in placebo and nocebo responses and implications for drug trials. *Clinical Pharmacology and Therapeutics, 90*, 722–726. doi:10.1038/clpt.2011.204

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

Rojas-Mirquez, J. C., Rodriguez-Zuñiga, M. J., Bonilla-Escobar, F. J., Garcia-Perdomo, H. A., Petkov, M., Becerra, L., . . . Linnman, C. (2014). Nocebo effect in randomized clinical trials of antidepressants in children and adolescents: Systematic review and meta-analysis. *Frontiers in Behavioral Neuroscience, 8*, 375. doi:10.3389/fnbeh.2014.00375

Rutherford, B. R., & Roose, S. P. (2013). A model of placebo response in antidepressant clinical trials. *American Journal of Psychiatry, 170*, 723–733. doi:10.1176/appi.ajp.2012.12040474

Rutherford, B. R., Sneed, J., Devanand, D., Eisenstadt, R., & Roose, S. (2010). Antidepressant study design affects patient expectancy: A pilot study. *Psychological Medicine, 40*, 781–788. doi:10.1017/S0033291709991085

Rutherford, B. R., Sneed, J. R., & Roose, S. P. (2009). Does study design influence outcome? The effects of placebo control and treatment duration in antidepressant trials. *Psychotherapy and Psychosomatics, 78*, 172–181. doi:10.1159/000209348

Rutherford, B. R., Sneed, J. R., Tandler, J. M., Rindskopf, D., Peterson, B. S., & Roose, S. P. (2011). Deconstructing pediatric depression trials: An analysis of the effects of expectancy and therapeutic contact. *Journal of the American Academy of Child and Adolescent Psychiatry, 50*, 782–795. doi:10.1016/j.jaac.2011.04.004

Rutherford, B. R., Tandler, J., Brown, P. J., Sneed, J. R., & Roose, S. P. (2014). Clinic visits in late-life depression trials: Effects on signal detection and therapeutic outcome. *American Journal of Geriatric Psychiatry, 22*, 1452–1461. doi:10.1016/j.jagp.2013.09.003

Rutherford, B. R., Wall, M. M., Brown, P. J., Choo, T. H., Wager, T. D., Peterson, B. S., . . . Roose, S. P. (2017). Patient expectancy as a mediator of placebo effects in antidepressant clinical trials. *American Journal of Psychiatry, 174*, 135–142. doi:10.1176/appi.ajp.2016.16020225

Sandler, A. D., & Bodfish, J. W. (2008). Open-label use of placebos in the treatment of ADHD: A pilot study. *Child: Care, Health and Development, 34*, 104–110. doi:10.1111/j.1365-2214.2007.00797.x

Sandler, A. D., Glesne, C. E., & Bodfish, J. W. (2010). Conditioned placebo dose reduction: A new treatment in attention-deficit hyperactivity disorder? *Journal of Developmental and Behavioral Pediatrics, 31*, 369–375. doi:10.1097/DBP.0b013e3181e121ed

Schaefer, M., Harke, R., & Denke, C. (2016). Open-label placebos improve symptoms in allergic rhinitis: A randomized controlled trial. *Psychotherapy and Psychosomatics, 85*, 373–374. doi:10.1159/000447242

Schedlowski, M., & Pacheco-López, G. (2010). The learned immune response: Pavlov and beyond. *Brain, Behavior, and Immunity, 24*, 176–185. doi:10.1016/j.bbi.2009.08.007

Schmidt, L., Braun, E. K., Wager, T. D., & Shohamy, D. (2014). Mind matters: Placebo enhances reward learning in Parkinson's disease. *Nature Neuroscience, 17*, 1793–1797. doi:10.1038/nn.3842

Shankman, S. A., Klein, D. N., Tenke, C. E., & Bruder, G. E. (2007). Reward sensitivity in depression: A biobehavioral study. *Journal of Abnormal Psychology, 116*, 95–104. doi:10.1037/0021-843X.116.1.95

Sharma, T., Guski, L. S., Freund, N., & Gøtzsche, P. C. (2016). Suicidality and aggression during antidepressant treatment: Systematic review and meta-analyses based on clinical study reports. *BMJ, 352*, i65. doi:10.1136/bmj.i65

Sharpe, D. (1997). Of apples and oranges, file drawers and garbage: Why validity issues in meta-analysis will not go away. *Clinical Psychology Review, 17*, 881–901. doi:10.1016/S0272-7358(97)00056-1

Simmons, K., Ortiz, R., Kossowsky, J., Krummenacher, P., Grillon, C., Pine, D., & Colloca, L. (2014). Pain and placebo in pediatrics: A comprehensive review of laboratory and clinical findings. *Pain, 155*, 2229–2235. doi:10.1016/j.pain.2014.08.036

Sneed, J. R., Culang, M. E., Keilp, J. G., Rutherford, B. R., Devanand, D. P., & Roose, S. P. (2010). Antidepressant medication and executive dysfunction: A deleterious interaction in late-life depression. *American Journal of Geriatric Psychiatry, 18*, 128–135. doi:10.1097/JGP.0b013e3181c796d2

Sneed, J. R., Roose, S. P., Keilp, J. G., Krishnan, K. R., Alexopoulos, G. S., & Sackeim, H. A. (2007). Response inhibition predicts poor antidepressant treatment response in very old depressed patients. *American Journal of Geriatric Psychiatry, 15*, 553–563. doi:10.1097/JGP.0b013e3180302513

Sneed, J. R., Rutherford, B. R., Rindskopf, D., Lane, D. T., Sackeim, H. A., & Roose, S. P. (2008). Design makes a difference: A meta-analysis of antidepressant response rates in placebo-controlled versus comparator trials in late-life depression. *American Journal of Geriatric Psychiatry, 16*, 65–73. doi:10.1097/JGP.0b013e3181256b1d

Spoerri, A., Zwahlen, M., Bopp, M., Gutzwiller, F., & Egger, M. (2010). Religion and assisted and non-assisted suicide in Switzerland: National Cohort Study. *International Journal of Epidemiology, 39*, 1486–1494. doi:10.1093/ije/dyq141

Stewart-Williams, S., & Podd, J. (2004). The placebo effect: Dissolving the expectancy versus conditioning debate. *Psychological Bulletin, 130*, 324–340. doi:10.1037/0033-2909.130.2.324

Stone, D. A., Kerr, C. E., Jacobson, E., Conboy, L. A., & Kaptchuk, T. J. (2005). Patient expectations in placebo-controlled randomized clinical trials. *Journal of Evaluation in Clinical Practice, 11*, 77–84. doi:10.1111/j.1365-2753.2004.00512.x

Stone, M. B. (2014). The FDA warning on antidepressants and suicidality—Why the controversy? *New England Journal of Medicine, 371*, 1668–1671. doi:10.1056/NEJMp1411138

Sugarman, M. A., Loree, A. M., Baltes, B. B., Grekin, E. R., & Kirsch, I. (2014). The efficacy of paroxetine and placebo in treating anxiety and depression: A meta-analysis of change on the Hamilton Rating Scales. *PLoS ONE, 9*, e106337. doi:10.1371/journal.pone.0106337

't Jong, G. W., Vulto, A. G., de Hoog, M., Schimmel, K. J., Tibboel, D., & van den Anker, J. N. (2000). Unapproved and off-label use of drugs in a children's hospital. *New England Journal of Medicine, 343*, 1125. doi:10.1056/NEJM200010123431515

Taylor, W. D. (2014). Clinical practice. Depression in the elderly. *New England Journal of Medicine, 371*, 1228–1236. doi:10.1056/NEJMcp1402180

Taylor, W. D., & Doraiswamy, P. M. (2004). A systematic review of antidepressant placebo-controlled trials for geriatric depression: Limitations of current data and directions for the future. *Neuropsychopharmacology, 29*, 2285–2299. doi:10.1038/sj.npp.1300550

Tedeschini, E., Levkovitz, Y., Iovieno, N., Ameral, V. E., Nelson, J. C., & Papakostas, G. I. (2011). Efficacy of antidepressants for late-life depression: A meta-analysis and meta-regression of placebo-controlled randomized trials. *Journal of Clinical Psychiatry, 72*, 1660–1668. doi:10.4088/JCP.10r06531

Tham, A., Jonsson, U., Andersson, G., Söderlund, A., Allard, P., & Bertilsson, G. (2016). Efficacy and tolerability of antidepressants in people aged 65 years or older with major depressive disorder - A systematic review and a meta-analysis. *Journal of Affective Disorders, 205*, 1–12. doi:10.1016/j.jad.2016.06.013

Trachsel, M., & Gaab, J. (2016). Disclosure of incidental constituents of psychotherapy as a moral obligation for psychiatrists and psychotherapists. *Journal of Medical Ethics, 42*, 493–495. doi:10.1136/medethics-2015-102986

Tsapakis, E. M., Soldani, F., Tondo, L., & Baldessarini, R. J. (2008). Efficacy of antidepressants in juvenile depression: Meta-analysis. *British Journal of Psychiatry, 193*, 10–17. doi:10.1192/bjp.bp.106.031088

Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A., & Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *New England Journal of Medicine, 358*, 252–260. doi:10.1056/NEJMsa065779

Turner, E. H., & Rosenthal, R. (2008). Efficacy of antidepressants. *BMJ, 336*, 516–517. doi:10.1136/bmj.39510.531597.80

Tuttle, A. H., Tohyama, S., Ramsay, T., Kimmelman, J., Schweinhardt, P., Bennett, G. J., & Mogil, J. S. (2015). Increasing placebo responses over time in U.S. clinical trials of neuropathic pain. *Pain, 156*, 2616–2626. doi:10.1097/j.pain.0000000000000333

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

Voudouris, N. J., Peck, C. L., & Coleman, G. (1985). Conditioned placebo responses. *Journal of Personality and Social Psychology, 48*, 47–53. doi:10.1037/0022-3514.48.1.47

Wager, T. D., & Roy, M. (2010). Separate mechanisms for placebo and opiate analgesia? *Pain, 150*, 8–9. doi:10.1016/j.pain.2010.03.010

Walker, L. S., Williams, S. E., Smith, C. A., Garber, J., Van Slyke, D. A., & Lipani, T. A. (2006). Parent attention versus distraction: Impact on symptom complaints by children with and without chronic functional abdominal pain. *Pain, 122*, 43–52. doi:10.1016/j.pain.2005.12.020

Walsh, B. T., Seidman, S. N., Sysko, R., & Gould, M. (2002). Placebo response in studies of major depression: Variable, substantial, and growing. *JAMA, 287*, 1840–1847. doi:10.1001/jama.287.14.1840

Walsh, B. T., & Sysko, R. (2005). Placebo control groups in trials of major depressive disorder among older patients. *Journal of Clinical Psychopharmacology, 25*(Suppl. 1), S29–S33.

Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work* (2nd ed.). New York, NY: Routledge.

Wampold, B. E., Minami, T., Tierney, S. C., Baskin, T. W., & Bhati, K. S. (2005). The placebo is powerful: Estimating placebo effects in medicine and psychotherapy from randomized clinical trials. *Journal of Clinical Psychology, 61*, 835–854. doi:10.1002/jclp.20129

Wechsler, M. E., Kelley, J. M., Boyd, I. O. E., Dutile, S., Marigowda, G., Kirsch, I., . . . Kaptchuk, T. J. (2011). Active albuterol or placebo, sham acupuncture, or no intervention in asthma. *New England Journal of Medicine, 365*, 119–126. doi:10.1056/NEJMoa1103319

Wendler, D., & Miller, F. G. (2004). Deception in the pursuit of science. *Archives of Internal Medicine, 164*, 597–600. doi:10.1001/archinte.164.6.597

Whalley, B., & Hyland, M. E. (2013). Placebo by proxy: The effect of parents' beliefs on therapy for children's temper tantrums. *Journal of Behavioral Medicine, 36*, 341–346. doi:10.1007/s10865-012-9429-x

Wiech, K. (2016). Deconstructing the sensation of pain: The influence of cognitive processes on pain perception. *Science, 354*, 584–587. doi:10.1126/science.aaf8934

World Medical Association. (2013). *WMA declaration of Helsinki: Ethical principles for medical research involving human subjects*. Retrieved from http://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/

Zimmerman, M., Multach, M. D., Clark, H. L., Walsh, E., Rosenstein, L. K., & Gazarian, D. (2016). Inclusion/exclusion criteria in late life depression antidepressant efficacy trials. *International Journal of Geriatric Psychiatry*. Advance online publication. doi:10.1002/gps.4560

Appendix A

Study I:

Locher, C., Koechlin, H., Zion, S. R., Werner, C., Pine, S. D., Kirsch, I., Kessler, R.C., & Kossowsky, J. (2017). *Efficacy and safety of SSRIs, SNRIs, and placebo in common psychiatric disorders: A comprehensive meta-analysis in children and adolescents*. Manuscript submitted for publication.

Efficacy and Safety of SSRIs, SNRIs, and Placebo in Common Psychiatric Disorders: A Comprehensive Meta-Analysis in Children and Adolescents

Cosima Locher, MSc., Helen Koechlin, MSc.*, Sean R Zion, MA, Christoph Werner, BSc, Daniel S. Pine, MD, Irving Kirsch, PhD, Ronald C. Kessler, PhD, Joe Kossowsky, PhD*

* These authors contributed equally to this study

Running Title: Meta-Analysis of Antidepressants in Children

Author Affiliations: Ms. Locher is affiliated with the Department of Clinical Psychology & Psychotherapy, University of Basel, Switzerland. Mrs. Koechlin is affiliated with the Department of Anesthesiology, Perioperative and Pain Medicine, Boston Children's Hospital, Harvard Medical School (HMS), Boston, MA, USA, and the Department of Clinical Psychology & Psychotherapy, University of Basel, Switzerland. Mr. Zion is affiliated with the Department of Psychology, Stanford University, Stanford, CA. Mr. Werner is affiliated with the Department of Clinical Psychology & Psychotherapy, University of Basel, Switzerland. Dr. Pine is affiliated with the National Institute of Mental Health (NIMH) Intramural Research Program (IRP). Dr. Kirsch is a member of the Program in Placebo Studies, BIDMC, HMS, Boston, MA, USA. Dr. Kessler is affiliated with the Department of Health Care Policy, Harvard Medical School, Boston, MA, USA. Dr. Kossowsky is affiliated with the Department of Anesthesiology, Perioperative and Pain Medicine, Boston Children's Hospital, Harvard Medical School (HMS), Boston, MA, USA and the Department of Clinical Psychology & Psychotherapy, University of Basel, Switzerland,

Word count: 3733 words

Corresponding Author:
Dr. Joe Kossowsky
Department of Anesthesiology, Perioperative, and Pain Medicine
Boston Children's Hospital
333 Longwood Avenue, Boston, MA, 02115
Email: Joe.Kossowsky@childrens.harvard.edu

Keywords: SSRI, SNRI, Placebo, Children, Meta-analysis

Abstract

Importance: Depressive disorders (DD), anxiety disorders (AD), obsessive-compulsive disorder (OCD), and posttraumatic stress disorder (PTSD) are among the most common mental disorders in children and adolescents.

Objective: To examine the relative efficacy and safety of SSRIs, SNRIs and placebo for the treatment of DD, AD, OCD, and PTSD in children and adolescents.

Data Sources: PubMed, Embase, PsycINFO, Web of Science, and Cochrane through August 2016.

Study Selection: Published and unpublished randomized, double-blind, placebo controlled studies of SSRIs or SNRIs in youth diagnosed with DD, AD, OCD, or PTSD were included. Trials that used other antidepressants (e.g. tricyclic antidepressants, MAOIs) were excluded.

Data Extraction and Synthesis: Effect sizes (ES) were summarized as standardized mean differences (Hedges'g) in a random-effects model.

Main Outcome(s) and Measure(s): Primary outcomes as defined by authors on pre- and post-intervention data or mean change data and side effect data were extracted independently by multiple observers following PRISMA guidelines.

Results: We deemed 36 studies eligible, including 6778 participants; 17 studies for DD, 10 for AD, 8 for OCD and one study for PTSD. Overall, SSRIs and SNRIs were significantly more effective compared to placebo, yielding a small effect size (g=0.323, $p<0.001$). AD (g=0.557, $p<0.001$) showed significantly larger between group ES than DD (g=0.201, $p<0.001$). This difference was driven primarily by the placebo response: patients with DD exhibited significantly larger placebo responses (g=1·569, $p<0·001$) compared to those with AD (g=1.023, $p<0.001$). Compared to placebo, patients taking either SSRIs or SNRIs reported significantly more serious adverse events (SSRI: 6.80% and SNRI: 7.59% vs. placebo: 3.32%; $ps\leq0.05$), but showed no significant difference in treatment emergent adverse events ($p=0.73$). No moderator was significant in the multivariate meta-regression.

Conclusion and Relevance: SSRIs and SNRIs are more effective than placebo, however, the effect is small and disorder-specific, yielding a larger effect for AD than for other conditions. Response to placebo is large, especially in DD. Adverse event profiles appear to be disorder-dependent, and serious adverse events are significantly more common in SSRIs and SNRIs compared to placebo.

Introduction

Depressive disorders (DD), anxiety disorders (AD), obsessive-compulsive disorder (OCD), and posttraumatic stress disorder (PTSD) are among the most common mental disorders in children and adolescents [1]. All these disorders are major public health concerns and predict long-term risk for various adverse outcomes [2]. Thus, early diagnosis and proper treatment is of critical importance. Selective serotonin reuptake inhibitors (SSRIs) and serotonin and norepinephrine reuptake inhibitors (SNRIs) are first line pharmaceutical treatments for these disorders [3]. This meta-analysis aims to contribute to the current literature by comparing the efficacy of these drugs across the disorders for which they are primarily prescribed in a pediatric population, focusing on differences in the response to placebo as well as differences in side effects.

Since the release of fluoxetine in the mid 1980's, the number of SSRIs and SNRIs has grown dramatically. However, their use in children and adolescents is still debated. This relates to the need for more research into their safety and efficacy as well as questions about comparative efficacy for the newer SNRIs versus SSRIs [5]. Current data suggests that fluoxetine has the most favorable risk-benefit ratio in pediatric DD [6], but recent meta-analyses generate many questions about the overall benefits versus costs of using SSRIs to treat major depression in children and adolescents [7]. The small amount of research on SNRIs for pediatric DD has had mixed results [8]. One meta-analysis on pediatric depression found that while SSRIs differed significantly from placebo, SNRIs and tricyclic antidepressants did not [9].

Although most prior reviews and meta-analyses of the effects of SSRIs and SNRIs focused on pediatric DD, considerable data also exist on pediatric AD and OCD. The latter studies suggest that most SSRIs (fluoxetine, fluvoxamine, sertraline and paroxetine) have a favorable risk-benefit ratio, while there are insufficient data for the remaining SSRIs (citalopram, escitalopram) [8]. There have been relatively few studies on SNRIs for pediatric AD, despite the fact that the only FDA-approved agent for pediatric AD, duloxetine, is an SNRI. No double-blind RCTs of SNRIs for pediatric OCD have been conducted as of 2016, and only very limited data have been reported for either SSRIs or SNRIs in pediatric PTSD [10].

In addition to disorder-specific and drug-specific analyses, another segment of the literature on SSRIs/SNRIs among pediatric cases has focused on safety and tolerability. Research indicates a high risk of developing treatment emergent adverse effects (TEAEs), most prominently headache and nausea, during treatment with an antidepressant, including SSRIs and

SNRIs, in pediatric DD [9]. More severe side effects such as an increased risk of suicidal thoughts and behavior in adults and youth taking antidepressants have also been reported [11], leading to the implementation of a "black box" warning on the labels of all antidepressants for pediatric use by the FDA in 2004. However, this remains controversial due to conflicting results of re-analyses of the data [12]. Additionally, no recent meta-analyses have focused on how pediatric side effect profiles of SSRIs, SNRIs, and placebo might differ across disorders.

Finally, there is a growing body of literature that aims to consider the role of placebo effects in studies of SSRIs and SNRIs, based on large placebo response rates in studies of antidepressants in both adult and pediatric samples [13]. Factors such as the contact with research staff may lead to high placebo response rates in pediatric depression [14] and may in fact explain much of the variability in pediatric antidepressant trials [15]. For adult patients with DD, a genuine placebo effect is discussed, as the combination of placebo and supportive care has been shown to be more effective than supportive care alone [16]. Conversely, patients in the placebo group also demonstrate treatment emergent adverse events [9]. However, how response to placebo differs across disorders or other study design features in pediatrics is relatively understudied.

To our knowledge, only one other review or meta-analysis has examined the use of SSRIs and SNRIs across DD, AD, OCD, and PTSD [17]. However, that earlier study is now nearly a decade old and predates eleven primary studies (n=2068) that fulfill our inclusion criteria. The earlier review also did not include any studies on duloxetine, which is currently the only medication approved for pediatric AD. We therefore conducted an updated and extended review to assess the efficacy and safety of these drugs for DD, AD, OCD, and PTSD along with between-disorder variation in these drug and placebo responses. Psychological therapies will not be part of this meta-analysis, however, a recent review has compared psychological therapies alone and in combination with antidepressant medication for depression in children and adolescents [4].

Methods

Search Strategy and Study Selection:

For this meta-analysis, we searched PubMed, Embase, PsycInfo, Cochrane, Web of Science, Clinicaltrials.gov and fda.gov and checked references of originally identified papers and reviews. For additional information on search terms, see Supplemental Material (S1). In total, this

returned 4899 articles, which were reviewed by three authors (CL, HK, and SZ) (sFigure 1). We included randomized, double blind, placebo-controlled trials of SSRIs and SNRIs in children and adolescents < 18 years of age published through August 2016, including studies that examined drug vs. placebo in the context of a psychosocial intervention (i.e., drug+CBT vs. placebo+CBT), in which case the combination group was extracted only if no comparison of drug and placebo alone was given. Subjects were required to have a diagnosis of a depressive disorder, an anxiety disorder, obsessive-compulsive disorder, or posttraumatic stress disorder based on DSM-III, DSM-III-R, or DSM-IV-TR criteria. Comorbidity was allowed and any information about comorbid disorders was extracted.

Case reports, comments, letters, gray literature, and reviews were excluded. Further, non-second-generation antidepressants (e.g., monoamine oxidase inhibitors, tricyclic antidepressants) were excluded since they are not recommended as first-line medication for children and adolescents [18].

Methodological Quality Assessment

Two authors (CL and SZ) independently rated the quality of included papers according to the Cochrane Risk of Bias Assessment Tool [19], with final quality ratings based on consensus. Risk of Bias was assessed in individual studies (Table 1) and across studies (sFigure 2).

Outcome Measures and Data Extraction:

The primary outcome as defined by authors was chosen as the sole outcome measure for each individual study. Pre- and post-intervention data or mean change data had to be available. Outcomes had to be reported on a well-validated disorder specific scale (e.g., CDRS-R, MASC, CY-BOCS) or on a general severity scale (i.e., CGI-S). We included only continuous outcome data, since dichotomizing continuous scores into categorical outcome data leads to a loss of information, reduces power, and creates an artificial boundary [20,21]. We did not extract data from improvement scales such as the CGI-I. If unavailable, data were requested directly from the authors. Some studies did not include SDs or SEs and they had to be imputed [22,23].

Data were extracted independently by three authors (CL, HK, and SZ). Any discrepancies were resolved by consensus. To fit with our a priori hypotheses, data extraction concentrated on demographic information, dropout rates, adverse events, safety information, and baseline and

endpoint assessment points. Data from open label extensions or follow up after the pre-designated endpoint was not extracted.

Data Analysis:

Comprehensive Meta-Analysis V3 (Biostat, Englewood, NJ) and R 3.2.1 (R Foundation; Vienna, Austria) were used for calculations and analyses. Effect sizes were calculated as Hedges' g [24]. We chose to use random-effects models rather than fixed-effects models as the studies we included were heterogeneous and the number of studies for the sub-analyses was relatively small [25]. Three effect sizes were calculated for each included study. First, differences in mean change scores between groups were evaluated. We also calculated within-group pre-post effect sizes for both drug and placebo. Heterogeneity was assessed by calculating the Q statistic [26], the τ^2, and the I^2, a transformation of Q that indicates the proportion of observed variance that can be attributed to heterogeneity rather than sampling error [27]. The τ^2 offers an estimate of the variance among true effect sizes [28]. Effect size differences between subgroups were analyzed using a mixed-effects model [29].

Moderator analyses were conducted for six continuous moderators (treatment duration, publication year, illness duration, age of onset, number of sites, and baseline severity) and four categorical moderators (placebo lead-in, comorbidity, region, and primary funding source) for both the combined disorders group and individual disorders groups. We examined whether specific characteristics of the studies were related to the effect sizes (i.e., drug-placebo differences) in univariate analyses and multivariate regression analyses. Details of the applied statistical approaches are provided in the Supplemental Material (S5).

Publication bias was assessed visually by means of funnel plots [30] and formally by means of the fail-safe N [31] and the Begg adjusted-rank correlation test [32]. We estimated the sensitivity of publication bias using the trim-and-fill method [33].

This study is registered with PROSPERO, number CRD42016048552.

Role of the funding source:

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing the report.

Results

Our search identified thirty-six randomized, double blind trials [10,22,23,34-66][1] including 6778 participants that compared an SSRI or an SNRI against placebo in patients <18 years with a diagnosis of AD, DD[2], OCD, or PTSD (sFigure 1). Characteristics of the thirty-six included trials are shown in Table 1. Details regarding heterogeneity and publication bias can be found in the Supplemental Material (S3).

We conducted three main pre-post analyses: a between group analysis stratified by disorder and by drug, a within antidepressant group analysis stratified by disorder and by drug, and a within placebo group analysis stratified by disorder. The combined analysis between group across all disorders yielded a small effect size (g=0.32, CI=0.25-0.40, p<.001). In the between group analysis stratified by disorder, AD (g=0.56, CI=0.40-0.72, p<.001) and OCD (g=0.39; CI=0.25-0.54, p<.001) did not differ from each other (p=.14) but both were significantly higher (AD vs. DD: p<.001 and OCD vs. DD: p=.02) than the DD group (g=0.20, CI=0.13-0.27, p<.001) (Figure 1). Between drug analysis yielded the smallest effect sizes for citalopram (g=0.18, CI=-0.18-0·54, p=.33) and escitalopram (g=0.18, CI=0.01-0.34, p=.03) and the largest effect size for fluvoxamine (g=0.68, CI=-0.05-1.41, p=.07). However, due to the large 95% CI, fluvoxamine did not yield significance.

In the between group analysis stratified by drug category, SSRIs and SNRIs did not differ significantly for the DD group (p=.51), but SSRIs were significantly better (p=.04) compared to SNRIs in the AD group. No studies investigated the use of SNRIs in OCD.

The within antidepressant group analysis stratified by disorder yielded no significant difference (p=0.06) between studies of AD (g=1.58, CI=1.35-1.81, p<.001) and DD (g=1.85, CI=1.7-2.0, p<.001), yet both yielded significantly larger effect sizes (ps<.001) than studies of OCD (g=1.01, CI=0.88-1.14, p<.001). When stratified by drug, duloxetine yielded the largest effect size (g=1.95, CI=1.73-2.18, p<.001) and fluvoxamine the smallest (g=1.22, CI= 0.41-2.02, p=.003), however, the difference between those two was not significant (p=.08). The combined analysis across all disorders for the within group analysis yielded an antidepressant effect size of

---

[1] One study reported two trials that were treated independently for analyses [53]

[2] We use the abbreviation DD rather than MDD, as one study included MDD, Dysthymia, and Depressive Disorder Not Otherwise Specified [45].

g=1.62 (CI=1.48-1.76, p<.001). SSRIs and SNRIs did not differ significantly in both the DD group (p=.13) and in the AD group (p=.40)

The within placebo group analysis stratified by disorder yielded a large placebo response for studies of DD (*g*=1.57, CI=1.36-1.78, p<.001), which was significantly larger (p<.001) than the effect size for studies of AD (*g*=1.02, CI=0.85-1.20, p<.001). The moderate placebo response in the OCD group (*g*=0.63, CI=0.47-0.79, p<.001) was significantly lower than both the DD (p<.001) and AD (p=.001) groups (Figure 2). The combined analysis across all disorders for the within group analysis yielded a placebo effect size of *g*=1.22 (CI=1.06-1·38, p<.001).

Side Effect Analysis:

Treatment emergent adverse events (TEAEs) and serious adverse events (SAEs) were examined and tabulated across diagnosis categories to determine if emergent events were dependent on diagnosis or intervention type (Table 2).

Data on individual TEAEs and SAEs were available for a total of 2,542 patients taking an SSRI/SNRI and 2,294 patients taking placebo (sTable 2).

TEAE and SAE data revealed a diagnosis-dependent main effect. In patients treated with SSRIs or SNRIs, those with DD were significantly less likely to report TEAEs than those with AD (DD: 60.5% vs. AD: 80.85%; p=.001). The SAE data revealed the opposite pattern: Depressed patients treated with either SSRI or SNRI were significantly more likely to experience SAEs than AD (DD: 6.77% vs. AD: 2.30%; p=.007). The TEAEs (82.61%) and SAEs (2.42%) rates in the OCD group were almost identical to those in the AD group (Table 2). However, OCD side effect data differed significantly from DD with regard to TEAEs (p=.009), but not with regard to SAEs (p=.44).

Compared to placebo, patients taking both SSRIs and SNRIs reported significantly more SAEs (SSRI: 6.80%, SNRI: 7.59%) compared to placebo (3.32%; p=.01), but reported no significant difference in TEAEs. No difference in rates of SAEs and TEAEs was found between SSRIs and SNRIs.

Moderator Analysis:

The results of the univariate moderator analyses are presented in the Supplemental Material (S5, sTables 3-4). Notably, none of the categorical or continuous moderators was found to be

significant in a multivariate meta-regression with weighted effect sizes to adjust for multiple comparisons (S5, sTable 5).

Discussion

Our meta-analysis addresses the effectiveness and the safety profile of SNRIs and SSRIs in pediatric depression, anxiety disorders, OCD, and PTSD. We undertook both between group and within group analyses to determine effect sizes of SSRIs, SNRIs, and placebo treatments. The results indicate that SSRIs and SNRIs are more effective than placebo at treating several commonly diagnosed conditions in children and adolescents. However, it should be noted that this effect is only small. The magnitude of the drug versus placebo difference varies significantly by disorder, with a larger effect in AD than DD in between group analyses. This is surprising, given that only one SNRI, duloxetine, is currently FDA-approved for pediatric AD. However, it should be noted that DD and AD yield similar effect sizes in the within antidepressant group analysis. Further, patients with OCD exhibit a significantly smaller response to both drug treatment and placebo treatment compared to AD and DD.

The small effect size between SSRIs/SNRIs and placebo in pediatric DD might be due to lack of a clear depression phenotype. This was apparent in DSM-5 Field Trials on MDD, which found a low test-retest reliability (kappa: 0·28) for children, adolescents and adults [67]. This is further complicated by the high comorbidity between the disorders. A recent review on the use of SSRIs and SNRIs in pediatric populations reported that around 25% of patients with MDD had a comorbid AD [68]. In the studies included in our meta-analysis, while not all studies report comorbidity rates, those doing so report rates ranging between 6-56% of depressive patients having a comorbid anxiety disorder. Yet, attempts by the DSM-5 work group to create a "mixed anxiety and depression disorder" resulted in an unacceptable rate of test-retest reliability (kappa: -0·04) when tested in the DSM-5 Field Trials [67].

While it appears that the response to placebo is robust in pediatric DD, children and adolescents with anxiety disorders, who respond to pharmacological treatment to the same degree as those with DD, do not appear to exhibit such a robust placebo response. While in line with older reviews in children [69], this is in contrast to adult studies that found no significant differences in placebo effect size between depression and anxiety [70]. This contrast is not unique: placebo responses between children and adults differ significantly for binary outcomes across a

wide variety of diseases [71]. One explanation might be that children and adolescents with MDD may be more demoralized than patients with anxiety disorders and are therefore more sensitive to changes in hope and favorable meanings [69]. However, as no pediatric trial included a no-treatment arm, that could serve as a control for the natural course of the disorders, the difference in placebo response may also reflect differences in the probability of spontaneous improvement between the two pediatric disorders rather than differences in the placebo effect. Due to the small amount of studies in children, we could not estimate the drug and placebo response for the individual anxiety disorders, yet a recent adult study found drug-placebo effects size to be roughly equivalent across anxiety disorders [72] . In pediatric patients, however, those with panic disorder seem to experience a greater placebo response compared to patients with GAD or social phobia[73].

The substantial placebo response in MDD indicates that children and adolescents with depression might benefit from innovative treatment modalities that attempt to harness the power of the placebo effect in an ethical fashion, as these children and adolescents seem to respond well to clinician contact and social support that promote the therapeutic alliance [74] and other common factors such as the patients' expectations of improvement, their desire for relief, and the exposure to treatment rituals. It also offers several implications for research designs in antidepressant trials. Alternative designs such as a discontinuation design [75] or n of 1 trials [76,77] might be recommended when establishing efficacy [78], yet also have their individual shortcomings. The former is comprised of an acute treatment phase (i.e. patients receive their medication in an open fashion), followed by a continuation treatment phase (i.e. patients who had an adequate response are randomly assigned to medication or placebo), though this design might be prone to breaking blind [79]. Differences between the two medication groups could provide information about the magnitude of expectancy effects [80]. In this regard, response and remission rates to antidepressants have been shown to be significantly higher in comparator trials compared to placebo-controlled trials [81]. Future instructive studies could incorporate designs in which people who respond to placebo are kept on placebo.

With regard to side effects, our meta-analysis found that a similar percentage of patients' experience at least one side effect (TEAE), irrespective of being assigned to SSRIs (71.46%), SNRIs (65.92%), or to placebo (68.59%). Patients taking SSRIs and SNRIs did not report more TEAEs compared to placebo. This is in line with previous research showing that antidepressant

and placebo groups present with a similar risk to develop adverse events in both depressive children and adolescents [9]. With regard to TEAEs, it has been shown that negative expectations from investigators and patients can influence adverse effect reporting [82]. Further, some depressed patients may attribute pre-existing symptoms, which are highly common in the general population (e.g., headache, abdominal discomfort) to the antidepressant under investigation [83]. However, these mechanisms might not apply to serious adverse events. Accordingly, the serious side effects profiles for both SSRIs and SNRIs significantly exceeded that of the placebo arm in our meta-analysis. This is in line with other meta-analyses reporting increased suicidality (Odds Ratio=2.39, CI=1.31- 4.33) [84], suicidal ideation, and suicide attempts (risk difference: antidepressant vs. placebo: 0.7%, CI=0.1%-1.3%) [17] in children and adolescents receiving SSRIs and SNRIs compared to placebo. In conclusion, our results regarding serious side effects profiles support concerns about the safety of antidepressants in children and adolescents.

Perhaps the most remarkable finding regarding moderators in our meta-analysis is that we did not find a single predictor that was significantly related to the effect size in the multivariate analyses. However, this must be considered with caution. First, the absolute amount of statistical heterogeneity was small to moderate between studies, therefore making it very unlikely to detect moderators. Second, the multivariate analyses may have lacked the power to reveal statistically significant interaction effects [85]. A randomized trial in which participants are stratified by the predictor under investigation [86] or a patient-level meta-analysis [87] might be better suited to investigate the association between predictors and effect sizes than the methodology applied in this meta-analysis.

Limitations:

First, none of the RCTs included directly compared effectiveness across disorders. Accordingly, we could only make indirect conclusions with regard to disorder specificity. We only looked at observational comparisons, as no randomization across studies was possible. Second, although our meta-analysis included unpublished trials, reporting bias could lead to an overly positive representation of findings in the literature [88]. Third, restrictive inclusion criteria of clinical trials such as non-inclusion of comorbidity and a higher symptom severity threshold make it difficult to generalize results to real-world populations [89]. Fourth, there was significant heterogeneity in how side effects were reported and few studies explain how TEAE data are elicited from the

patients (i.e., open questions versus structured questionnaire). Fifth, for PTSD, only one study fit our inclusion criteria[10]. Therefore, we were unable to include a categorical analysis of SSRIs/SNRIs for the treatment of pediatric PTSD.

Future directions:

The main findings of this analysis present multiple avenues for further analyses. First, the nearly identical response rate for pediatric DD and AD deserves further investigation and perhaps the revision of federal prescribing guidelines for these two conditions. While several SSRIs and SNRIs have been approved for the treatment of pediatric DD and OCD, only one – duloxetine – has recently received FDA approval for pediatric ADs [90]. Second, the substantial differential response to both drug treatment and placebo treatment between OCD compared to AD and DD highlights underlying differences in the etiologies and pathogeneses of the disorders that may require additional interventions for pediatric patients with OCD (see for example [91]). It is our hope that an RDoC approach [92] will help elucidate the above-mentioned points and could lead to better treatment outcomes in the future. Third, our results point to the need for additional research into the factors that moderate the efficacy of SSRIs and SNRIs in children and highlight the need for more comprehensive reporting of population and illness details (such as age of onset and duration of illness) in clinical and pragmatic trials. Similarly, additional research in understanding the developmental ontogeny of the placebo response is warranted. Finally, our results highlight the need for a standardized method of reporting TEAEs and SAEs. Given the potential for life threatening events in young children and adolescents, understanding the extent to which these medications pose a genuine risk to youth is of extreme urgency. This would allow future research to deviate from the current line of research estimating the magnitude and differences between drug and placebo effects and focus more on precision medicine driven questions, such as which treatment (or combination thereof) may be most advantageous for certain patient subgroups in certain clinical settings.

Contributors

CL, SZ, and JK conceived and designed the study. CL, HK, and SZ, selected the articles and extracted the data. CL, HK, JK, SZ and CW analyzed the data. CL, JK, SZ and HK wrote the first draft of the manuscript. CL, HK, JK, RK, DP, CW and IK interpreted the data and wrote the final version. CL and HK contributed equally to this study. All authors read and met the ICMJE criteria for authorship and agree with the results and conclusions of this Article.

Declaration of interests

In the past 3 years, Dr. Kessler received support for his epidemiological studies from Sanofi Aventis, was a consultant for Johnson & Johnson Wellness and Prevention, and served on an advisory board for the Johnson & Johnson Services Inc. Lake Nona Life Project. Kessler is a co-owner of DataStat, Inc., a market research firm that carries out healthcare research.

Acknowledgments

This study was supported by NLM grant # T15LM007092 and grant project (P300P1_158427) awarded to JK by the Swiss National Science Foundation. Dr. Pine's work is supported by NIMH-IRP project ZIAMH-002781. We thank Argyris Stringaris and Michael Sugarman for their comments on the manuscript.

References

1. Merikangas KR, He JP, Burstein M, et al. Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication--Adolescent Supplement (NCS-A). *J Am Acad Child Adolesc Psychiatry* 2010; **49**(10): 980-9.
2. Rutter M, Kim-Cohen J, Maughan B. Continuities and discontinuities in psychopathology between childhood and adult life. *Journal of Child Psychology and Psychiatry* 2006; **47**(3-4): 276-95.
3. Vaswani M, Linda FK, Ramesh S. Role of selective serotonin reuptake inhibitors in psychiatric disorders: a comprehensive review. *Progress in neuro-psychopharmacology and biological psychiatry* 2003; **27**(1): 85-102.
4. Cox GR, Callahan P, Churchill R, et al. Psychological therapies versus antidepreaant medication, alone and in combination for depression in children and adolescents. *Cochrane Database of Systematic Reviews* 2014; (11).
5. Machado M, Einarson T. Comparison of SSRIs and SNRIs in major depressive disorder: a meta-analysis of head-to-head randomized clinical trials. *Journal of clinical pharmacy and therapeutics* 2010; **35**(2): 177-88.
6. Whittington CJ, Kendall T, Fonagy P, Cottrell D, Cotgrove A, Boddington E. Selective serotonin reuptake inhibitors in childhood depression: systematic review of published versus unpublished data. *The Lancet* 2004; **363**(9418): 1341-5.
7. Ciprani A, Zhou X, Del Giovane C, et al. Comparative efficacy and tolerability of antidepressants for major depressive disorder in children and adolescents: a network meta-analysis. *The Lancet* 2016; **388**: 881-90.
8. Garland EJ, Kutcher S, Virani A, Elbe D. Update on the use of SSRIs and SNRIs with children and adolescents in clinical practice. *Journal of the Canadian Academy of Child and Adolescent Psychiatry* 2016; **25**(1): 4.
9. Rojas-Mirquez JC, Rodriguez-Zuniga MJ, Bonilla-Escobar FJ, et al. Nocebo effect in randomized clinical trials of antidepressants in children and adolescents: systematic review and meta-analysis. *Front Behav Neurosci* 2014; **8**: 375.
*10. Robb AS, Cueva JE, Sporn J, Yang R, Vanderburg DG. Sertraline treatment of children and adolescents with posttraumatic stress disorder: a double-blind, placebo-controlled trial. *Journal of child and adolescent psychopharmacology* 2010; **20**(6): 463-71.

11. Mann JJ, Emslie G, Baldessarini RJ, et al. ACNP task force report on SSRIs and suicidal behavior in youth. *Neuropsychopharmacology* 2006; **31**: 473-92.
12. Stone MB. The FDA Warning on Antidepressants and Suicidality — Why the Controversy? *New England Journal of Medicine* 2014; **371**(18): 1668-71.
13. Walsh BT, Seidman SN, Sysko R, Gould M. Placebo response in studies of major depression: variable, substantial, and growing. *JAMA* 2002; **287**(10): 1840-7.
14. Rutherford BR, Sneed JR, Tandler JM, Rindskopf D, Peterson BS, Roose SP. Deconstructing pediatric depression trials: an analysis of the effects of expectancy and therapeutic contact. *J Am Acad Child Adolesc Psychiatry* 2011; **50**(8): 782-95.
15. Bridge JA, Birmaher B, Iyengar S, Barbe RP, Brent DA. Placebo response in randomized controlled trials of antidepressants for pediatric major depressive disorder. *Am J Psychiatry* 2009; **166**(1): 42-9.
16. Leuchter AF, Hunter AM, Tartter M, Cook IA. Role of pill-taking, expectation and therapeutic alliance in the placebo response in clinical trials for major depression. *British Journal of Psychiatry* 2014; **205**(6): 443-9.
17. Bridge JA, Iyengar S, Salary CB, et al. Clinical response and risk for reported suicidal ideation and suicide attempts in pediatric antidepressant treatment: a meta-analysis of randomized controlled trials. *JAMA* 2007; **297**(15): 1683-96.
18. Birmaher B, Brent D, Issues AWGoQ, et al. Practice parameter for the assessment and treatment of children and adolescents with depressive disorders. *J Am Acad Child Adolesc Psychiatry* 2007; **46**(11): 1503-26.
19. Cochrane CT. Cochrane Handbook for Systematic Reviews of Interventions; 2011.
20. Altman DG, Royston P. The cost of dichotomising continuous variables. *British Medical Journal* 2006; **332**(7549): 1080.
21. Moncrieff J, Kirsch I. Efficacy of antidepressants in adults. *British Medical Journal* 2005; **331**(7509): 155-7.
*22. Wagner KD, Robb AS, Findling RL, Jin J, Gutierrez MM, Heydorn WE. A randomized, placebo-controlled trial of citalopram for the treatment of major depression in children and adolescents. *Am J Psychiatry* 2004a; **161**(6): 1079-83.

*23. Wagner KD, Jonas J, Findling RL, Ventura D, Saikali K. A double-blind, randomized, placebo-controlled trial of escitalopram in the treatment of pediatric depression. *J Am Acad Child Adolesc Psychiatry* 2006; **45**(3): 280-8.

24. Hedges LV, Olkin I. Statistical methods for meta-analysis: Academic press; 2014.

25. Cooper GL. The safety of fluoxetine--an update. *Br J Psychiatry Suppl* 1988; (3): 77-86.

26. Cochran WG. The comparison of percentages in matched samples. *Biometrika* 1950; **37**: 256-66.

27. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *British Medical Journal* 2003; **327**: 557-60.

28. Higgins JP. Commentary: Heterogeneity in meta-analyses should be expected and appropriately quantified. *International Journal of Epidemiology* 2008; **37**(5): 1158-60.

29. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods* 2010; **1**(2): 97-111.

30. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *Bmj* 1997; **315**(7109): 629-34.

31. Rosenthal R. Meta-analytic procedures for social research: Sage; 1991.

32. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994: 1088-101.

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

*34. Simeon JG, Dinicola VF, Ferguson HB, Copping W. Adolescent depression: a placebo-controlled fluoxetine treatment study and follow-up. *Prog Neuropsychopharmacol Biol Psychiatry* 1990; **14**(5): 791-5.

*35. Emslie GJ, Rush AJ, Weinberg WA, Kowatch RA, Hughes CW, Carmody T. A double-blind placebo controlled trial of fluoxetine in depressed children and adolescents. *Sixth World Congress of Biological Psychiatry; 1997 Jun 22-27; Nice, France* 1997.

*36. Keller MB, Ryan ND, Strober M, et al. Efficacy of paroxetine in the treatment of adolescent major depression: a randomized, controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry* 2001; **40**(7): 762-72.

*37. Emslie GJ, Heiligenstein JH, Wagner KD, et al. Fluoxetine for acute treatment of depression in children and adolescents: a placebo-controlled, randomized clinical trial. *J Am Acad Child Adolesc Psychiatry* 2002; **41**(10): 1205-15.

*38. Wagner KD, Ambrosini P, Rynn M, et al. Efficacy of sertraline in the treatment of children and adolescents with major depressive disorder: two randomized controlled trials. *JAMA* 2003; **290**(8): 1033-41.

*39. March J, Silva S, Petrycki S, et al. Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression - Treatment for adolescents with depression study (TADS) randomized controlled trial. *Jama-Journal of the American Medical Association* 2004; **292**(7): 807-20.

*40. Berard R, Fong R, Carpenter DJ, Thomason C, Wilkinson C. An international, multicenter, placebo-controlled trial of paroxetine in adolescents with major depressive disorder. *J Child Adolesc Psychopharmacol* 2006; **16**(1-2): 59-75.

*41. Emslie GJ, Wagner KD, Kutcher S, et al. Paroxetine treatment in children and adolescents with major depressive disorder: a randomized, multicenter, double-blind, placebo-controlled trial. *J Am Acad Child Adolesc Psychiatry* 2006; **45**(6): 709-19.

*42. von Knorring A-L, Olsson GI, Thomsen PH, Lemming OM, Hultén A. A randomized, double-blind, placebo-controlled study of citalopram in adolescents with major depressive disorder. *Journal of Clinical Psychopharmacology* 2006; **26**(3): 311-5.

*43. Emslie GJ, Findling RL, Yeung PP, Kunz NR, Li Y. Venlafaxine ER for the treatment of pediatric subjects with depression: results of two placebo-controlled trials. *J Am Acad Child Adolesc Psychiatry* 2007; **46**(4): 479-88.

*44. Emslie GJ, Ventura D, Korotzer A, Tourkodimitris S. Escitalopram in the treatment of adolescent depression: a randomized placebo-controlled multisite trial. *J Am Acad Child Adolesc Psychiatry* 2009; **48**(7): 721-9.

*45. Findling RL, Pagano ME, McNamara NK, et al. The short-term safety and efficacy of fluoxetine in depressed adolescents with alcohol and cannabis use disorders: a pilot randomized placebo-controlled trial. *Child Adolesc Psychiatry Ment Health* 2009; **3**(1): 11.

*46. PIR112487. A randomized, double-blind, placebo controlled, parallel group, flexible dose study to evaluate the efficacy and safety of Paxil® Tablets in children and adolescents with major depressive disorder. Japan: GlaxoSmithKline; 2011. p. 4.

*47. Atkinson SD, Prakash A, Zhang Q, et al. A double-blind efficacy and safety study of duloxetine flexible dosing in children and adolescents with major depressive disorder. *J Child Adolesc Psychopharmacol* 2014; **24**(4): 180-9.

*48. Emslie GJ, Prakash A, Zhang Q, Pangallo BA, Bangs ME, March JS. A double-blind efficacy and safety study of duloxetine fixed doses in children and adolescents with major depressive disorder. *J Child Adolesc Psychopharmacol* 2014; **24**(4): 170-9.

*49. Psychopharmacology RUoP, Walkup JT, Labellarte MJ, et al. Fluvoxamine for the Treatment of Anxiety Disorders in Children and Adolescents. *New England Journal of Medicine* 2001; **344**(17): 1279-85.

*50. Rynn MA, Siqueland L, Rickels K. Placebo-controlled trial of sertraline in the treatment of children with generalized anxiety disorder. *Am J Psychiatry* 2001; **158**(12): 2008-14.

*51. Birmaher B, Axelson DA, Monk K, et al. Fluoxetine for the treatment of childhood anxiety disorders. *J Am Acad Child Adolesc Psychiatry* 2003; **42**(4): 415-23.

*52. Wagner KD, Berard R, Stein MB, et al. A multicenter, randomized, double-blind, placebo-controlled trial of paroxetine in children and adolescents with social anxiety disorder. *Arch Gen Psychiatry* 2004b; **61**(11): 1153-62.

*53. March JS, Entusah AR, Rynn M, Albano AM, Tourian KA. A Randomized controlled trial of venlafaxine ER versus placebo in pediatric social anxiety disorder. *Biol Psychiatry* 2007; **62**(10): 1149-54.

*54. Rynn MA, Riddle MA, Yeung PP, Kunz NR. Efficacy and safety of extended-release venlafaxine in the treatment of generalized anxiety disorder in children and adolescents: two placebo-controlled trials. *Am J Psychiatry* 2007; **164**(2): 290-300.

*55. Walkup JT, Albano AM, Piacentini J, et al. Cognitive behavioral therapy, sertraline, or a combination in childhood anxiety. *N Engl J Med* 2008; **359**(26): 2753-66.

*56. da Costa CZ, de Morais RM, Zanetta DM, et al. Comparison among clomipramine, fluoxetine, and placebo for the treatment of anxiety disorders in children and adolescents. *J Child Adolesc Psychopharmacol* 2013; **23**(10): 687-92.

*57. Strawn JR, Prakash A, Zhang Q, et al. A randomized, placebo-controlled study of duloxetine for the treatment of children and adolescents with generalized anxiety disorder. *Journal of the American Academy of Child & Adolescent Psychiatry* 2015; **54**(4): 283-93.

*58. Melvin GA, Dudley AL, Gordon MS, et al. Augmenting cognitive behavior therapy for school refusal with fluoxetine: A randomized controlled trial. *Child Psychiatry & Human Development* 2016: 1-13.

*59. Riddle MA, Scahill L, King RA, et al. Double-blind, crossover trial of fluoxetine and placebo in children and adolescents with obsessive-compulsive disorder. *J Am Acad Child Adolesc Psychiatry* 1992; **31**(6): 1062-9.

*60. March JS, Biederman J, Wolkow R, et al. Sertraline in children and adolescents with obsessive-compulsive disorder: a multicenter randomized controlled trial. *JAMA* 1998; **280**(20): 1752-6.

*61. Geller DA, Hoog SL, Heiligenstein JH, et al. Fluoxetine treatment for obsessive-compulsive disorder in children and adolescents: a placebo-controlled clinical trial. *J Am Acad Child Adolesc Psychiatry* 2001; **40**(7): 773-9.

*62. Riddle MA, Reeve EA, Yaryura-Tobias JA, et al. Fluvoxamine for children and adolescents with obsessive-compulsive disorder: a randomized, controlled, multicenter trial. *J Am Acad Child Adolesc Psychiatry* 2001; **40**(2): 222-9.

*63. Liebowitz MR, Turner SM, Piacentini J, et al. Fluoxetine in children and adolescents with OCD: a placebo-controlled trial. *J Am Acad Child Adolesc Psychiatry* 2002; **41**(12): 1431-8.

*64. Team TPOSP, March JS, Foa E, et al. Cognitive behavior therapy, sertraline, and their combination for children and adolescents with obsessive-compulsive disorder. *JAMA* 2004; **292**(16): 1969-76.

*65. Geller DA, Wagner KD, Emslie G, et al. Paroxetine treatment in children and adolescents with obsessive-compulsive disorder: a randomized, multicenter, double-blind, placebo-controlled trial. *J Am Acad Child Adolesc Psychiatry* 2004; **43**(11): 1387-96.

*66. Storch EA, Bussing R, Small BJ, et al. Randomized, placebo-controlled trial of cognitive-behavioral therapy alone or combined with sertraline in the treatment of pediatric obsessive-compulsive disorder. *Behav Res Ther* 2013; **51**(12): 823-9.

67. Regier DA, Narrow WE, Clarke DE, et al. DSM-5 Field Trials in the United States and Canada, part II: Test-retest reliability of selected categorical diagnoses. *The American Journal of Psychiatry* 2013; **170**(1): 59-70.

68. Garland EJ, Kutcher S, Virani A, Elbe D. Position Paper: Update on the use of SSRIs and SNRIs with children and adolescents in clinical practice. *Journal of the Canadian Academy of Child and Adolescent Psychiatry* 2016; **25**(1): 4-10.

69. Cohen D, Deniau E, Maturana A, et al. Are child and adolescent responses to placebo higher in major depression than in anxiety disorders? A systematic review of placebo-controlled trials. *PLoS One* 2008; **3**(7): e2632.

70. Sugarman MA, Loree AM, Baltes BB, Grekin ER, Kirsch I. The efficacy of paroxetine and placebo in treating anxiety and depression: a meta-analysis of change on the Hamilton Rating Scales. *PLoS One* 2014; **9**(8): e106337.

71. Janiaud P, Cornu C, Lajoinie A, Djemli A, Cucherat M, Kassai B. Is the perceived placebo effect comparable between adults and children? A meta-regression analysis. *Pediatric Research* 2016; **doi:10.1038/pr.2016.181**: 1-7.

72. Sugarman MA, Kirsch I, Huppert JD. Obsessive-compulsive disorder has a reduced placebo (and antidepressant) response compared to other anxiety disorders. *Manuscript submitted for publication* 2016.

73. Dobson ET, Strawn JR. Placebo response in pediatric anxiety disorders: Implications for clinical trial design and interpretation. *Journal of Child and Adolescent Psychopharmacology* 2016; **26**(8): 686-93.

74. Rutherford BR, Sneed JR, Tandler JM, Rindskopf D, Peterson BS, Roose SP. Deconstructing pediatric depression trials: An analysis of the effect of expectancy and theurapeutic contact. *Journal of the American Academy of Child & Adolescent Psychiatry* 2011; **50**(8): 782-95.

75. Emslie GJ, Kennard BD, Mayes TL, et al. Fluoxetine versus placebo in preventing relapse of major depression in children and adolescents. *The American Journal of Psychiatry* 2008; **165**(4): 459-67.

76. Scuffham PA, Nikles J, Mitchell GK, et al. Using N-of-1 trials to improve patient management and save costs. *Journal of General Internal Medicine* 2010; **25**(9): 906-13.

77. Kent MA, Camfield CS, Camfield PR. Double-bling metylphenidate trials: practical, useful, and highly endorsed by families. *Archives of pediatrics and adolescent medicine* 1999; **153**(12): 1292-6.

78. Temple R, Ellenberg SS. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: ethical and scientific issues. *Annals of internal Medicine* 2000; **133**(6): 455-63.

79. Kopec JA, Abrahamowicz M, Esdaile JM. Randomized discontinuation trials: utility and efficiency. *J Clin Epidemiol* 1993; **46**(9): 959-71.

80. Deshauer D, Grimshaw J. Can discontinuation trials inform the use of antidepressants in depressed children? *The American Journal of Psychiatry* 2008; **165**(9): 1205-.

81. Rutherford BR, Sneed JR, Roose SP. Does study design influence outcome? The effects of placebo control and treatment duration in antidepressant trials. *Psychotherapy and Psychosomatics* 2009; **78**: 172-81.

82. Rief W, Bingel U, Schedlowski M, Enck P. Mechanisms involved in placebo and nocebo responses and implications for drug trials. *Clinical Pharmacology and Therapeutics* 2011; **90**(5): 722-6.

83. Rief W, Nestoriuc Y, von Lilienfeld-Toal A, et al. Differences in adverse effect reporting in placebo groups in SSRI and tricyclic antidepressant trials: A systematic review and meta-analysis. *Drug Safety* 2009; **32**(11): 1041-56.

84. Sharma T, Guski LS, Freund N, Gøtzsche PC. Suicidality and aggression during antidepressant treatment: systematic review and meta-analyses based on clinical study reports. *BMJ* 2016; **352**.

85. Gerger H, Hlavica M, Gaab J, Munder T, Barth J. Does it matter who provides psychological interventions for medically unexplained symptoms? A meta-analysis. *Psychotherapy and Psychosomatics* 2015; **84**(4): 17-26.

86. Cuijpers P, van Straten A, Warmerdam L, Smits N. Characteristics of effective psychological treatment of depression: a metaregression analysis. *Psychotherapy Research* 2008; **18**(2): 225-36.

87. Fournier JC, DeRubeis RJ, Hollon SD, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *Jama* 2010; **303**(1): 47-53.

88. Roest AM, de Jonge P, Williams CD, de Vries A, Schoevers RA, Turner EH. Reporting bias in clinical trials investigating the efficacy of second-generation antidepressants in the treatment of anxiety disorders. A report of 2 meta-analyses. *JAMA Psychiatry* 2015; **72**(5): 500-10.

89. Zimmerman M, Clark HL, Multach MD, Walsh E, Rosenstein LK, Gazarian D. Have treatment studies of depression become even less generalizable? A review of the inclusion and exclusion criteria used in placebo-controlled antidepressant efficacy trials published during the past 20 years. *Mayo Clinic Proceedings* 2015; **90**(9): 1180-6.
90. Strawn JR, Welge JA, Wehry AM, Keeshin B, Rynn MA. Efficacy and tolerability of antidepressants in pediatric anxiety disorders: a systematic review and meta-analysis. *Depress Anxiety* 2015; **32**(3): 149-57.
91. Weinberg A, Kotov R, Proudfit GH. Neural indicators of error processing in generalized anxiety disorder, obsessive-compulsive disorder, and major depressive disorder. *Journal of Abnormal Psychology* 2014; **124**(1): 172-85.
92. Insel T, Cuthbert B, Garvey M, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry* 2010; **167**(7): 748-51.

Figure Legends

Figure 1. Forest Plot of Between Group Analyses (Stratified by Disorder)

Figure 2. Drug and Placebo Effect Size by Disorder Category

Table 1. Demographics and Study Characteristics

Source	Diagnosis	Tx Length wk	No. of Patients[c]	Primary Outcome	Intervention	FDA Approval	Mean Age, y	Female %	Length of Illness, m	Age of Onset, y	Quality	Funding	Location
Depressive Disorder													
Simeon et al,[34] 1990	MDD[a]	7	40	HAM-D	Fluoxetine (20 - 60 mg/d)	≥ 8 years	16	55	-	-	0.33	Industry	North America
Emslie et al,[35] 1997	MDD[a, b]	8	96	CDRS-R	Fluoxetine (20 mg/d)	≥ 8 years	12.35 (2.65)	46	18.40	10.80 (2.65)	0.78	Public	North America
Keller et al,[36] 2001[d]	MDD	8	180	HAM-D	Paroxetine (20 - 40 mg/d)	N/A	14.95 (1.60)	64	13.52	13.29 (2.56)	0.89	Industry	North America
Emslie et al,[37] 2002	MDD[b]	9	219	CDRS-R	Fluoxetine (20 mg/d)	≥ 8 years	12.70 (2.57)	49	14	10.33 (3.02)	0.89	Industry	North America
Wagner et al,[38] 2003	MDD	10	376	CDRS-R	Sertraline (50 - 200 mg/d)	N/A	6-17[e]	51	23.07	10.04	0.78	Industry	Inter-national
March et al,[39] 2004[d, h]	MDD	12	328	CDRS-R	Fluoxetine (10 - 40 mg/d)	≥ 8 years	14.60 (1.5)[f]	55[f]	40[a]	-	1.00	Public[i]	North America
Wagner et al,[22] 2004a	MDD	8	178	CDRS-R	Citalopram (20 - 40 mg/d)	N/A	12.10 (2.95)	63	27	9.80 (3.15)	0.78	Industry	North America
Berard et al,[40] 2006	MDD	12	286	MADRS	Paroxetine (20 - 40 mg/d)	N/A	15.60 (1.60)	67	-	-	0.89	Industry	Inter-national
Emslie et al,[41] 2006	MDD[a]	8	206	CDRS-R	Paroxetine (10 - 50 mg/d)	N/A	12.00 (2.97)	47	25.90	9.80 (3.30)	0.78	Industry	North America
von Knorring et al,[42] 2006	MDD	12	244	K-SADS-P	Citalopram (10 - 40 mg/d)	N/A	16 (1.00)	-	-	-	0.44	Industry	Europe

Table 1. Demographics and Study Characteristics (cont.)

Source	Diagnosis	Tx Length wk	No. of Patients[c]	Primary Outcome	Intervention	FDA Approval	Mean Age, y	Female %	Length of Illness, m	Age of Onset, y	Quality	Funding	Location
Wagner et al,[23] 2006	MDD	8	268	CDRS-R	Escitalopram (10 - 20 mg/d)	≥ 12 years	12.30 (3.00)	52	25.80	10.15 (3.25)	0.67	Industry	North America
Emslie et al,[43] 2007	MDD	8	367	CDRS-R	Venlafaxine ER (37.50 - 225 mg/d)	N/A	12.25 (2.60)	46	21.13	-	0.78	Industry	North America
Emslie et al,[44] 2009	MDD	8	316	CDRS-R	Escitalopram (10 - 20 mg/d)	≥ 12 years	14.60 (1.55)	59	16.10	12.35 (2.55)	0.67	Industry	North America
Findling et al,[45] 2009	MDD, Dys-thymia, DDNOS, SUD	8	34	CDRS-R	Fluoxetine (10 - 20 mg/d)	≥ 8 years	16.46 (1.08)	15	49.04	11.41 (2.50)	0.89	Industry	North America
PIR-112487,[46] 2011	MDD[b]	8	56	CDRS-R	Paroxetine (10 - 40 mg/d)	N/A	14.59 (2.33)	61	-	-	0.56	Industry	Asia
Atkinson et al,[47] 2014	MDD[a, b]	10	337	CDRS-R	Duloxetine (60 - 120 mg/d) Fluoxetine (20 - 40 mg/d)	N/A ≥ 8 years	13.20 (3.14)	52	-	11.60	0.89	Industry	Inter-national
Emslie et al,[48] 2014	MDD[a, b]	10	463	CDRS-R	Duloxetine (60 mg/d) Duloxetine (30 mg/d) Fluoxetine (20 mg/d)	N/A N/A ≥ 8 years	12.98 (2.98)	51	-	-	0.78	Industry	Inter-national

Table 1. Demographics and Study Characteristics (cont.)

Source	Diagnosis	Tx Length wk	No. of Patients[c]	Primary Outcome	Intervention	FDA Approval	Mean Age, y	Female %	Length of Illness, m	Age of Onset, y	Quality	Funding	Location
Anxiety Disorder													
RUPP,[49] 2001	SP, SAD, GAD	8	128	PARS	Fluvoxamine (50 - 300 mg/d)	N/A	10.30 (2.95)	49	-	-	0.67	Public / Industry	North America
Rynn et al,[50] 2001	GAD	9	22	HAM-A	Sertraline (25 - 50 mg/d)	N/A	11.70 (3.90)	23	>12	-	0.67	Public	North America
Birmaher et al,[51] 2003	SAD, SP, GAD, SM, PH, PD	12	74	PARS[f]	Fluoxetine (10 - 20 mg/d)	N/A	11.80 (2.80)	54	62.70	-	0.78	Public[i]	North America
Wagner et al,[52] 2004b	SP	16	322	LSAS-CA[f]	Paroxetine (10 - 50 mg/d)	N/A	13.10 (2.77)	49	-	-	0.89	Industry	Inter-national
March et al,[53] 2007	SP	16	293	SAS-CA	Venlafaxine ER (37.50 - 225 mg/d)	N/A	13.60 (2.55)	57	58	-	1.00	Industry	North America
Rynn et al,[54] 2007[g]	GAD	8	323	K-SADS-GA	Venlafaxine (37.50 - 225 mg/d)	N/A	11.30 (2.86)	41	39.12	-	0.78	Industry	North America
Walkup et al,[55] 2008[d, h]	SAD, GAD, SP	12	349	PARS	Sertraline (25 - 200 mg/d)	N/A	10.70 (2.80)	50	-	-	1.00	Public[i]	North America
da Costa et al,[56] 2013[d]	GAD, SAD, SP	12	21	MASC[f]	Fluoxetine (10 - 60 mg/d)	N/A	11.50	48	-	-	0.56	Public	South America
Strawn et al,[57] 2015	GAD	10	272	PARS	Duloxetine (30 - 120 mg/d)	N/A	12.40 (2.95)	53	52.20		1.00	Public / Industry	Inter-national
Melvin et al,[58] 2016	SP, PH, GAD, SAD, PD	10	42	RCMAS[f]	Fluoxetine + CBT (10 - 60 mg/d)	≥ 8 years	13.60 (1)	45	-	-	0.89	Public[i]	Australia

Table 1. Demographics and Study Characteristics (cont.)

Source	Diagnosis	Tx Length wk	No. of Patients[c]	Primary Outcome	Intervention	FDA Approval	Mean Age, y	Female %	Length of Illness, m	Age of Onset, y	Quality	Funding	Location
Obsessive-Compulsive Disorder													
Riddle et al,[59] 1992	OCD	8	14	CY-BOCS[f]	Fluoxetine (20 mg/d)	≥ 7 years	11.80 (2.30)	57	-	-	0.44	Public[i]	North America
March et al,[60] 1998	OCD	12	189	CY-BOCS[f]	Sertraline (25 - 200 mg/d)	≥ 6 years	12.60 (6-17)[e]	-	45.60	-	0.78	Industry	North America
Geller et al,[61] 2001	OCD	13	103	CY-BOCS	Fluoxetine (20 - 60 mg/d)	≥ 7 years	11.40 (2.90)	37	>6	-	0.89	Industry	North America
Riddle et al,[62] 2001[d]	OCD	10	120	CY-BOCS	Fluvoxamine (50 - 200 mg/d)	≥ 8 years	13.03	47	43.20	-	0.67	Industry	North America
Liebowitz et al,[63] 2002[d]	OCD	8	43	CY-BOCS	Fluoxetine (20 - 80 mg/d)	≥ 7 years	12.65 (2.19)	42	≥12	-	1.00	Public / Industry	North America
POTS,[64] 2004[h]	OCD	12	84	CY-BOCS	Sertraline (25-200 mg/d)	≥ 6 years	12 (2.70)	45	-	-	1.00	Public[i]	North America
Geller et al,[65] 2004	OCD	10	207	CY-BOCS	Paroxetine (10 - 50 mg/d)	N/A	11.30 (3.00)	42	50.40	7.50 (3.09)	1.00	Industry	North America
Storch et al,[66] 2013	OCD	18	47	CY-BOCS	Sertraline (Reg) + CBT (25 - 200 mg/d) Sertraline (Slow) + CBT (25 - 200 mg/d)	≥ 6 years ≥ 6 years	11.90 (3.47)	38	-	-	0.67	Public[i]	North America

Table 1. Demographics and Study Characteristics (cont.)

Source	Diagnosis	Tx Length wk	No. of Patients[c]	Primary Outcome	Intervention	FDA Approval	Mean Age, y	Female %	Length of Illness, m	Age of Onset, y	Quality	Funding	Location
Posttraumatic Stress Disorder													
Robb et al,[10] 2010	PTSD	10	131	UCLA-PTSD	Sertraline (50 - 200 mg/d)	N/A	10.98 (1.75)	60	29.40	-	0.78	Industry	North America

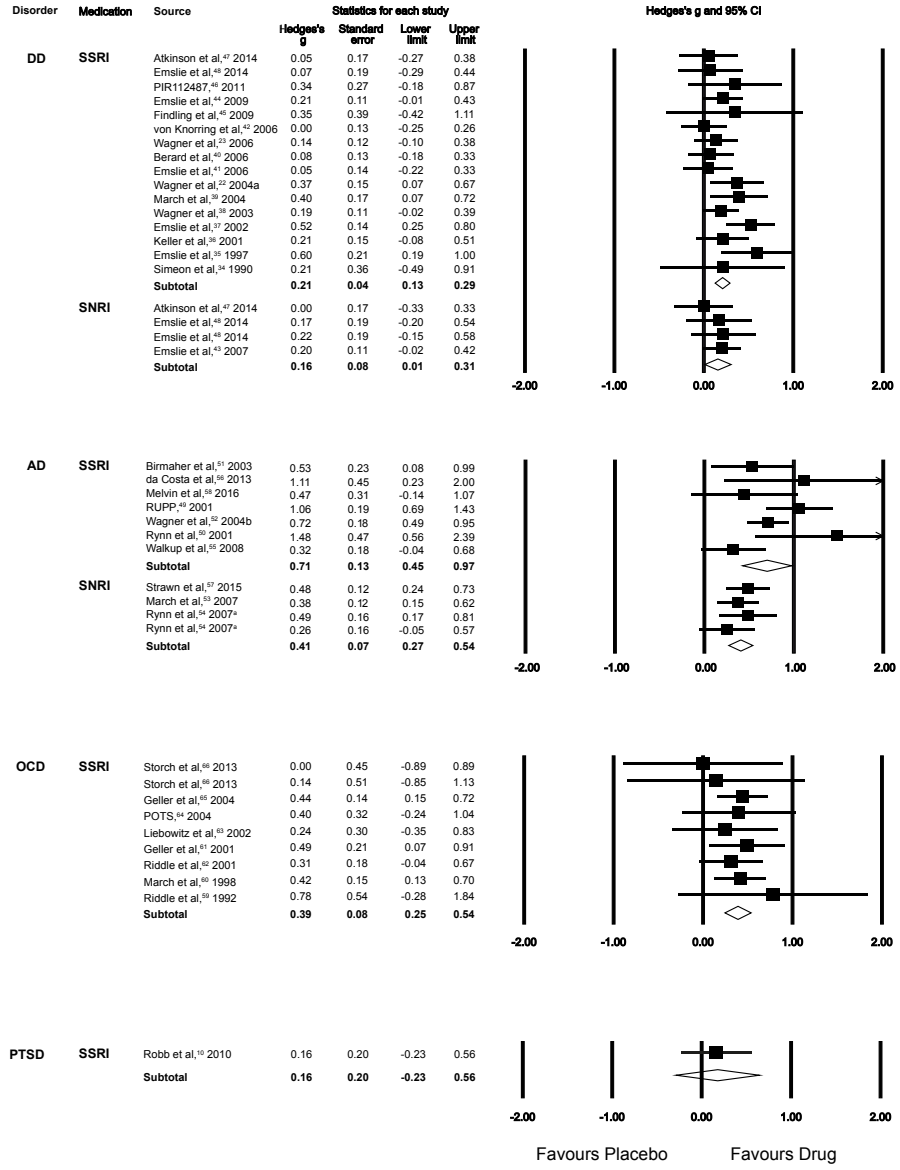
Abbreviations: OCD, Obsessive Compulsive Disorder; MDD, Major Depressive Disorder; DDNOS, Depressive Disorder Not Otherwise Specified; SUD, Substance Use Disorder; SAD, Separation Anxiety Disorder; SP, Social Phobia; GAD, Generalized Anxiety Disorder; SM, Selective Mutism; PH, Specific Phobia; PD, Panic Disorder.
[a]Without psychotic features.
[b]Single episode or recurrent.
[c]N included in our analysis. Some studies included additional arms (i.e., tricyclic antidepressant, CBT alone, etc.) that were not extracted.
[d]Only SSRI or SNRI treatment arms from acute treatment phase / pre-crossover phase extracted as per protocol.
[e]Range. Mean (SD) not available.
[f]Primary outcome not specified or data not usable; most common measure for which data was available was chosen.
[g]Two trials. Analyses broken down by study.
[h]Additional arm (antidepressant + CBT) was not extracted.
[i]Study medication and matching placebo provided by the drug manufacturer.
[j]Based on specification of a main outcomes, selective reporting, observer-rated outcomes, blinding of outcome assessors, selective attrition, generation of allocation sequence, concealment, randomization, and ITT-analyses. Scores denote means over all items ranging from 0 to 1, with higher scores implicating greater methodological quality.

Table 2. Treatment Emergent Adverse Events (TEAE) and Serious Adverse Events (SAE)

Disorder & Intervention		TEAE		SAE	
		Total No.	Patients with ≥1 TEAE, %	Total No.	Patients with ≥1 SAE, %
In Drug and Placebo Groups by Disorder					
DD					
	SSRI	1099	65.70	1120	8.66
	SNRI	510	51.76	169	13.61
	Placebo	999	60.16	1280	4.22
	DD Overall	2608	60.85	2569	6.77
AD					
	SSRI	163	88.34	436	1.38
	SNRI	429	82.75	292	4.11
	Placebo	604	77.48	532	2.07
	AD Overall	1196	80.85[a]	1260	2.30[a]
OCD					
	SSRI	254	85.04	136	3.68
	SNRI	-	-	-	-
	Placebo	229	79.91	112	0.89
	OCD Overall	483	82.61[a]	248	2.42
In SSRI, SNRI and Placebo Groups					
SSRI					
	Citalopram	121	75.21	121	14.88
	Escitalopram	286	73.78	286	2.10
	Fluoxetine	479	62.21	223	17.49
	Fluvoxamine	58	81.03	-	-
	Paroxetine	573	75.74	666	6.76
	Sertraline	67	76.12	556	3.24
	SSRI Overall	1584	71.46	1852	6.80[b]
SNRI					
	Duloxetine	476	67.65	135	7.41
	Venlafaxine	463	64.15	326	7.67
	SNRI Overall	939	65.92	461	7.59[b]
Placebo					
	Placebo Overall	1894	68.59	1986	3.32

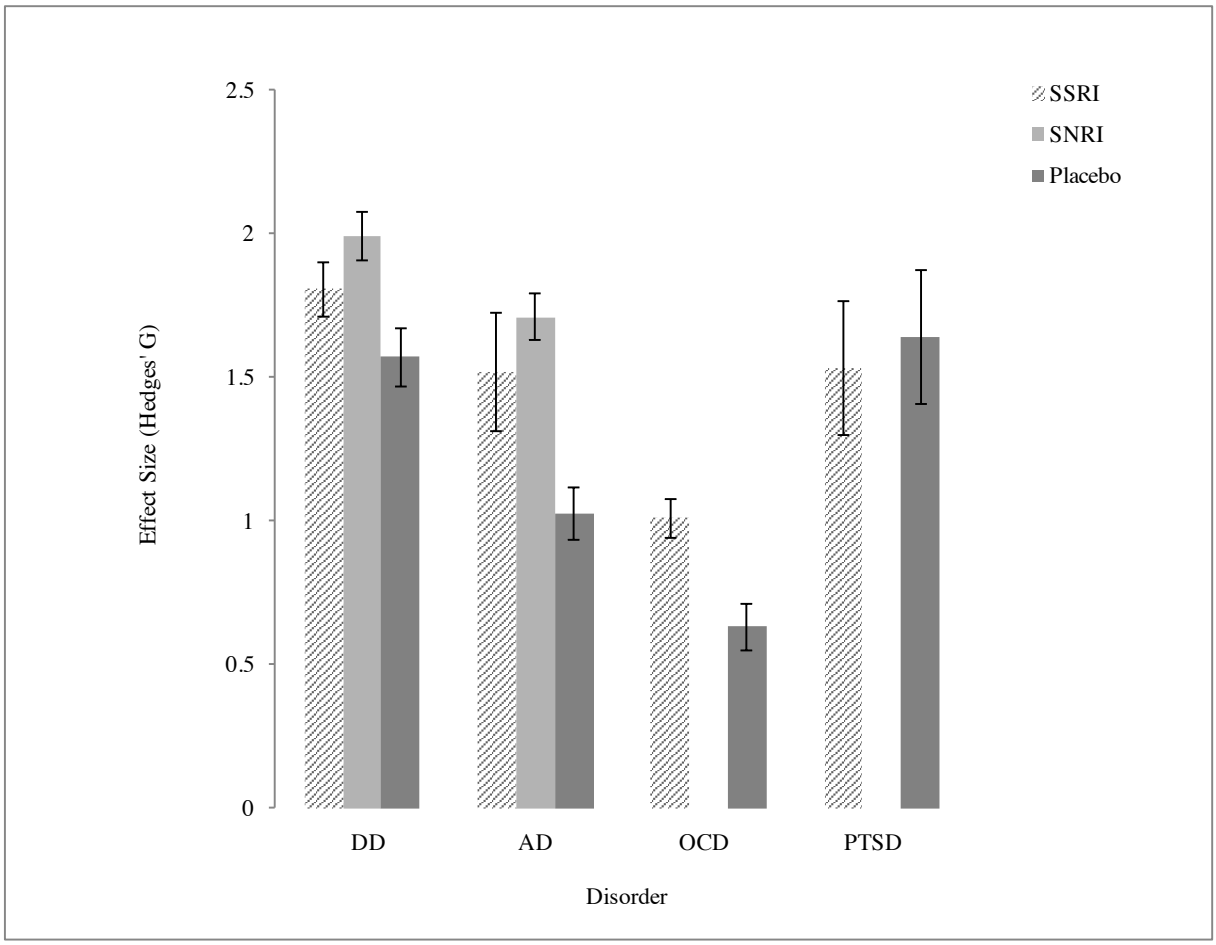
[a]Statistically different from corresponding DD values with *p*-value <.01.
[b]Statistically different from corresponding Placebo values with *p*-value <.05.

Figure 1. Forest Plot of Between Group Analyses (Stratified by Disorder)



Due to the low number of studies (N=1), PTSD is not included in the overall analysis. For the combined analysis across all disorders, see Result Section.
Abbreviations: DD, Depressive Disorders; AD, Anxiety Disorders; OCD, Obsessive Compulsive Disorder; PTSD, Posttraumatic Stress Disorder.
[a]One study reported two trials that were treated independently for analyses.

Figure 2. Drug and Placebo Effect Size by Disorder Category



SUPPLEMENTARY MATERIALS

Table of Contents

S1: Search Terms

S1.1. PubMed

"Depressive Disorder"[mesh] OR depression*[tiab] OR depressive[tiab] OR dysthymic[tiab] OR dysthymia*[tiab] OR "Anxiety Disorders"[Mesh] OR "Anxiety"[Mesh:noexp] OR anxiety[tiab] OR obsessive-compulsive[tiab] OR ocd[tiab] OR anankastic[tiab] OR phobic[tiab] OR phobia*[tiab] OR panic[tiab] OR stress disorder*[tiab] OR post traumatic stress[tiab] OR posttraumatic stress[tiab] OR post traumatic symptom*[tiab] OR posttraumatic symptom*[tiab] OR ptsd[tiab]

"Serotonin Uptake Inhibitors"[Mesh] OR "Serotonin Uptake Inhibitors"[pa] OR serotonin reuptake inhibitor*[tiab] OR serotonin uptake inhibitor*[tiab] OR SSRI*[tiab] OR SRI*[tiab] OR serotonin norepinephrine reuptake inhibitor*[tiab] OR serotonin norepinephrine uptake inhibitor*[tiab] OR SNRI* OR venlafaxin*[tiab] OR desvenlafaxin*[tiab] OR effexor[tiab] OR pristiq[tiab] OR milnacipran[tiab] OR levomilnacipran[tiab] OR fetzima[tiab] OR savella[tiab] OR duloxetin*[tiab] OR cymbalta[tiab] OR sibutramine[tiab] OR citalopram[tiab] OR celexa[tiab] OR escitalopram[tiab] OR lexapro[tiab] OR fluoxetin*[tiab] OR prozac[tiab] OR sarafem[tiab] OR symbyax[tiab] OR fluvoxamin*[tiab] OR luvox[tiab] OR paroxetin*[tiab] OR paxil[tiab] OR brisdelle[tiab] OR sertralin*[tiab] OR zoloft[tiab]

Child[MeSH Terms] OR Pediatrics[MeSH] OR child*[tiab] OR adolescen*[tiab] OR toddler*[tiab] OR teen*[tiab] OR boy[tiab] OR boys[tiab] OR girl*[tiab] OR pediatric[tiab] OR paediatric[tiab] OR puber*[tiab] OR pubescen*[tiab] OR prepubescen*[tiab] OR prepuberty*[tiab] OR schoolchild*[tiab] OR school age*[tiab] OR preschool*[tiab] OR kindergar*[tiab] OR primary school*[tiab] OR secondary school*[tiab] OR elementary school*[tiab] OR high school*[tiab] OR highschool*[tiab] OR youth*[tiab]

random*[tw] OR blind*[tiab] OR placebo*[tiab] OR trial[tiab] OR untreated[tiab] OR "not treated"[tiab] OR sham[tiab]

S1.2. Embase

'depression'/exp OR depression*:ab,ti OR depressive:ab,ti OR dysthymic:ab,ti OR dysthymia*:ab,ti OR 'anxiety disorder'/exp OR 'anxiety'/de OR anxiety:ab,ti OR obsessive-compulsive:ab,ti OR ocd:ab,ti OR anankastic:ab,ti OR phobic:ab,ti OR phobia*:ab,ti OR panic:ab,ti OR (stress NEXT/1 disorder*):ab,ti OR (('post traumatic' OR posttraumatic) NEXT/1 (stress OR symptom*)):ab,ti OR ptsd:ab,ti

'serotonin uptake inhibitor'/exp OR 'serotonin noradrenalin reuptake inhibitor'/exp OR (('serotonin reuptake' OR 'serotonin uptake' OR 'serotonin norepinephrine reuptake' OR 'serotonin norepinephrine uptake') NEXT/1 inhibitor*):ab,ti OR ssri*:ab,ti OR snri*:ab,ti OR venlafaxin*:ab,ti OR desvenlafaxin*:ab,ti OR effexor:ab,ti OR pristiq:ab,ti OR milnacipran:ab,ti OR levomilnacipran:ab,ti OR fetzima:ab,ti OR savella:ab,ti OR duloxetin*:ab,ti OR cymbalta:ab,ti OR sibutramine:ab,ti OR citalopram:ab,ti OR celexa:ab,ti OR escitalopram:ab,ti OR lexapro:ab,ti OR fluoxetin*:ab,ti OR prozac:ab,ti OR sarafem:ab,ti OR symbyax:ab,ti OR fluvoxamin*:ab,ti OR luvox:ab,ti OR paroxetin*:ab,ti OR paxil:ab,ti OR brisdelle:ab,ti OR sertralin*:ab,ti OR zoloft:ab,ti

'child'/exp AND 'pediatrics'/exp OR child*:ab,ti OR adolescen*:ab,ti OR toddler*:ab,ti OR teen*:ab,ti OR boy:ab,ti OR boys:ab,ti OR girl*:ab,ti OR pediatric:ab,ti OR paediatric:ab,ti OR puber*:ab,ti OR pubescen*:ab,ti OR prepubescen*:ab,ti OR prepuberty*:ab,ti OR schoolchild*:ab,ti OR (school NEXT/1 age*):ab,ti OR preschool*:ab,ti OR kindergar*:ab,ti OR ((primary OR secondary OR elementary OR high) NEXT/1 school*):ab,ti OR highschool*:ab,ti OR youth*:ab,ti

random*:ab,de,ti OR blind*:ab,ti OR placebo*:ab,ti OR trial:ab,ti OR untreated:ab,ti OR 'not treated':ab,ti OR sham:ab,ti

S1.3. PsycInfo

DE ("Major Depression" OR "Dysthymic Disorder" OR "Endogenous Depression" OR "Reactive Depression" OR "Recurrent Depression" OR "Treatment Resistant Depression" OR "Anxiety" OR "Acute Stress Disorder" OR "Generalized Anxiety Disorder" OR "Obsessive Compulsive Disorder" OR "Panic Disorder" OR "Phobias" OR "Posttraumatic Stress Disorder" OR "Panic Disorder" OR "Panic" OR "Panic Attack") OR TI (depression* OR depressive OR dysthymic OR dysthymia* OR anxiety OR "obsessive-compulsive" OR ocd OR anankastic OR phobic OR phobia* OR panic OR "stress disorder*" OR "post traumatic stress" OR "posttraumatic stress" OR "post traumatic symptom*" OR "posttraumatic symptom*" OR ptsd) OR AB (depression* OR depressive OR dysthymic OR dysthymia* OR anxiety OR "obsessive-compulsive" OR ocd OR anankastic OR phobic OR phobia* OR panic OR "stress disorder*" OR "post traumatic stress" OR "posttraumatic stress" OR "post traumatic symptom*" OR "posttraumatic symptom*" OR ptsd)

DE ("Serotonin Reuptake Inhibitors" OR "Chlorimipramine" OR "Citalopram" OR "Fluoxetine" OR "Fluvoxamine" OR "Paroxetine" OR "Zimeldine" OR "Serotonin Norepinephrine Reuptake Inhibitors" OR "Venlafaxine") OR TI ("serotonin reuptake inhibitor*" OR "serotonin uptake inhibitor*" OR SSRI* OR SRI* OR "serotonin norepinephrine reuptake inhibitor*" OR "serotonin norepinephrine uptake inhibitor*" OR SNRI* OR venlafaxin* OR desvenlafaxin* OR effexor OR pristiq OR milnacipran OR levomilnacipran OR fetzima OR savella OR duloxetin* OR cymbalta OR sibutramine OR citalopram OR celexa OR escitalopram OR lexapro OR fluoxetin* OR prozac OR sarafem OR symbyax OR fluvoxamin* OR luvox OR paroxetin* OR paxil OR brisdelle OR sertralin* OR zoloft) OR AB ("serotonin reuptake inhibitor*" OR "serotonin uptake inhibitor*" OR SSRI* OR SRI* OR "serotonin norepinephrine reuptake inhibitor*" OR "serotonin norepinephrine uptake inhibitor*" OR SNRI* OR venlafaxin* OR desvenlafaxin* OR effexor OR pristiq OR milnacipran OR levomilnacipran OR fetzima OR savella OR duloxetin* OR cymbalta OR sibutramine OR citalopram OR celexa OR escitalopram OR lexapro OR fluoxetin* OR prozac OR sarafem OR symbyax OR fluvoxamin* OR luvox OR paroxetin* OR paxil OR brisdelle OR sertralin* OR zoloft)

AG ("Childhood (birth-12 yrs)") OR TI (child* OR adolescen* OR toddler* OR teen* OR boy OR boys OR girl* OR pediatric OR paediatric OR puber* OR pubescen* OR prepubescen* OR prepuberty* OR schoolchild* OR "school age*" OR preschool* OR kindergar* OR "primary school*" OR "secondary school*" OR "elementary school*" OR "high school*" OR highschool* OR youth*) OR AB (child* OR adolescen* OR toddler* OR teen* OR boy OR boys OR girl* OR pediatric OR paediatric OR puber* OR pubescen* OR prepubescen* OR prepuberty* OR schoolchild* OR "school age*" OR preschool* OR kindergar* OR "primary school*" OR "secondary school*" OR "elementary school*" OR "high school*" OR highschool* OR youth*)

DE (random*) OR TI (random* OR placebo* OR trial OR untreated OR sham) OR AB (random* OR placebo* OR trial OR untreated OR sham)
Note: "not treated" is handled as a stop word so all records with treated are retrieved.

S1.4. Cochrane Central

TI ("serotonin reuptake inhibitor*" OR "serotonin uptake inhibitor*" OR SSRI* OR SRI* OR "serotonin norepinephrine reuptake inhibitor*" OR "serotonin norepinephrine uptake inhibitor*" OR SNRI* OR venlafaxin* OR desvenlafaxin* OR effexor OR pristiq OR milnacipran OR levomilnacipran OR fetzima OR savella OR duloxetin* OR cymbalta OR sibutramine OR citalopram OR celexa OR escitalopram OR lexapro OR fluoxetin* OR prozac OR sarafem OR symbyax OR fluvoxamin* OR luvox OR paroxetin* OR paxil OR brisdelle OR sertralin* OR zoloft) OR AB ("serotonin reuptake inhibitor*" OR "serotonin uptake inhibitor*" OR SSRI* OR SRI* OR "serotonin norepinephrine reuptake inhibitor*" OR "serotonin norepinephrine uptake inhibitor*" OR SNRI* OR venlafaxin* OR desvenlafaxin* OR effexor OR pristiq OR milnacipran OR levomilnacipran OR fetzima OR savella OR duloxetin* OR cymbalta OR sibutramine OR citalopram OR celexa OR escitalopram OR lexapro OR fluoxetin* OR prozac OR sarafem OR symbyax OR fluvoxamin* OR luvox OR paroxetin* OR paxil OR brisdelle OR sertralin* OR zoloft)

TI (depression* OR depressive OR dysthymic OR dysthymia* OR anxiety OR "obsessive-compulsive" OR ocd OR anankastic OR phobic OR phobia* OR panic OR "stress disorder*" OR "post traumatic stress" OR "posttraumatic stress" OR "post traumatic symptom*" OR "posttraumatic symptom*" OR ptsd) OR AB (depression* OR depressive OR dysthymic OR dysthymia* OR anxiety OR "obsessive-compulsive" OR ocd OR anankastic OR phobic OR phobia* OR panic OR "stress disorder*" OR "post traumatic stress" OR "posttraumatic stress" OR "post traumatic symptom*" OR "posttraumatic symptom*" OR ptsd)

TI (child* OR adolescen* OR toddler* OR teen* OR boy OR boys OR girl* OR pediatric OR paediatric OR puber* OR pubescen* OR prepubescen* OR prepuberty* OR schoolchild* OR "school age*" OR preschool* OR kindergar* OR "primary school*" OR "secondary school*" OR "elementary school*" OR "high school*" OR highschool* OR youth*) OR AB (child* OR adolescen* OR toddler* OR teen* OR boy OR boys OR girl* OR pediatric OR paediatric OR puber* OR pubescen* OR prepubescen* OR prepuberty* OR schoolchild* OR "school age*" OR preschool* OR kindergar* OR "primary school*" OR "secondary school*" OR "elementary school*" OR "high school*" OR highschool* OR youth*)

S1.5. Web of Science

TS=("serotonin reuptake inhibitor*" OR "serotonin uptake inhibitor*" OR SSRI* OR SRI* OR "serotonin norepinephrine reuptake inhibitor*" OR "serotonin norepinephrine uptake inhibitor*" OR SNRI* OR venlafaxin* OR desvenlafaxin* OR effexor OR pristiq OR milnacipran OR levomilnacipran OR fetzima OR savella OR duloxetin* OR cymbalta OR sibutramine OR citalopram OR celexa OR escitalopram OR lexapro OR fluoxetin* OR prozac OR sarafem OR symbyax OR fluvoxamin* OR luvox OR paroxetin* OR paxil OR brisdelle OR sertralin* OR zoloft)

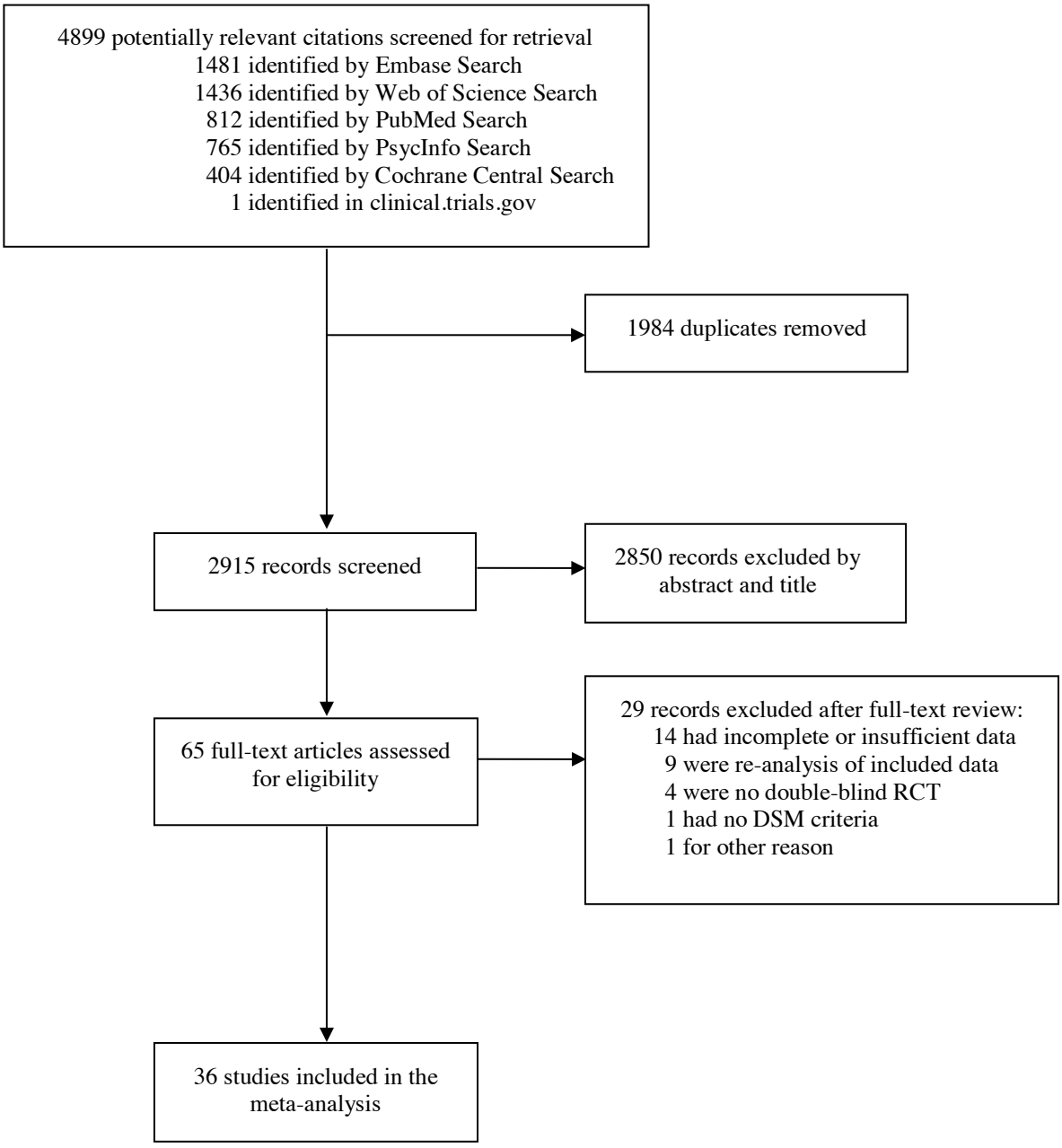
TS=(depression* OR depressive OR dysthymic OR dysthymia* OR anxiety OR "obsessive-compulsive" OR ocd OR anankastic OR phobic OR phobia* OR panic OR "stress disorder*" OR "post traumatic stress" OR "posttraumatic stress" OR "post traumatic symptom*" OR "posttraumatic symptom*" OR ptsd)

TS=(child* OR adolescen* OR toddler* OR teen* OR boy OR boys OR girl* OR pediatric OR paediatric OR puber* OR pubescen* OR prepubescen* OR prepuberty* OR schoolchild* OR "school age*" OR preschool* OR kindergar* OR "primary school*" OR "secondary school*" OR "elementary school*" OR "high school*" OR highschool* OR youth*)

TS=(random* OR placebo* OR trial OR untreated OR sham)

S2: Study Selection

sFigure 1. Flow Chart



S3: Details on Heterogeneity and Publication Bias

sTable 1. Heterogeneity

Drug	Treatment Arms	Hedges g	95% CI	SE	p-Value	Q-value	p-Value	I^2	Tau^2
Stratified Between Drug									
Citalopram	2	0.18	-0.18 - 0.54	0.18	.33	N/A[a]	N/A[a]	N/A[a]	N/A[a]
Escitalopram	2	0.18	0.01 - 0.34	0.08	.03	N/A[a]	N/A[a]	N/A[a]	N/A[a]
Fluoxetine	13	0.38	0.26 - 0.51	0.06	<0.001	13.17	.36	8.90	0.01
Fluvoxamine	2	0.68	-0.05 - 1.41	0.37	.07	N/A[a]	N/A[a]	N/A[a]	N/A[a]
Paroxetine	6	0.31	0.07 - 0.54	0.12	.01	19.83	.001	74.78	0.06
Sertraline	8	0.31	0.15 - 0.47	0.08	<.001	9.38	.23	25.37	0.01
Venlafaxine	4	0.31	0.18 - 0.44	0.07	<.001	2.66	.45	0.00	0.00
Duloxetine	4	0.24	0.06 - 0.46	0.16	.04	5.91	.12	49.20	0.03
Stratified Within Drug									
Citalopram	2	1.78	1.56 - 2.04	0.12	<.001	N/A[a]	N/A[a]	N/A[a]	N/A[a]
Escitalopram	2	1.68	1.48 - 1.87	0.10	<.001	N/A[a]	N/A[a]	N/A[a]	N/A[a]
Fluoxetine	13	1.73	1.32 - 2.13	0.21	<.001	100.36	<.001	88.05	0.45
Fluvoxamine	2	1.22	0.41 - 2.02	0.41	.003	N/A[a]	N/A[a]	N/A[a]	N/A[a]
Paroxetine	6	1.46	1.31 - 1.61	0.08	<.001	7.19	.21	30.45	0.01
Sertraline	8	1.38	1.02-1.73	0.18	<.001	37.54	<.001	81.35	0.18

sTable 1. Heterogeneity (cont.)

Drug	Treatment Arms	Hedges g	95% CI	SE	p-Value	Q-value	p-Value	I^2	Tau^2
Venlafaxine	4	1.77	1.59-1.95	0.09	<.001	3.71	.29	19.15	0.01
Duloxetine	4	1.95	1.73-2.18	0.11	<.001	5.17	.16	41.97	0.02

[a]Heterogeneity was not assessed due to the low number of studies.

S3.1. Stratified by Disorder

OCD: The eight studies exhibited no heterogeneity (Q=2.28, p=.07, I^2=0.00 τ^2=0.00). There was no evidence of publication bias in a funnel plot. Neither the Begg's test nor the Egger's test yielded a significant result. The fail-safe N indicated that 43 unpublished null studies would be needed to remove the significance from the findings. The trim-and-fill method lead to a very slight adjustment of Hedges's g (g=0.41, CI=0.26-0.55).

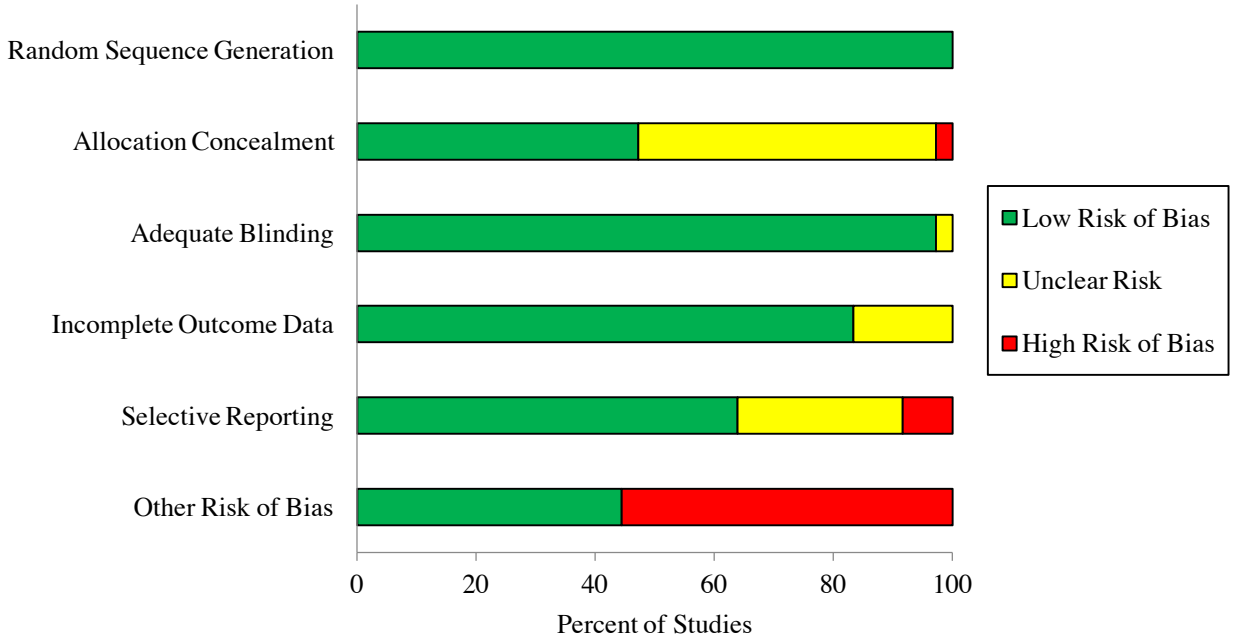
DD: The seventeen studies exhibited moderate heterogeneity (Q=20.28, p=.38, I^2=6.31, τ^2=0.00). There was no evidence of publication bias in a funnel plot. Neither the Begg's test nor the Egger's test yielded a significant result. The fail-safe N indicated that 165 unpublished null studies would be needed to remove the significance from the findings. The trim-and-fill method did not lead to an adjustment of Hedges's g.

AD: The ten studies exhibited moderate heterogeneity (Q=22.93, p=.01, I^2=56.40, τ^2=0.04). There was some evidence of publication bias in a funnel plot. Both the Begg's test and the Egger's test yielded a non-significant result (2-tailed p > .05). The fail-safe N indicated that 308 unpublished null studies would be needed to remove the significance from the findings. The trim-and-fill method lead to a slight adjustment of the standard mean difference (g=0.53, CI=0.36-0.70).

Across all studies: The combined analysis yielded low to moderate heterogeneity (Q=76.62, p<.001, I^2=47.79, τ^2=0.03).

S3.2. Risk of Bias Assessment

sFigure 2



Note: The large amount of high risk in the "other risk of bias" category was mainly due to per protocol analysis rather than intent-to-treat analysis.

S4. Side Effects

sTable 2. Individual Side Effects by Drug and Placebo

	Drugs (No.=2542)		Placebo (No.=2294)		
Side Effect	**No. Reported**	**Percent**	**No. Reported**	**Percent**	**p-Value**
Headache	321	12.63	266	11.60	.58
Nausea	205	8.06	136	5.93	.16
Insomnia	203	7.99	193	4.05	.04[a]
Abdominal Pain	187	7.36	139	6.06	.38
Agitation	119	4.68	20	0.87	.11
Diarrhea	111	4.37	49	2.14	.03[a]
Pharyngitis	103	4.05	96	4.18	.55
Vomiting	99	3.89	33	1.44	.18
Asthenia	87	3.42	47	2.05	.21
Respiratory Illness	84	3.30	71	3.10	.42
Hyperkinesia	79	3.11	25	1.09	.03[a]
Rhinitis	73	2.87	52	2.27	.21
Decreased Appetite	63	2.48	19	0.83	.002[a]
Anorexia	62	2.44	13	0.57	.08
Fatigue	57	2.24	24	1.05	.05[a]
Somnolence	56	2.20	25	1.09	.19

Individual side effects are reported across all drugs (SSRI and SNRI) due to insufficient data in the studies of SNRIs. Percent indicates total percent of all reported side effects that were the specific side effect in question.
[a]Significant.

S5. Moderator Analysis

S5.1. Methods and Results for the Univariate Analyses – Continuous Variables

Methods: Continuous variables were analyzed with a meta-regression analysis using method-of-moments analyses in a random-effects model. The Z-statistic was used to test the significance of the slope. As various scales were used to assess baseline severity, we standardized the baseline and outcome values by dividing the mean values by the SD.

Results: The relationship between effect size and publication year was significant in the combined analyses (Z=-2.36, p=.02), as well as in the DD subgroup analyses (Z=-2.26, p=.02), with recently published studies yielding smaller antidepressant-placebo differences. Further, the relationship between effect size and illness duration was significant in the combined analyses (Z=2.89, p=.004), indicating that children with a longer duration of illness exhibit greater response to antidepressants compared to placebo. Finally, number of sites was found to be significantly correlated to effect size in the combined analyses (Z=-2.98, p=.003), as well as in the DD subgroup analyses (Z=-2.16, p<.03), and the OCD subgroup analyses (Z=-2.16, p=.03), with number of study sites negatively associated with magnitude of differences between antidepressants and placebo.

S5.2. Methods and Results for the Univariate Analyses – Categorical Variables

Methods: Categorical variables were analyzed using a mixed-effects model.

Results: The relationship between effect size and primary funding source was significant in the combined analyses (p = .02), as well as in the DD subgroup analyses (p = .02). In both cases, studies that were funded by industry yielded significantly smaller effect sizes than those that reported public sources of funding only (e.g., NIMH).

S5.3. Methods for the moderator analysis: Multivariate metaregression analysis

Given the relatively large number of moderator analyses, we decided to conduct a multivariate meta-regression in order to adjust for multiple comparisons. Effect sizes (i.e., dependent variable) were weighted by the sample size divided by s2 (i.e., n/var; (1)). Multivariate regression analyses were conducted in SPSS (Version 21.0.0.2).

This approach is in line with the methods adopted by Cuijpers (2-5). Besides adjusting for multiple comparisons, the model indicates the significance of each potential moderator while controlling for the others. To avoid collinearity among the predictors of the regression model, we first tested whether high correlations (i.e., correlations higher than 0.60) were found among the moderators that could be entered into the model. Three variables were found to have correlations higher than 0.60: the funding source correlated high with the number of sites (r = .698), treatment duration correlated high with illness duration (r = 0.62), and comorbidity correlated high with the number of sites (r = -0.75). We decided to use the number of sites (not funding source or comorbidity) and treatment duration (not illness duration) as predictors in the model. All remaining variables (i.e., treatment duration, publication year, baseline severity number of sites, age of onset, placebo lead-in, and study location) were included as predictors in the model.

sTable 3. Continuous Moderator Analyses: treatment duration, publication year, baseline severity, number of sites, illness duration, and age of onset.

Moderator	Point Estimate	Standard Error	95 % CI	Z-Value	p-Value
Overall					
Treatment Duration	0.01	0.02	-0.02 – 0.04	0.57	.57
Publication Year	-0.02	0.01	-0.03 – -0.00	-2.36	.02
Baseline Severity	0.00	0.02	-0.03 – 0.03	0.20	.84
Number of Sites	-0.01	0.00	-0.01 – -0.00	-2.10	.003
Illness Duration	0.01	0.00	0.00 – 0.01	2.89	.004
Age of Onset	-0.04	0.03	-0.10 – 0.03	-1.11	.27
Depressive Disorder					
Treatment Duration	-0.03	0.02	-0.07 – 0.02	-1.12	.26
Publication Year	-0.02	0.01	-0.03 – -0.00	-2.26	.02
Baseline Severity	0.02	0.02	-0.01 – 0.05	1.21	.23
Number of Sites	-0.00	0.00	-0.01 – -0.00	-2.16	.03
Illness Duration	0.00	0.01	-0.01 – 0.01	-0.00	1.00
Age of Onset	-0.01	0.04	-0.10 – 0.07	-0.31	.76
Obsessive-Compulsive Disorder					
Treatment Duration	-0.02	0.04	-0.09 – 0.05	-0.47	.64
Publication Year	-0.01	0.02	-0.06 – 0.02	-0.88	.38
Baseline Severity	-0.05	0.15	-0.35 – 0.25	-0.33	.74
Number of Sites	-0.00	0.00	-0.01 – -0.00	-2.16	.03
Illness Duration	0.01	0.03	-0.05 – 0.07	0.45	.65
Age of Onset	*N/A*[a]				
Anxiety Disorder					
Treatment Duration	-0.02	0.04	-0.09 – 0.05	-0.55	.58
Publication Year	-0.04	0.02	-0.08 – -0.00	-1.90	.06
Baseline Severity	-0.03	0.03	-0.08 – 0.02	-1.18	.24
Number of Sites	-0.01	0.00	-0.02 – 0.00	-1.84	.07
Illness Duration	0.01	0.01	-0.00 – 0.02	1.62	.11
Age of Onset	*N/A*[b]				

[a]Only 1 Study
[b]No Studies

sTable 4. Categorical Moderator Analyses: placebo lead-in, comorbidity, study location, primary funding source

Moderator	Number of included studies	Hedges g	95% CI	Q-value	I^2	p-Value
Overall						
Placebo lead-in				0.23		.63
No	28	0.36	0.24 - 0.48		64.30	
Yes	13	0.29	0.21 - 0.38		8.66	
Comorbidity				2.47		.12
No	6	0.24	0.04 - 0.43		29.90	
Yes	28	0.41	0.31 - 0.51		60.61	
Study location				1.94		.16
US only	27	0.38	0.28 - 0.48		50.10	
Not US only	14	0.26	0.13 - 0.39		62.97	
Primary funding source				5.42		.02[a]
Industry only	27	0.26	0.19 - 0.33		37.91	
Public only	11	0.48	0.31 - 0.64		2.97	
Depressive Disorder						
Placebo lead-in				2.71		.10
No	11	0.15	0.06 - 0.24		0.00	
Yes	9	0.26	0.16 - 0.35		23.51	
Comorbidity				1.98		.16
No	3	0.12	-0.04 - 0.28		0.00	
Yes	11	0.25	0.16 - 0.35		27.54	
Study location				2.61		.11
US only	10	0.25	0.16 - 0.35		15.15	
Not US only	10	0.15	0.05 - 0.24		0.00	
Primary funding source				5.64		.02[a]
Industry only	18	0.18	0.11 - 0.25		0.00	
Public only	2	0.46	0.24 - 0.68		0.00	
Obsessive-Compulsive Disorder						
Placebo lead-in				0.05		.83
No	7	0.41	0.22 - 0.60		0.00	
Yes	2	0.38	0.15 - 0.60		0.00	
Comorbidity	*N/A*[b]					
Study location	*N/A*[b]					
Primary funding source				0.07		.79
Industry only	4	0.41	0.25 - 0.58		0.00	
Public only	4	0.36	-0.03 - 0.74		0.00	
Anxiety Disorder						
Placebo lead-in				1.69		.19
No	9	0.69	0.47 - 0.91		71.87	
Yes	2	0.37	-0.06 - 0.80		4.13	
Comorbidity				3.32		0.07
No	3	0.37	0.06 - 0.69		0.00	
Yes	8	0.74	0.51 - 0.97		71.38	
Study location				0.00		0.96
US only	7	0.62	0.37 - 0.88		78.68	
Not US only	4	0.63	0.28 - 0.99		9.85	
Primary funding source				0.31		0.58
Industry only	4	0.47	0.24 - 0.70		55.44	
Public only	5	0.57	0.28 - 0.87		48.85	

[a]Significant.
[b]Not enough variance.

sTable 5. Multivariate Metaregression Analyses

	B	95% CI	p
Placebo lead-in	0.09	-6377.27 – 7630.54	.83
Study location	0.11	-6719.62 – 8175.84	.89
Treatment Duration	0.41	-818.31 – 1766.67	.40
Publication Year	-0.38	-830.60 – 437.03	.48
Age of Onset	0.01	-2421.64 – 2444.84	.99
Number of Sites	0.39	-113.43 – 223.59	.45
Baseline Severity	-0.00	-1179.76 – 1177.66	1.00

S6. PRISMA Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	p. 1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	p. 2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	p.3-4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	p.4
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	p.6
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	p.5
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	p.4 sFig 1
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	S1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	p.5
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	p.5-6

Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	p.5
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	p.5
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	p.6
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	p.6

References

1. Lipsey MW, Wilson DB. Practical meta-analysis: Sage publications Thousand Oaks, CA; 2001.
2. Cuijpers P, van Straten A, Warmerdam L, Smits N. Characteristics of effective psychological treatment of depression: a metaregression analysis. Psychotherapy Research. 2008;18(2):225-36.
3. Cuijpers P, van Straten A, Bohlmeijer E, Hollon SD, Andersson G. The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. Psychological Medicine. 2010;40:211-23.
4. Cuijpers P, Huibers M, Ebert DD, Koole SL. How much psychotherapy is needed to treat depression? A metaregression analysis. Journal of Affective Disorders. 2013;149:1-13.
5. Cuijpers P, Sijbrandij M, Koole SL, Andersson G, Beekman AT, Reynolds CFI. The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: a meta-analysis of direct comparisons. World Psychiatry. 2013;12:137-48.

Appendix B

Study II:

Locher, C., Kossowsky, J., Gaab, J., Kirsch, I., Bain, P., & Krummenacher, P. (2015). Moderation of antidepressant and placebo outcomes by baseline severity in late-life depression: A systematic review and meta-analysis. *Journal of Affective Disorders, 181*, 50–60. doi:10.1016/j.jad.2015.03.062

Journal of Affective Disorders 181 (2015) 50–60

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research report

Moderation of antidepressant and placebo outcomes by baseline severity in late-life depression: A systematic review and meta-analysis

Cosima Locher [a,*,1], Joe Kossowsky [a,b,c,1], Jens Gaab [a], Irving Kirsch [c], Paul Bain [d], Peter Krummenacher [a,e]

[a] Department of Clinical Psychology and Psychotherapy, University of Basel, Basel, Switzerland
[b] Department of Anesthesiology, Perioperative and Pain Medicine, Boston Children's Hospital, Harvard Medical School, Boston, USA
[c] Program in Placebo Studies and the Therapeutic Encounter, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, USA
[d] Countway Library of Medicine, Harvard Medical School, Boston, USA
[e] Collegium Helveticum, University of Zurich and ETH Zurich, Zurich, Switzerland

ARTICLE INFO

Article history:
Received 12 February 2015
Received in revised form
31 March 2015
Accepted 31 March 2015
Available online 15 April 2015

Keywords:
Meta-analysis
Antidepressants
Placebo
Depression
Elderly
Baseline Severity

ABSTRACT

Background: Baseline severity is a crucial moderator of trial outcomes in adult depression, with the advantage of antidepressants over placebo increasing as severity increases. However, this relationship has not been examined in late-life depression.

Methods: PubMed, Embase, Web of Science, PsycINFO, and Cochrane were searched for studies published through September 2014. Randomized, acute phase, and double-blind studies comparing an antidepressant group with a placebo group in depressed elderly patients were included.

Results: Nineteen studies met all inclusion criteria. Within-group effect sizes revealed significant improvement in antidepressant groups ($g=1.35$, $p<.000$), as well as in placebo groups ($g=.96$, $p<.000$). Change in depressive symptoms assessed by Hamilton Depression Rating Scale (HDRS) was moderated by baseline severity in antidepressant groups ($Z=2.67$, $p=.008$) and placebo groups ($Z=4.46$, $p<.000$). However, this would be expected as a result of regression toward the mean, and mean differences between groups did not increase ($r=.19$, $p=.469$) as a function of baseline severity.

Limitations: Limited to published data and information was only analyzed at the level of treatment groups.

Conclusion: Baseline severity was not associated with an antidepressant–placebo difference and placebo responses are large in the treatment of depressed elderly people. We propose a stepwise approach, i.e., to initially offer elderly depressed patients psychosocial interventions and only consider antidepressants if patients do not respond.

1. Introduction

Although the placebo effect and its moderators have been examined extensively in adult populations with major depressive disorder (MDD) (Brunoni et al., 2009; Kirsch et al., 2008), comparable studies for late-life depression are scarce. There is no agreement upon definition of late-life depression; the term may be used to refer to patients with symptoms that fall on a continuum from subthreshold to clinically significant, and a minimum age criterion in the range 55–65 years (Rodda et al., 2011). MDD is the most common psychiatric disorder in elderly people, showing a point prevalence of 4.6–9.3% (Meeks et al., 2011). In addition, subclinical symptoms such as minor depression and dysthymia are more common in old age, with a point prevalence of 10% (Pinquart et al., 2006). All of these forms of depression have been found to have a negative influence on the quality of life (Nelson et al., 2013). Late-life depressive disorders also increase disability (Nelson et al., 2013), are associated with poorer outcomes in clinically significant illnesses (Jiang et al., 2001), and a higher suicide rate (Conwell et al., 2002).

With regard to effective treatment of depression in elderly patients, practice guidelines identifies both antidepressants and psychotherapeutic interventions as a first line treatment for MDD, especially for mild to moderate depression, and a combination thereof or antidepressants alone for severe depression (American Psychiatric Association, 2010). Given that psychotherapy and pharmacotherapy did not show strong differences in effect sizes in elderly patients in a direct comparison (Pinquart et al., 2006), the authors recommend that treatment choice should be based on other criteria, such as contraindications, treatment access, or

* Corresponding author.
E-mail address: cosima.locher@unibas.ch (C. Locher).
[1] Contributed equally to the article and should both be considered first authors.

http://dx.doi.org/10.1016/j.jad.2015.03.062
0165-0327/

patient preferences. For neuropharmacological practice, selective serotonin reuptake inhibitors (SSRIs) and other second-generation antidepressants medications should be considered over monoamine oxidase inhibitors or tricyclic antidepressants (American Psychiatric Association, 2010; Rodda et al., 2011). Moreover, antidepressant use in elderly people with depression increased over the last years, mainly due to a growing SSRI-use (Sonnenberg et al., 2008). SSRIs have been shown to be superior to a placebo pill in controlled clinical trials and meta-analyses investigating late-life depression (Kok et al., 2012; Mittmann et al., 1997; Nelson et al., 2008). However, overall drug effects in elderly patients with symptoms of depression are only modest, with an odds ratio (OR)=1.40 (95% CI: 1.24–1.57) for response (i.e., ≥50% improvement from baseline on mood scales), and OR=1.27 (95% CI: 1.12–1.44) for remission (i.e., no longer meeting diagnostic criteria) versus placebo in a meta-analysis of 10 trials (Nelson et al., 2008).

With regard to possible moderators of pharmacological and placebo outcomes in depression, mixed-age studies have repeatedly shown that the mean differences between groups treated with antidepressant medication and placebo become larger as baseline severity increases (Fournier et al., 2010; Khan et al., 2002; Kirsch et al., 2008). It is unclear whether the increasing benefits, as severity increases, of drug treatment over placebo treatment are due to a decrease in the response to placebo treatment or an increase in the response to pharmacological intervention. The data reported by Kirsch et al. (2008) indicated that the increased benefit of drug treatment for severely depressed patients is related to a decrease in responsiveness to placebos, with no change in responsiveness to the drug. However, two meta-analyses have shown that initial severity predicted symptom improvement in adult patients who took antidepressant medication (Fournier et al., 2010; Khan et al., 2002). In the Khan et al. (2002) analysis, improvement as a function of baseline severity increased in drug groups but decreased in placebo groups. In Fournier et al. (2010), improvement as a function of severity increased significantly in both drug and placebo groups (as would be predicted by regression toward the mean), but the increase was significantly larger in the drug group. It should be noted that a re-analysis of the Kirsch et al. (2008) data set, which controlled for the effect of structural coupling (this occurs when baseline values and change score are coupled algebraically, thus possibly leading to an inflated association between the variables; Tu et al., 2004) concluded that baseline severity did not influence treatment outcome (Fountoulakis et al., 2013).

Studies looking at predictors of treatment outcome in elderly patients with depression are limited and most studies in this field do not focus on baseline depression severity. To date, symptom severity at baseline has not been shown to be a moderator of outcome in depressed elderly people. A meta-analysis by Gibbons et al. (2012) found that in a geriatric subgroup, baseline severity was not related to a positive treatment outcome for fluoxetine compared with placebo. Another meta-analysis found an association between initial severity and drug over placebo efficacy in elderly patients who had suffered from depression for at least 10 years, but not in the majority of patients, who had a shorter disease history (Nelson et al., 2013). However, there are several limitations to the reported meta-analyses. First, they rely on a limited number of studies, thus Gibbons et al. (2012) included 4 geriatric studies, whereas Nelson et al. (2013) included 10 trials of second-generation antidepressants in patients with late-life depression. Second, the authors included only a restricted range of baseline severity scores as they focused on MDD. However, only a minority of significantly depressed elderly patients fulfill the diagnostic criteria for depression, yet the rate of sub-threshold late-life depression rises with age and is responsible for comparable disability and distress (Pinquart et al., 2006).

Consequently, to assess treatment effects in late-life depression, a meta-analysis including a broader range of studies and taking minor depression and dysthymia into account is of a high relevance. With this background, we undertook a systematic review and meta-analysis to test the assumption that mean differences between antidepressant and placebo interventions become larger as baseline severity increases in a geriatric population.

2. Method

2.1. Search strategy and eligibility criteria

We performed searches in Cochrane, Embase, PsycINFO, PubMed, and Web of Science on studies published through September 30, 2014. Search terms were adapted to the electronic bibliographic databases and consisted of keyword combinations based on the inclusion criteria (for details see Appendix). In addition to the systematic search, the references of all included articles were reviewed.

We included peer-reviewed randomized, double-blind, placebo-controlled clinical trials reported in English or German comparing depressed elderly individuals in a placebo group with depressed elderly individuals in an intervention group receiving second-generation antidepressants (i.e., SSRI's and other novel atypical antidepressants). We classified antidepressants according to the Anatomical Therapeutic Chemical (ATC)[2] classification system of the World Health Organization as an internationally accepted standard of defining whether a drug counts as an antidepressant or not. Moreover, we grouped antidepressants as SSRI's or other novel atypical antidepressants in accordance with other meta-analyses (Anderson, 2000; Kok et al., 2012). The minimum age criterion was set at a mean or median age of 55 years, or described as elderly, geriatric or older adults.

Outcomes had to be reported as mean change in depressive symptoms on a continuous mood scale, such as the Hamilton Depression Rating Scale (HDRS; Hamilton, 1967) or Montgomery-Asberg Depression Rating Scale (MADRS; Montgomery and Asberg, 1979). We included only continuous outcome data, since dichotomizing continuous scores into categorical outcome data leads to a loss of information, reduces power and creates an artificial boundary (Altman and Royston, 2006; Moncrieff and Kirsch, 2005). Pre- and post-intervention data had to be available. We included studies investigating patients with MDD or subclinical depressive symptoms (i.e., minor depressive disorder or dysthymia) according to explicit, reliable, and reproducible diagnostic criteria, which were based on DSM-III, DSM-III-R, DSM-IV or DSM-IV-TR. However, we included one study where diagnostic criteria were not explicitly stated (Gerner et al., 1980). Medical comorbidities such as diabetes (Paile-Hyvärinen, Wahlbeck, & Eriksson, 2007), diagnosis of heart failure (Fraguas et al., 2010), or age-related macular degeneration (Brody et al., 2011) were not grounds for exclusion, as they are not neurological disorders.

Studies in which patients had depression following cerebrovascular disease (i.e., vascular depression and post-stroke depression), a cognitive impairment (i.e., moderate to severe dementia), or Parkinson's disease were excluded. We excluded studies investigating these neurological disorders because executive dysfunction and associated learning impairments in older patients with depression have been associated with a lower probability of antidepressant and placebo response (Alexopoulos et al., 2005; Benedetti et al., 2006a, 2006b). However, we included patients with mild cognitive impairment according to the Mini Mental-State Examination (MMSE > 19; Folstein et al., 1975) and two papers, which had not explicitly

[2] Available at: www.whocc.no. Accessed January 27, 2015.

excluded patients with dementia (Fraguas et al., 2010; Gerner et al., 1980; Paile-Hyvärinen et al., 2007). Case reports, comments, letters and reviews were excluded as well. Using these criteria, 19 studies were identified and included in our analysis.

2.2. Data extraction and study outcomes

Two independent investigators (C.L. and J.K.) screened citations from the former databases and reviews. Identified abstracts were reviewed twice for eligibility by two independent investigators (C.L. and J.K.). Inconsistencies were resolved in consensus between the authors and confirmed with a third reviewer (P.K.) when necessary. The reported variables were depression diagnosis (i.e., MDD, minor depressive disorder, dysthymia), diagnostic criteria (i.e., DSM-III, DSM-IV), minimum age at entry, illness severity score at entry, MMSE score at entry, trial duration, type of antidepressant treatment (i.e., SSRIs or novel atypical antidepressants), dropout rate, and sociodemographic characteristics (see Table 1).

The primary outcome was mean change in depressive symptoms, reported either on a continuous self-rated mood scale (i.e., Beck Depression Inventory (BDI; Beck et al., 1961), Geriatric Depression Scale (GDS; Yesavage et al., 1983), Hospital Anxiety and Depression Scale (HADS; Zigmond and Snaith, 1983), 20-item Hopkins Symptom Checklist Depression Scale (HSCL-D-20; Derogatis et al., 1974), or on a clinical rating of depression (i.e., HDRS, MADRS). With respect to baseline depression severity, we first converted the different mood scales (i.e., BDI, GDS, HADS, HSCL-D-20, HDRS, MADRS) to a standardized scale (range 0–100), using the largest point of each mood scale as 100%. Where a study used more than one mood scale, all mood scales were converted and used independently in analyses. We conducted separate subgroup analyses for each mood scale. We pro-rated mean HDRS scores onto a 17-item scale (HDRS17) where studies had used other versions of the HDRS. For example, a 24-item HDRS score would be pro-rated as: HDRS17 = 17 × HDRS24/24 (Heo et al., 2007).

Studies differed in the effort made to minimize placebo responses. In two studies, subjects considered placebo responders in the single-blind, placebo lead-in phase (improvement of at least 20% on the outcome scale) were excluded from protocol (Fraguas et al., 2010; Tollefson et al., 1995).

2.3. Data analysis

Comprehensive Meta-Analysis V2 (CMA)[3] was used for calculations and analyses. Two main analyses were performed. First, effect sizes were calculated for the continuous outcome (i.e., mean change in depressive symptoms). Differences in mean change scores between groups were evaluated with differences in means (Hedges's *g*). Moreover, we calculated within-group pre-post effect sizes (Hedges's *g*). They inform about whether a small difference between groups is explained by a small change in either group or a meaningful change in both groups over treatment time. Second, moderator analyses were conducted to investigate the relationship between baseline depression severity and subsequent continuous outcomes of the trial.

For the analyses we chose to use random-effects models rather than fixed-effects models. A fixed-effects model assumes that there is one true effect size for all studies and any variations are due to sampling error, whereas a random-effects model assumes that variations in effect sizes for the samples are a combination of sampling error and true variance in effect size (Borenstein et al., 2011). Random effect sizes were preferable for this meta-analysis as the studies we included were heterogeneous and we had relatively small numbers of studies for the sub-analyses.

To assess the moderating effect of baseline depression severity on outcome measure (i.e., mean change in depressive symptoms), we conducted within-group and between-group comparisons. To test the moderating effect of baseline depression severity on outcome measures within each group, we performed meta-regression using method-of-moments analyses (random-effects model). The *Z* statistic was used to test the significance of the slope. In the case that data conform to the null hypothesis, *Z* has a normal distribution. A significant *Z* would indicate that the slope is probably not zero, and hence that baseline depression severity moderates the outcomes within the group (Borenstein et al., 2011). Further, we analyzed the mean difference scores to test the hypothesis that between-groups mean differences increase as a function of baseline depression severity. We calculated the overall baseline severity for each study (i.e., mean of antidepressant and placebo baseline severity, weighted by number of participants) and the antidepressant-placebo difference in improvement. Pearson's correlation between those two variables was calculated.

To assess heterogeneity between studies, we calculated the *Q* statistic. A statistically significant *Q* indicates a heterogeneous distribution, meaning that systematic differences between studies are present and it rejects the null-hypothesis that all the variation in effects is due to random error. Similarly, the higher the *Q* value, the more variation in the studies can be explained by a true variance of effects between studies (Cochran, 1954). In addition, the I^2 statistic was used to quantify inconsistency. It measures the proportion of observed variance across studies, that is a result of real heterogeneity rather than chance. An I^2 value of 0% indicates no heterogeneity, a value of 25% is classified as low, 50% as moderate and 75% as high (Higgins et al.,2003).

We conducted additional sensitivity analyses to explore the effects of possible sources of bias and artifacts on the results: First, the presence of publication bias was illustrated by the funnel plot (Egger et al., 1997), and formally calculated by the fail-safe *N* method (Rosenthal, 1984) and the Begg adjusted-rank correlation test (Begg and Mazumdar, 1994). We estimated the sensitivity of publication bias by the trim-and-fill method (Duval and Tweedie, 2000). Second, we conducted subgroup analyses to test for significant differences between outcome data in different categories of studies. We focused on type of antidepressant and mood scale (i.e., categorical variables). Differences between Hedges's *g* were calculated using a one-way ANOVA, whereas the effect sizes were weighted by the sample size divided by s^2 (i.e., n/var; Lipsey and Wilson, 2001). Four studies included two treatment groups and one placebo group (Katona et al., 2012; Rapaport et al., 2003, 2009; Schatzberg and Roose, 2006). To deal with the resulting dependency in these cases, we included both comparisons using the same mean for each placebo sub-group but used half the sample size for *n* when weighting (n/var) the means of the placebo group in each comparison.

Moderator analyses, heterogeneity, and publication bias were only assessed for mood scales used in more than three trials. In cases where both intention-to-treat and on-treatment data were available, we used the intention-to-treat data for calculations. We used the Cochrane Collaboration's tool for assessing the risk of bias (Higgins and Green, 2011).

3. Results

3.1. Study selection and study characteristics

The study selection procedure is shown in Fig. 1. In total, 19 studies met inclusion criteria and provided relevant data for the

[3] Available at: www.meta-analysis.com. Accessed June 27, 2014.

Table 1
Selected characteristics of studies investigating the association between baseline depression severity and treatment outcomes.

Study	Depression Diagnosis	Diagnostic Criteria	Minimum Age (Years)	Illness Severity Score at Entry	MMSE score	Trial Duration (Weeks)	Group characteristics				
							Group	N	Age (years)	Sex (% female)	Dropout *N*, %
Bose et al. (2008)	MDD	DSM-IV	60	MADRS ≥ 22	≥ 24	12	Escitalopram	129	68.1 (6.7)	58.9%	36, 27.9%
							Placebo	134	68.5 (7.1)	67.5%	26, 19.4%
Brody et al. (2011)	MDD or MiDD	DSM-IV	–	$HDRS_{17} \geq 10$	d.e.	2 × 8	Escitalopram	7	78.7 (6.6)	42.9%	2, 12.5%
							Placebo	9	79.8 (2.3)	88.9%	(total)
Chen et al. (2011)	MDD	DSM-IV	65	GDS ≥ 20	d.e.	8	Escitalopram	29	68.9 (6.1)	61.8%	2, 6.9%
							Placebo	26	(total)	(total)	2, 7.7%
Devanand et al. (2005)	DD	DSM-IV	60	$HDRS_{24} \geq 8$ and ≤ 25	> 24	12	Fluoxetine	44	69.0 (6.0)	32.6%	12, 27.3%
							Placebo	46	70.8 (6.3)	40.9%	7, 15.2%
Fraguas et al. (2010)	MDD	DSM-IV	65	$HDRS_{31} \geq 18$	-	8	Citalopram	19	74.4 (6.0)	47.4%	3, 15.7%
							Placebo	18	72.6 (4.6)	55.6%	7, 38.9%
Gerner et al. (1980)	unipolar depression	-	60	HDRS ≥ 18	-	4	Trazodone	19	68.4 (60-90) (total)	38.3% (total)	7, 36.8%
							Placebo	2			7, 35.0%
Heun et al. (2013)	MDD	DSM-IV-TR	65	$HDRS_{17} \geq 22$	≥ 22	8	Agomelatine	151	71.9 (5.1)	69.5%	26, 17.2%
							Placebo	71	71.7 (4.8)	64.8%	21, 29.6%
Hewett et al. (2010)	MDD	DSM-IV	65	$HDRS_{17} \geq 18$	> 24	10	Bupropion	211	70.9 (5.6)	74.4%	49, 23.0%
							Placebo	207	71.3 (5.9)	69.6%	46, 22.0%
Katona et al. (2012)	MDD	DSM-IV-TR	65	MADRS ≥ 26	≥ 24	8	Vortioxetine	156	70.5 (4.8)	68.6%	20, 12.8%
							Duloxetine	151	70.9 (5.5)	66.2%	23, 15.2%
							Placebo	145	70.3 (4.4)	62.1%	17, 11.7%
Paile-Hyvärinen et al. (2007)	mild MDD	DSM-IV	50	-	-	24	Paroxetine	23	59.2 (5.4)	26.1%	0, 0%
							Placebo	20	59.5 (6.0)	20.0%	6, 30.0%
Rapaport et al. (2009)	MDD	DSM-IV	60	$HDRS_{17} \geq 18$	> 24	10	$Paroxetine_{12.5mg}$	164	67.0 (6.1)	60.0%	39, 23.8%
							$Paroxetine_{25mg}$	173	67.0 (6.6)	60.0%	39, 22.5%
							Placebo	179	68.0 (6.7)	63.0%	53, 29.6%
Rapaport, et al. (2003)	MDD	DSM-IV	60	$HDRS_{17} \geq 18$	> 24	12	$Paroxetine_{CR}$	104	70.4 (5.9)	48.1%	23, 22.1%
							$Paroxetine_{IR}$	106	70.1 (6.6)	56.6%	30, 28.3%
							Placebo	109	69.4 (5.4)	63.3%	25, 22.9%
Robinson et al. (2014)	MDD	DSM-IV-TR	65	MADRS ≥ 20	≥ 20	12	Duloxetine	249	73.0 (6.3)	78.4%	70, 28.1%
							Placebo	121	73.1 (5.6)	73.7%	43, 35.5%
Roose et al. (2004)	MDD	DSM-IV	75	$HDRS_{24} \geq 20$	> 19	8	Citalopram	84	79.8 (4.0)	53.6%	18, 21.4%
							Placebo	90	79.3 (4.7)	62.2%	11, 12.2%
Schatzberg and Roose (2006)	MDD	DSM-IV	65	$HDRS_{21} \geq 20$	> 19	8	Venlafaxine	104	71.0	56.0%	36, 34.6%
							Fluoxetine	100	71.0	45.0%	30, 30.0%
							Placebo	96	71.0	46.0%	24, 25.0%
Schneider et al. (2003)	MDD	DSM-IV	60	$HDRS_{17} \geq 18$	≥ 24	8	Sertraline	371	70.0 (6.8)	53.6%	87, 23.5%
							Placebo	376	69.6 (6.5)	58.4%	65, 17.3%
Sheikh et al. (2004)	MDD	DSM-IV	60	$HDRS_{17} \geq 18$	≥ 24	8	Sertraline	360	70.0	53.9%	46, 12.8%
							Placebo	368	69.6	57.9%	42, 11.4%
Tollefson et al. (1995)	MDD	DSM-III-R	60	$HDRS_{17} \geq 16$	≥ 24	6	Fluoxetine	335	67.4 (5.4)	53.7%	72, 21.5%
							Placebo	336	68.1 (5.9)	55.6%	65, 19.3%
Williams et al. (2000)	MiDD or DD	DSM-III-R	60	$HDRS_{17} \geq 10$	> 24	11	Paroxetine	137	71.0 (6.8)	38.7%	43, 31.4%
							Placebo	140	71.0 (7.2)	45.0%	31, 22.1%

Note. DD = Dysthymic Disorder; d.e. = dementia excluded; MDD = Major Depressive Disorder; MiDD = Minor Depressive Disorder.

meta-analysis. The trials included a total of 5737 elderly depressed patients, of whom 3226 received active drug and 2511 received placebo. Sample sizes of included studies were between $N = 16$ and $N = 747$. Publication year ranged from 1980 to 2014. Most trials were based on a parallel design, except one study, which used a crossover design (Brody et al., 2011). We therefore only analyzed the first 8-week period prior to the crossover. Studies investigated different types of second-generation antidepressants. SSRIs (i.e., fluoxetine, escitalopram, paroxetine, sertraline, and citalopram) were examined in 14 trials, novel atypical antidepressants (i.e., duloxetine, trazodone, agomelatine, bupropion, vortioxetine, and venlafaxine) in another 6 trials. Trial duration varied from 4 to 24 weeks (see Table 1). Based on HDRS17, classification of baseline depression severity ranged from mild to very severe (see Table 2). Table 3 shows an assessment of the risk of bias for each study using a tool developed by the Cochrane Collaboration (Higgins and Green, 2011). The risk of bias across the included studies was generally low or unclear and is summarized in Fig. 2.

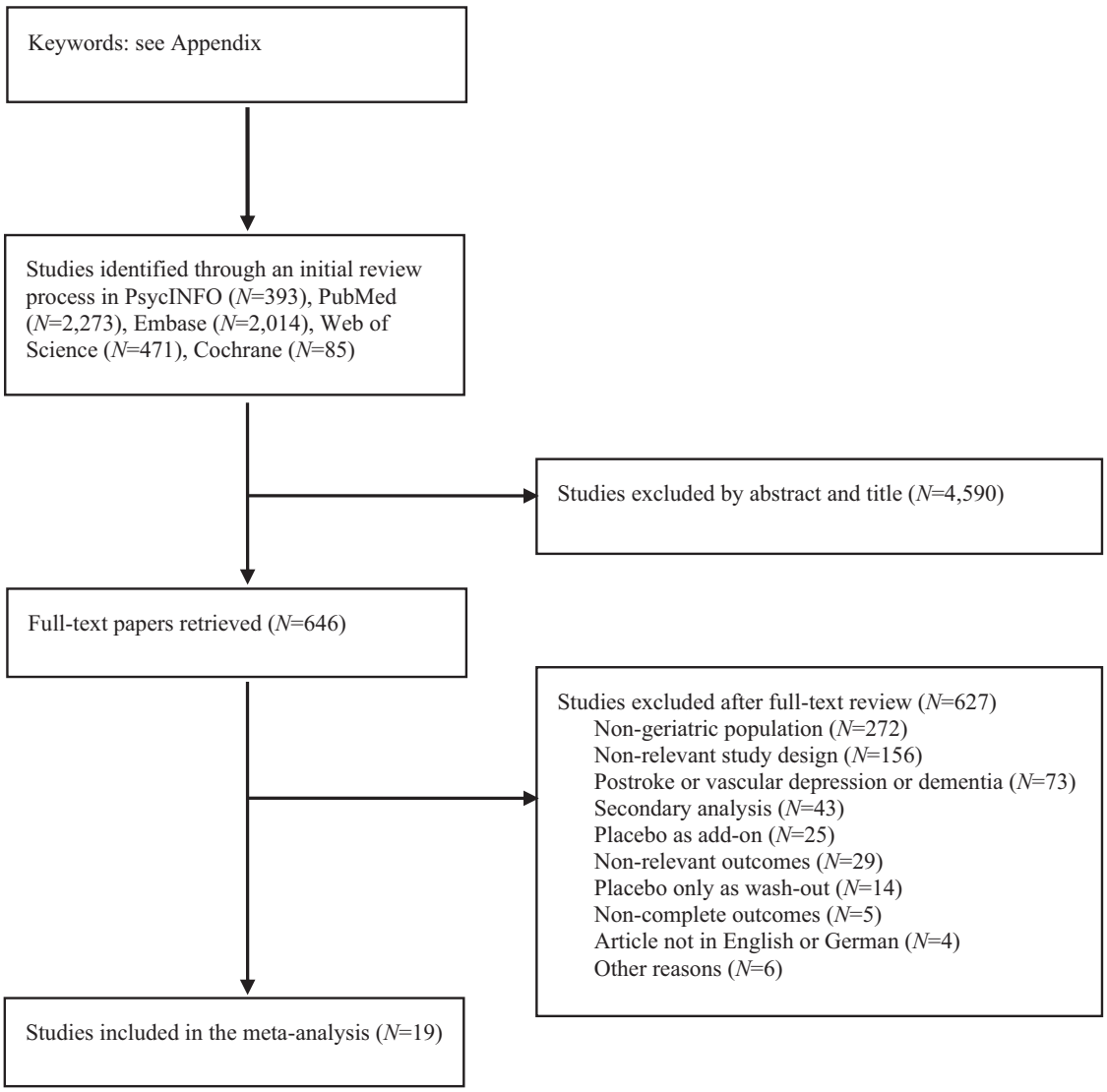


Fig. 1. Study selection procedure.

3.2. Effect sizes for antidepressant and placebo treatments

Combined over all mood scales, patients in the treatment groups showed a significantly higher mean change in depressive symptoms than patients in the placebo groups ($g=.37$, 95% CI: .27 – .46, $p<.001$; see Fig. 3). One study differed considerably from the others (Chen et al., 2011). Excluding the study led to a slightly lower, yet still significant difference between antidepressant and placebo treatments ($g=.32$, 95% CI: .25–.40, $p<.001$). With respect to HDRS scores, studies exhibited moderate, yet significant between-studies heterogeneity. There was an evident publication bias in a funnel plot. The trim-and-fill test led to an adjustment of the Hedges's g, yet still reached statistical significance (see Table 4). Regarding GDS scores, studies showed high and significant between-studies heterogeneity. There was evidence of some possible publication bias in a funnel plot. The trim-and-fill method led to a corrected Hedges's g, which then did not reach statistical significance (see Table 4).

To analyze the mean change in depressive symptoms within each group (see Table 3), we calculated pre-post effect sizes. Findings revealed that there is a significant treatment improvement in antidepressant groups ($g=1.35$, 95% CI: 1.14–1.57, $p<.000$), as well as in placebo groups ($g=.96$, 95% CI: .79–1.13, $p<.000$).

3.3. Association between baseline severity and mean change in depressive symptoms

The slope representing the overall relationship between baseline severity and change in symptoms was not significant in either antidepressant groups ($Z=1.47$, $p=.142$) or placebo groups ($Z=1.38$, $p=.168$), nor was baseline severity significantly correlated with drug-placebo differences ($r=.17$, $p=.392$).

Subgroup analyses of various mood scales were only assessed for mood scales used in more than three trials (i.e., GDS and HDRS). Separate analysis of self-rated scales produced comparable results: change in mean total GDS score was not related to baseline severity in elderly patients taking antidepressants ($Z=1.15$, $p=.251$), or elderly patients taking placebos ($Z=-.04$, $p=.971$), nor was baseline severity significantly correlated with drug–placebo differences ($r=.69$, $p=.199$). In contrast, studies using the clinician-rated HDRS mood scale indicated that mean change in depressive symptoms increased significantly in antidepressant trials ($Z=2.67$, $p=.008$, $R^2=.40$) and placebo trials ($Z=4.46$, $p<.000$, $R^2=.50$) as a function of HDRS baseline severity, which would be expected as a result of regression toward the mean. As displayed in Fig. 4, the slope of the regression lines increased within each group. Nevertheless, the overall baseline and the

Table 2
Outcome characteristics of studies investigating the association between baseline depression severity and treatment outcomes.

Study	Mood scale	Baseline depression severity drug (mean)	Baseline depression severity placebo (mean)	Mean change in depressive symptoms (Drug)	Mean change in depressive symptoms (Placebo)	Classification (based on HDRS17)
Bose et al. (2008)	$HDRS_{17}$	20.3	19.6	7.5	7.1	Severe
Brody et al. (2011)	$HDRS_{17}$	17.1	15.2	6.1	2.9	Mild/moderate
Chen et al. (2011)	GDS	23.4	24.0	12.7	2.9	
Devanand et al. (2005)	$HDRS_{17}$ (modified)	10.6	10.0	4.1	1.3	Mild/moderate
	BDI	12.7	13.1	1.1	.0	
Fraguas et al. (2010)	$HDRS_{17,}$	15.4	17.3	9.7	9.6	Mild/moderate
	MADRS	21.9	20.1	15.1	9.4	
Gerner et al. (1980)	$HDRS_{17}$ (modified)	22.1	20.2	14.8	4.1	Severe
	BDI	15.1	13.4	5.4	.4	
Heun et al. (2013)	$HDRS_{17}$	26.9	26.8	13.5	10.7	Very severe
Hewett et al. (2010)	MADRS	29.5	29.8	16.6	13.6	
Katona et al. (2012)	$HDRS_{17}$	Vortioxetine: 22.7 Duloxetine: 22.3	22.7	14.7 17.0	10.8	Severe
Paile-Hyvärinen et al. (2007)	HADS	7.3	8.4	1.8	2.2	
Rapaport et al. (2009)	$HDRS_{17}$	$Paroxetine_{12.5mg}$: 22.6	22.7	10.7	8.9	Severe
	GDS-short	$Paroxetine_{25mg}$: 23.1		12.1		
		$Paroxetine_{12.5mg}$: 8.9 $Paroxetine_{25mg}$: 9.1	8.7	3.2 3.5	2.2	
Rapaport et al. (2003)	$HDRS_{17}$	$Paroxetine_{CR}$: 22.1 $Paroxetine_{IR}$: 22.3	22.1	14.4 13.9	10.5	Severe
Robinson et al. (2014)	$HDRS_{17,}$	19.4	19.3	6.0	3.9	Severe
	GDS	18.5	17.6	4.3	4.5	
Roose et al. (2004)	$HDRS_{17}$ (modified)	17.3	17.1	9.6	8.2	Mild/moderate
	MADRS	24.4	25.0	9.7	8.5	
Schatzberg and Roose (2006)	$HDRS_{17}$ (modified)	Venlafaxine: 19.4 Fluoxetine: 19.4	18.6	8.4 6.7	7.5	Severe
Schneider et al. (2003)	$HDRS_{17}$	21.4	21.4	7.4	6.6	Severe
Sheikh et al. (2004)	$HDRS_{17}$	21.4	21.4	7.9	6.4	Severe
Tollefson et al. (1995)	$HDRS_{17}$	22.2	22.1	8.2	6.4	Severe
	GDS	19.1	18.7	3.3	2.1	
Williams et al. (2000)	HSCL-D-20	1.4	1.4	.6	.4	Mild/moderate

Table 3
Assessment of risk of bias.

Study	Random sequence generation	Allocation concealment	Adequate blinding	Incomplete outcome data	Selective reporting	Other risks of bias
Bose et al. (2008)	Low risk	Unclear	Unclear	Unclear	Low risk	Low risk
Brody et al. (2011)	Unclear	Low risk	Unclear	Unclear	Low risk	Unclear
Chen et al. (2011)	Low risk	Unclear	Unclear	Low risk	Low risk	Low risk
Devanand et al. (2005)	Low risk	Unclear	Low risk	Low risk	Low risk	Low risk
Fraguas et al. (2010)	Unclear	Unclear	Unclear	High risk	Low risk	High risk
Gerner et al. (1980)	Unclear	Unclear	Unclear	High risk	Low risk	Unclear
Heun et al. (2013)	Low risk	Unclear	Low risk	Unclear	Low risk	Low risk
Hewett et al. (2010)	Low risk	Unclear	Unclear	Low risk	Low risk	Low risk
Katona et al. (2012)	Low risk	Low risk	Low risk	Unclear	Low risk	Low risk
Paile-Hyvärinen et al. (2007)	Low risk	Low risk	Unclear	High risk	Low risk	Low risk
Rapaport et al. (2009)	Low risk	Unclear	Unclear	Low risk	Unclear	Low risk
Rapaport et al. (2003)	Unclear	Unclear	Unclear	Low risk	Low risk	Low risk
Robinson et al. (2014)	Low risk	Low risk	Unclear	Unclear	Unclear	Unclear
Roose et al. (2004)	Low risk	Low risk	Unclear	Low risk	Low risk	High risk
Schatzberg and Roose (2006)	Low risk	Unclear	Unclear	Low risk	High risk	Low risk
Schneider et al. (2003)	Low risk	Unclear	Low risk	Low risk	Low risk	Low risk
Sheikh et al. (2004)	Unclear	Unclear	Unclear	Low risk	Low risk	Unclear
Tollefson et al. (1995)	Unclear	Unclear	Unclear	Unclear	Low risk	Low risk
Williams et al. (2000)	Low risk	Low risk	Low risk	High risk	Low risk	Low risk

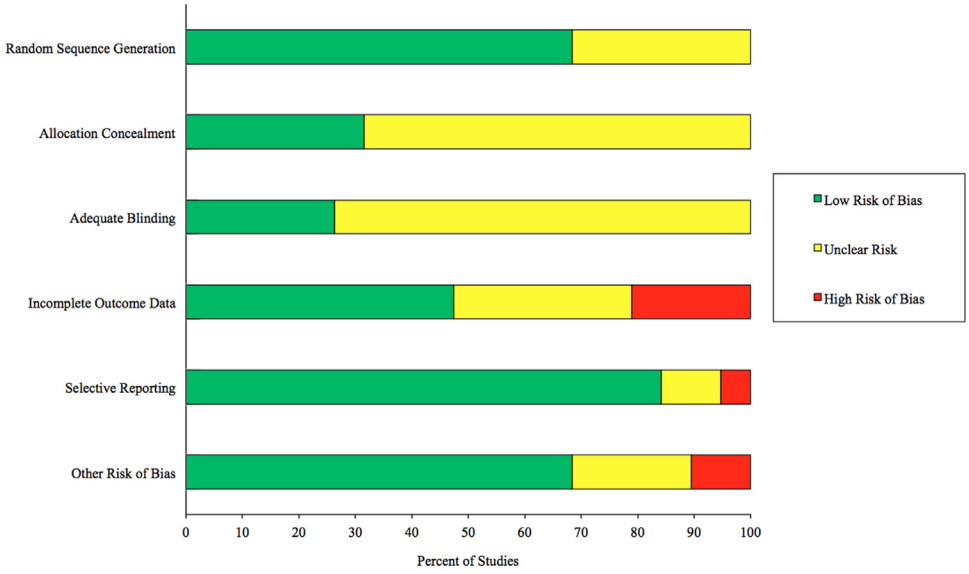


Fig. 2. Risk of bias assessment.

antidepressant–placebo difference in improvement were not significantly correlated ($r=.19$, $p=.469$).

To clarify further the clinical significance of the differences between the two treatments, we refer to the National Institute for Clinical Excellence guidelines (NICE, 2004). These suggest that a standardized mean difference (d) of .5 or a three-point difference in HDRS17 scores should be used as the threshold for clinical significance. For elderly patients with mild to moderate depression (HDRS17 score of ≤ 18), Cohen's d was .24 (95% CI: .02–.47, $p=.031$), for patients with an HDRS17 score in the severe range (19–22), Cohen's d was .39 (95% CI: .25–.52, $p<.001$), and for patients with HDRS17 score in the very severe range (≥ 23), we found an effect size of $d=.37$ (95% CI: .09–.66, $p=.011$). In summary, none of the values reached the proposed cutoff of $d=.5$ for clinical significance. However, the criterion of a difference of ≥ 3 points on the HDRS was met for baseline HDRS17 scores of ≥ 21, indicated by a red line in Fig. 4.

3.4. Categorical moderator variables

Moderator analyses examined whether several categories of studies were related to improvement. Different types of antidepressant (i.e., agomelatine, bupropion, citalopram, duloxetine, escitalopram, fluoxetine, paroxetine, sertraline, trazodone, vortioxetine, and venlafaxine) were unrelated to effect sizes ($F(10,20)=1.92$, $p=.104$). Similarly, there was no significant association between different mood scales (i.e., BDI, GDS, HADS, HDRS, HSCL-D-20, MADRS) and improvement ($F(5,25)=.77$, $p=581$).

4. Discussion

The purpose of the present meta-analysis was to investigate the moderating effects of baseline severity on mean outcome measures in depressed elderly patients treated with antidepressants or placebos.

Concerning HDRS scores, we did find an increase in mean change in depressive symptoms with increasing baseline severity within antidepressant and placebo interventions. However, one must be careful in interpreting relations between baseline severity and within-group changes, as they can be strongly influenced by regression toward the mean (Calati et al., 2013). Consequently, we analyzed between-group data, looking at the antidepressant-placebo difference as a function of baseline. The overall baseline and the antidepressant–placebo difference in improvement were not significantly correlated. Therefore, studies in late-life depression do not confirm the severity hypotheses found in mixed-aged studies, in which an increasing advantage of antidepressants over placebos was reported with increasing baseline severity (Fournier et al., 2010; Khan et al., 2002; Kirsch et al., 2008). Our failure to find an association between initial severity and drug over placebo efficacy is similar to previous geriatric studies (Gibbons et al., 2012; Nelson et al., 2013). Nevertheless, our study differed from these reviews as we included a broader range of studies, including studies looking at minor depression and dysthymia.

Our results indicate that placebo responses are important in the treatment of depressed elderly people. First, within-group pre-post effect sizes revealed that there is a treatment improvement for depressed elderly participants taking placebos. Second, clinically meaningful differences between antidepressant and placebo interventions were only observable in patients with severe depressive symptoms (i.e., HDRS17 ≥ 21) at baseline. On the one hand, it is possible that placebo effects might even be larger and more important than reported. Analgesia trials indicate that the way in which instructions are given influences the magnitude of placebo analgesia; placebo effects were larger in experimental studies that explicitly investigated mechanisms of placebo analgesia than in RCTs where placebo was only used as a control condition (Vase et al., 2002). Studies also indicate that the tests of the blind in double-blind designs usually show that it is penetrated and thus susceptible to the researcher's assumption that the active drug will prove to be more effective than the placebo (Fisher and Greenberg, 1993). Moreover, a recent study demonstrated that a belief in the effectiveness of the antidepressant and a supportive therapeutic alliance were crucial elements in determining the treatment response (Leuchter et al., 2014). On the other hand, the observed placebo response can be attributed to a number of factors, including spontaneous remission, measurement factors (i.e., regression to the mean, rater bias and response bias), unreported co-interventions, and clinical characteristics of participating patients (Benedetti, 2008; Rutherford and Roose, 2013). Moreover, the additivity assumption, that the difference between antidepressant response and placebo response is attributable to the pharmacological effect of the antidepressant, may be incorrect (Kirsch, 2000). It is therefore possible that antidepressant effects are underestimated in RCTs (Lund et al., 2014). Nevertheless, the

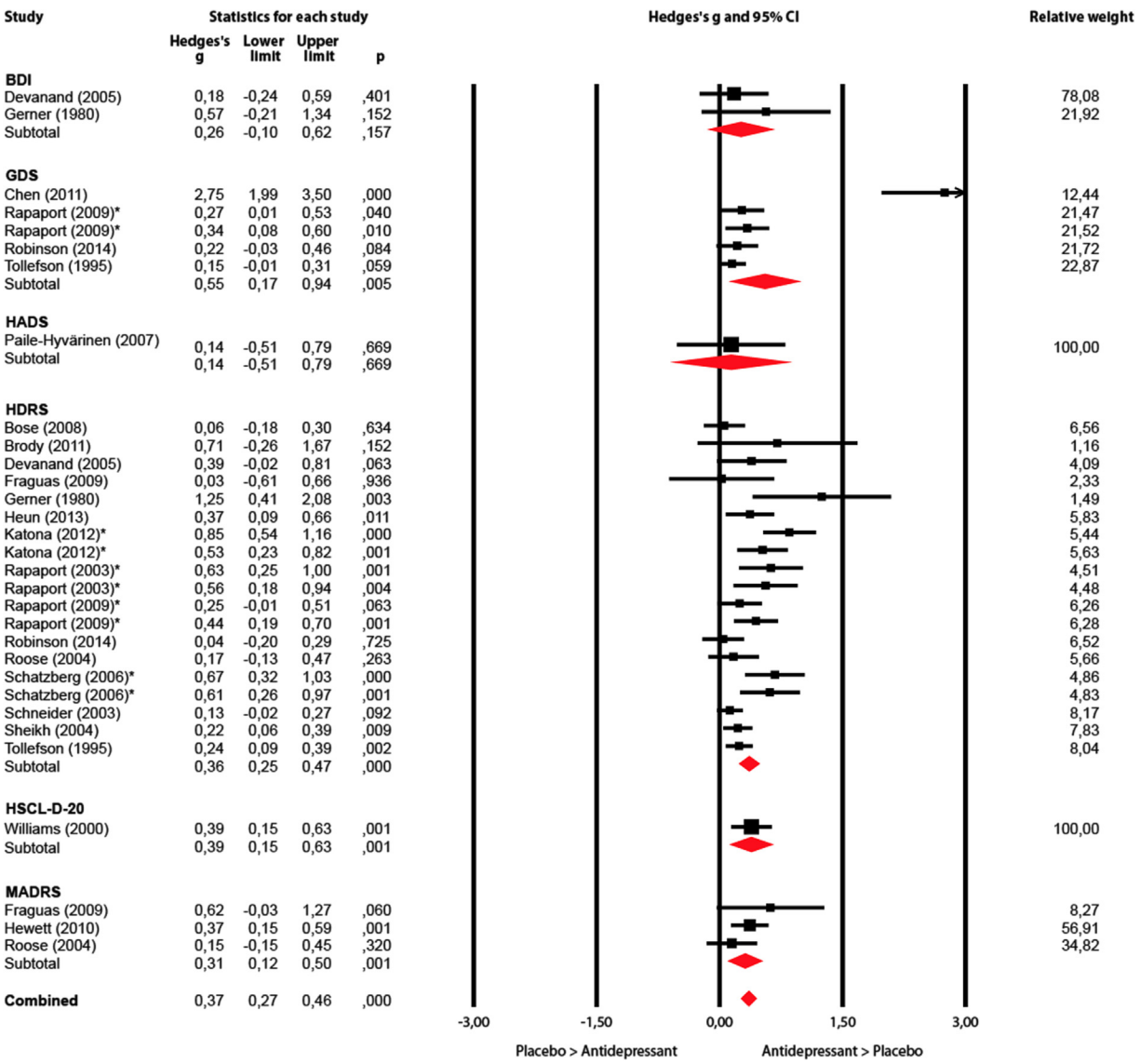


Fig. 3. Mean change in depressive symptoms of antidepressant treatment for depression compared to placebo treatment (random effects).
The red diamonds indicate the combined effect sizes for studies sorted by mood scale, as well as the overall effect size of the meta-analysis (top to bottom). An asterisk indicates that the study included two separate antidepressant samples.

finding of a meaningful placebo effect needs to be considered with respect to its therapeutic implications.

Over time, the conceptualization of the placebo has shifted from an inert sugar pill (i.e., a deceptive technique to increase outcome expectations) to the emphasis of contextual factors (Wampold et al., 2005), including implicit or explicit psychosocial stimulation of a therapeutic procedure (Finniss et al., 2010), and a therapeutic alliance (Kaptchuk et al., 2008). In the elderly, psychosocial support is described as a highly relevant part of placebo treatment (Alexopoulos et al., 2007). All therapeutic treatment is administered in the context of a complex sociocultural system, resulting in the perception of a meaningful therapeutic encounter (Di Blasi et al., 2001). Contextual cues may induce positive expectations, hope, trust, security, and hence a placebo response in elderly patients (Bingel et al., 2011). In clinical trials with depressed elderly participants, perceived social support is associated with subsequent decreases in depression (Oxman and Hull, 2001) and increased probability of recovery (Bosworth et al., 2002). Similarly, control treatments can have positive outcomes when they involve social contact. Hence, a study comparing placebo, paroxetine, and problem-solving treatments in subclinically depressed elderly participants concluded that the minimal between-groups differences were due to the impact of clinical management, which included social contact and was additionally given to all patients (Oxman and Hull, 2001). Similarly, depressed nursing home residents under a treatment condition involving exercise training and a control condition involving social conversation both showed improvement (Williams and Tappen, 2008). It is assumed that the impact of social support on treatment outcome is especially relevant for elderly patients, as they often live alone and may have little social contact (Bingel et al., 2011). Correspondingly, a recent study reported that self-rated reclusiveness predicted response in a supportive patient–practitioner relationship (Conboy et al., 2011).

Our failure to find that baseline severity is associated with an antidepressant–placebo difference in study outcome may be influenced by the initial severity grades of the examined trials. We included only one study of very severely depressed patients, showing a HDRS initial severity of 26.9 (Heun et al., 2013). Similarly, Kok et al. (2012) concluded that only a few studies have focused on severely depressed older people. In contrast, other

meta-analyses of mixed-age patients examined more strongly affected patients with HDRS baseline scores of over 30.0 in several trials (Fournier et al., 2010; Khan et al., 2002; Kirsch et al., 2008). One reason for our lack of studies investigating the most severe cases is due to our exclusion of elderly patients with executive dysfunctions, as severe depression has been shown to be associated with Alzheimer Disease (Gracia-García et al., 2013), all-cause dementia (Chen et al., 2008), and other memory deficits (Boeker et al., 2012).

Our meta-analysis has several limitations. First, the investigated studies had substantial differences regarding outcome measures, resulting in heterogeneity across studies. Second, our meta-analysis was limited to published data, which may have resulted in a considerable bias towards studies reporting a positive outcome (Turner et al., 2008). There was evidence of such a bias in mean difference outcome data in our study, which we attempted to rectify using statistical adjustment procedures, namely the trim-and-fill method. Third, our study analyzed information only at the level of treatment groups, yet contained no data for a patient-level analysis (Fournier et al., 2010). This may have resulted in ecologically fallacious findings, where the cumulative association fails to reproduce associations at the individual level (Spoerri et al., 2010). Finally, stable physical illness and comorbid disorders were common in our sample. Nevertheless, elderly patients with a number of age-related disorders in addition to depression are representative of the population of elderly patients with depression (Nelson et al., 2013).

Despite these limitations, we found clear indications that placebo responses are large and meaningful in the treatment of depressed elderly people, irrespective of baseline depression severity. Further studies should investigate antidepressant and placebo reactions in severely depressed old patients without executive dysfunctions – assessed by more comprehensive and valid tests than the MMSE – in order to make final conclusions about the possible moderating effect of initial depression severity. In addition, further research into the determinants of the effect of psychological interventions in the treatment of late-life depression is needed.

Nevertheless, our placebo responses findings should remind healthcare practitioners that the therapeutic environment and social support are of particular importance in elderly patients. In accordance with Bingel et al. (2011), we propose that it is essential to make use of supportive psychosocial and environmental mechanisms to optimize treatment with antidepressants. Moreover, social support and increased attention to patients has been shown to improve compliance with medication regimes (Packer, 1990). Finally, we note that this knowledge may be particularly important in elderly patients, as they are more likely to have serious medical conditions and thus receiving polypharmacotherapy, which often leads to adverse drug reactions and interactions between medications (Taylor and Doraiswamy, 2004). We propose a stepwise approach, i.e., to initially offer elderly depressed patients psychosocial interventions and only consider antidepressants if patients do not respond. Given the propensity to multiple adverse drug reactions noticed in elderly patients, psychosocial interventions may represent a safer alternative (Andreescu and Reynolds, 2011).

Table 4
Results of publication bias and heterogeneity.

Variable	Mean difference	df or 95% CI
Measured with GDS		
Publication bias		
Funnel plot distribution	Asymmetrical	
Begg's adjusted-rank correlation (p value)	.043	
Classic fail-safe N	57	
Trim-and-fill test (Hedges's g or RR)	.20	−.21–.62
Heterogeneity		
I^2 statistic[a]	90.79***	4
tau^2 statistic[b]	.17	
Measured with HDRS		
Publication bias		
Funnel plot distribution	Asymmetrical	
Begg's adjusted-rank correlation (p value)	.010	
Classic fail-safe N	520	
Trim-and-fill test (Hedges's g or RR)	.25	.13–.37
Heterogeneity		
I^2 statistic[a]	64.24***	18
tau^2 statistic[b]	.03	

[a] The data represent the variance between studies as a proportion of the total variance; heterogeneity was tested using the I^2 statistic (low heterogeneity=25%; moderate heterogeneity=50%; high heterogeneity=75%). The P values refer to significance of the Q statistic (the I^2 statistic does not include a test of significance).

[b] Heterogeneity was tested using the tau^2 statistic, which estimates the between-study variance.

*** $p \leq .001$.

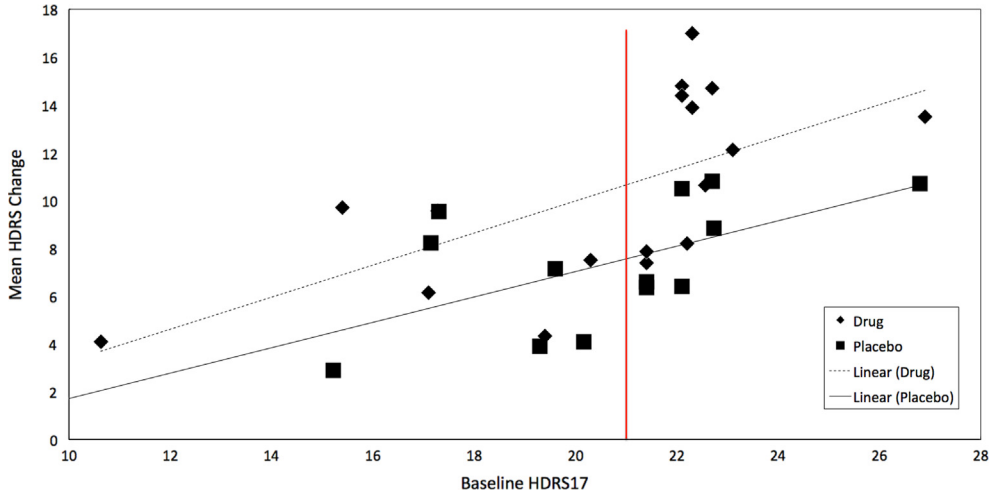


Fig. 4. Relationship between baseline severity and mean change in HDRS17 score among the antidepressant and placebo groups. The NICE threshold for clinical significance (HDRS mean difference ≥ 3) was met for initial HDRS17 scores of 21 or greater, visualized by the red line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

Role of funding source

The authors received no specific funding for this study.

Conflict of interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). She is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author and which has been configured to accept email from cosima.locher@unibas.ch.

Acknowledgment

This study was supported by Grant project (P300P1_158427) awarded to Joe Kossowsky by the Swiss National Science Foundation. The authors thank Ted Kaptchuk for provided mentoring, which was supported by NIH Grant #2K24AT004095.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.jad.2015.03.062.

References

Alexopoulos, G.S., Kanellopoulos, D., Murphy, C., Gunning-Dixon, F., Katz, R., Heo, M., 2007. Placebo response and antidepressant response. Am. J. Geriatr. Psychiatry 15, 149–158.

Alexopoulos, G.S., Kiosses, D.N., Heo, M., Murphy, C.F., Shanmugham, B., Gunning-Dixon, F., 2005. Executive dysfunction and the course of geriatric depression. Biol. Psychiatry 58, 204–210.

Altman, D.G., Royston, P., 2006. The cost of dichotomising continuous variables. Br. Med. J. 332, 1080.

American Psychiatric Association, 2010. Practice Guideline For The Treatment Of Patients With Major Depressive Disorder 2015, from: ⟨http://psychiatryonline.org/pb/assets/raw/sitewide/practice_guidelines/guidelines/mdd.pdf⟩.

Anderson, I.M., 2000. Selective serotonin reuptake inhibitors versus tricyclic antidepressants: a meta-analysis of efficacy and tolerability. J. Affect. Disord. 58, 19–36.

Andreescu, C., Reynolds, C.F., 2011. Late-life depression: evidence-based treatment and promising new directions for research and clinical practice. Psychiatr. Clin. North Am. 34, 335–355.

Beck, A.T., Ward, C., Mendelson, M., 1961. Beck depression inventory (BDI). Arch. Gen. Psychiatry 4, 561–571.

Begg, C.B., Mazumdar, M., 1994. Operating characteristics of a rank correlation test for publication bias. Biometrics 50, 1088–1101.

Benedetti, F., 2008. Mechanisms of placebo and placebo-related effects across diseases and treatments. Annu. Rev. Pharmacol. Toxicol. 48, 33–60.

Benedetti, F., Arduino, C., Costa, S., Vighetti, S., Tarenzi, L., Rainero, I., Asteggiano, G., 2006a. Loss of expectation-related mechanisms in Alzheimer's disease makes analgesic therapies less effective. Pain 121, 133–144.

Benedetti, F., Bernasconi, A., Pontiggia, A., 2006b. Depression and neurological disorders. Curr. Opin. Psychiatry 19, 14–18.

Bingel, U., Colloca, L., Vase, L., 2011. Mechanisms and clinical implications of the placebo effect: is there a potential for the elderly? A mini-review. Gerontology 57, 354–363.

Boeker, H., Schulze, J., Richter, A., Nikisch, G., Schuepbach, D., Grimm, S., 2012. Sustained cognitive impairments after clinical recovery of severe depression. J. Nerv. Ment. Dis. 200, 773–776.

Borenstein, M., Hedges, L.V., Higgins, J.P., Rothstein, H.R., 2011. Introduction to Meta-Analysis. John Wiley & Sons, Chichester.

Bose, A., Li, D., Gandhi, C., 2008. Escitalopram in the acute treatment of depressed patients aged 60 years or older. Am. J. Geriatr. Psychiatry 16, 14–20.

Bosworth, H.B., Hays, J.C., George, L.K., Steffens, D.C., 2002. Psychosocial and clinical predictors of unipolar depression outcome in older adults. Int. J. Geriatr. Psychiatry 17, 238–246.

Brody, B.L., Field, L.C., Roch-Levecq, A.C., Moutier, C.Y., Edland, S.D., Brown, S.I., 2011. Treatment of depression associated with age-related macular degeneration: a double-blind, randomized, controlled study. Ann. Clin. Psychiatry 23, 277–284.

Brunoni, A.R., Lopes, M., Kaptchuk, T.J., Fregni, F., 2009. Placebo response of non-pharmacological and pharmacological trials in major depression: a systematic review and meta-analysis. PLOS One 4, e4824.

Calati, R., Signorelli, M.S., Balestri, M., Marsano, A., De Ronchi, D., Aguglia, E., Serretti, A., 2013. Antidepressants in elderly: metaregression of double-blind, randomized clinical trials. J. Affect. Disord. 147, 1–8.

Chen, R., Hu, Z., Wei, L., Qin, X., McCracken, C., Copeland, J.R., 2008. Severity of depression and risk for subsequent dementia: cohort studies in China and the UK. Br. J. Psychiatry 193, 373–377.

Chen, Y.M., Huang, X.M., Thompson, R., Zhao, Y.B., 2011. Clinical features and efficacy of escitalopram treatment for geriatric depression. J. Int. Med. Res. 39, 1946–1953.

Cochran, W.G., 1954. The combination of estimates from different experiments. Biometrics 10, 101–129.

Conboy, L.A., Macklin, E., Kelley, J., Kokkotou, E., Lembo, A., Davis, R., Kaptchuk, T., 2011. Which patients improve: characteristics increasing sensitivity to a supportive patient–practitioner relationship. Soc. Sci. Med. 72, 1409 –1409.

Conwell, Y., Duberstein, P.R., Caine, E.D., 2002. Risk factors for suicide in later life. Biol. Psychiatry 52, 193–204.

Derogatis, L.R., Lipman, R.S., Rickels, K., Uhlenhuth, E.H., Covi, L., 1974. The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. Behav. Sci. 19, 1–15.

Devanand, D.P., Nobler, M.S., Cheng, J., Turret, N., Pelton, G.H., Roose, S.P., Sackeim, H.A., 2005. Randomized, double-blind, placebo-controlled trial of fluoxetine treatment for elderly patients with dysthymic disorder. Am. J. Geriatr. Psychiatry 13, 59–68.

Di Blasi, Z., Harkness, E., Ernst, E., Georgiou, A., Kleijnen, J., 2001. Influence of context effects on health outcomes: a systematic review. Lancet 357, 757–762.

Duval, S., Tweedie, R., 2000. A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. J. Am. Stat. Assoc. 95, 89–98.

Egger, M., Davey Smith, G., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. Br. Med. J. 315, 629–634.

Finniss, D.G., Kaptchuk, T.J., Miller, F., Benedetti, F., 2010. Biological, clinical, and ethical advances of placebo effects. Lancet 375, 686–695.

Fisher, S., Greenberg, R.P., 1993. How sound is the double-blind design for evaluating psychotropic drugs? J. Nerv. Ment. Dis. 181, 345–350.

Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J. Psychiatr. Res. 12, 189–198.

Fountoulakis, K.N., Veroniki, A.A., Siamouli, M., Möller, H.J., 2013. No role for initial severity on the efficacy of antidepressants: results of a multi-meta-analysis. Ann. Gen. Psychiatry 12, 26.

Fournier, J.C., DeRubeis, R.J., Hollon, S.D., Dimidjian, S., Amsterdam, J.D., Shelton, R.C., Fawcett, J., 2010. Antidepressant drug effects and depression severity: a patient-level meta-analysis. J. Am. Med. Assoc. 303, 47–53.

Fraguas, R., da Silva Telles, R.M., Alves, T.C., Andrei, A.M., Rays, J., Iosifescu, D.V., Wajngarten, M., 2010. A double-blind, placebo-controlled treatment trial of citalopram for major depressive disorder in older patients with heart failure: the relevance of the placebo effect and psychological symptoms. Contemp. Clin. Trials 30, 205–211.

Gerner, R., Estabrook, W., Steuer, J., Jarvik, L., 1980. Treatment of geriatric depression with trazodone, imipramine, and placebo: a double-blind study. J. Clin. Psychiatry 41, 216–220.

Gibbons, R.D., Hur, K., Brown, C.H., Davis, J.M., Mann, J.J., 2012. Benefits from antidepressants: synthesis of 6-week patient-level outcomes from double-blind placebo-controlled randomized trials of fluoxetine and venlafaxine. Arch. Gen. Psychiatry 69, 572–579.

Gracia-García, P., de-la-Cámara, C., Santabárbara, J., Lopez-Anton, R., Quintanilla, M.A., Ventura, T., Marcos, G., Campayo, A., Saz, P., Lyketsos, C., Lobo, A., 2013. Depression and incident alzheimer disease: the impact of disease severity. Am. J. Geriatr. Psychiatry 23, 119–129.

Hamilton, M., 1967. Development of a rating scale for primary depressive illness. Br. J. Soc. Clin. Psychol. 6, 278–296.

Heo, M., Murphy, C.F., Meyers, B.S., 2007. Relationship between the Hamilton depression rating scale and the Montgomery–Asberg depression rating scale in depressed elderly: a meta-analysis. Am. J. Geriatr. Psychiatry 15, 899–905.

Heun, R., Ahokas, A., Boyer, P., Giménez-Montesinos, N., Pontes-Soares, F., Olivier, V., 2013. The efficacy of agomelatine in elderly patients with recurrent major depressive disorder: a placebo-controlled study. J. Clin. Psychiatry 74, 587–594.

Hewett, K., Chrzanowski, W., Jokinen, R., Felgentreff, R., Shrivastava, R.K., Gee, M.D., Wightman, D.S., O'Leary, M.C., Millen, L.S., Leon, M.C., Briggs, M.A., Krishen, A., Modell, J.G., 2010. Double-blind, placebo-controlled evaluation of extended-release bupropion in elderly patients with major depressive disorder. J. Psychopharmacol. 24, 521–529.

Higgins, J.P., Green, S., 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. Retrieved from ⟨http://www.cochrane-handbook.org⟩: The Cochrane Collaboration.

Higgins, J.P., Thompson, S.G., Deeks, J.J., Altman, D.G., 2003. Measuring inconsistency in meta-analyses. Br. Med. J. 327, 557–560.

Jiang, W., Alexander, J., Christopher, E., Kuchibhatla, M., Gaulden, L.H., Cuffe, M.S., Blazing, M.A., Davenport, C., Califf, R.M., Krishnan, R.R., O'Connor, C.M., 2001. Relationship of depression to increased risk of mortality and rehospitalization in patients with congestive heart failure. Arch. Intern. Med. 161, 1849–1856.

Kaptchuk, T.J., Kelley, J.M., Conboy, L.A., Davis, R.B., Kerr, C.E., Jacobson, E.E., Kirsch, I., Schyner, R.N., Nam, B.H., Nguyen, L.T., Park, M., Rivers, A.L., McManus, C., Kokkotou, E., Drossman, D.A., Goldman, P., Lembo, A.J., 2008. Components of placebo effect: randomised controlled trial in patients with irritable bowel syndrome. Br. Med. J. 336, 999–1003.

Katona, C., Hansen, T., Olsen, C.K., 2012. A randomized, double-blind, placebo-controlled, duloxetine-referenced, fixed-dose study comparing the efficacy and safety of Lu AA21004 in elderly patients with major depressive disorder. Int. Clin. Psychopharmacol. 27, 215–223.

Khan, A., Leventhal, R.M., Khan, S.R., Brown, W.A., 2002. Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. J. Clin. Psychopharmacol. 22, 40–45.

Kirsch, I., 2000. Are drug and placebo effects in depression additive? Biol. Psychiatry 47, 733–735.

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, e45.

Kok, R.M., Nolen, W.A., Heeren, T.J., 2012. Efficacy of treatment in older depressed patients: a systematic review and meta-analysis of double-blind randomized controlled trials with antidepressants. J. Affect. Disord. 141, 103–115.

Leuchter, A.F., Hunter, A.M., Tartter, M., Cook, I.A., 2014. Role of pill-taking, expectation and therapeutic alliance in the placebo response in clinical trials for major depression. Br. J. Psychiatry 205, 443–449.

Lipsey, M.W., Wilson, D.B., 2001. Practical Meta-Analysis. California Sage Publications, Thousand Oaks p. 247.

Lund, K., Vase, L., Petersen, G.L., Jensen, T.S., Finnerup, N.B., 2014. Randomised controlled trials may underestimate drug effects: balanced placebo trial design. PLOS One 9, e84104.

Meeks, T.W., Vahia, I.V., Lavretsky, H., Kulkarni, G., Jeste, D.V., 2011. A tune in "a minor" can "b major": a review of epidemiology, illness course, and public health implications of subthreshold depression in older adults. J. Affect. Disord. 129, 126–142.

Mittmann, N., Herrmann, N., Einarson, T.R., Busto, U.E., Lanctôt, K.L., Liu, B.A., Shulman, K.I., Silver, I.L., Narango, C.A., Shear, N.H., 1997. The efficacy, safety and tolerability of antidepressants in late life depression: a meta-analysis. J. Affect. Disord. 46, 191–217.

Moncrieff, J., Kirsch, I., 2005. Efficacy of antidepressants in adults. Br. Med. J. 331, 155–157.

Montgomery, S.A., Asberg, M., 1979. A new depression scale designed to be sensitive to change. Br. J. Psychiatry 134, 382–389.

National Institute for Clinical Excellence, 2004. Depression: management of depression in primary and secondary care. Clinical practice guideline No. 23. London, England: National Institute for Clinical Excellence.

Nelson, J.C., Delucchi, K.L., Schneider, L.S., 2008. Efficacy of second generation antidepressants in late-life depression: a meta-analysis of the evidence. Am. J. Geriatr. Psychiatry 16, 558–567.

Nelson, J.C., Delucchi, K.L., Schneider, L.S., 2013. Moderators of outcome in late-life depression: a patient-level meta-analysis. Am. J. Psychiatry 170, 651–659.

Oxman, T.E., Hull, J.G., 2001. Social support and treatment response in older depressed primary care patients. J. Gerontol. B. Psychol. Sci. Soc. Sci. 56, 35–45.

Packer, M., 1990. The placebo-effect in heart-failure. Am. Heart J. 120, 1579–1582.

Paile-Hyvärinen, M., Wahlbeck, K., Eriksson, J.G., 2007. Quality of life and metabolic status in mildly depressed patients with type 2 diabetes treated with paroxetine: a double-blind randomised placebo controlled 6-month trial. BMC. Fam. Pract. 8, 34.

Pinquart, M., Duberstein, P.R., Lyness, J.M., 2006. Treatments for later-life depressive conditions: a meta-analytic comparison of pharmacotherapy and psychotherapy. Am. J. Psychiatry 163, 1493–1501.

Rapaport, M.H., Lydiard, R.B., Pitts, C.D., Schaefer, D., Bartolic, E.I., Iyengar, M., Carfagno, M., Lipschitz, A., 2009. Low doses of controlled-release paroxetine in the treatment of late-life depression: a randomized, placebo-controlled trial. J. Clin. Psychiatry 70, 46–57.

Rapaport, M.H., Schneider, L.S., Dunner, D.L., Davies, J.T., Pitts, C.D., 2003. Efficacy of controlled-release paroxetine in the treatment of late-life depression. J. Clin. Psychiatry 64, 1065–1074.

Robinson, M., Oakes, T.M., Raskin, J., Liu, P., Shoemaker, S., Nelson, J.C., 2014. Acute and long-term treatment of late-life major depressive disorder: duloxetine versus placebo. Am. J. Geriatr. Psychiatry 22, 34–45.

Rodda, J., Walker, Z., Carter, J., 2011. Depression in older adults. Br. Med. J. 343, d5219.

Roose, S.P., Sackeim, H.A., Krishnan, K.R., Pollock, B.G., Alexopoulos, G., Lavretsky, H., Katz, I.R., Hakkarainen, H., 2004. Antidepressant pharmacotherapy in the treatment of depression in the very old: a randomized, placebo-controlled trial. Am. J. Psychiatry 161, 2050–2059.

Rosenthal, R., 1984. Meta-Analytic Procedures for Social Research. Sage Press, Beverly Hills.

Rutherford, B.R., Roose, S.P., 2013. A model of placebo response in antidepressant clinical trials. Am. J. Psychiatry 170, 723–733.

Schatzberg, A., Roose, S., 2006. A double-blind, placebo-controlled study of venlafaxine and fluoxetine in geriatric outpatients with major depression. Am. J. Geriatr. Psychiatry 14, 361–370.

Schneider, L.S., Nelson, J.C., Clary, C.M., Newhouse, P., Krishnan, K.R., Shiovitz, T., Weihs, K., 2003. An 8-week multicenter, parallel-group, double-blind, placebo-controlled study of sertraline in elderly outpatients with major depression. Am. J. Psychiatry 160, 1277–1285.

Sheikh, J.I., Cassidy, E.L., Doraiswamy, P.M., Salomon, R.M., Hornig, M., Holland, P.J., Mandel, F.S., Clary, C.M., Burt, T., 2004. Efficacy, safety, and tolerability of sertraline in patients with late-life depression and comorbid medical illness. J. Am. Geriatr. Soc. 52, 86–92.

Sonnenberg, C.M., Deeg, D.J.H., Comijs, H.C., Van Tilburg, W., Beekman, A.T.F., 2008. Trends in antidepressant use in the older population: results from the LASA-study over a period of 10 years. J. Affect. Disord. 111, 299–305.

Spoerri, A., Zwahlen, M., Bopp, M., Gutzwiller, F., Egger, M., 2010. Religion and assisted and non-assisted suicide in Switzerland: national Cohort Study. Int. J. Epidemiol. 39, 1486–1494.

Taylor, W.D., Doraiswamy, P.M., 2004. A systematic review of antidepressant placebo-controlled trials for geriatric depression: limitations of current data and directions for the future. Neuropsychopharmacology 29, 2285–2299.

Tollefson, G.D., Bosomworth, J.C., Heiligenstein, J.H., Potvin, J.H., Holman, S., 1995. A double-blind, placebo-controlled clinical trial of fluoxetine in geriatric patients with major depression. Int. Psychogeriatr. 7, 89–104.

Tu, Y.K., Maddick, I.H., Griffiths, G.S., Gilthorpe, M.S., 2004. Mathematical coupling can undermine the statistical assessment of clinical research: illustration from the treatment of guided tissue regeneration. J. Dent. 32, 133–142.

Turner, E.H., Matthews, A.M., Linardatos, E., Tell, R.A., Rosenthal, R., 2008. Selective publication of antidepressant trials and its influence on apparent efficacy. N. Engl. J. Med. 358, 252–260.

Vase, L., Riley, J.L., Price, D.D., 2002. A comparison of placebo effects in clinical analgesic trials versus studies of placebo analgesia. Pain 99, 443–452.

Wampold, B.E., Minami, T., Tierney, S.C., Baskin, T.W., Bhati, K.S., 2005. The placebo is powerful: estimating placebo effects in medicine and psychotherapy from randomized clinical trials. J. Clin. Psychol. 61, 835–854.

Williams, C.L., Tappen, R.M., 2008. Exercise training for depressed older adults with Alzheimer's disease. Aging Men Health 12, 72–80.

Williams, J.W., Barrett, J., Oxman, T., Frank, E., Katon, W., Sullivan, M., Cornell, J., Sengupta, A., 2000. Treatment of dysthymia and minor depression in primary care: a randomized controlled trial in older adults. J. Am. Med. Assoc. 284, 1519–1526.

Yesavage, J.A., Brink, T.L., Rose, T.L., Lum, O., Huang, V., Adey, M., Leirer, V.O., 1983. Development and validation of a geriatric depression screening scale: a preliminary report. J. Psychiatr. Res. 17, 37–49.

Zigmond, A.S., Snaith, R.P., 1983. The hospital anxiety and depression scale. Acta Psychiatr. Scand. 67, 361–370.

Appendix C

Study III:

Locher, C., Frey Nascimento, A., Kossowsky, J., Meyer, A., & Gaab, J. (2017). *Is the rationale more important than deception? A randomized controlled trial of open-label placebo analgesia*. Manuscript submitted for publication.

Is the Rationale More Important than Deception? A Randomized Controlled Trial of Open-Label Placebo Analgesia

Cosima Locher[1,5], Antje Frey Nascimento[1], Irving Kirsch[2], Joe Kossowsky[1,3], Andrea Meyer[4], Jens Gaab[1]

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

[2] Program in Placebo Studies, Beth Israel Deaconess Medical Center, Harvard Medical School (HMS), Boston, MA, USA

[3] Department of Anesthesiology, Perioperative and Pain Medicine, Boston Children's Hospital, Harvard Medical School (HMS), Boston, MA, USA

[4] Department of Psychology, Clinical Psychology and Epidemiology, University of Basel, Switzerland

[5] Address for correspondence: Locher, Cosima, M.Sc., Department of Clinical Psychology and Psychotherapy, University of Basel, Missionsstrasse 62a, 4055 Basel, Switzerland. Email: cosima.locher@unibas.ch

Number of text pages of the entire manuscript: 31

Number of tables: 2

Number of figures: 2

Abstract

Research on open-label placebos questions whether deception is a necessary characteristic of placebo effects. Yet, comparisons between open-label and deceptive placebos are lacking. We therefore assessed effects of open-label and deceptive placebos in comparison to no treatment with a standardized experimental heat pain paradigm in a RCT in healthy participants. Participants ($N = 160$) were randomly assigned to no treatment (NT), open-label placebo without rationale (OPR^-), open-label placebo with rationale (OPR^+) and deceptive placebo (DP). We conducted baseline and posttreatment measurements of heat pain threshold and tolerance. Apart from the NT, all groups received an application of a placebo cream. Primary outcomes were planned comparisons of heat pain tolerance and the corresponding intensity and unpleasantness ratings. Objective posttreatment pain tolerance did not differ among groups. However, for subjective heat pain ratings at the posttreatment tolerance level, groups with a rationale (OPR^+ and DP) reported diminished heat pain intensity ($t(146) = -2.15, p = .033, d = 0.43$) and unpleasantness ratings ($t(146) = -2.43, p = .016, d = 0.49$) compared to the OPR^- group. Interestingly, the OPR^+ and the DP groups did not differ in heat pain intensity ($t(146) = -1.10, p = .272$) or unpleasantness ratings ($t(146) = -0.05, p = .961$) at posttreatment tolerance level. Our findings reveal that placebos with a plausible rational are more effective than placebos without a rationale. Even more, open-label placebos do not differ in their effects from deceptive placebos. Therefore, we question the ubiquitously assumed necessity of concealment in placebo administration.

Keywords

Pain; open-label placebos; TSA II; rationale; deception; heat pain paradigm

Introduction

A vast body of research corroborates the substantial benefit of placebos on healthy participants [24,56,63] as well as on certain clinical conditions [25,32,35,39], and for some disorders placebo effects are even as effective as active medication [68]. Yet, the implementation of deceptive placebos in clinical practice is ethically unfeasible and incompatible with key principles of openness and patient autonomy [6]. However, recent evidence generally questions whether deception is indeed a necessary characteristic of the placebo effect and suggests the possibility of openly prescribed placebos with full transparency [1,9,30,31,34,46,49,58].

Several open-label placebo studies have been conducted with full disclosure and the provision of a scientific rationale, i.e. explanations of the effects and mechanisms of placebos [30,31,34], and thereby aimed to alter subjective expectation by the means of verbal suggestions [50]. Open-label placebo administration led to symptom reduction of irritable bowel syndrome [31] and juvenile ADHD [58,59] when compared to no treatment, and two further studies underpinned the effectiveness of open-label placebos compared to treatment as usual conditions in patients suffering from rhinitis [60] and chronic low back pain [9]. In contrast, a pilot open-label study with patients suffering from major depression did not find any significant improvements compared to a waiting list control group; yet a medium effect size for open-label placebos was found, exceeding standardized drug-placebo differences found in antidepressant RCT's [34].

Despite these promising results, open-label placebos with full disclosure have not been directly compared to deceptive placebo administration and they have yet to be studied in an experimental analgesia paradigm although pain is the best examined condition in placebo research [3] and a current meta-analysis emphasizes the high susceptibility of pathological pain to placebo effects [14]. We therefore set out to examine the effects of open-label placebos in a standardized heat pain experiment [22,37,40] with healthy participants. We

compared open-label placebo administration with a rationale (OPR^+) and without a rationale (OPR^-) to deceptive placebo (DP) administration and a no treatment (NT). Primary outcomes were heat pain tolerance and the corresponding intensity and unpleasantness ratings. The following hypotheses were tested. First, we predicted that participants' heat pain analgesia (i.e., an increase in heat pain tolerance and a decrease of corresponding intensity and unpleasantness ratings) is enhanced after the application of a placebo cream in groups with a convincing rationale (OPR^+ and DP) compared to subjects receiving a placebo cream without any rationale (OPR^-). Second, we hypothesized that deception (DP) promotes placebo analgesia more efficiently than an open-label placebo administration with a rationale (OPR^+). Third, we hypothesized that the three groups with an application of a placebo cream (OPR^-, OPR^+, and DP) would show an enhanced heat pain analgesia when compared to the NT group.

Methods

Study Design

Between January 2016 and July 2016, we conducted a randomized-controlled trial (RCT) at the Division of Clinical Psychology and Psychotherapy at the University of Basel, Switzerland. Written informed consent was obtained from each participant before participation in the study. The Local Ethics Committee of the Canton Basel, Switzerland, approved the design and informed consent of the study. The study is registered at ClinicalTrials.gov: NCT02578420.

Study Population

160 healthy adults from the general population were recruited via advertisements for "a novel mind-body management study of individual pain perception" [31]. They had to be healthy by self-report, right-handed, aged between 18 and 65 years and have sufficient German language skills. Exclusion criteria were any acute or chronic diseases (e.g., chronic

pain, hypertension, heart disease, renal disease, liver disease, diabetes) as well as skin pathologies, neuropathies or nerve entrapment symptoms, or any other sensory abnormalities affecting the tactile or thermal modality. Participants were also excluded if they took medications (e.g., psychoactive medication, narcotics, or intake of analgesics), were in psychological or psychiatric treatment, reported current or regular drug consumption, or consumed more than three alcoholic standard beverages per day. Finally, we did not include Psychology or Medicine students, since they potentially have a previous knowledge of placebo mechanisms and effects. All participants were paid 50 Swiss francs for their participation in the study.

Study Procedure

Participants were randomly assigned to no treatment (NT; $N = 40$), open-label placebo without a rationale (OPR^-; $N = 40$), open-label placebo with a rationale (OPR^+; $N = 40$), and deceptive placebo (DP; $N = 40$). Upon arrival, all participants performed an objective baseline assessment of heat pain, as well as subjective heat pain ratings (for a description of the heat pain assessments, see below). After baseline measurements, the treatment phase was conducted.

Participants in the NT group did not receive any treatment and were told that they are in the "no treatment group". All participants in the three other groups (OPR^-, OPR^+, and DP) received an inert white placebo cream. However, the provided rationale in the three groups differed. In the OPR^- group, participants were informed that they are receiving an inert placebo cream and no additional information regarding placebo mechanisms was provided. In the OPR^+ group, participants were informed that they are receiving an inert placebo cream. In accordance with Kaptchuk et al. [31], the investigator explained that (a) the placebo effect is powerful, (b) the body can automatically respond to placebos like Pavlov's dogs who salivated when they heard a bell, and (c) a positive attitude towards placebos can be helpful

but is not necessary. We did not mention the importance of compliance (i.e., taking the placebo faithfully is critical) as in Kaptchuk et al. [31], since our treatment consisted of a single application of the placebo cream. In the DP group, participants were told that they are receiving an analgesic cream – named "Antidolor", containing the active substance Lidocaine – while in fact they received an inert placebo cream.

After the treatment phase, heat pain measurement procedures were performed again (posttreatment), regardless of group allocation.

Randomization and Blinding

The random allocation sequence was created using the built-in random number generator in Microsoft Excel for Mac, version 15 (Microsoft Corp, Redmond, WA). Regarding the location of the heat pain stimuli on the left forearm, block randomization was used such that an equal number of participants followed the same location sequence in each group (see below for detailed description). Participants were enrolled and assigned to treatments by the first author (CL). In order to implement the random allocation sequence, investigators received the participant number, group allocation and patch position sequence before the start of the trial.

Due to the study design, only participants from the deceptive group were blinded. Study investigators knew the allocation code at the start of the trial.

Objective Heat Pain Threshold and Tolerance and Corresponding Subjective Heat Pain Ratings

Pain sensation was assessed using the suprathreshold method of the Thermo Sensory Analyser (TSA-II). The TSA-II is a commonly used and safe device to study analgesic effects. To prevent physical injuries, the measurement stops automatically at a maximum temperature of 52°C. The thermode of the TSA-II was fixed on two different locations (location A and B; determined by using a positioning device) on the left volar forearm [45], applying a randomly

counterbalanced order within each group. Half of the participants started with the location A for the baseline measurements of heat pain stimuli, followed by the location B for the posttreatment measurements (for the other half of the participants, it was the exact opposite). The thermode of the TSA-II was moved to different locations in order to prevent effects of sensitization or habituation [16].

Prior to the actual measurements, participants were made familiar with the heat stimuli and the handling of the controlling device. Heat pain threshold was measured to determine the point when the sensation went from being warm to feeling painful using the method of limits, starting at 32 °C. Participants were asked to stop the heat stimulus at the point they feel it changes from "hot" to "painful" with a rise of 0.5 °C every second. Heat pain thresholds were assessed three times [55]. Heat pain tolerance was also determined using the method of limits: participants were asked to stop the increasing heat stimulus at the moment they could not stand the heat any longer. Three measurements were taken, each starting at 32 °C, with a rise of 0.5 °C every second [26]. Heat pain threshold was assessed before measuring heat pain tolerance in order to minimize interference between heat pain threshold and tolerance [38].

Following each heat pain threshold and tolerance stimulation, participants were asked to rate pain intensity and unpleasantness on Visual Analogue Scales (VAS) [53]. The intensity VAS with a range from 0 up to 100 was titled at the left by the descriptors "no pain sensation" and at the right by "the most intense pain sensation imaginable". Similarly, the unpleasantness VAS (ranged from 0-100) was anchored by the descriptors "not at all unpleasant" and "the most unpleasant imaginable". Subjective pain intensity and unpleasantness are commonly assessed pain dimensions in heat pain paradigm studies [48]. Intensity entails the cognitive dimensions of pain, whereas unpleasantness comprises the affective dimension of pain [51]. Our design allows for the assessment of individual heat pain threshold and tolerance temperatures for each of the three baseline and posttreatment stimulations as well as for the corresponding subjective heat pain intensity and unpleasantness ratings [40].

Measures and Questionnaires

At screening, demographic variables (i.e., age, gender, nationality, family status, educational level, employment situation, and income) were assessed.

Furthermore, after the treatment phase and before the second heat pain measurement procedures (posttreatment), pain expectancy and desire for relief were assessed. In order to measure expectancy of relief, we deployed a VAS to assess the expected pain intensity ("What do you expect the pain intensity to be after the application of the cream?") and pain unpleasantness ("What do you expect the pain unpleasantness to be after the application of the cream?"). Pain expectancy ratings were made on the same VAS (ranging from 0-100) as those for pain intensity and pain unpleasantness [53]. The Desire for Relief Scale (DRS) assessed the participants desire for relief on a VAS (ranging from 0-100) with the anchors "no desire for relief" on the left and "the most intense desire for relief imaginable" on the right [53].

After the posttreatment phase, the credibility of the treatment was measured. Participants from the DP group were asked to rate whether they believed they had received an analgesic cream (Likert scale from 1 = "I was sure that I received an analgesic cream", 2 = "I doubted whether I received an analgesic cream" and 3 = "I did not belief that I received an analgesic cream"), whereas participants from the OPR^- and OPR^+ group had to report whether they believed they had received a placebo (Likert scale from 1 = "I was sure that I received a placebo cream", 2 = "I doubted whether I received a placebo cream" and 3 = "I did not believe that I received a placebo cream"). As in previous studies [55], participants who disbelieved that they had received a placebo cream or an analgesic cream, respectively (i.e., rating 3 on the Likert scale) were excluded from analyses. Further, all participants had to fill out the placebo interventions questionnaire (Likert scale from 1 = "I know the term placebo and I can describe it in my own words", 2 = "I have heard about the term placebo but I do not know what it is" and 3 = "I have never heard the term placebo before") [17]. Participants who

were randomized to the OPR^- or OPR^+ group, but could not define the term placebo at the end of treatment (i.e., rating 2 or 3 on the Likert scale) were also excluded from analyses.

Statistical Analyses

Primary outcomes were objective heat pain tolerance and the corresponding subjective intensity and unpleasantness ratings. Objective heat pain threshold and the corresponding subjective intensity and unpleasantness ratings, as well as pain expectancy and desire for relief were chosen as secondary outcomes. We decided to use heat pain tolerance and the corresponding subjective intensity and unpleasantness ratings as primary outcomes since heat pain tolerance has been shown to be more connected with affective and motivational aspects than heat pain threshold which has been shown to be more associated with a sensory discrimination of nociceptive quality [22,27,42]. Hence, heat pain tolerance entails the experience of maximal discomfort, leading to enhanced subjective distress [22] and has further been shown to be linked to pathological pain [15].

For our objective primary outcome, i.e., posttreatment heat pain tolerance, we calculated one one-way analysis of covariance (ANCOVA), using the treatment group (NT, OPR^-, OPR^+, and DP) as independent between-subject factor and baseline heat pain tolerance as covariate. For our subjective primary outcomes, i.e., intensity and unpleasantness ratings for heat pain tolerance at posttreatment, we conducted two separate one-way ANCOVAs with treatment group as between-subject factor and the corresponding outcome variable measured at baseline as covariate. For all three primary outcome ANCOVAs, we tested three orthogonal planned contrasts (two-tailed): (c1) NT group vs groups with a cream application (OPR^-, OPR^+, and DP) which is in line with the recommendation to compare the NT group against all treatment groups [18]; (c2) OPR^- group vs groups with a rationale (OPR^+ and DP) in order to test the significance of the rationale; and (c3) OPR^+ group vs DP group in order to evaluate the significance of deception. Here, we decided to define the contrasts a priori since it has

been recommended to use planned contrasts instead of post hoc tests: they reduce the risk of type I errors, derive from specific hypotheses and complex comparisons can be accommodated [53,57].

Regarding primary outcomes, two sensitivity analyses were calculated: First, completer analyses were applied, whereby all participants who were randomized and who finished the experiment were included in the analyses ($N = 159$). Second, we tested whether there is a significant correlation between objective posttreatment heat pain tolerance and the corresponding subjective intensity and unpleasantness ratings. Then, we calculated additional ANCOVAs for subjective heat pain ratings by including the objective heat pain tolerance as a further covariate [18]. A detailed description of the statistical procedure used for the secondary outcomes (i.e. objective heat pain threshold and the corresponding subjective intensity and unpleasantness ratings) can be found in the Supplement (S1).

Regarding subjective expectancy ratings, we calculated two one-way ANCOVAs, with expectancy ratings of intensity and unpleasantness for heat pain tolerance, respectively, as outcome, defining the treatment group as an independent between-subject factor. Corresponding subjective ratings for heat pain tolerance at baseline were included as covariates. Further, desire for relief ratings were subjected to one one-way ANCOVA with treatment group as independent between-subject factor. No orthogonal planned contrasts were defined for all secondary outcomes. However, as the ANCOVA cannot provide detailed information on differences between particular pairs of treatment groups, post hoc tests were conducted for analysis with a significant omnibus test for treatment group, applying the Benjamini-Hochberg (BH) multiple testing correction [28].

Cohen's d was computed to provide an effect size estimate and were interpreted as small ($d = 0.2$), medium ($d = 0.5$), and large ($d = 0.8$) based on benchmarks suggested by Cohen [11]. All hypotheses were tested with an alpha-level of $p \leq 0.05$. On the basis of an omnibus test in a one-way analysis of variance with four groups and 5% error level, we

estimated that the total sample size of $N = 160$ would provide 99% power to detect a large effect ($f = 0.4$, effect size calculation based on Kaptchuk et al. [31]) and 75% power to detect a medium effect ($f = 0.25$, effect size estimation based on Kam-Hansen et al. [30]).

All statistical analyses were computed using R for Mac, version 3.3.2. (R Foundation; Vienna, Austria).

Results

Sample Characteristics

Participants had a mean age of 27.15 (SD 9.51) years and 68% of the participants were female (see Table 1 for more descriptive details). In total, we included 151 participants: One participant from the OPR^- group did not have sufficient German language skills and had to be excluded during the trial by the investigator. Two participants in the OPR^- and three participants in the OPR^+ group had to be excluded since they could not define the term placebo at the end of treatment. Further, three participants in the DP group who did not believe that they received an analgesic cream were not included in the analyses (see Supplement SF1).

Objective Heat Pain Tolerance

Planned contrasts indicated that the groups did not differ regarding their objective heat pain tolerance at posttreatment (NT vs. OPR^-, OPR^+, and DP: $t(146) = 0.35$, $p = .724$; OPR^- vs. OPR^+ and DP: $t(146) = 1.15$, $p = .254$; OPR^+ vs. DP: $t(146) = 0.37$, $p = .711$). The results did not change in the additional completer analyses (see Supplement ST1).

Subjective Intensity and Unpleasantness Ratings for Heat Pain Tolerance

In contrast to objective heat pain tolerance, the corresponding subjective intensity and unpleasantness ratings did differ among groups (see Table 2 for group means). First, planned contrasts indicated that the NT group and the other three groups did not differ in heat pain

intensity ratings at posttreatment (c1: $t(146) = -0.44$, $p = .658$). However, the two groups with a rationale (OPR^+ and DP) showed significantly lower ratings at posttreatment when compared to the OPR^- group (c2: $t(146) = -2.15$, $p = .033$, $d = 0.43$). Further, the OPR^+ and DP group did not differ from each other (c3: $t(146) = -1.10$, $p = .272$; see Figure 1).

Results for heat pain unpleasantness ratings at posttreatment were similar. No difference was found between the NT group and the three other groups (c1: $t(146) = -1.38$, $p = .169$). Participants in the two groups with a rationale (OPR^+ and DP) reported lower ratings at posttreatment compared to participants from the OPR^- group (c2: $t(146) = -2.43$, $p = .016$, $d = 0.49$), and the OPR^+ and DP group did not differ from each other (c3: $t(146) = -0.05$, $p = .961$; see Figure 2).

The results of the primary outcomes did not change significantly in the additional sensitivity analyses. First, the completer analyses revealed comparable results ($N = 159$) (see Supplement ST1). Second, the results of the subjective heat pain intensity and unpleasantness ratings did not change when objective heat pain tolerance was included as an additional covariate (see Supplement ST2). Also, the correlations between objective heat pain tolerance and the corresponding subjective intensity ($p = .906$), as well as unpleasantness ($p = .462$) ratings, were not significant.

Objective Heat Pain Threshold and Corresponding Subjective Intensity and Unpleasantness Ratings

Secondary analyses of objective heat pain threshold and the corresponding subjective intensity and unpleasantness ratings did not reveal distinct significant findings; detailed outcomes can be found in the Supplement (S3, S4, ST3).

Expectancy and Desire for Relief Ratings

For all groups, expectancy ratings of intensity and unpleasantness for posttreatment heat pain tolerance and threshold can be found in Table 2 and ST3, respectively. The

ANCOVA revealed that the treatment groups differed in expectancy ratings of intensity for posttreatment heat pain tolerance (omnibus test $F(3, 146) = 5.41, p = .001$). Post hoc comparisons revealed that the difference between the NT and the DP groups was significant, indicating that participants in the NT group expected a significantly higher heat pain intensity than the DP group ($p = .038$). Findings for expectancy ratings of unpleasantness for posttreatment heat pain tolerance were similar, again resulting in a significant omnibus test for treatment group ($F(3, 146) = 5.00, p = .002$), whereby post hoc comparisons revealed that the NT and DP group significantly differed from each other ($p = .039$). The ANCOVA for desire of relief, however, did not indicate a significant treatment group effect (omnibus test, $F(3, 147) = 0.70, p = .555$).

Discussion

Despite their clinical potential, very little is known about open-label placebos as basic research on their analgesic effects as well as comparisons to deceptive placebos is lacking. We addressed this dearth and conducted – to the best of our knowledge – the first RCT comparing open-label placebos with and without a rationale directly to a deceptive group in an experimental standardized heat pain paradigm. We found that healthy participants given open-label placebos with a persuasive rationale showed a decline in subjective intensity and unpleasantness ratings for heat pain tolerance which, surprisingly, did not differ significantly from deceptive placebo administration. This is in line with an older study, showing that for a verum (i.e., naproxen) as well as for a placebo, the analgesic effect is significantly better in the informed-consent group when compared to the uninformed group [4]. Accordingly, the necessity of deception in placebo application – at least in healthy participants – needs to be reconsidered.

In line with our prediction, the provision of a convincing rationale, either open or deceptive (OPR^+ and DP group), did outperform the mere cream application without any

rationale (OPR$^-$ group) with regard to subjective ratings for heat pain tolerance. The impact of the comprehensible theoretical embedding, which was offered to the participants in both groups, emphasizes the importance of plausibility and conviction of treatment rationales [7].

Further, we did not find group differences between the no treatment group and the combined effect of the three other groups which may be due to the conceptual heterogeneity between the three groups receiving the placebo cream. This finding, which stands in contrast to our hypothesis, may be due to the fact that the OPR$^-$ group had numerically higher subjective ratings of heat pain tolerance than the NT group and in contrast to the other groups, the OPR$^-$ group did not report pain analgesia in any of the subjective outcomes but rather displayed higher subjective ratings of heat pain tolerance at posttreatment compared to baseline scores. This stands in contrast to a study by Kam-Hansen et al. [30], where an open application of placebos without additional information lead to a 14.5% pain reduction in patients with episodic migraine, which differed significantly from the untreated attacks, where patients reported a pain increase of 15.4%. In our case, it is possible that participants in the OPR$^-$ were disappointed to "only receive a placebo." In fact, the experience of disappointment is an issue in control groups as well as in in experiments with healthy participants [41,62]. Further, the application of the cream may have led to an increased focus on the forearm in combination with the absence of any cognitive processing of a rationale [65].

We only found significant group differences in subjective heat pain ratings and not on objective heat pain tolerance. Whether placebo effects are merely detected in subjective measurements or also on objective parameters seems to depend on the object of investigation as well as on further conceptual aspects of studies. Thus, several studies detected placebo effects only on a subjective level [29,61,69], whereas placebo responses are indeed measurable with neuroimaging [36] and the impact of placebo on immune and endocrine processes has been reported [3]. Our results are in line with the view that placebos primarily affect subjective self-report and self-appraisal symptoms [33]. It is noteworthy that Kam-

Hansen et al. [30] also only found an impact of openly prescribed placebos on subjective measures and not on pain freedom – which would be the absolute absence of pain and is more objective than subjective pain sensation ratings. Interestingly, a former account of a heat pain study in children showed reversed effects – placebo treatment responses were detected on an objective level only (i.e. heat pain threshold and tolerance), however not on subjective heat pain ratings [37].

Regarding heat pain expectancy ratings, participants' expectations differed substantially according to group allocation, indicating that expectancy manipulations occurred due to the differing treatment procedures. In particular, participants in the DP group expected reduced heat pain intensity and unpleasantness for posttreatment compared to the NT group. These findings are in line with evidence of a link between placebo effects and induced expectations [47] and the general association between enhanced analgesia and induced positive expectations [5], besides showing that the deception in the DP group was successful.

To sum up, placebo analgesia in healthy participants may be achieved with an adequate and convincing rationale whereby deception may not be necessary as long as participants find the explanation at hand meaningful and plausible [44].

Strengths and Limitations

In any case, our study corroborates previous findings of open-label placebo effectiveness and shows that verbal constructs and the rationales, in particular play a fundamental role in altering expectancies and, hence, induce a placebo response [50]. Remarkably, this placebo response is achievable even with full disclosure. It has to be considered that we examined open-label placebo analgesia in a healthy population, hence the effect may be higher than in a comparable clinical setting [66,67]. However, a recent meta-analysis suggests that patients with clinical conditions benefit even more from analgesic placebo treatments than healthy participants, and that clinical pain conditions and pain

induced in experiments may respond equally to placebo application [20]. In this regard, the increased desire for pain relief that significantly contributes to placebo analgesia in patients [52] has been shown to be lower in healthy volunteers [10]. Also, the value of social support regarding positive physical health outcomes [23] turns out to be superfluous in experimental pain.

This RCT has several limitations. Importantly, subjective intensity and unpleasantness ratings are not independent of the corresponding objective heat pain tolerance. Differences in subjective heat pain ratings could be suppressed since assessing heat pain tolerance allows participants to stop the pain stimulus at different points in time. However, in our study, objective heat pain tolerance and the corresponding subjective ratings were not correlated. Further, additional sensitivity analyses of the subjective primary outcomes in which we statistically controlled for participants' individual heat pain tolerance did not lead to different results (see Supplement ST2)

We found medium effect sizes for subjective heat pain ratings. However, it is well known that placebo analgesic effects are smaller in studies that use short-term stimuli (as did our study) when compared to long-term pain stimuli (> 20s) [64,66]. Finally, advertisement of a "novel mind–body study of individual pain perception" may have selectively attracted individuals who are open to new approaches and concepts [9]. Nevertheless, selective attraction to the advertised study is present in almost all experimental trials [31].

Implications and Future Studies

Given the long-held belief of an inextricable interconnection between deceit and placebo usage [19,43], our findings offer an empirical starting point for a conceptual rethinking of the necessity of deception in placebo application. The ethically problematic aspect of placebos – a spurious rationale [6] – may under certain circumstances be comparable to a transparent and scientific rationale – at least in terms of their effects on

subjective outcomes. In this sense, our study also affirms that open-label placebos bare the potential to be harnessed in clinical practice, especially concerning the alleviation of subjective ailments. Authors of earlier studies already argued that open-label placebos could be prescribed with a "wait and watch" strategy before the administration of drugs [31] or to be prescribed after repeated administration of active drugs to achieve drug-like effects [13]. Moreover, open-label placebos may have the potential to work in treatment resistant patients, assumedly due to a form of empowerment [9], and therefore offers a unique approach to patients with chronic diseases without conventional treatment response.

Most intriguingly, the non-effect of our placebo group without any theoretical embedding indicates the special significance of the rationale itself. Clinicians should be aware that a convincing story behind an intervention leads to better outcomes - at least concerning openly prescribed placebos. The importance of a certain rationale, i.e. a verbal suggestion, is also of relevance regarding the augmentation of nocebo effects [54] and in other domains such as in psychotherapy [7,21]. Therefore, our findings emphasise that the power of verbal suggestions should not be underestimated and deserves further scrutiny in relation to future placebo research. This is also in line with the recommendation that physicians may best benefit from placebo effects by enhancing patients' expectations through communication [2,8]. Here, the claim that the proceeding study of placebo effects must go hand in hand with ethical debates to avoid misuse [12] must be picked up again.

Conclusion

Open-label placebos can lead to relevant changes in subjective experiences in healthy participants whereby the rationale itself is the vehicle which transports the meaning: An open application of a placebo cream with a convincing rationale had an impact on subjective pain relief, whereas an open application of a placebo cream without any rationale or theoretical embedding showed no effects. The observation that open-label placebos have the same effects

on our subjective primary outcome as deceptive placebos indicates that the rationale in fact might bear more weight than the deception in placebo application.

Acknowledgements

This study was supported by a grant project (325130_170117) awarded to JG by the Swiss National Science Foundation. The authors wish to thank Nafisa Müller, Larissa Tellini, and Christoph Werner for providing excellent help with the recruitment. We also want to thank Dr. Markus Messerli for his helpful suggestions regarding the design of the placebo cream. The authors are also very grateful for the support of Sebastian Hasler.

Declaration of Interests

The authors declare no financial interest or potential conflicts of interest.

References

[1] Aulas JJ, Rosner I. [Efficacy of a non blind placebo prescription]. Encephale 2003;29:68–71.

[2] Barrett B, Muller D, Rakel D, Rabago D, Marchand L, Scheder JC. Placebo, meaning, and health. Perspect Biol Med 2006;49:178–98.

[3] Benedetti F. Placebo effects (2nd ed). Oxford, UK: Oxford University Press, 2014.

[4] Bergmann JF, Chassany O, Gandiol J, Deblois P, Kanis JA, Segrestaa JM, Caulin C, Dahan R. A randomised clinical trial of the effect of informed consent on the analgesic activity of placebo and naproxen in cancer pain. Clin Trials Metaanal 1994;29:41–7.

[5] Bingel U, Wanigasekera V, Wiech K, Ni Mhuircheartaigh R, Lee MC, Ploner M, Tracey I. The effect of treatment expectation on drug efficacy: imaging the analgesic benefit of the opioid remifentanil. Sci Transl Med 2011;3:70ra14.

[6] Blease C, Colloca L, Kaptchuk TJ. Are open-label placebos ethical? Informed consent and ethical equivocations. Bioethics 2016;30:407–14.

[7] Borkovec TD, Nau SD. Credibility of analogue therapy rationales. J Behav Ther Exp Psychiatry 1972;3:257–60.

[8] Bystad M, Bystad C, Wynn R. How can placebo effects best be applied in clinical practice? A narrative review. Psychol Res Behav Manag 2015;8:41–5.

[9] Carvalho C, Caetano JM, Cunha L, Rebouta P, Kaptchuk TJ, Kirsch I. Open-label placebo treatment in chronic low back pain: a randomized controlled trial. Pain 2016;157:2766–72.

[10] Charron J, Rainville P, Marchand S. Direct comparison of placebo effects on clinical and experimental pain. Clin J Pain 2006;22:204–11.

[11] Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.

[12] Colloca L, Benedetti F. Placebos and painkillers: is mind as real as matter? Nat Rev Neurosci 2005;6:545–52.

[13] Colloca L, Enck P, DeGrazia D. Relieving pain using dose-extending placebos: a scoping review. Pain 2016;157:1590–8.

[14] Cragg JJ, Warner FM, Finnerup NB, Jensen MP, Mercier C, Richards JS, Wrigley P, Soler D, Kramer JL. Meta-analysis of placebo responses in central neuropathic pain: impact of subject, study, and pain characteristics. Pain 2016;157:530–40.

[15] Edwards RR, Doleys DM, Fillingim RB, Lowery D. Ethnic differences in pain tolerance: clinical implications in a chronic pain population. Psychosom Med 2001;63:316–23.

[16] Emerson NM, Zeidan F, Lobanov OV, Hadsel MS, Martucci KT, Quevedo AS, Starr CJ, Nahman-Averbuch H, Weissman-Fogel I, Granovsky Y, Yarnitsky D, Coghill RC. Pain sensitivity is inversely related to regional grey matter density in the brain. Pain 2014;155:566–73.

[17] Fässler M, Gnädinger M, Rosemann T, Biller-Andorno N. Placebo interventions in practice: a questionnaire survey on the attitudes of patients and physicians. Br J Gen Pract 2011;61:101–7.

[18] Field A, Jeremy M, Field Z. Discovering statistics using R. London, UK: SAGE Publications, 2012.

[19] Foddy B. A duty to deceive: placebos in clinical practice. Am J Bioeth 2009;9:4–12.

[20] Forsberg JT, Martinussen M, Flaten MA. The placebo analgesic effect in healthy individuals and patients: a meta-analysis. Psychosom Med 2017;79:388–94.

[21] Gaab J, Blease C, Locher C, Gerger H. Go open: a plea for transparency in psychotherapy. Psychol Conscious (Wash D C) 2016;3:175–98.

[22] Gaab J, Jiménez J, Voneschen L, Oschwald D, Meyer AH, Nater UM, Krummenacher P. Psychosocial stress-induced analgesia: an examination of effects on heat pain

threshold and tolerance and of neuroendocrine mediation. Neuropsychobiology 2017;74:87–95.

[23] Geers AL, Helfer SG, Kosbab K, Weiland PE, Landry SJ. Reconsidering the role of personality in placebo effects: dispositional optimism, situational expectations, and the placebo response. J Psychosom Res 2005;58:121–7.

[24] Geers AL, Weiland PE, Kosbab K, Landry SJ, Helfer SG. Goal activation, expectations, and the placebo effect. J Pers Soc Psychol 2005;89:143–59.

[25] Goetz CG, Wuu J, McDermott MP, Adler CH, Fahn S, Freed CR, Hauser RA, Olanow WC, Shoulson I, Tandon PK, Parkinson Study Group, Leurgans S. Placebo response in Parkinson's disease: comparisons among 11 trials covering medical and surgical interventions. Mov Disord 2008;23:690–9.

[26] Granot M, Sprecher E, Yarnitsky D. Psychophysics of phasic and tonic heat pain stimuli by quantitative sensory testing in healthy subjects. Eur J Pain 2003;7:139–43.

[27] Harris G, Rollman GB. The validity of experimental pain measures. Pain 1983;17:369–76.

[28] Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. Stat Med 1990;9:811–8.

[29] Hróbjartsson A, Gøtzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. N Engl J Med 2001;344:1594–602.

[30] Kam-Hansen S, Jakubowski M, Kelley JM, Kirsch I, Hoaglin DC, Kaptchuk TJ, Burstein R. Altered placebo and drug labeling changes the outcome of episodic migraine attacks. Sci Transl Med 2014;6:218ra5.

[31] Kaptchuk TJ, Friedlander E, Kelley JM, Sanchez MN, Kokkotou E, Singer JP, Kowalczykowski M, Miller FG, Kirsch I, Lembo AJ. Placebos without deception: a randomized controlled trial in irritable bowel syndrome. PLoS One 2010;5:e15591.

[32] Kaptchuk TJ, Kelley JM, Conboy LA, Davis RB, Kerr CE, Jacobson EE, Kirsch I, Schyner RN, Nam BH, Nguyen LT, Park M, Rivers AL, McManus C, Kokkotou E, Drossman DA, Goldman P, Lembo AJ. Components of placebo effect: randomised controlled trial in patients with irritable bowel syndrome. BMJ 2008;336:999–1003.

[33] Kaptchuk TJ, Miller FG. Placebo effects in medicine. N Engl J Med 2015;373:8–9.

[34] Kelley JM, Kaptchuk TJ, Cusin C, Lipkin S, Fava M. Open-label placebo for major depressive disorder: a pilot randomized controlled trial. Psychother Psychosom 2012;81:312–4.

[35] Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. PLoS Med 2008;5:e45.

[36] Krummenacher P, Candia V, Folkers G, Schedlowski M, Schönbächler G. Prefrontal cortex modulates placebo analgesia. Pain 2010;148:368–74.

[37] Krummenacher P, Kossowsky J, Schwarz C, Brugger P, Kelley JM, Meyer A, Gaab J. Expectancy-induced placebo analgesia in children and the role of magical thinking. J Pain 2014;15:1282–93.

[38] Kut E, Schaffner N, Wittwer A, Candia V, Brockmann M, Storck C, Folkers G. Changes in self-perceived role identity modulate pain perception. Pain 2007;131:191–201.

[39] Macedo A, Baños JE, Farré M. Placebo response in the prophylaxis of migraine: a meta-analysis. Eur J Pain 2008;12:68–75.

[40] Maeoka H, Hiyamizu M, Matsuo A, Morioka S. The influence of repeated pain stimulation on the emotional aspect of pain: a preliminary study in healthy volunteers. J Pain Res 2015;8:431–6.

[41] McCambridge J, Sorhaindo A, Quirk A, Nanchahal K. Patient preferences and performance bias in a weight loss trial with a usual care arm. Patient Educ Couns 2014;95:243–7.

[42] Merskey H. Psychological and psychiatric aspects of pain control. In: Smith WL, Merskey H, Gross SC, editors. Pain: Meaning and Management. New York, NY: SP Medical and Scientific Books, 1980. pp. 105–117.

[43] Miller FG, Wendler D, Swartzman LC. Deception in research on the placebo effect. PLoS Med 2005;2:e262.

[44] Moerman DE, Jonas WB. Deconstructing the placebo effect and finding the meaning response. Ann Intern Med 2002;136:471–6.

[45] Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 1971;9:97–113.

[46] Park LC, Covi L. Nonblind placebo trial: an exploration of neurotic patients' responses to placebo when its inert content is disclosed. Arch Gen Psychiatry 1965;12:36–45.

[47] Petersen GL, Finnerup NB, Grosen K, Pilegaard HK, Tracey I, Benedetti F, Price DD, Jensen TS, Vase L. Expectations and positive emotional feelings accompany reductions in ongoing and evoked neuropathic pain following placebo interventions. Pain 2014;155:2687–98.

[48] Petersen GL, Finnerup NB, Nørskov KN, Grosen K, Pilegaard HK, Benedetti F, Price DD, Jensen TS, Vase L. Placebo manipulations reduce hyperalgesia in neuropathic pain. Pain 2012;153:1292–300.

[49] Petkovic G, Charlesworth JEG, Kelley J, Miller F, Roberts N, Howick J. Effects of placebos without deception compared with no treatment: protocol for a systematic review and meta-analysis. BMJ Open 2015;5:e009428.

[50] Pollo A, Amanzio M, Arslanian A, Casadio C, Maggi G, Benedetti F. Response expectancies in placebo analgesia and their clinical relevance. Pain 2001;93:77–84.

[51] Price DD. Psychological and neural mechanisms of the affective dimension of pain. Science 2000;288:1769–72.

[52] Price DD, Finniss DG, Benedetti F. A comprehensive review of the placebo effect: recent advances and current thought. Annu Rev Psychol 2008;59:565–90.

[53] Price DD, Milling LS, Kirsch I, Duff A, Montgomery GH, Nicholls SS. An analysis of factors that contribute to the magnitude of placebo analgesia in an experimental paradigm. Pain 1999;83:147–56.

[54] Quinn VF, Livesey EJ, Colagiuri B. Latent inhibition reduces nocebo nausea, even without deception. Ann Behav Med 2017 [Epub ahead of print]. http://dx.doi.org/10.1007/s12160-016-9867-8

[55] Rief W, Glombiewski JA. The hidden effects of blinded, placebo-controlled randomized trials: an experimental investigation. Pain 2012;153:2473–7.

[56] Rosenzweig P, Brohier S, Zipfel A. The placebo effect in healthy volunteers: influence of experimental conditions on the adverse events profile during phase I studies. Clin Pharmacol Ther 1993;54:578–83.

[57] Ruxton GD, Beauchamp G. Time for some a priori thinking about post hoc testing. Behav Ecol 2008;19:690-3.

[58] Sandler AD, Bodfish JW. Open-label use of placebos in the treatment of ADHD: a pilot study. Child Care Health Dev 2008;34:104–10.

[59] Sandler AD, Glesne CE, Bodfish JW. Conditioned placebo dose reduction: a new treatment in attention-deficit hyperactivity disorder? J Dev Behav Pediatr 2010;31:369–75.

[60] Schaefer M, Harke R, Denke C. Open-label placebos improve symptoms in allergic rhinitis: a randomized controlled trial. Psychother Psychosom 2016;85:373–4.

[61] Schwarz KA, Büchel C. Cognition and the placebo effect--dissociating subjective perception and actual performance. PLoS One 2015;10:e0130492.

[62] Skingley A, Bungay H, Clift S, Warden J. Experiences of being a control group: lessons from a UK-based randomized controlled trial of group singing as a health promotion initiative for older people. Health Promot Int 2014;29:751–8.

[63] Turi Z, Mittner M, Paulus W, Antal A. Placebo intervention enhances reward learning in healthy individuals. Sci Rep 2017;7:41028.

[64] Tuttle AH, Tohyama S, Ramsay T, Kimmelman J, Schweinhardt P, Bennett GJ, Mogil JS. Increasing placebo responses over time in U.S. clinical trials of neuropathic pain. Pain 2015;156:2616–26.

[65] van Laarhoven AI, Kraaimaat FW, Wilder-Smith OH, Evers AW. Role of attentional focus on bodily sensations in sensitivity to itch and pain. Acta Derm Venereol 2010;90:46–51.

[66] Vase L, Petersen GL, Riley JL III, Price DD. Factors contributing to large analgesic effects in placebo mechanism studies conducted between 2002 and 2007. Pain 2009;145:36–44.

[67] Vase L, Riley JL III, Price DD. A comparison of placebo effects in clinical analgesic trials versus studies of placebo analgesia. Pain 2002;99:443–52.

[68] Wampold BE, Minami T, Tierney SC, Baskin TW, Bhati KS. The placebo is powerful: estimating placebo effects in medicine and psychotherapy from randomized clinical trials. J Clin Psychol 2005;61:835–54.

[69] Wechsler ME, Kelley JM, Boyd IO, Dutile S, Marigowda G, Kirsch I, Israel E, Kaptchuk TJ. Active albuterol or placebo, sham acupuncture, or no intervention in asthma. N Engl J Med 2011;365:119–26.

Figure Captions

Figure 1. Posttreatment scores of subjective intensity ratings for heat pain tolerance of participants in the no treatment (NT), open-label placebo without rationale (OPR^-), open-label placebo with rationale (OPR^+), and deceptive placebo (DP) group. Positive values indicate placebo analgesia. Scores were adjusted for baseline ratings.

Figure 2. Posttreatment scores of subjective unpleasantness ratings for heat pain tolerance of participants in the no treatment (NT), open-label placebo without rationale (OPR^-), open-label placebo with rationale (OPR^+), and deceptive placebo (DP) group. Positive values indicate placebo analgesia. Scores were adjusted for baseline ratings.

Table 1

Sociodemographic Characteristics of Study Participants

Group	N (included)	Age (SD)	N (%) Female	Family Status N(%)	Highest Educational Level N(%)	Employment Level N(%)
NT	40	27.9 (8.52)	29 (73%)	single: 37 (92.5%) married: 0 (0%) registered partnership: 1 (2.5%) divorced: 2 (5%)	primary school: 0 (0%) secondary school: 6 (15%) high school: 15 (37.5%) university: 19 (47.5%)	full-time: 5 (12.5%) part-time: 21 (52.5%) none or student: 14 (35%)
OPR$^-$	37	28.27 (11.34)	24 (65%)	single: 32 (86.5%) married: 5 (13.5%) registered partnership: 0 (0%) divorced: 0 (0%)	primary school: 0 (0%) secondary school: 4 (10.8%) high school: 23 (62.2%) university: 10 (27%)	full-time: 3 (8.1%) part-time: 15 (40.5%) none or student: 19 (51.4%)
OPR$^+$	37	25.7 (7.76)	27 (73%)	single: 34 (91.9%) married: 2 (5.4%) registered partnership: 0 (0%) divorced: 1 (2.7%)	primary school: 0 (0%) secondary school: 3 (8.1%) high school: 23 (62.2%) university: 11 (29.7%)	full-time: 6 (16.2%) part-time: 19 (51.4%) none or student: 12 (32.4%)
DP	37	26.65 (10.25)	23 (62 %)	single: 35 (94.6%) married: 1 (2.7%) registered partnership: 0 (0%) divorced: 1 (2.7%)	primary school: 1 (2.7%) secondary school: 5 (13.5%) high school: 19 (51.4%) university: 12 (32.4%)	full-time: 7 (18.9%) part-time: 14 (37.8%) none or student: 16 (43.3%)

Note. DP = deceptive placebo; NT = no treatment group; OPR$^-$ = open-label placebo without rationale; OPR$^+$ = open-label placebo with rationale.

Table 2

Objective Heat Pain Tolerance and Corresponding Subjective Intensity and Unpleasantness Ratings

Group	Baseline			Posttreatment[a]			Expectancy Ratings[b]	
	Objective Heat Pain Tolerance	Subjective Heat Pain Intensity	Subjective Heat Pain Unpleasantness	Objective Heat Pain Tolerance	Subjective Heat Pain Intensity	Subjective Heat Pain Unpleasantness	Heat Pain Intensity	Heat Pain Unpleasantness
NT (*N*=40)								
Outcomes (mean, se)	48.11 (0.27)	62.24 (3.80)	64.26 (4.10)	48.07 (0.14)	60.35 (1.36)	63.80 (1.48)	60.26 (2.21)	62.42 (2.44)
OPR$^-$ (*N*=37)								
Outcomes (mean, se)	48.15 (0.28)	59.02 (3.94)	64.28 (4.26)	47.99 (0.14)	62.13 (1.42)	64.47 (1.54)	53.38 (2.29)	54.52 (2.53)
OPR$^+$ (*N*=37)								
Outcomes (mean, se)	48.80 (0.28)	62.48 (3.94)	64.48 (4.26)	48.15 (0.14)	59.51 (1.42)	59.93 (1.54)	54.99 (2.29)	53.71 (2.53)
DP (*N*=37)								
Outcomes (mean, se)	48.32 (0.28)	59.11 (3.94)	60.77 (4.26)	48.23 (0.14)	57.30 (1.42)	59.83 (1.55)	47.52 (2.29)	49.09 (2.54)

Note: a=posttreatment means are adjusted for the corresponding baseline mean; b=expectancy means are adjusted for the corresponding baseline mean; DP = deceptive placebo; NT = no treatment group; OPR$^-$ = open-label placebo without rationale; OPR$^+$ = open-label placebo with rationale.

Figure 1.

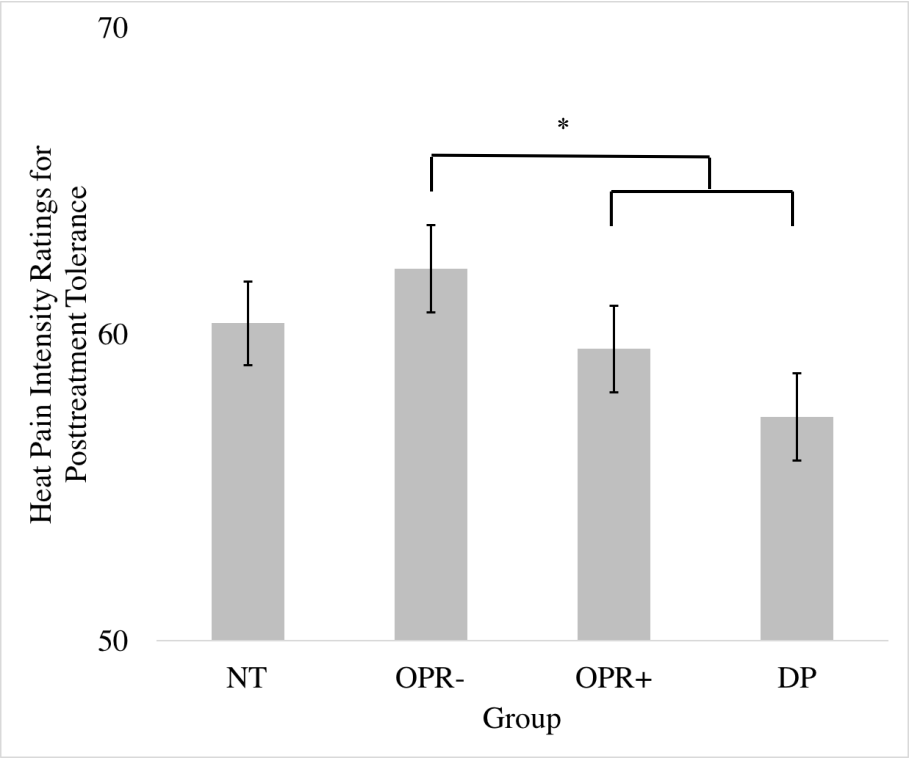
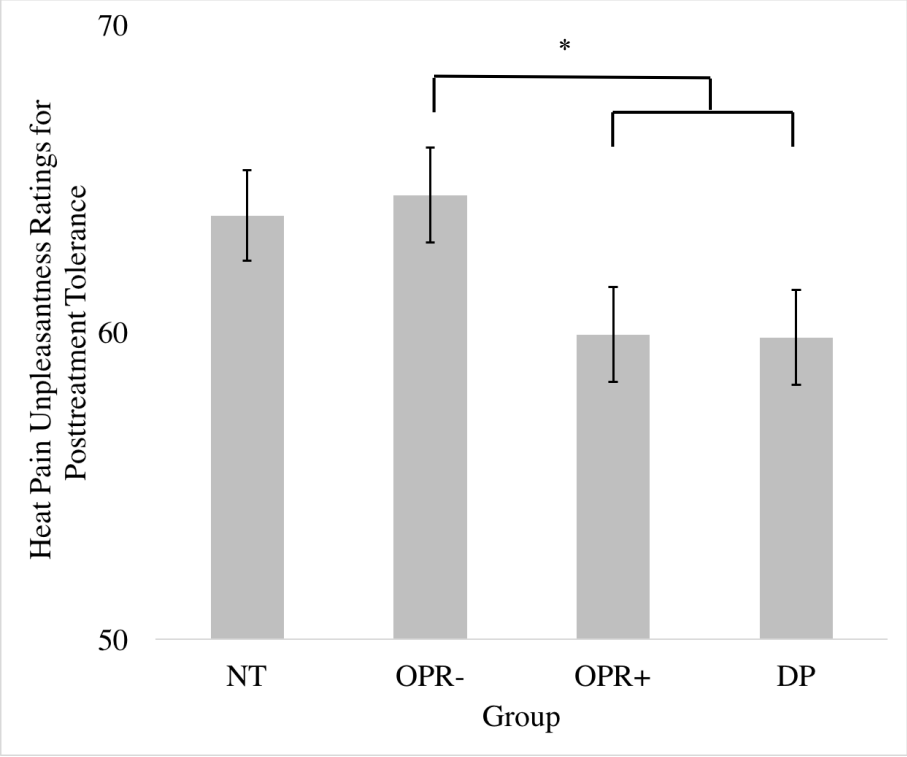


Figure 2.



SUPPLEMENTARY MATERIALS

Table of Contents

S1: Methods: Objective Heat Pain Threshold

For posttreatment heat pain threshold, we calculated one separate one-way ANCOVAs, using the treatment group as independent between-subject factor and baseline heat pain threshold as covariate.

S2: Methods: Subjective Ratings for Heat Pain Threshold

Regarding intensity and unpleasantness ratings for heat pain threshold at posttreatment, we conducted two separate one-way ANCOVAs with treatment group (NT vs OPR^- vs OPR^+ vs DP) as between-subject factor and the corresponding outcome variable measured at baseline as covariate.

S3: Results: Objective Heat Pain Threshold

The one-way ANCOVA for posttreatment heat pain threshold did not show a significant main effect for treatment group ($F(3, 146) = 0.17, p = .918$).

S4: Results: Subjective Ratings for Heat Pain Threshold

In the intensity ratings for heat pain threshold at posttreatment, we found no differences among the four groups (omnibus test, $F(3, 146) = 1.35, p = 0.259$). Similarly, the groups did not differ in the unpleasantness ratings for heat pain threshold (omnibus test, $F(3, 146) = 2.19, p = .091$).

ST1
Sensitivity Analyses for Primary Outcomes: Completer Analyses (N = 159)

	Objective Heat Pain Tolerance	**Subjective Heat Pain Intensity**	**Subjective Heat Pain Unpleasantness**
Statistical Analyses			
Contrast 1 (NT vs. OPR-, OPR+, and DP)	$t(154) = 0.28$, $p = .776$	$t(154) = -0.42$, $p = .676$	$t(154) = -1.35$, p = .178
Contrast 2 (OPR- vs. OPR+ and DP)	$t(154) = 1.00$, $p = .321$	$t(154) = -1.89$, $p = .061$	$t(154) = -2.19$, $p = .03$
Contrast 3 (OPR+ group vs DP)	$t(154) = -0.39$, $p = .699$	$t(154) = -1.24$, $p = .217$	$t(154) = -0.19$, $p = .853$

ST2
Sensitivity Analyses for Primary Outcomes: Subjective Intensity and Unpleasantness Ratings Adjusted for Objective Heat Pain Tolerance

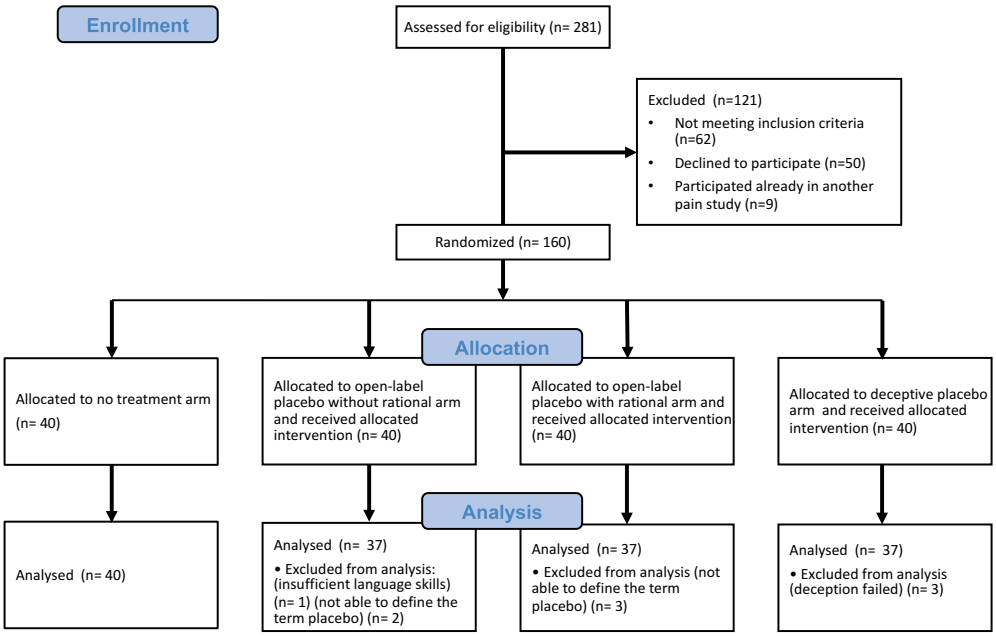
Tolerance	**Intensity**			**Unpleasantness**		
Statistical Analyses	**Adjusted for baseline heat pain tolerance**	**Adjusted for mean heat pain tolerance**	**Adjusted for posttreatment heat pain tolerance**	**Adjusted for baseline heat pain tolerance**	**Adjusted for mean heat pain tolerance**	**Adjusted for posttreatment heat pain tolerance**
Contrast 1 (NT vs. OPR-, OPR+, and DP)	$t(145) = -0.43$, $p = .665$	$t(145) = -0.45$, $p = .65$	$t(145) = -0.47$, p = .637	$t(145) = -1.44$, $p = .152$	$t(145) = -1.47$, $p = .143$	t(145) = -1.49, p = .138
Contrast 2 (OPR- vs. OPR+ and DP)	$t(145) = -2.12$, $p = .035$	$t(145) = -2.14$, $p = .034$	$t(145) = -2.17$, $p = .032$	$t(145) = -2.49$, $p = .014$	$t(145) = -2.54$, $p = .012$	$t(145) = -2.58$, $p = .011$
Contrast 3 (OPR+ group vs DP)	$t(145) = -1.10$, $p = .273$	$t(145) = -1.08$, $p = .281$	$t(145) = -1.07$, $p = .286$	$t(145) = -0.02$, $p = .984$	$t(145) = 0.03$, $p = .973$	$t(145) = -0.03$, $p = .976$

ST3
Objective Heat Pain Threshold and Corresponding Subjective Intensity and Unpleasantness Ratings

Group	Baseline			Posttreatment[a]			Expectancy Ratings[b]	
	Objective Heat Pain Tolerance	Subjective Heat Pain Intensity	Subjective Heat Pain Unpleasantness	Objective Heat Pain Tolerance	Subjective Heat Pain Intensity	Subjective Heat Pain Unpleasantness	Heat Pain Intensity	Heat Pain Unpleasantness
NT (*N*=40)								
Outcomes (mean, se)	44.25 (0.44)	29.89 (3.48)	33.09 (3.63)	43.49 (0.33)	23.57 (2.11)	23.34 (2.19)	41.98 (3.33)	41.62 (3.35)
OPR$^-$ (*N*=37)								
Outcomes (mean, se)	44.81 (0.46)	29.11 (3.61)	35.31 (3.77)	43.64 (0.34)	28.74 (2.19)	30.68 (2.28)	37.18 (3.47)	38.76 (3.49)
OPR$^+$ (*N*=37)								
Outcomes (mean, se)	45.04 (0.46)	30.78 (3.61)	32.46 (3.77)	43.69 (0.34)	26.23 (2.19)	28.09 (2.28)	34.10 (3.47)	33.07 (3.48)
DP (*N*=37)								
Outcomes (mean, se)	45.13 (0.46)	31.94 (3.61)	30.79 (3.77)	43.39 (0.34)	23.38 (2.19)	24.62 (2.28)	27.12 (3.47)	28.50 (3.49)

Note: a = posttreatment means are adjusted for the corresponding baseline mean; b = expectancy means are adjusted for the corresponding baseline mean; DP = deceptive placebo; NT = no treatment group; OPR$^-$ = open-label placebo without rationale; OPR$^+$ = open-label placebo with rationale.

SF1: Flow Chart



Appendix D

Curriculum Vitae

Curriculum Vitae

Personal Details

Name	Cosima Locher
Date of Birth	15.05.1989
Place of birth	Basel BS, Switzerland
Nationality	Swiss
E-Mail	cosima.locher@unibas.ch
h-Index	3

Academic Education

Since 09.2014	**PhD Candidate** at the University of Basel, Switzerland „Placebos and their Consequences in Clinical and Basic Research: Effects and Possible Ways to Harness Them“
08.2012-07.2014	**Master of Science** in Clinical Psychology and Neuroscience at the University of Basel, Switzerland (Summa Cum Laude) **Master thesis**: “Baseline severity: A moderator of antidepressant and placebo outcomes in late-life depression”
08.2009-07.2012	**Bachelor of Science** in Psychology at the University of Basel, Swizerland (Insigni Cum Laude) **Bachelor thesis**: Geschlechtsunterschiede chronisch depressiver Patienten: Ansprechen auf eine CBASP-Therapie und interpersonelles Verhalten”

Professional experience

12.2011-06.2014	**Research assistant**, University of Basel, Department of Clinical Psychology and Psychotherapy
11.2013-12.2013	**Research assistant** of the study "Genetische Polymorphismen in Assoziation mit Placebo Analgesie" at the University of Basel
05.2013-09.2013	**Research assistant** for the SSP-SGP Congress at the University of Basel
02.2013-05.2013	**Internship** at the Universitäts-Kinderspital beider Basel (UKBB): "Diabetes na und? Eine Studie zur Befindlichkeit bei Kindern mit Diabetes und deren Familie"
09.2012-10.2012	**Research assistant** of the study "Ambulatory monitoring in PD and PTSD" at the University of Basel
10.2011-12.2011	**Research assistant**, University of Basel, Department of Cognitive and Decision Sciences
06.2011-09.2011	**Internship** at the Universitätsklinikum Freiburg, Germany: station for affective disorders, specialized for chronic depression
09.2010-06.2011	**Tutor for statistics**, University of Basel

Research Visits

02.2017	Personal meeting with Prof. Dr. Michael Hyland in Plymouth, UK
06.2016	Research Visit at the Program in Placebo Studies & Therapeutic Encounter (PiPS) at Harvard Medical Scool, USA
05.2016	Association for Psychological Science (APS) Congress in Chicago, USA
04.2016	Placebo Workshop, Held in Manchester Metropolitan University, MMU CREWE Campus
11.2014	Personal meeting with Prof. Dr. Irving Kirsch in Zürich, Switzerland
09.2014	European Association for Behavioural and Cognitive Therapies (EABCT) Congress in The Hague, The Netherlands
06.2014	Placebo Congress at the Foundation Brocher, Hermance, Switzerland

Academic Training

02.2016-11.2016	Antelope@university – successful career program, University of Basel
05.2016	Workshop: How to publish in peer-reviewed Journals
09.2016	Workshop: Hot to present at International Conferences
05.2015	Workshop: Self-Branding and Self-Promotion (Petra Wüst)
01.2015	Workshop: Writing (PD Dr. Jürgen Barth)
04.2013-05.2013	Meta Analysis course at Statistics.com (Dr. Michael Borenstein & Dr. Hannah Rothstein)

Teaching

02.2016-06.2016	FS 2016 Lecture "Klinische Somatopsychologie" (with Prof. Dr. Jens Gaab)
02.2016-06.2016	FS 2016 Seminar "Kolloquium Klinische Psychologie und Psychotherapie"
08.2015-12.2015	HS 2015 Seminar "Einführung in die Placebo- und Noceboforschung"
08.2015-12.2015	HS 2015 Seminar "Kolloquium Klinische Psychologie und Psychotherapie"

Supervision

Since 02.2016	Masterproject "Subjective narratives of the placebo - qualitative approach" (Supervision of 2 students)
Since 08.2015	Masterproject "Open-label placebo- experimental study" (Supervision of 3 students)
Since 08.2015	Supervision of Multiple Bachelor Theses

Additional Skills

Languages

German:	native
English:	fluent in spoken and written
French:	good knowledge
Latin:	good knowledge

Technical Skills

Microsoft, Mac OS, SPSS, EndNote, CMA, "R"

Talks and Poster Presentations

Talks

04.2017	Talk at the "Society for Interdisciplinary Placebo Studies" (SIPS) Congress at Leiden, the Netherlands: "Open-label Placebos in Healthy Participants: An Experimental Pain Investigation"
03.2017 & 04.2016 & 05.2015	Talk for the seminar "Resilienz und Ressourcen" at the Zürcher Hochschule für Angewandte Psychologie, ZHAW, Switzerland: "Placebo und die Bedeutung für die Therapie
05.2015	Talk for the lecture "Forschungsethik in der Psychologie" at the University of Basel: "Placebo and Ethics"
06.2014	Talk at the Placebo Congress at the Foundation Brocher, Hermance, Switzerland: "Placebo Response in Geriatric Depression"
09.2013	Talk at the SSP-SGP Congress in Basel, Switzerland: "The Separation Anxiety Hypothesis of Panic Disorder Revisited: A Meta-Analysis"

Poster Presentations

04.2017	Poster at the SIPS Congress at Leiden, the Netherlands: "Inpatients' perspectives on therapeutic moments"
06.2016	Poster Presentation at the Association for Psychological Science (APS) Congress in Chicago, USA: "Efficacy of Psychological Interventions in Improving Highly Active Antiretroviral Therapy Adherence. A meta-analysis of controlled studies"
09.2014	Poster Presentation at the European Association for Behavioural and Cognitive Therapies (EABCT) Congress in The Hague, the Netherlands: "Baseline Severity: A Moderator of Antidepressant and Placebo Outcomes in Late-Life Depression" Abstract retrieved from http://www.crpitalia.eu/psychomed.html

List of publications

Articles

In preparation

Locher, C., Messerli, M., Gaab, J., & Gerger, H. (in prep). The Challenge of Behaviour Change in Psychological Adherence Enhancing Interventions: A Review and Meta-Analysis

Locher, C., Meier, S., & Gaab, J. (in prep). Psychotherapy: The World of Meanings

Submitted

Locher, C., Koechlin, H., Zion, S., Werner, C., Pine, D.S., Kirsch, I., Kessler, R.C., & Kossowsky, J. (submitted). Efficacy and Safety of SSRIs, SNRIs, and Placebo in Common Psychiatric Disorders: A Comprehensive Meta-Analysis in Children and Adolescents

Locher, C., Frey, A., Kossowsky, J., & Gaab, J. (submitted). Is the Rationale More Important than Deception? A Randomized Controlled Trial of Open-Label Placebo Analgesia

Tondorf, T., Kaufmann, L. K., Degel, A., **Locher, C**., Birkhäuer, J., Gerger, H., Ehlert, U., & Gaab, J. (submitted). Randomized-Controlled Open-Hidden Design Evaluation of Expressive Writing

2017

Gaab, J, **Locher, C**. & Gerger, H. (2017). Placebo und Psychotherapie - Selbstabschaffung oder Erkenntnisgewinn? Stellungnahme der Autoren zum Leserbrief von Harald Walach. Verhaltenstherapie, 27:59–60, DOI:10.1159/000453056, IF 0.6

2016

Locher, C., Hasler, S., & Gaab, J. (2016). When Do Placebos in Psychotherapeutic Research Work? A Systematic Review on the Example of Systematic Desensitization. *Verhaltenstherapie*. doi: 10.1159/00443464

2015

Locher, C., Kossowsky, J., Gaab, J., Kirsch, I., Bain, P., & Krummenacher, P. (2015). Moderation of antidepressant and placebo outcomes by baseline severity in late-life depression: A systematic review and meta-analysis. *Journal of Affective Disorders, 181*, 50-60. http://dx.doi.org/10.1016/j. jad.2015.03.062

Gaab, J., Blease, C., **Locher, C.**, & Gerger, H. (2015). Go open - A plea for transparency in psychotherapy. *Psychology of Consciousness: Theory, Research, and Practice*. http://dx.doi.org/10.1037/cns0000063.

2013

Kossowsky, J., Pfaltz, M. C., Schneider, S., Taeymans, J., **Locher, C.**, & Gaab, J. (2013). The separation anxiety hypothesis of panic disorder revisited: A meta-analysis. *American Journal of Psychiatry, 170*, 768–781. doi:10.1176/appi.ajp.2012.12070893