

From Randomized Controlled Trials of Antidepressant Drugs to the Meta-Analytic Synthesis of Evidence: Methodological Aspects Lead to Discrepant Findings

Konstantinos N. Fountoulakis^{1,*}, Roger S. McIntyre² and André F. Carvalho³

¹3rd Department of Psychiatry, School of Medicine, Aristotle University of Thessaloniki, Greece; ²Departments of Psychiatry and Pharmacology; Mood Disorders Psychopharmacology Unit (MDPU), Toronto, ON, Canada; ³Translational Psychiatry Research Group and Department of Clinical Medicine, Faculty of Medicine, Federal University of Ceará, Fortaleza, CE, Brazil



K.N. Fountoulakis

Abstract: During the last decade, several meta-analytic studies employing different methodological approaches have had inconsistent conclusions regarding antidepressant efficacy. Herein, we aim to comment on methodological aspects that may have contributed to disparate findings. We initially discuss methodological inconsistencies and limitations related to the conduct of individual antidepressant randomized controlled trials (RCTs), including differences in allocated samples, limitations of psychometric scales, possible explanations for the heightened placebo response rates in antidepressant RCTs across the past two decades as well as the reporting of conflicts of interest. In the second part of this article, we briefly describe the various meta-analyses techniques (e.g., simple random effects meta-analysis and network meta-analysis) and the application of these methods to synthesize evidence related to antidepressant efficacy. Recently published antidepressant meta-analyses often provide discrepant results and similar results often lead to different interpretations. Finally, we propose strategies to improve methodology considering real-world clinical scenarios.

Keywords: Antidepressants, conflicts of interest, depression, meta-analysis, network meta-analysis, placebo, pooled analysis, randomized controlled trial, rating scales.

INTRODUCTION

Meta-analysis is regarded as the ‘gold standard’ approach for the evaluation and ranking of evidence in healthcare [1, 2]. However, important concerns related to the conduct of meta-analyses have emerged in the literature [3, 4]. Notwithstanding meta-analysis as a relevant method to synthesize and rank research data [5], a number of methodological issues and the presence of inherent biases (e.g., heterogeneity and discrepant methodologies across included trials) often lead to erroneous meta-analytic results and interpretations [4, 6, 7].

Currently, there is a significant number of meta-analysis concerning antidepressants and overall, they reported the presence of significant publication bias [8], a relatively small effect size in comparison to placebo [9-16], while there is a controversy on the role of initial severity [12, 13, 15]. Notwithstanding a new network meta-analysis affirms the efficacy of antidepressants for mild depression [17]. A number of meta-analyses also support the efficacy of manual-based psychotherapies (e.g., cognitive-behavioral therapy) [18-33], Notwithstanding replicated meta analytic data positive with antidepressants important methodological

issues have raised concerning the validity of the results with conclusions of non-efficacy in some analysis [12, 34]. The publication of meta-analysis concluding non-efficacy of antidepressants has fuelled skepticism among stakeholders involved in the major depressive disorder (MDD) ecosystem [35].

The overarching aims of the present narrative review are three-fold: (1) to discuss methodological limitations concerning antidepressant RCTs; (2) to briefly overview the strengths and shortcomings of the main meta-analytic techniques; and (3) to describe antidepressant meta-analyses published in the past decade along with a critical methodological appraisal. Lastly, we propose strategies for improving the conduct and interpretation of antidepressant RCTs on a clinically informative basis, providing guiding principles or a systematic approach to meta-analysis to enhance consistency and rigor.

LIMITATIONS OF ANTIDEPRESSANT RCTs

Participant Characteristics

The recruitment and screening of participants for antidepressant RCTs are often problematic. The reasons behind this problem are mainly the combination of financial benefits for researchers in combination with pressure to meet deadlines. As a result, it has been reported that often the initial ratings are inflated (i.e., patients recruited later for

*Address correspondence to this author at the 6, Odysseos str (1st Parodos Ampelonn str.), 55535 Pylaia Thessaloniki, Greece; Tel: +30 2310 435702; Fax: +30 2310 266570; E-mails: kfount@med.auth.gr; kostasfountoulakis@gmail.com

antidepressant RCTs may have their symptom severity artificially inflated) [36, 37].

Furthermore, in general patients recruited in RCTs do not correspond to the average real-life patient (*i.e.*, the representativeness of included participants is often limited) [38-44]. It is also interesting that there is a large variability between sites and countries. North America and Western Europe experience great difficulties in the recruitment of patients while on the contrary, eastern European countries and China recruit participants much easier and faster [4]. The presence of medical comorbidities as well as the history of response to medications in the past do not seem to be consistently reported in most RCTs [45, 46]. Failed trials have negative consequences both in terms of some kind of tolerance towards antidepressants [47], but also psychological consequences (*e.g.*, demoralization) [45]. Thus, it seems necessary to radically change our approach and possibly the complete RCT paradigm (*vide infra*) [48].

TRIAL DESIGN CHARACTERISTICS AND RATING SCALES

Larger RCTs are considered to be more reliable and methodologically superior in comparison to smaller studies [49]. Most meta-analyses conclude that there are important basic flaws within RCTs, therefore it seems more important to focus on the improvement of the RCTs basic design and structure and utilize the knowledge on the issue which has been accumulated during the recent years [50, 51], rather than to improve meta-analytical methods so as to be able to analyze flawed data in a quasi-omnipotent way. A radically different approach suggests that treatments can better be evaluated by a series of smaller but very well designed trials of high quality especially concerning study sample characteristics [4, 37, 52]. Notwithstanding the belief that increments in sample size would improve the signal-to-noise ratio in psychiatric RCTs, in certain circumstances, the quality of data declines with larger sample sizes [37]. For example, in large-scale multicenter RCTs there often is a financial compensation both for researchers and patients. The combination of financial interests and ethical and other administrative restrictions, with a significant pressure to complete the recruitment within strict deadlines and also with the competition among study centers to recruit larger samples might result in the violation (*p.e.*, inflation of severity scores) of inclusion criteria for participants recruited later in the trial [37]. Therefore, noise is no longer random, but it is systematically related to sample size and eventually this may lead to a deterioration of the signal-to-noise ratio.

The psychometric scales constitute an additional methodological problem. In addition to the fact that most of them do not correspond to the modern concept of depression, their scores are not rated as continuous variables, but rather as ordinal categorical ones, with unequal distances between score levels. In essence, depression rating scales derive scores from the accumulated number of qualitatively different questions/items. Thus, similar scores might correspond to radically different clinical profiles. The Hamilton depression rating scale (HDRS), which is the most commonly employed rating instrument across antidepressant RCTs has a number of serious drawbacks. We know today

that these drawbacks limit its utility and in extreme cases they might make it even inappropriate for use in RCTs. The HDRS includes items reflecting core symptoms of depression, however most of the items reflect either non-specific symptoms like anxiety or sleep disturbances, or medication side effects (*e.g.*, gastrointestinal) [53-55]. Additionally, a cut-off point > 7 for the diagnosis of depression is generally suggested [56], but a score of 15 or 20 is often required for inclusion in an RCT [57, 58]. Moreover, a number of agents, including but not limited to benzodiazepines, second generation antipsychotics or antihistamines could have a significant effect on HDRS scores, which could be erroneously attributed to 'true' antidepressant effects (Table 1). Considering that in many antidepressant RCTs, benzodiazepines or similar agents are allowed either in the placebo or in both arms, the final score might reflect an add-on effect of benzodiazepines rather than the actual drug vs. placebo effect. For example, in an RCT of bipolar depression olanzapine promoted a significant reduction in Montgomery-Åsberg Depression Rating Scale (MADRS) scores because it improved sleep, agitation and appetite, in spite of the fact it did not improve core depressive symptoms [59]. Moreover, common side effects of antidepressants (*e.g.*, headaches or gastrointestinal symptoms) could artificially inflate HDRS scores, thereby 'masking' antidepressant effects on HDRS items related to core depression dimensions. This 'masking' effect may be more substantial in the case of mild depression where a 'floor' effect cases the numerical improvement of core items to be small. A report on the analysis of the change in core HDRS items seems to support this hypothesis. This analysis suggested that when only the core HDRS items were considered, standardized mean difference (SMD) was impressively higher. When only core HDRS items were utilized, the SMD values reached 6 or were higher in comparison to the arbitrary NICE (National Institute for Health and Care Excellence) criteria for efficacy (*i.e.*, a standardized mean difference - SMD above 0.5) [54, 55, 60-62]. Finally, the measurement of other domains may provide useful information in the clinical usefulness of antidepressants. For example, cognitive function has been related to psychosocial functioning, notably work performance [63], it may be considered an alternative outcome for antidepressant RCTs [64].

THE PLACEBO RESPONSE

In antidepressant trials for adults with MDD, the mean response rate is 31% in the placebo group vs. 50% in the medication group, and it has increased at a 7% rate per decade over the last 30 years [65]. Thus, high placebo response rates have been regarded as a culprit to the fact that less than half of antidepressant efficacy trials submitted to the US Food and Drug Administration for regulatory approval found the active drug superior to placebo [66]. In recent years, significant efforts have been directed to pursue a better comprehension of variables related to the high placebo response rates in antidepressant trials (reviewed in [67]).

First, it is important to differentiate a "placebo response" from a placebo *effect*. A "placebo response" usually refers to

Table 1. Hamilton depression rating scale (HDRS) items and their possible relationship to side effects and response to various agents.

HDRS Item	Side Effect	RESPONDS TO				
		BZD	AHis	OLZ	AP	MIRT
Loss of libido	+	-	-	-	-	-
Gastroenterological	+	+	-	-	+/-	-
Weight loss	+/-	-	+	+	+	+
Insomnia	+	+	+	+	+	+
General somatic symptoms	+	+				
Agitation	(+)	+	+	+	+	+
Anxiety	(+)	+	+	+	+	+

Abbreviations: BZD: benzodiazepines; AHis: anti-histamine; OLZ: olanzapine; AP: antipsychotics; MIRT: mirtazapine.

the percentage of participants randomized to placebo who achieve at least a 50% reduction in baseline depressive symptoms, while a “placebo effect” refers to the therapeutic effect during the receiving of a substance or undergoing a procedure that is not caused by inherent powers of the specific substance or procedure [67, 68]. Different factors may influence the magnitude of the placebo response in antidepressant trials, which can be grouped as therapeutic factors, measurement factors, natural history of the illness and participant characteristics (Fig. 1).

Treatment Effects

Two theories for understanding the mechanisms of placebo effects have been proposed, namely expectancy theory and classical conditioning [69]. The expectancy theory hypothesizes that placebo treatment promotes a conscious expectation by the patient that drive symptomatic improvement. On the other hand, classic conditioning theorists attribute placebo responses to unconscious learning processes in which the individual patient associates the improvement in symptoms (unconditioned response) with neutral stimuli including pills, treatment setting, etc. (conditioned stimulus). These stimuli by itself is capable of inducing a therapeutic effect (conditioned response). It is likely that both mechanisms may contribute to the observed placebo effects in antidepressant drug trials. Importantly, placebo treatment may influence neurobiological mechanisms involved in depression pathophysiology (*e.g.*, dopaminergic neurotransmission) (see reference [68] for a review).

Khan *et al.* was the first to report that the higher the number of treatment arms in an antidepressant RCT the lower would be the “success” of the trial [70]. A greater number of active medication arms may increase the probability of receiving the ‘active treatment’, which might enhance patient expectations and in this way it can generate higher placebo response rates.

Consistent with this hypothesis, in MDD trials, the mean response rates in head-to-head comparator trials are significantly higher in the medication group in comparison to placebo [71]. Papakostas and Fava confirmed in a meta-

analysis that in a clinical trial the probability of receiving placebo was negatively associated with both antidepressant and placebo responses. Interestingly, for each 10% decrease in this probability, the antidepressant response increased by 1.8%, while the of placebo response increased by 2.6% [72].

Several lines of evidence also indicate that the amount of therapeutic contact that participants receive throughout a trial may influence placebo response rates [67, 73, 74]. For example, Posternak and Zimmerman calculated changes in HDRS scores in 41 6-week RCTs of MDD as a function of the number of study visits [73]. A cumulative effect of increasing study visits on the placebo response rates was consistently demonstrated: between weeks 2 and 6, the mean improvement on HDRS scores was 4.24 points in those patients who had weekly visits vs. 3.33 points in those patients with one visit less vs. 2.49 points in those with two visits less. An analysis of antidepressant clinical trials in children and adolescents provided interesting results [74]. In contrast to the large differences in placebo response between the various study types in adults with MDD, there were no significant differences in placebo response rates between comparator and placebo-controlled studies in children and adolescents. The amount of therapeutic contact participants received appeared to influence treatment response rather than increased expectancy: a greater number of study visits was correlated with higher placebo response rates among adolescents. In summary, participant expectations and the amount of therapeutic contact they receive throughout a trial seem to play a role in placebo response rates. The magnitude of each of these effects appears to be influenced by features related to trial design as well as patient characteristics. The high placebo response rates may significantly decrease the likelihood of detecting medication-placebo differences. One approach to dealing with expectancy-related placebo effects has been the conduct of a single-blind placebo lead-in phase in which patients with a high placebo response rate are prematurely excluded from the trial. Notwithstanding this, previous reports indicate that this approach may not be effective in reducing placebo response [75, 76]. It is important to note that one study argued that double-blind lead-in periods may be more effective [77].

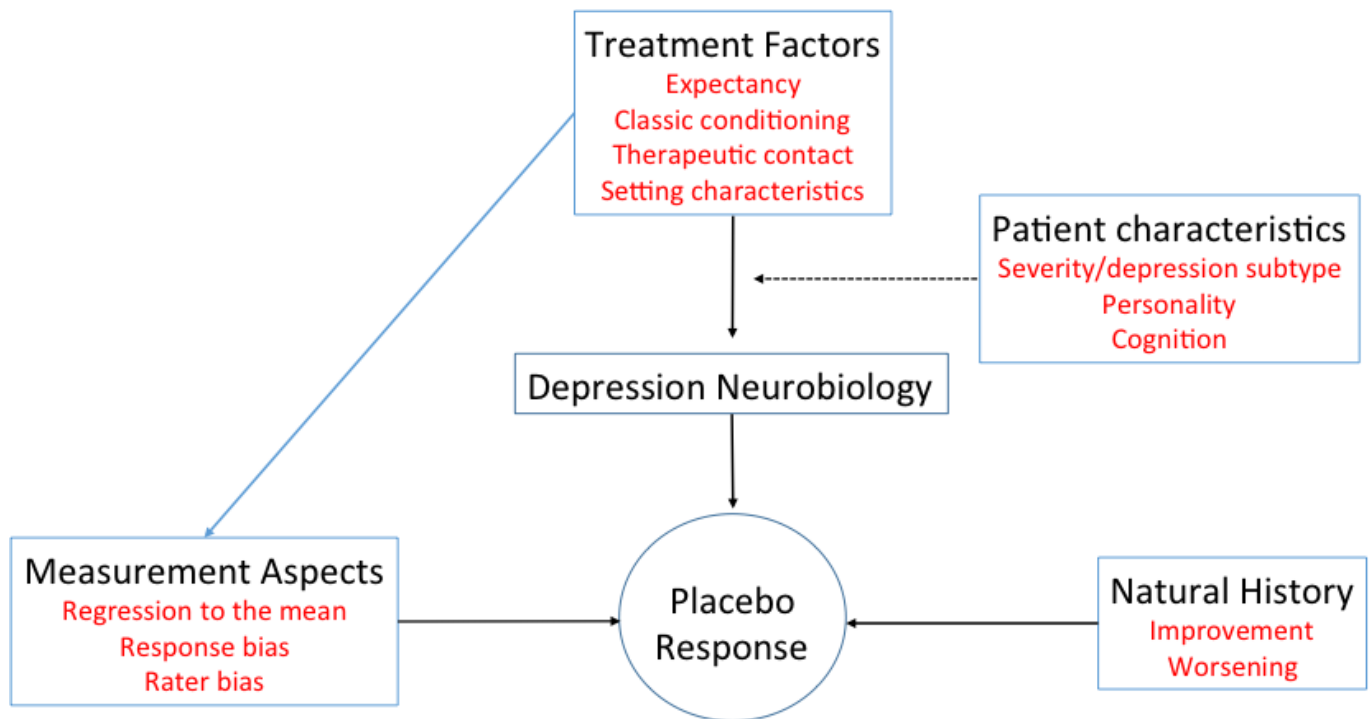


Fig. (1). A heuristic model for the placebo response in antidepressant clinical trials.

Measurement Factors

In most antidepressant clinical trials, investigators rate participants' depressive symptoms based on changes in depressive severity that are either self-reported by participants or elicited by trained interviewers. Measurements of depressive symptoms are subjected to random error the same way any other measure is. However unlike objective measures (e.g., cholesterol levels), the measurement of depressive symptoms may be associated with additional sources of bias.

Regression to the Mean

Regression to the mean occurs when repeated measurements subject to random error are obtained from the same individual over time. For example, imagine that the criteria for inclusion in an antidepressant trial requires a HDRS score > 16. Some included participants may have 'true' means <16 and the statistical tendency of the scores of these patients to decrease on repeated measures will provide the appearance that depressive symptoms improved, when in reality no true therapeutic effect occurred.

Sources of Bias

Rater bias occurs when the measurement of depressive symptoms is influenced by underlying beliefs of the drugs under study. Furthermore, the recruitment of participants for multicenter trials is a competitive process. Thus, the financial and professional returns related to an enrollment of a participant (instead of screening-out a patient) may lead an inflation of baseline severity scores [37, 67, 78]. Conversely, a response bias refers to the systematic tendency participants may have to respond to questionnaire items in accordance to the expectation of researchers (i.e., "on demand") [67].

"Howthorne effects" are a phenomenon whereby participants in a given experiment modify their behavior under study exactly because they know that the specific behavior is being measured. Therefore, response bias may be more problematic in antidepressant trials due to the inherently subjective nature of rating symptoms based on patients' reports [67]. Mancini and colleagues performed a patient-level analysis of duloxetine (≥ 60 mg/day) RCTs obtained from Lilly [37]. Lower effect sizes were found for participants in the lowest baseline HDRS depression severity and in patients in the last category of the recruitment period, whereas a higher effect size was obtained for subjects recruited in centers equal or lower in size than 2.5 times the average site-size for the trial.

The methodological shortcomings posed by regression to the mean and rater bias (i.e. baseline score inflation and low inter and intra-rater reliabilities) have been explored in different ways. For example, one strategy involves setting a minimum baseline score for enrollment in a trial, but then including in the final analysis participants with a priori defined higher score thresholds. Another strategy has been the use of centralized (and highly-trained) raters but this is often not possible at individual study sites. However, a recent report demonstrated no significant benefits of enhancing interviews with the Structured Interview Guide for the Montgomery- Åsberg depression rating scale (SIGMA), audiotaping of patients' interviews and "central" appraisal with Rater Applied Performance Scale (RAPS) [79].

Natural History of the Illness

The impact of the natural course of depression on trial outcomes is better appreciated in psychotherapy trials, which

commonly enroll a waiting list control group. A meta-analysis found that patients allocated to waiting control group experience an average improvement of 4 points on the HDRS over a mean follow-up duration of 4 weeks [80]. It seems reasonable to assume that the natural history features play a progressively important role in outcomes of depression trials over time as the population enrolled in trials change. For example, in the 1960s and 1970s, most trials enrolled inpatients with more severe depression compared to more recent trials which usually enroll participants with less severe depression. Arguably, individuals with less severe depression may present higher fluctuation in their symptoms (vide infra). Notwithstanding the recruitment of participants of longer illness duration may mitigate the influence of natural history factors, this issue seems to less dependent on investigator behavior than are measurement factors (Table 2).

Characteristics of Enrolled Subjects

Several characteristics of enrolled subjects may influence the placebo response, namely prior exposure to antidepressant treatments, severity (vide infra), duration of illness, personality characteristics, degree of refractoriness, depression subtype (eg atypical versus melancholic), and comorbid psychiatric and medical conditions.

The Nocebo Effect

Nocebo refers to adverse events (AEs) related to the negative expectations that a treatment may harm instead of ameliorate the underlying medical condition. Nocebo effects may be evaluated in RCTs. A recent meta-analysis demonstrated that 44.7% of participants enrolled to placebo experienced a at least one AE, while one out of 20 placebo-treated patients is reported to had discontinued treatment due to AEs [81]. Furthermore, there were quantitative and qualitative associations between active and placebo AEs [81]. Thus, some strategies may prevent nocebo effects in antidepressant RCTs. For example, informed consents for the active treatments under investigation may be modified; the nocebo effect should clearly discussed with the participant; and the proper blinding of raters who measure AEs in antidepressant RCTs may be an important step.

The Additive Model

The additivity thesis of pharmacological efficacy is crucial since it suggests that the specific or ‘true’ size of the pharmacological treatment effect is limited to the difference between the drug and placebo responses [82]. Although this is a convenient and practical model and does not implies the presence of a similar neurobiological mode of therapeutic action, it is important to note that at the end of the day this theory does indirectly imply such a similarity. This method is purely quantitative and thus demands similar ‘quality’. This method does not take into account that participants allocated to the placebo arm often receive additional treatments which may influence several HDRS items. Furthermore, this model has never been confirmed by neurobiological research. On the contrary, antidepressant and placebo responses could be distinct phenomena even if some degree of overlap exists. Four types of response patterns may exist: (i) placebo-only responders; (ii) treatment-only responders; (iii) placebo and treatment responders; and (iv) never responders. Kirsch[83] had proposed a modified version of the balanced placebo design to answer this question. According to this proposal, half of the study participants would be given medication and half would be given placebos. However, informed consents are obtained for participants receiving either drugs or placebo, and participants are informed (or misinformed) after this consent has been given. All subjects are debriefed by the end of the investigation. Since this design would induce deception in a distressed population, serious ethical concerns have been raised [82].

META-ANALYSIS METHODS

Grossly, there are two meta-analyses methods which have been employed to evaluate to efficacy of treatments in psychiatry. Standard pairwise meta-analysis allows the direct comparisons of two treatments. For example, some antidepressant meta-analyses determined the relative efficacy and/ or safety of specific antidepressants over placebo [84, 85], whereas other meta-analyses compared one antidepressant over another agent [86] or even over an antidepressant class [87] through the inclusion of head-to-head randomized trials.

Table 2. Variables influencing placebo response rates in antidepressant clinical trials.

Factor	Influences Placebo Response	Related to Depression Neurobiology	Passive of Modification
Treatment factors			
Expectancy-related placebo effects	+	+	+
Therapeutic setting	+	+	+
Measurement effects			
Rater bias	+	-	+
Response bias	+	-	+
Natural history factors	+	+/-	+
Participant characteristics	+	+	+

The overall estimation of effect sizes is influenced by methodological quality, publication bias as well as the heterogeneity across studies. A fundamental assumption of all meta-analysis is that either the true treatment effect is constant across trials (fixed effects model) or that the trial-specific treatment differences follow a common distribution (random effects model). More recently, a meta-analysis method referred to as network meta-analysis (also referred to as mixed treatments comparisons meta-analysis and multiple treatments meta-analysis) have gained increasing popularity in psychiatry [88-90]. Network meta-analysis (NMA) allows the comparison of different treatments on a Bayesian framework through the incorporation of indirect evidences. Head-to-head (*i.e.*, comparator) trials are relatively uncommon in medicine, including psychiatry [91]. Notwithstanding NMA has a strong potential to rank evidences in psychiatry and, therefore, to influence public policies several assumptions and limitations need to be addressed.

While comparing a treatment A versus a treatment B, NMA incorporates both direct (A versus B comparisons) and indirect comparisons (for example, the combination of trials A versus C and B versus C) to estimate the AB difference in efficacy. For example in the hypothetical Fig. 2, treatments A and C have not been compared directly, however there is indirect evidence contrasting the effect size from the direct AB evidence from the effect size of the direct BC evidence. Importantly, indirect comparisons are built on an assumption of transitivity, which is of legitimate importance for a NMA [92]. The transitivity assumption requires that studies making distinct direct comparisons must be similar in all aspects other than the treatments. When both direct and indirect evidences are available in a network we state that there is mixed evidence. For example, in Fig. 2, there is indirect evidence only concerning the comparison BD, while there is mixed evidences for comparisons AB, AD, AC, DC, and BC. Multiple treatments meta-analysis relies on the circumstances of each set of trials (eg, inclusion criteria, randomization, baseline depression severity, etc.), thus clinical judgment is important [93].

Several pairwise meta-analyses are not sufficiently powered [94] and similar concerns may extend to NMA [95]. Underpowered RCTs tend to be more prone to bias (*e.g.*, spurious and exaggerated effect estimates and selective reporting of results). Combination of biased data may give rise to unreliable estimates on a network. Network meta-analysis constitutes a unique methodological approach to investigate whether heterogeneity exists in the pairwise comparisons it encompasses. Statistical heterogeneity occurs when estimates of treatment effects (*e.g.*, odds ratios or relative risks) that were obtained from different trials may vary more often than what would be expected by chance. Clinical heterogeneity occurs whenever there are differences between individual studies in terms of characteristics of included participants. Furthermore, NMA allows the determination of whether coherence or consistency is present in the results of different clinical trials that constitute indirect comparisons vs. the available evidence from direct contrasts between treatments [96]. Box 1 depicts the advantages and limitations of NMA when compared to standard pairwise meta-analysis.

It is of high importance to decide on which method to use and which is the most appropriate way to express changes and effect sizes. Most analyses to date use the Raw Mean Difference (RMD) as the measure of effect size, except for a few reports which employed both RMD and SMD (standardized mean differences) [12, 17, 97]. This choice is very important because it leads to different results and subsequently to different interpretations. Adopting the RMD does not take into account the variability within studies, whereas SMD to a certain extent controls for floor and ceiling effects.

THE ISSUE OF BASELINE SEVERITY

A basic problem is that the concept of ‘severity’ not adequately studied and it is poorly defined. It should be noted that some items including ‘depressed mood’ manifest a ceiling effect as severity increases, while others including

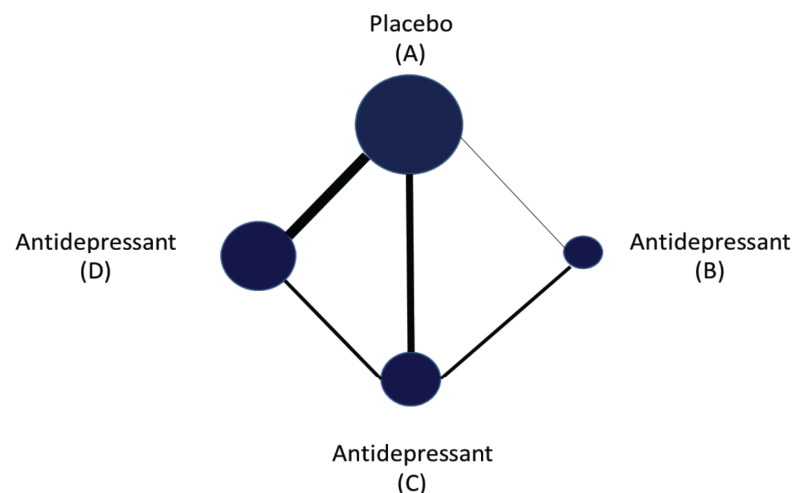


Fig. (2). Hypothetical network of antidepressants and placebo for the treatment of depression. The width of the lines is proportional to the number of trials comparing each pair of treatments, and the size of each node is proportional to the number of randomized participants (sample size).

Box 1. Advantages and limitations (i.e., risks) of network meta-analysis.

Advantages
<ul style="list-style-type: none"> • Compared to conventional pairwise meta-analysis, NMA allows the incorporation of both direct as well as indirect sources of evidence; • Network meta-analysis allows a probability-based rank order of different treatments in terms of safety and efficacy; • Network meta-analysis may inform future research directions by graphically illustrating existing direct and indirect comparisons across treatments; • Network meta-analysis can accommodate complex research questions by simultaneously incorporating several outcomes or by adding expert opinions in the form of probability-based prior distributions; • Indirect comparisons may in certain circumstances eradicate trial specific-biases that are sometimes not properly identified in comparator (i.e. head-to-head) trials.
Limitations
<ul style="list-style-type: none"> • Statistical heterogeneity; • Clinical heterogeneity; • Between-studies methodological inconsistencies in the context of a NMA may affect several pooled effect estimates; • Incoherence (i.e., significant differences between evidences provided from direct versus indirect comparisons) may limit a consistent ranking of different treatments.

‘suicidality’ manifest a floor effect with lower severity [53, 98-107]. Severity of the acute episode does not necessarily reflect overall severity of the illness. The latter should rely upon the long-term course of the illness, burden and outcome. Unfortunately, the HDRS and the MADRS both describe a concept of depression which does not corresponds to modern ideas and classification criteria [53, 104, 108]. The real correlation of HDRS scores and depression severity is a matter of debate. It is believed among clinicians that patients with higher disease severity at baseline respond better to treatment. This relation of baseline disease severity with treatment has a generic name in the statistical literature: ‘the relation between change and initial value’ [109]. In psychology, it is also well-known as the ‘law of initial value’ [110].

In this frame, the concept of ‘mathematical coupling’ [111] suggests that there is a strong structural (mathematical) correlation (~0.71) between the baseline values and change after treatment. This correlation is present, even when ‘change’ is calculated on the basis of two columns of random numbers [112]. Mathematical coupling leads to an artificially inflated association between initial value and change scores [113]. Therefore, in every medical field and every intervention, it is expected that initial severity is related to treatment outcome. This is the result of a mathematical structural characteristic, which is intrinsic to methodology. Bayesian methods, which are able to partially control for this artifact, are not routinely applied in meta-analytic research [114-116]. The problem is that, even Bayesian methods are not completely free from this phenomenon.

The issue of initial severity is very important because eventually this is the reason why many treatment guidelines are reluctant to recommend pharmacotherapy for milder forms of major depression. During the last decade, many authors argued that antidepressants act only in severe depression [12]. They also have argued that alternative treatment approaches are more suitable for mild cases. As a

consequence patients suffering from mild depression are not deprived from the right to receive treatment with antidepressants. However, this is an incorrect assumption based of inappropriate methods of analysis. A recent meta-analysis of data at the patient level suggested that initial severity plays no role [117]. Furthermore, a careful multiple treatments meta-analysis of the Kirsch [12] data set rejected initial severity as a factor that should dictate the treatment options [17]. However, another individual-level meta-analysis suggests that initial severity plays a major role in antidepressant response rates, with patients with mild depression having unclear therapeutic benefits following antidepressant treatment [13]. As suggested by a recent meta-analysis [118] the therapeutic efficacy of antidepressants for mild depression remains to be established. Clearly, future research should focus on resolving the issue of baseline severity.

RANKING ANTIDEPRESSANTS

Cipriani and colleagues [88] published an influential network meta-analysis of head-to-head antidepressant randomized trials of second generation antidepressants. These authors found escitalopram to have the best balance between efficacy and safety. However, the authors suggested that sertraline should be regarded as the first-line choice based on the fact that sertraline would have lower costs. This conclusion seems peculiar as the authors did not perform a formal cost-effectiveness analysis. This meta-analysis has been extensively criticized elsewhere [119-124]. In brief, we believe that the authors overstated their findings and did not acknowledge several methodological pitfalls of their meta-analysis. For example, the exclusion of placebo-controlled comparisons represents a significant source of bias (vide supra). Furthermore, there is a significant selective reporting of antidepressant trial results (i.e., publication bias) [8]. A significant proportion of negative trials submitted to the US Food and Drugs Administration (FDA) are either not published or published in a way conveying a positive

outcome [8]. The primary outcome for this meta-analysis (*i.e.*, treatment response) is binary in nature and may artificially inflate differences between treatments [125]. Thus, methodological heterogeneity between included studies, lack of full representativeness of the studied dataset, problematic analyses, conflicts of interests, and shortcoming in data analysis preclude firm conclusions of different efficacies between newer generation antidepressants. Interestingly, a similar network meta-analysis did not identify meaningful differences in efficacy between second-generation antidepressants [126]. These authors updated their meta-analysis and continued to find no evidences for recommending a particular second generation antidepressant on the basis of differences in efficacy [127].

More recently, Naudet and colleagues performed a network meta-analysis comparing different placebos from fluoxetine, venlafaxine and fluoxetine/venlafaxine versus placebo trials [123]. Notwithstanding the authors did not find significant differences in response/remission rates between the three placebos (*i.e.*, fluoxetine-placebo, venlafaxine-placebo, and venlafaxine/fluoxetine placebo) in terms of response/remission rates, the authors argue that due to publication bias, a firm conclusion that 'sucrose equals sucrose' could not be established. In their, epistemologically sound analysis, they suggested that the field should focus in improving trial methodology instead of attempting to prematurely rank available antidepressants regarding efficacy.

CONCLUSIONS

Meta-analysis complements primary research by distilling the raw data and by providing more specific answers. It is however dangerous to over-analyze data or utilize problematic methods of analysis. Also the risk of over-interpretation is high. The significant number of meta-analysis performed so far made antidepressants maybe the class of drugs best meta-analytically studied in the whole of medicine. The non-harmonization of meta-analytic techniques and methodological inconsistencies in included trials (*i.e.*, clinical heterogeneity) has unintentionally fostered inconsistent results that have belied our wish to arrive at true evaluations of drug efficacy compared to placebo and each other. These inconsistent results resulted in a negativistic way of conceptualization the treatment with antidepressants by the lay press, but this was also the case with prominent medical scientists (reviewed in reference [35]). This is essentially a new type of stigma for depressed patients[35].

It is clear that meta-analysis has the potential to be at the highest level of evidence concerning the evaluation of interventions in health care. On the other hand, methodological inconsistencies across trials and in the inclusion criteria for different meta-analyses pose a significant concern. For example, two 23 network meta-analyses that ranked efficacies of second generation antidepressants failed to demonstrate differential efficacies between drugs [126, 127]. Conversely, the meta-analysis performed by Cipriani and colleagues [88] which had studied the same antidepressants reported mirtazapine and venlafaxine as the most efficacious antidepressants, and duloxetine, fluvoxamine, paroxetine, and reboxetine as the

least efficacious. Considering that most available evidences regarding antidepressant efficacy are derived from placebo-controlled trials, it is possible that the exclusion of placebo comparisons from the later NMA [88] might have altered the results.

To conclude it is important to establish transparent consensus-based standards for the design and conduction of more well-designed and homogeneous antidepressant RCTs. This initiative has the potential to allow the establishment of more clinically informed and sound evidences of 'true' antidepressant effects, which could be more suitable for the synthesis of evidence. Furthermore, the inclusion criteria and conduction of NMA of antidepressant efficacy is open to debate. At the current state of knowledge in the field, it seems premature to rank different antidepressants in terms of efficacy and safety.

SOURCES OF FUNDING

No funding was available for the current study from any source

CONFLICTS OF INTEREST

KNF has received support concerning travel and accommodation expenses from various pharmaceutical companies in order to participate in medical congresses. He has also received honoraria for lectures from Astra-Zeneca, Janssen-Cilag, Eli-Lilly and research grants from Pfizer Foundation. RSM declares that he has been on advisory boards and/or received honoraria for educational activities and/or research grants from AstraZeneca, Bristol-Myers Squibb, Janssen-Ortho, Eli Lilly, Forest, Lundbeck, Pfizer, Shire, Merck, Sepracor and Otsuka. KNF has received support concerning travel and accommodation expenses from various pharmaceutical companies to participate in medical congresses. He has also received honoraria for lectures and participated in advisory boards from Astra-Zeneca, Janssen-Cilag and Eli-Lilly. KNF also received two research grants from Pfizer Foundation. AFC declares no relationships with financial interest.

ACKNOWLEDGEMENTS

AFC is supported by a research fellowship award from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; level II; Brazil).

REFERENCES

- [1] Lyman, G.H.; Kuderer, N.M. The strengths and limitations of meta-analyses based on aggregate data. *BMC Med. Res. Methodol.*, **2005**, *5*, 14. <http://dx.doi.org/10.1186/1471-2288-5-14>
- [2] Patsopoulos, N.A.; Analatos, A.A.; Ioannidis, J.P. Relative citation impact of various study designs in the health sciences. *JAMA*, **2005**, *293*, 2362-2366. <http://dx.doi.org/10.1001/jama.293.19.2362>
- [3] Feinstein, A.R. Meta-analysis: statistical alchemy for the 21st century. *J. Clin. Epidemiol.*, **1995**, *48*, 71-79. [http://dx.doi.org/10.1016/0895-4356\(94\)00110-C](http://dx.doi.org/10.1016/0895-4356(94)00110-C)
- [4] Fountoulakis, K.N.; Samara, M.T.; Siamouli, M. Burning issues in the meta-analysis of pharmaceutical trials for depression. *J. Psychopharmacol.*, **2014**, *28*, 106-117. <http://dx.doi.org/10.1177/0269881113504014>
- [5] Maier, W.; Moller, H.J. Meta-analyses: a method to maximise the evidence from clinical studies? *Eur. Arch. Psychiatry Clin.*

- Neurosci.*, **2009**, *260*, 17-23. <http://dx.doi.org/10.1007/s00406-009-0068-3>
- [6] Pereira, T.V.; Ioannidis, J.P. Statistically significant meta-analyses of clinical trials have modest credibility and inflated effects. *J. Clin. Epidemiol.*, **2011**, *64*, 1060-1069. <http://dx.doi.org/10.1016/j.jclinepi.2010.12.012>
- [7] Borm, G.F.; Donders, A.R. Updating meta-analyses leads to larger type I errors than publication bias. *J. Clin. Epidemiol.*, **2009**, *62*, 825-830 e810. <http://dx.doi.org/10.1056/NEJMsa065779>
- [8] Turner, E.H.; Matthews, A.M.; Linardatos, E. Selective publication of antidepressant trials and its influence on apparent efficacy. *N. Engl. J. Med.*, **2008**, *358*, 252-260. <http://dx.doi.org/10.1111/j.1399-5618.2008.00639.x>
- [9] Ghaemi, S.N. Why antidepressants are not antidepressants: STEP-BD, STAR*D, and the return of neurotic depression. *Bipolar Disord.*, **2008**, *10*, 957-968. <http://dx.doi.org/10.1111/j.1399-5618.2008.00639.x>
- [10] Bech, P.; Cialdella, P.; Haugh, M.C. Meta-analysis of randomised controlled trials of fluoxetine v. placebo and tricyclic antidepressants in the short-term treatment of major depression. *Br. J. Psychiatry*, **2000**, *176*, 421-428. <http://dx.doi.org/10.1192/bjp.176.5.421>
- [11] Moncrieff, J.; Wessely, S.; Hardy, R. Active placebos versus antidepressants for depression. *Cochrane Database Syst. Rev.*, **2004**, CD003012. <http://dx.doi.org/10.1002/14651858.cd003012.pub2>
- [12] Kirsch, I.; Deacon, B.J.; Huedo-Medina, T.B. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med.*, **2008**, *5*, e45. <http://dx.doi.org/10.1371/journal.pmed.0050045>
- [13] Fournier, J.C.; DeRubeis, R.J.; Hollon, S.D. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA*, **2010**, *303*, 47-53. <http://dx.doi.org/10.1001/jama.2009.1943>
- [14] Barbui, C.; Furukawa, T.A.; Cipriani, A. Effectiveness of paroxetine in the treatment of acute major depression in adults: a systematic re-examination of published and unpublished data from randomized trials. *CMAJ*, **2008**, *178*, 296-305. <http://dx.doi.org/10.1097/00004714-200202000-00007>
- [15] Khan, A.; Leventhal, R.M.; Khan, S.R. Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. *J. Clin. Psychopharmacol.*, **2002**, *22*, 40-45. <http://dx.doi.org/10.1038/npp.2011.306>
- [16] Undurraga, J.; Baldessarini, R.J. Randomized, placebo-controlled trials of antidepressants for acute major depression: thirty-year meta-analytic review. *Neuropsychopharmacology*, **2012**, *37*, 851-864. <http://dx.doi.org/10.1038/npp.2011.306>
- [17] Fountoulakis, K.N.; Veroniki, A.A.; Siamouli, M. No role for initial severity on the efficacy of antidepressants: results of a multi-meta-analysis. *Ann. Gen. Psychiatry*, **2013**, *12*, 26. <http://dx.doi.org/10.1186/1744-859X-12-26>
- [18] Ekers, D.; Webster, L.; Van Straten, A. Behavioural activation for depression; an update of meta-analysis of effectiveness and subgroup analysis. *PLoS One*, **2014**, *9*, e100100. <http://dx.doi.org/10.1371/journal.pone.0100100>
- [19] Cuijpers, P.; Karyotaki, E.; Weitz, E. The effects of psychotherapies for major depression in adults on remission, recovery and improvement: a meta-analysis. *J. Affect Disord.*, **2014**, *159*, 118-126. <http://dx.doi.org/10.1016/j.jad.2014.02.026>
- [20] Renner, F.; Cuijpers, P.; Huibers, M.J. The effect of psychotherapy for depression on improvements in social functioning: a meta-analysis. *Psychol. Med.*, **2014**, *44*, 2913-2926. <http://dx.doi.org/10.1017/S0033291713003152>
- [21] Cuijpers, P.; Berking, M.; Andersson, G. A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. *Can. J. Psychiatry*, **2013**, *58*, 376-385.
- [22] Barth, J.; Munder, T.; Gerger, H. Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. *PLoS Med.*, **2013**, *10*, e1001454. <http://dx.doi.org/10.1371/journal.pmed.1001454>
- [23] Cuijpers, P.; Turner, E.H.; Mohr, D.C. Comparison of psychotherapies for adult depression to pill placebo control groups: a meta-analysis. *Psychol. Med.*, **2014**, *44*, 685-695. <http://dx.doi.org/10.1017/S0033291713000457>
- [24] Cuijpers, P.; Huibers, M.; Ebert, D.D. How much psychotherapy is needed to treat depression? A meta-regression analysis. *J. Affect. Disord.*, **2013**, *149*, 1-13. <http://dx.doi.org/10.1016/j.jad.2013.02.030>
- [25] Bower, P.; Kontopantelis, E.; Sutton, A. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. *BMJ*, **2013**, *346*, f540. <http://dx.doi.org/10.1136/bmj.f540>
- [26] Cuijpers, P.; Andersson, G.; Donker, T. Psychological treatment of depression: results of a series of meta-analyses. *Nord. J. Psychiatry*, **2011**, *65*, 354-364. <http://dx.doi.org/10.3109/08039488.2011.596570>
- [27] Cuijpers, P.; Clignet, F.; van Meijel, B. Psychological treatment of depression in inpatients: a systematic review and meta-analysis. *Clin. Psychol. Rev.*, **2011**, *31*, 353-360. <http://dx.doi.org/10.1016/j.cpr.2011.01.002>
- [28] Cuijpers, P.; Geraedts, A.S.; van Oppen, P. Interpersonal psychotherapy for depression: a meta-analysis. *Am. J. Psychiatry*, **2011**, *168*, 581-592. <http://dx.doi.org/10.1176/appi.ajp.2010.10101411>
- [29] Cuijpers, P.; van Straten, A.; van Oppen, P. Comparing psychotherapy and pharmacotherapy for adult depression: adjusting for differential dropout rates. *J. Clin. Psychiatry*, **2010**, *71*, 1246. <http://dx.doi.org/10.4088/JCP.09105891blu>
- [30] Cuijpers, P.; Smit, F.; Bohlmeijer, E. Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias. *Br. J. Psychiatry*, **2010**, *196*, 173-178. <http://dx.doi.org/10.1192/bjp.bp.109.066001>
- [31] Cuijpers, P.; van Straten, A.; Bohlmeijer, E. The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. *Psychol. Med.*, **2010**, *40*, 211-223. <http://dx.doi.org/10.1037/a0013075>
- [32] Cuijpers, P.; van Straten, A.; Andersson, G. Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. *J. Consult. Clin. Psychol.*, **2008**, *76*, 909-922. <http://dx.doi.org/10.1037/a0013075>
- [33] Cuijpers, P.; van Straten, A.; Warmerdam, L. Psychotherapy versus the combination of psychotherapy and pharmacotherapy in the treatment of depression: a meta-analysis. *Depress. Anxiety*, **2009**, *26*, 279-288. <http://dx.doi.org/10.1002/da.20519>
- [34] Ioannidis, J.P. Effectiveness of antidepressants: an evidence myth constructed from a thousand randomized trials? *Philos. Ethics. Humanit. Med.*, **2008**, *3*, 14. <http://dx.doi.org/10.1186/1747-5341-3-14>
- [35] Fountoulakis, K.N.; Hoschl, C.; Kasper, S. The media and intellectuals' response to medical publications: the antidepressants' case. *Ann. Gen. Psychiatry*, **2013**, *12*, 11. <http://dx.doi.org/10.1186/1744-859X-12-11>
- [36] Kobak, K.A.; Leuchter, A.; DeBrot, D. Site versus centralized raters in a clinical depression trial: impact on patient selection and placebo response. *J. Clin. Psychopharmacol.*, **2010**, *30*, 193-197.
- [37] Mancini, M.; Wade, A.G.; Perugi, G. Impact of patient selection and study characteristics on signal detection in placebo-controlled trials with antidepressants. *J. Psychiatr. Res.*, **2014**, *51*, 21-29.
- [38] Licht, R.W.; Gouliava, G.; Vestergaard, P. Generalisability of results from randomised drug trials. A trial on antimanic treatment. *Br. J. Psychiatry*, **1997**, *170*, 264-267. <http://dx.doi.org/10.1192/bjp.170.3.264>
- [39] Talamo, A.; Baldessarini, R.J.; Centorrino, F. Comparison of mania patients suitable for treatment trials versus clinical treatment. *Hum. Psychopharmacol.*, **2008**, *23*, 447-454. <http://dx.doi.org/10.1002/hup.952>
- [40] Zetin, M.; Hoepner, C.T. Relevance of exclusion criteria in antidepressant clinical trials: a replication study. *J. Clin. Psychopharmacol.*, **2007**, *27*, 295-301. <http://dx.doi.org/10.1097/JCP.0b013e318058263f>
- [41] Zimmerman, M.; Mattia, J.I.; Posternak, M.A. Are subjects in pharmacological treatment trials of depression representative of patients in routine clinical practice? *Am. J. Psychiatry*, **2002**, *159*, 469-473. <http://dx.doi.org/10.1176/appi.ajp.159.3.469>
- [42] Partonen, T.; Sihvo, S.; Lonnqvist, J.K. Patients excluded from an antidepressant efficacy trial. *J. Clin. Psychiatry*, **1996**, *57*, 572-575. <http://dx.doi.org/10.4088/JCP.v57n1203>
- [43] Wisniewski, S.R.; Rush, A.J.; Nierenberg, A.A. Can phase III trial results of antidepressant medications be generalized to clinical practice? A STAR*D report. *Am. J. Psychiatry*, **2009**, *166*, 599-607. <http://dx.doi.org/10.1176/appi.ajp.2008.08071027>

- [44] van der Lem, R.; van der Wee, N.J.; van Veen, T. The generalizability of antidepressant efficacy trials to routine psychiatric out-patient practice. *Psychol. Med.*, **2010**, *41*, 1353-1363. <http://dx.doi.org/10.1017/S0033291710002175>
- [45] Wolf, S.; Pinsky, R.H. Effects of placebo administration and occurrence of toxic reactions. *J. Am. Med. Assoc.*, **1954**, *155*, 339-341. <http://dx.doi.org/10.1001/jama.1954.03690220013004>
- [46] Tomba, E. Nowhere patients. *Psychother. Psychosom.*, **2012**, *81*, 69-72. <http://dx.doi.org/10.1159/000334112>
- [47] Fava, G.A.; Offidani, E. The mechanisms of tolerance in antidepressant action. *Prog. Neuropsychopharmacol. Biol. Psychiatry*, **2010**, *35*, 1593-1602. <http://dx.doi.org/10.1016/j.pnpbp.2010.07.026>
- [48] Fava, G.A.; Tomba, E.; Tossani, E. Innovative trends in the design of therapeutic trials in psychopharmacology and psychotherapy. *Prog. Neuropsychopharmacol. Biol. Psychiatry*, **2013**, *40*, 306-311. <http://dx.doi.org/10.1016/j.pnpbp.2012.10.014>
- [49] Trikalinos, T.A.; Churchill, R.; Ferri, M. Effect sizes in cumulative meta-analyses of mental health randomized trials evolved over time. *J. Clin. Epidemiol.*, **2004**, *57*, 1124-1130. <http://dx.doi.org/10.1016/j.jclinepi.2004.02.018>
- [50] Fava, G.A.; Belaise, C.A. A discussion on the role of clinimetrics and the misleading effects of psychometric theory. *J. Clin. Epidemiol.*, **2005**, *58*, 753-756. <http://dx.doi.org/10.1016/j.jclinepi.2004.12.006>
- [51] Fava, G.A.; Tomba, E.; Sonino, N. Clinimetrics: the science of clinical measurements. *Int. J. Clin. Pract.*, **2011**, *66*, 11-15. <http://dx.doi.org/10.1111/j.1742-1241.2011.02825.x>
- [52] Borm, G.F.; Lemmers, O.; Fransen, J. The evidence provided by a single trial is less reliable than its statistical analysis suggests. *J. Clin. Epidemiol.*, **2009**, *62*, 711-715 e711.
- [53] Bech, P. Rating scales in depression: limitations and pitfalls. *Dialogues Clin. Neurosci.*, **2006**, *8*, 207-215.
- [54] Bech, P.; Fava, M.; Trivedi, M.H. Outcomes on the pharmacopsychometric triangle in bupropion-SR vs. bupropion augmentation of citalopram in the STAR*D trial. *Acta Psychiatr. Scand.*, **2011**, *125*, 342-348. <http://dx.doi.org/10.1111/j.1600-0447.2011.01791.x>
- [55] Bech, P.; Fava, M.; Trivedi, M.H. Factor structure and dimensionality of the two depression scales in STAR*D using level 1 datasets. *J. Affect. Disord.*, **2011**, *132*, 396-400. <http://dx.doi.org/10.1016/j.jad.2011.03.011>
- [56] Romera, I.; Perez, V.; Menchon, J.M. Optimal cutoff point of the Hamilton Rating Scale for Depression according to normal levels of social and occupational functioning. *Psychiatry Res.*, **2010**, *186*, 133-137. <http://dx.doi.org/10.1016/j.psychres.2010.06.023>
- [57] Goldstein, D.J.; Mallinckrodt, C.; Lu, Y. Duloxetine in the treatment of major depressive disorder: a double-blind clinical trial. *J. Clin. Psychiatry*, **2002**, *63*, 225-231. <http://dx.doi.org/10.4088/JCP.v63n0309>
- [58] Tourian, K.A.; Padmanabhan, S.K.; Groark, J. Desvenlafaxine 50 and 100 mg/d in the treatment of major depressive disorder: an 8-week, phase III, multicenter, randomized, double-blind, placebo-controlled, parallel-group trial and a post hoc pooled analysis of three studies. *Clin. Ther.*, **2009**, *31 Pt 1*, 1405-1423. <http://dx.doi.org/10.1016/j.clinthera.2009.07.006>
- [59] Tohen, M.; Vieta, E.; Calabrese, J.. Efficacy of olanzapine and olanzapine-fluoxetine combination in the treatment of bipolar I depression. *Arch. Gen. Psychiatry*, **2003**, *60*, 1079-1088. <http://dx.doi.org/10.1001/archpsyc.60.11.1079>
- [60] Maier, W.; Philipp, M. Comparative analysis of observer depression scales. *Acta Psychiatr. Scand.*, **1985**, *72*, 239-245. <http://dx.doi.org/10.1111/j.1600-0447.1985.tb02601.x>
- [61] Bech, P. Is the antidepressive effect of second-generation antidepressants a myth? *Psychol. Med.*, **2010**, *40*, 181-186. <http://dx.doi.org/10.1017/S0033291709006102>
- [62] Bech, P.; Tanghoj, P.; Cialdella, P. Escitalopram dose-response revisited: an alternative psychometric approach to evaluate clinical effects of escitalopram compared to citalopram and placebo in patients with major depression. *Int. J. Neuropsychopharmacol.*, **2004**, *7*, 283-290. <http://dx.doi.org/10.1017/S1461145704004365>
- [63] McIntyre, R.S.; Cha, D.S.; Soczynska, J.K. Cognitive deficits and functional outcomes in major depressive disorder: determinants, substrates, and treatment interventions. *Depress. Anxiety*, **2013**, *30*, 515-527. <http://dx.doi.org/10.1002/da.22063>
- [64] McIntyre, R.S. Using measurement strategies to identify and monitor residual symptoms. *J. Clin. Psychiatry*, **2013**, *74 Suppl 2*, 14-18. <http://dx.doi.org/10.4088/JCP.12084su1c.03>
- [65] Walsh, B.T.; Seidman, S.N.; Sysko, R. Placebo response in studies of major depression: variable, substantial, and growing. *Jama* **2002**, *287*, 1840-1847. <http://dx.doi.org/10.1001/jama.287.14.1840>
- [66] Khan, A.; Khan, S.; Brown, W.A. Are placebo controls necessary to test new antidepressants and anxiolytics? *Int. J. Neuropsychopharmacol.*, **2002**, *5*, 193-197. <http://dx.doi.org/10.1017/S1461145702002912>
- [67] Rutherford, B.R.; Roose, S.P. A model of placebo response in antidepressant clinical trials. *Am. J. Psychiatry*, **2013**, *170*, 723-733. <http://dx.doi.org/10.1176/appi.ajp.2012.12040474>
- [68] Jubbs, J.; Bensing, J.M. The sweetest pill to swallow: how patient neurobiology can be harnessed to maximise placebo effects. *Neurosci. Biobehav. Rev.*, **2013**, *37*, 2709-2720. <http://dx.doi.org/10.1016/j.neubiorev.2013.09.006>
- [69] Stewart-Williams, S.; Podd, J. The placebo effect: dissolving the expectancy versus conditioning debate. *Psychol. Bull.*, **2004**, *130*, 324-340. <http://dx.doi.org/10.1037/0033-2909.130.2.324>
- [70] Khan, A.; Kolts, R.L.; Thase, M.E. Research design features and patient characteristics associated with the outcome of antidepressant clinical trials. *Am. J. Psychiatry*, **2004**, *161*, 2045-2049. <http://dx.doi.org/10.1176/appi.ajp.161.11.2045>
- [71] Rutherford, B.R.; Sneed, J.R.; Roose S.P. Does study design influence outcome?. The effects of placebo control and treatment duration in antidepressant trials. *Psychother. Psychosom.*, **2009**, *78*, 172-181. <http://dx.doi.org/10.1159/000209348>
- [72] Papakostas, G.I.; Fava, M. Does the probability of receiving placebo influence clinical trial outcome? A meta-regression of double-blind, randomized clinical trials in MDD. *Eur. Neuropsychopharmacol.*, **2009**, *19*, 34-40. <http://dx.doi.org/10.1016/j.euroneuro.2008.08.009>
- [73] Posternak, M.A.; Zimmerman, M. Therapeutic effect of follow-up assessments on antidepressant and placebo response rates in antidepressant efficacy trials: meta-analysis. *Br. J. Psychiatry*, **2007**, *190*, 287-292. <http://dx.doi.org/10.1192/bjp.bp.106.028555>
- [74] Rutherford, B.R.; Sneed, J.R.; Tandler, J.M. Deconstructing pediatric depression trials: an analysis of the effects of expectancy and therapeutic contact. *J. Am. Acad. Child. Adolesc. Psychiatry*, **2011**, *50*, 782-795. <http://dx.doi.org/10.1016/j.jaac.2011.04.004>
- [75] Rutherford, B.R.; Marcus, S.M.; Wang, P. A randomized, prospective pilot study of patient expectancy and antidepressant outcome. *Psychol. Med.*, **2013**, *43*, 975-982. <http://dx.doi.org/10.1017/S0033291712001882>
- [76] Trivedi, M.H.; Rush, H. Does a placebo run-in or a placebo treatment cell affect the efficacy of antidepressant medications? *Neuropsychopharmacology*, **1994**, *11*, 33-43. <http://dx.doi.org/10.1038/npp.1994.63>
- [77] Faries, D.E.; Heiligenstein, J.H.; Tollefson, G.D. The double-blind variable placebo lead-in period: results from two antidepressant clinical trials. *J. Clin. Psychopharmacol.*, **2001**, *21*, 561-568. <http://dx.doi.org/10.1097/00004714-200112000-00004>
- [78] Kobak, K.A.; Kane, J.M.; Thase, M.E. Why do clinical trials fail? The problem of measurement error in clinical trials: time to test new paradigms? *J. Clin. Psychopharmacol.*, **2007**, *27*, 1-5. <http://dx.doi.org/10.1097/JCP.0b013e31802eb4b7>
- [79] Khan, A.; Faucett, J.; Brown, W.A. Magnitude of placebo response and response variance in antidepressant clinical trials using structured, taped and appraised rater interviews compared to traditional rating interviews. *J. Psychiatr. Res.*, **2014**, *51*, 88-92. <http://dx.doi.org/10.1016/j.jpsychires.2014.01.005>
- [80] Rutherford, B.R.; Mori, S.; Sneed, J.R. Contribution of spontaneous improvement to placebo response in depression: a meta-analytic review. *J. Psychiatr. Res.*, **2012**, *46*, 697-702. <http://dx.doi.org/10.1016/j.jpsychires.2012.02.008>
- [81] Mitsikostas, D.D.; Mantonakis, L.; Chalarakis, N. Nocebo in clinical trials for depression: a meta-analysis. *Psychiatry Res.*, **2014**, *215*, 82-86. <http://dx.doi.org/10.1016/j.psychres.2013.10.019>
- [82] Waring, D.R. The antidepressant debate and the balanced placebo trial design: an ethical analysis. *Int. J. Law. Psychiatry*, **2008**, *31*, 453-462. <http://dx.doi.org/10.1016/j.ijlp.2008.09.001>
- [83] Kirsch, I. Are drug and placebo effects in depression additive? *Biol. Psychiatry*, **2000**, *47*, 733-735. [http://dx.doi.org/10.1016/S0006-3223\(00\)00832-5](http://dx.doi.org/10.1016/S0006-3223(00)00832-5)
- [84] Beasley, C.M., Jr.; Nilsson, M.E.; Koke, S.C. Efficacy, adverse events, and treatment discontinuations in fluoxetine clinical studies of major depression: a meta-analysis of the 20-mg/day dose. *J.*

- Clin. Psychiatry.*, **2000**, *61*, 722-728. <http://dx.doi.org/10.4088/JCP.v61n1003>
- [85] Pae, C.U.; Wang, S.M.; Han, C. Vortioxetine: a meta-analysis of 12 short-term, randomized, placebo-controlled clinical trials for the treatment of major depressive disorder. *J. Psychiatry Neurosci.*, **2014**, *39*, 1401-20.
- [86] Montgomery, S.; Hansen, T.; Kasper, S. Efficacy of escitalopram compared to citalopram: a meta-analysis. *Int. J. Neuropsychopharmacol.*, **2011**, *14*, 261-268. <http://dx.doi.org/10.1017/S146114571000115X>
- [87] Thase, M.E.; Haight, B.R.; Richard, N. Remission rates following antidepressant therapy with bupropion or selective serotonin reuptake inhibitors: a meta-analysis of original data from 7 randomized controlled trials. *J. Clin. Psychiatry*, **2005**, *66*, 974-981. <http://dx.doi.org/10.4088/JCP.v66n0803>
- [88] Cipriani, A.; Furukawa, T.A.; Salanti, G. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet*, **2009**, *373*, 746-758. [http://dx.doi.org/10.1016/S0140-6736\(09\)60046-5](http://dx.doi.org/10.1016/S0140-6736(09)60046-5)
- [89] Leucht, S.; Cipriani, A.; Spineli, L. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. *Lancet*, **2013**, *382*, 951-962. [http://dx.doi.org/10.1016/S0140-6736\(13\)60733-3](http://dx.doi.org/10.1016/S0140-6736(13)60733-3)
- [90] Yildiz, A.; Nikodem, M.; Vieta, E. A network meta-analysis on comparative efficacy and all-cause discontinuation of antimanic treatments in acute bipolar mania. *Psychol. Med.*, **2014**, 1-19.
- [91] Song, F.; Loke, Y.K.; Walsh, T. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. *BMJ*, **2009**, *338*, b1147. <http://dx.doi.org/10.1136/bmj.b1147>
- [92] Salanti, G.; Del Giovane, C.; Chaimani, A. Evaluating the quality of evidence from a network meta-analysis. *PLoS One*, **2014**, *9*, e99682. <http://dx.doi.org/10.1371/journal.pone.0099682>
- [93] Lu, G.; Ades, A.E. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat. Med.*, **2004**, *23*, 3105-3124. <http://dx.doi.org/10.1002/sim.1875>
- [94] Guyatt, G.H.; Oxman, A.D.; Kunz, R. GRADE guidelines 6. Rating the quality of evidence--imprecision. *J. Clin. Epidemiol.*, **2011**, *64*, 1283-1293. <http://dx.doi.org/10.1016/j.jclinepi.2011.01.012>
- [95] Thorlund, K.; Mills, E.J. Sample size and power considerations in network meta-analysis. *Syst. Rev.*, **2012**, *1*, 41. <http://dx.doi.org/10.1186/2046-4053-1-41>
- [96] Mills, E.J.; Thorlund, K.; Ioannidis, J.P. Demystifying trial networks and network meta-analysis. *BMJ*, **2013**, *346*, f2914. <http://dx.doi.org/10.1136/bmj.f2914>
- [97] Fountoulakis, K.N.; Moller, H.J. Efficacy of antidepressants: a re-analysis and re-interpretation of the Kirsch data. *Int. J. Neuropsychopharmacol.*, **2011**, *14*, 405-412. <http://dx.doi.org/10.1017/S1461145710000957>
- [98] Bech, P. Rating scales for affective disorders: their validity and consistency. *Acta. Psychiatr. Scand. Suppl.*, **1981**, *295*, 1-101.
- [99] Bech, P. Assessment scales for depression: the next 20 years. *Acta. Psychiatr. Scand. Suppl.*, **1983**, *310*, 117-130.
- [100] Bech, P. The instrumental use of rating scales for depression. *Pharmacopsychiatry*, **1984**, *17*, 22-28. <http://dx.doi.org/10.1055/s-2007-1017402>
- [101] Bech, P. Rating scales in psychopharmacology. Statistical aspects. *Acta. Psychiatr. Belg.*, **1988**, *88*, 291-302.
- [102] Bech, P. Rating scales for mood disorders: applicability, consistency and construct validity. *Acta. Psychiatr. Scand. Suppl.*, **1988**, *345*, 45-55. <http://dx.doi.org/10.1111/j.1600-0447.1988.tb08567.x>
- [103] Bech, P. Psychometric developments of the Hamilton scales: the spectrum of depression, dysthymia, and anxiety. *Psychopharmacol. Ser.*, **1990**, *9*, 72-79. http://dx.doi.org/10.1007/978-3-642-75373-2_9
- [104] Bech, P. Modern psychometrics in clinimetrics: impact on clinical trials of antidepressants. *Psychother. Psychosom.*, **2004**, *73*, 134-138. <http://dx.doi.org/10.1159/000076448>
- [105] Bech, P. Applied psychometrics in clinical psychiatry: the pharmacopsychometric triangle. *Acta. Psychiatr. Scand.*, **2009**, *120*, 400-409. <http://dx.doi.org/10.1111/j.1600-0447.2009.01445.x>
- [106] Bech, P.; Allerup, P.; Gram, L.F. The Hamilton depression scale. Evaluation of objectivity using logistic models. *Acta. Psychiatr. Scand.*, **1981**, *63*, 290-299. <http://dx.doi.org/10.1111/j.1600-0447.1981.tb00676.x>
- [107] Bech, P.; Gram, L. F.; Dein, E. Quantitative rating of depressive states. *Acta. Psychiatr. Scand.*, **1975**, *51*, 161-170. <http://dx.doi.org/10.1111/j.1600-0447.1975.tb00002.x>
- [108] Bagby, R.M.; Ryder, A.G.; Schuller, D.R. The Hamilton Depression Rating Scale: has the gold standard become a lead weight? *Am. J. Psychiatry*, **2004**, *161*, 2163-2177. <http://dx.doi.org/10.1176/appi.ajp.161.12.2163>
- [109] Blomqvist, N. On the relation between change and initial value. *J. Am. Stat. Assoc.*, **1977**, *72*, 746-749.
- [110] Jin P. Toward a reconceptualization of the law of initial value. *Psychol. Bull.*, **1992**, *111*, 176-184. <http://dx.doi.org/10.1037/0033-2909.111.1.176>
- [111] Fountoulakis, K.N.; Kontis, D. Mathematical coupling and the true role of baseline severity in acute mania trials. *Neuropsychopharmacology*, **2012**, *37*, 850. <http://dx.doi.org/10.1038/npp.2011.234>
- [112] Oldham, P.A. A note on the analysis of repeated measurements of the same subjects. *J. Chronic. Dis.*, **1962**, *15*, 969-977. [http://dx.doi.org/10.1016/0021-9681\(62\)90116-9](http://dx.doi.org/10.1016/0021-9681(62)90116-9)
- [113] Tu, Y.K.; Maddick, I.H.; Griffiths, G.S. Mathematical coupling can undermine the statistical assessment of clinical research: illustration from the treatment of guided tissue regeneration. *J. Dent.*, **2004**, *32*, 133-142. <http://dx.doi.org/10.1016/j.jdent.2003.10.001>
- [114] Goodman, S.N. Toward evidence-based medical statistics. 2: The Bayes factor. *Ann. Intern. Med.*, **1999**, *130*, 1005-1013. <http://dx.doi.org/10.7326/0003-4819-130-12-199906150-00019>
- [115] Goodman, S.N. Toward evidence-based medical statistics. 1: The P value fallacy. *Ann. Intern. Med.*, **1999**, *130*, 995-1004. <http://dx.doi.org/10.7326/0003-4819-130-12-199906150-00008>
- [116] Johnson, S.R.; Tomlinson, G.A.; Hawker, G.A. Methods to elicit beliefs for Bayesian priors: a systematic review. *J. Clin. Epidemiol.*, **2009**, *63*, 355-369. <http://dx.doi.org/10.1016/j.jclinepi.2009.06.003>
- [117] Gibbons, R.D.; Hur, K.; Brown, C.H. Benefits from antidepressants: synthesis of 6-week patient-level outcomes from double-blind placebo-controlled randomized trials of fluoxetine and venlafaxine. *Arch. Gen. Psychiatry*, **2012**, *69*, 572-579. <http://dx.doi.org/10.1001/archgenpsychiatry.2011.2044>
- [118] Cameron, I.M.; Reid, I.C.; MacGillivray, S.A. Efficacy and tolerability of antidepressants for sub-threshold depression and for mild major depressive disorder. *J. Affect. Disord.*, **2014**, *166*, 48-58. <http://dx.doi.org/10.1016/j.jad.2014.04.078>
- [119] Gartlehner, G.; Gaynes, B.N.; Hansen, R.A. Ranking antidepressants. *Lancet*, **2009**, *373*, 1761; author reply 1761-1762.
- [120] Turner, E.; Moreno, S.G.; Sutton, A.J. Ranking antidepressants. *Lancet*, **2009**, *373*, 1760; author reply 1761-1762.
- [121] Seyringer, M.E.; Kasper, S. Ranking antidepressants. *Lancet*, **2009**, *373*, 1760-1761; author reply 1761-1762.
- [122] Ioannidis, J.P. Ranking antidepressants. *Lancet*, **2009**, *373*, 1759-1760; author reply 1761-1762.
- [123] Naudet, F.; Millet, B.; Charlier, P. Which placebo to cure depression? A thought-provoking network meta-analysis. *BMC Med.*, **2013**, *11*, 230. <http://dx.doi.org/10.1186/1741-7015-11-230>
- [124] Naudet, F.; Falissard, B. Does reductio ad absurdum have a place in evidence-based medicine? *BMC Med.*, **2014**, *12*, 106. <http://dx.doi.org/10.1186/1741-7015-12-106>
- [125] Deeks, J.J. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat. Med.*, **2002**, *21*, 1575-1600. <http://dx.doi.org/10.1002/sim.1188>
- [126] Gartlehner, G.; Gaynes, B.N.; Hansen, R.A. Comparative benefits and harms of second-generation antidepressants: background paper for the American College of Physicians. *Ann. Intern. Med.*, **2008**, *149*, 734-750. <http://dx.doi.org/10.7326/0003-4819-149-10-200811180-00008>
- [127] Gartlehner, G.; Hansen, R.A.; Morgan, L.C. Comparative benefits and harms of second-generation antidepressants for treating major depressive disorder: an updated meta-analysis. *Ann. Intern. Med.*, **2011**, *155*, 772-785. <http://dx.doi.org/10.7326/0003-4819-155-11-201112060-00009>