Antidepressant drugs and the response in the placebo group: the real problem lies in our understanding of the issue

Konstantinos N Fountoulakis and Hans-Jürgen Möller

J Psychopharmacol 2012 26: 744 originally published online 17 September 2011
DOI: 10.1177/0269881111421969

Short Report

Antidepressant drugs and the response in the placebo group: the real problem lies in our understanding of the issue

Journal of Psychopharmacology
26(5) 744–750

DOI: 10.1177/0269881111421969
jop.sagepub.com

Konstantinos N Fountoulakis[1] and Hans-Jürgen Möller[2]

Abstract

In a recent paper, Horder and colleagues (Horder et al., 2010, *J Psychopharmacol* 25: 1277–1288) have suggested that the main problem in the Kirsch analysis is methodological. We argue that the results are similar irrespective of the method used. In our opinion the data suggest that placebo and drug effects are non-additive: antidepressants act independently of depression severity, while the placebo effect is present only in milder cases. While the response in the placebo group is due to unstable 'noise' and 'artefacts', the medication effect is reliable, valid and stable.

Keywords

Placebo, active drug, randomized controlled trial, model, depression

Introduction

During the last decade, the usefulness of antidepressants in clinical practice has been questioned, with meta-analytic studies suggesting their effect size is rather small (Bech et al., 2000; Ghaemi, 2008; Moncrieff et al., 2004). Another issue is that recently it has been documented that there is a significant bias in the publication of antidepressant trials (Turner et al., 2008). At the zenith of this discussion, the meta-analysis on the usefulness of antidepressants (Kirsch et al., 2008) attracted much attention both from scientists and from the general public. Kirsch and colleagues obtained data from the United States Food and Drug Administration (FDA) and thus tried to avoid publication bias. That meta-analysis reported that the effect size and the magnitude of change in Hamilton rating scale for depression (HRSD) score were small (effect size *d*-value below 0.50 and change in HRSD equal to 1.80 points) and thus antidepressants fell well below criteria for clinical relevance suggested by the National (UK) Institute for Clinical Excellence (NICE). Similar findings were reported by Barbui and colleagues for paroxetine alone (Barbui et al., 2008). Kirsch et al. (2008) also reported that efficacy reaches clinical relevance only in trials involving the most extremely depressed patients, and that this pattern is due to a decrease in the response to placebo rather than an increase in the response to medication. They also found no linear relation between severity and response to medication. More recently, Fournier et al. (2010) have reported similar results, but on a much different sample of randomized trials.

Kirsch went further and accused the FDA of having made an explicit decision to keep this information from the public and from prescribing physicians (Kirsch, 2009a). He also suggested that because they do not incur drug risks, alternative therapies (e.g. exercise and psychotherapy), showing equal benefits to those of antidepressants, may be a better treatment choice for depression (Kirsch, 2009a) and went on to author a book under the title *The Emperor's New Drugs: Exploding the Antidepressant Myth* (Kirsch, 2009b). This was also the picture painted in the media.

Several authors have criticized the above interpretation (Bech, 2010; Broich, 2009; Ghaemi, 2008; McAllister-Williams, 2008; Möller, 2008, 2009b, 2009c; Möller and Broich, 2010; Möller and Maier, 2010b) by focusing on the limitations of randomized controlled trials (RCTs) on clinical issues, and especially on the problematic properties of the HRSD (Bech, 2001, 2004, 2006, 2010; Bech et al., 2004) and on the fact that the effectiveness of antidepressants in clinical practice is normally optimized by sequential and combined therapy approaches (Möller, 2009a; Rush, 2007).The usefulness of 'alternative therapies' was also discussed. Overall, psychotherapy seems to be significantly less effective than pharmacotherapy, since the effect size (in RCTs of lower quality than that of drugs) is close to the placebo effect size (Cuijpers et al., 2010a, 2010b) and, in addition, it seems that the publication bias is more pronounced concerning non-pharmacological treatments (Cuijpers et al., 2010a). The results of

[1]Department of Psychiatry, School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece
[2]Department of Psychiatry, Ludwig-Maximilians-University, Munich, Germany

Corresponding author:
Konstantinos N Fountoulakis, 6 Odysseos str. (1st Parodos Ampelonon str.), 55535 Pylaia Thessaloniki, Greece
Email: kfount@med.auth.gr

psychotherapy studies cannot be directly compared to results of antidepressant trials; in spite of being based on RCT methodology, in essence psychotherapy studies are not blinded, in contrast to antidepressant trials (Möller and Maier, 2010a). The data on the efficacy of exercise and other alternative therapies, are either negative or do not exist.

Critical review of the issue

In spite of these critiques, the Kirsch conclusion was strong since it seemed to rely on solid data while detractors were using seemingly vague arguments. More recently, we performed a re-analysis of the Kirsch et al. (2008) data (Fountoulakis and Möller, 2010), reporting not only different results but also proposing a different way of interpreting the data. Also the paper 'Placebo, Prozac and PLos: significant lessons for psychopharmacology' has recently been published (Horder et al., 2010). That paper commented on the methods used by Kirsch et al. in their 2008 paper (Kirsch et al., 2008) and performed a limited re-analysis of the data.

Although the paper by Horder et al. (2010) agrees in many ways with us (namely, that the difference in the effectiveness of active drug versus placebo is between 2.18 and 2.68 HRSD points instead of 1.80), we sharply disagree on the interpretation of the situation. In their re-analysis, Horder and colleagues suggest that the main problem in the Kirsch analysis was methodological; that is, Kirsch et al. used unusual or unconventional methods to pool and analyse the data. They also suggest a number of conceptual problems in the interpretation (e.g. what 'severity' stands for, what the true placebo effect stands for, etc.) In our opinion these are not the main problems and they only turn our attention away from the core of the issue.

There are two main problems with the Kirsch analysis. The first concerns the method of analysis. We performed the re-analysis by using simple averaging, weighting by sample size and by the inverse variance. The differences in the results after these different approaches were not that important, since the correction of the effect size was minimal (Table 1).

Second, Kirsch et al. failed to report that (according to their method) the change in HRSD score was slightly below the threshold of 3 points suggested by the NICE for specific antidepressants. According both to our analysis and to Horder et al.'s analysis, the correct values are above 3 points for venlafaxine and for paroxetine. This also points to a high heterogeneity among antidepressants, making it problematic to view the results of their meta-analysis as being equally relevant to the whole class of antidepressants. Still, the respective values for fluoxetine and nefazodone are low. It is to be noted that it is impossible to calculate individual *d*-values without calculating the change scores first, so these were interim results in the process of the analysis and Kirsch et al. should have reported them, especially in the frame of the importance of a possible 'clinical difference' between drugs. Furthermore, four RCTs concerned the elderly, leading to lower *d*-values since it is known that the elderly constitute a refractory population (Fountoulakis et al., 2003, 2004). A previous comment suggested that at least the calculations were correct (e.g. 'Undoubtedly the findings in this analysis are robust, as far as the studies included in the analysis are concerned') and that all relevant results were published in the paper (McAllister-Williams, 2008).

Theoretical interpretation of the data

However, the real problem lies in our understanding and interpretation of the findings, and this constitutes a core issue. Horder et al. (2010) write '...it makes no sense to say that the increasing efficacy of antidepressants in more severe depression is not due to an increase the response to medication but is on the contrary due to decreasing response to placebo'. This statement is very close to the arguments of Kirsch et al. and may reflect the statement that 'the pharmacological effects are defined as [improvement with medication – improvement with placebo]' (Waring, 2008); however, it is essentially misleading.

Theoretically, the Kirsch approach predicts that in the chart of baseline HRSD (*x*-axis) versus effect size of intervention (*y*-axis), which Kirsch et al. included as Figure 3 in their paper (Kirsch et al., 2008), the placebo and the drug regression lines should have been more or less parallel (which they are not in any paper we know), thus reflecting that the effects are additive (which we argue are not). There is no way that the horizontal regression line of the drug can fit that approach and interpretation. Both the classical psychopharmacological approach in the design of RCTs and the Kirsch approach are disputed by these non-parallel regression lines, simply because they are not parallel. Indirectly, this assumption implies a similar biochemical mechanism underlying both the drug effect and the response in the placebo group, and also that the placebo effect is always present and what is under question is the drug effect (Figure 1).

However, if we step outside the usual assumptions, then there is no dead-end and the explanation is simple: the response in the placebo group and the drug effects are non-additive, and antidepressants act independently of depression severity, while the response in the placebo group is present predominantly in milder cases. Clinically this makes sense;

Table 1. Analysis of the Kirsch data with different methods.

	Active drug $N = 3292$				Placebo $N = 1841$					
	Baseline	Change	*d*-value	SD	Baseline	Change	*d*-value	SD	Diff. *d*	Difference in change
Kirsch's results	–	9.60	1.24	–	–	7.80	0.92	–	0.32	1.80
Simple averaging	25.63	10.46	1.27	8.32	25.45	7.53	0.92	8.37	0.35	2.93
Weighted by sample size	24.64	10.04	1.25	8.00	25.28	7.85	0.93	8.42	0.32	2.18
Weighted by inverse of the variance	24.64	10.16	1.28	8.00	25.28	7.48	0.93	8.42	0.35	2.68

melancholic patients respond less to placebo while milder depressive patients have an unstable clinical picture and spontaneous improvements which, in combination with the short duration of the studies, inflates the response in the placebo group. Of course there are other factors responsible for this response – for example, natural course of the illness, natural fluctuation, regression to the mean, and so on (Bland and Altman, 1994a, 1994b) – but in a similar way almost all are related to milder forms of depression and technically constitute unstable 'noise' and 'artefacts'. On the contrary the medication effect is reliable, valid and stable. A contrast of the composition of the drug and the response in the placebo group is shown in Table 2. Some of the responses in the placebo group components are fully present in the drug effect as well (e.g. the regression to the mean); others are clearly almost absent (e.g. expectancy). However, the rest might be partially present or absent and this depends on specific circumstances. For example, the natural fluctuation of the symptoms could be partially present in the drug effect, but also it is possible that the drug forces and levels out this fluctuation. The same could be true for concomitant medication. An additional problem which 'masks' the true difference in the effect size is the problematic

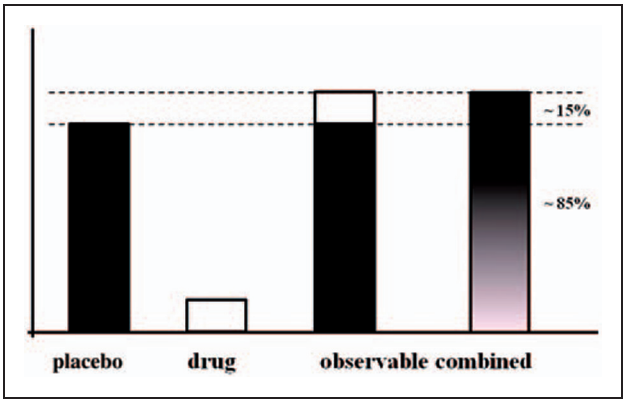

Figure 1. Graphic representation of the active drug effect versus the response in the placebo in group.

Table 2. A possible model of the components of the drug effect and the response in the placebo group. While the response in the placebo group could be equal in size to 85% of the active drug effect, the underlying mechanisms could be quite different. Figures are suggested based on Kaptchuk et al. (2008) and others; see the text for details.

	Drug effect	Placebo effect	% of effect
Regression to the mean	yes	yes	~40%
Natural course of illness	?	yes	
Fluctuation of symptoms (especially in less severe patients)	?	yes	
'Expectancy'	no	yes	~45%
Co-medication (e.g. benzodiazepines)	depends	yes	
Biochemical effect	yes	no	~60%

properties of the HRSD; some of its items assess core features of depression, others non-specific and transnosological symptoms that respond to a variety of agents (e.g. sleep, somatic anxiety) and others might reflect side effects (e.g. loss of libido).

Kirsch et al. (2008) published a figure (Figure 3 in their paper, reprinted as Figure 1 in Horder et al., 2010) showing that the medication effect *d*-values do not increase with increasing severity; however, Horder et al. in their Figure 2 suggested that the corresponding change in HRSD scores did increase. We suggest in Figures 1 and 2 in our re-analysis paper (Fountoulakis and Möller, 2010) that in this instance Kirsch et al. were right. Adopting a raw difference in HRSD scores for the *y*-axis leads to both regression lines having a positive slope; however, this is largely an artefact, since even with random sets of values for baseline and after-treatment scores, the difference correlates approximately with baseline, with a coefficient of 0.7. Also this method does not take into account the variability within studies. Thus, such a chart is not only mistaken, but also impossible to interpret. The use of effect size *d*-value has observable advantages since it controls for the variability within studies.

Since 2008, many commentators have suggested that antidepressants act only in severe depression; milder forms do not respond, hence alternative treatment approaches are more suitable for these patients. The industry strives to include only the more severe patients in the RCTs in order to secure a positive outcome. All of these are false approaches based on widespread misconception and misinterpretation of the data. An immediate consequence of this is that patients suffering from mild depression are deprived from receiving antidepressants, on the basis of this false interpretation of short-term clinical trial data and the overvaluation of 'alternative therapies'. Long-term observations suggest depression is a chronic relapsing disease.

The additivity thesis of pharmacological efficacy is central in RCT logic, being the assumption that the specific or 'true' magnitude of the pharmacological effect is limited to the difference between the drug and placebo responses (Waring, 2008). This is a convenient and practical way to prove a specific drug's efficacy, and does not necessarily demand an identical neurobiological mode of therapeutic action, although the theory behind this method implies such a similarity. This is because the method is 'purely quantitative' and thus demands 'similar quality'. Kirsch's scientific work is largely on 'response expectancy' and has been the focus of Kirsch's research for decades, especially concerning hypnosis, psychotherapy, placebo effects, etc. In a further exploration of the early Gelfand theory, response expectancy was found to be altered by previous experience, and even very small changes in the context of presentation could affect individual differences in the placebo response (entire placebo situation) (Gelfand et al., 1963), while the response in the placebo group was found to significantly associate with response expectancy (Whalley et al., 2008). Essentially this model does not take into consideration the fact that often placebo-arm patients receive additional treatment, usually with benzodiazepines which strongly affect several HRSD items. Patients treated with medication might not experience the same impact of this add-on therapy even if exposure were the same, because they already are receiving a treatment which might influence these symptoms; the effect

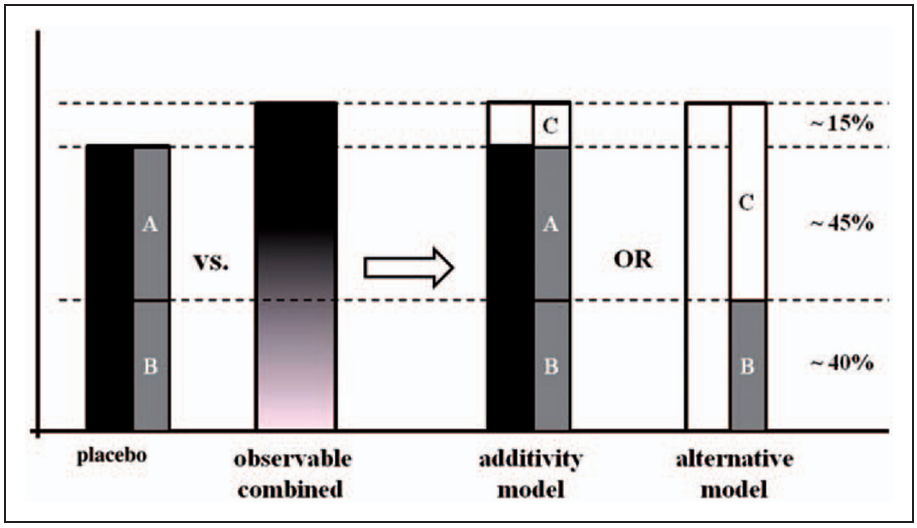

Figure 2. Graphic representation of the active drug effect versus the response in the placebo in group according to two different models, the additivity model and the alternative model. A: expectancy effect; B: regression to the mean, natural course of illness, fluctuation of symptoms; C: biochemical effect of the active drug.

might not be additive or a floor/ceiling effect could be present. Even worse, similar exposure to add-on therapy might increase the adverse events rate (e.g. sedation).

The Kirsch hypothesis concerning depression is that there is a response which lies on a continuum from no intervention at all (e.g. waiting lists) to neutral placebo, then to active and augmented placebo including psychotherapy, and finally to antidepressants which exert a slightly higher efficacy, probably because blinding is imperfect due to side effects (enhanced placebo) (Kirsch, 2004, 2005, 2008a, 2008b, 2009a; Kirsch and Johnson, 2008; Kirsch and Moncrieff, 2007). In order to confirm this theory, namely that all interventions including antidepressants work through 'response expectancy', one needs to prove that interventions with similar presentation characteristics have similar efficacy and that differences in efficacy can be explained by differences in the magnitude of 'response expectancy'.

Research so far suggests that there may not be one placebo response but several, and there are multiple mechanisms involved, which may differ as a function of the context in which the placebo is presented. Co-medication, usually with benzodiazepines according to clinical needs, is also responsible for a portion of the response in the placebo group. In this line of research, there was some data suggesting that a large component of the 'active' treatment was placebo-mediated (Moncrieff and Kirsch, 2005). The meta-analysis of antidepressant trials comparing those with a run-in period to those without one suggests that this method does not increase drug–placebo differences (Posternak et al., 2002; Trivedi and Rush, 1994), which means that maybe any patient could be a placebo-responder and this is not possible to predict. There are data suggesting that the proportion of patients who respond are indeed on a continuum (e.g. 28% on waiting list, 44% in limited group, and 62% in augmented group; Kaptchuk et al., 2008). The model, however, predicts that all augmented placebos should have similar efficacy and thus this theory does not explain why psychotherapy is inferior to pharmacotherapy (Cuijpers et al., 2010a, 2010b) since both are considered to have similar 'enhanced placebo' qualities according to this line of thinking. Kirsch seems to consider psychotherapy closer to active placebo and antidepressants closer to augmented placebo, partially because of methodological issues concerning the ability to blind (side effects make patients realize that they are taking drug instead of placebo, and this increases their expectation) (Kirsch, 2009a).

The final finding in support of the Kirsch theory was the results of his 2008 meta-analysis (Kirsch et al., 2008), which reported that the effect size and the magnitude of change in HDRS score were small (effect size *d*-value below 0.50 and change in HRSD equal to 1.80 points) and thus antidepressants fell well below criteria for clinical relevance suggested by the NICE; however, these criteria are not generally accepted (Möller, 2008). Similar findings were reported by Barbui et al. for paroxetine alone (Barbui et al., 2008). Kirsch et al. also reported that efficacy reaches clinical relevance only in trials involving the most extremely depressed patients, and that this pattern is due to a decrease in the response in the placebo group rather than an increase in the response to medication. They also found no linear relation between severity and response to medication. However, the reported 'efficacy' based on RCTs using a 'last observation carried forward' (LOCF) analysis is, in fact, a hybrid of both efficacy and tolerability.

As a final requirement for the model to be considered as correct, one needs to prove that antidepressants are not significantly better than placebo and that there are no differences among antidepressants. The supposed equal efficacy of all types of antidepressants supports this thesis – that is, SSRIs, tricyclics and monoamine oxidase inhibitors, despite the fact that they have different modes of operation; the supposed similar efficacy of some active drugs that are not considered antidepressants (amylobarbitone, lithium, liothyronine and adinazolam); and the high correlation between the placebo and the drug response (Kirsch, 2000), which implies a similar mechanism underlying response no matter the treatment

intervention. However, this argument is not strong, since in general medicine different classes of medication acting through different pathways might treat the same medical condition. It is to be noted that there are neuroimaging data suggesting that brain alterations in patients who improve are independent of treatment intervention (Konarski et al., 2009; Martin et al., 2001), while others suggest that the neurochemistry behind the drug response and the response in the placebo group are different (Mayberg et al., 2002).

However, the finding that antidepressants act at the same magnitude irrespective of initial severity while the response in the placebo group depends on it suggests a different mechanism underlying these two different interventions, with antidepressants being unrelated to response expectancy. Response expectancy is strongly related to severity of depression (according to cognitive theory), thus the regression lines in Kirsch et al.'s Figures 2 and 3 should have been parallel. This is the reason why, at higher severity, the response in the placebo group drops close to the levels of waiting list. These authors stress the finding that there was a negative relation between severity and the response in the placebo group, whereas there was no difference between those with relatively low and relatively high initial depression in their response to drug. Thus, the increased benefit for extremely depressed patients seems attributable to a decrease in responsiveness in the placebo group, rather than an increase in responsiveness to medication. However, they do not explain how this fits their position. Our analysis revealed that over the years, there was an increase in both the active drug and the placebo groups with regression lines being parallel (Figure 2 of the re-analysis paper of Fountoulakis and Möller, 2010), maybe suggesting that the techniques developed over the years to support and keep patients in the study also increase expectancy.

Other data against the Kirsch theory come from a recent meta-analysis suggesting escitalopram is the most effective agent with also the highest tolerability (Cipriani et al., 2009). This means that increased efficacy cannot be explained on the basis of unblinding because of side effects: escitalopram is better tolerated than several other antidepressants.

If we reject the additivity hypothesis and accept that there is a radically different composition between the drug and the response in the placebo group, then a different picture emerges. If we accept that the proportion of patients who respond are indeed on a continuum with around 28% on waiting list, 44% in the limited group, and 62% in the augmented group (Kaptchuk et al., 2008), and we apply these data in the list of Table 2 (rounded, suggesting an overall 75% response to active drug and the placebo effect to be 85% of the drug effect), we arrive at a rough model of qualitative and quantitative differences between the drug and the placebo response. According to such a model (Figure 2), although the placebo effect could be equal in size to 85% of the active drug effect, the qualitative composition is quite different and this difference is responsible for the long-term beneficial effect of pharmacotherapy. It is interesting that Kirsch seems to agree with this approach and has suggested that 'If there is a reasonably large pharmacologic effect, then it cannot be an addition to the placebo effect, in which case conventional double-blind studies are not appropriate for testing the drug effects' (Kirsch, 2000). We completely agree with this position. Either medications are similar to placebo – and there is a combined effect maybe only in the most severe cases of depression – or medications are superior to placebo: they act independently of initial severity and the additivity thesis is irrelevant. In essence we agree with the line of reasoning of Kirsch on this issue, but we arrive at the opposite conclusion.

Table 3. Conclusions of the current paper.

1. Antidepressant drugs are effective
2. Their efficacy is independent of initial severity
3. Their efficacy is unrelated to 'expectancy'
4. At least some antidepressants fulfill the NICE criteria for 'clinical relevance'
5. The response in the placebo group depends (directly or indirectly) on the initial severity
6. The additivity hypothesis is incorrect
7. Alternative therapies have no established efficacy
8. The Kirsch et al. (2008) meta-analysis reports false figures and interprets data in a false way

Conclusion

A summary of conclusions is shown in Table 3. We argue that the Kirsch results themselves suggest that although a large percentage of the response in the placebo group is due to expectancy, this is not true for the active drug, and that the drug and placebo effects are not additive. The drug effect is unrelated to depression severity while the placebo response is reduced in more severe depression. If this is confirmed in future research, then the value of the RCT as the major tool for investigating the efficacy of antidepressants may be doubtful. Unfortunately, the authors are currently in no position to propose an alternative.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict 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 a research grant from Pfizer Foundation. He is member of the board of Wyeth for desvenlafaxine and Bristol-Myers Squibb for aripirpazole in bipolar disorder.

HJM has received grants or is a consultant for and on the speakership bureaus of AstraZeneca, Bristol-Myers Squibb, Eisai, Eli Lilly, GlaxoSmithKline, Janssen Cilag, Lundbeck, Merck, Novartis, Organon, Pfizer, Sanofi-Aventis, Schering-Plough, Schwabe, Sepracor, Servier and Wyeth.

References

Barbui C, Furukawa TA and Cipriani A (2008) 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* 178: 296–305.

Bech P (2001) Meta-analysis of placebo-controlled trials with mirtazapine using the core items of the Hamilton Depression Scale as evidence of a pure antidepressive effect in the short-term treatment of major depression. *Int J Neuropsychopharmacol* 4: 337–345.

Bech P (2004) Modern psychometrics in clinimetrics: impact on clinical trials of antidepressants. *Psychother Psychosom* 73: 134–138.

Bech P (2006) Rating scales in depression: limitations and pitfalls. *Dialogues Clin Neurosci* 8: 207–215.

Bech P (2010) Is the antidepressive effect of second-generation antidepressants a myth? *Psychol Med* 40: 181–186.

Bech P, Cialdella P, Haugh MC, et al. (2000) Meta-analysis of randomised controlled trials of fluoxetine v. placebo and tricyclic antidepressants in the short-term treatment of major depression. *Br J Psychiatry* 176: 421–428.

Bech P, Tanghoj P, Cialdella P, et al. (2004) 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* 7: 283–290.

Bland J and Altman D (1994a) Regression towards the mean. *BMJ* 308: 1499.

Bland J and Altman D (1994b) Some examples of regression towards the mean. *BMJ* 309: 780.

Broich K (2009) Committee for Medicinal Products for Human Use (CHMP) assessment on efficacy of antidepressants. *Eur Neuropsychopharmacol* 19: 305–308.

Cipriani A, Furukawa TA, Salanti G, et al. (2009) Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet* 373: 746–758.

Cuijpers P, Smit F, Bohlmeijer E, et al. (2010a) Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias. *Br J Psychiatry* 196: 173–178.

Cuijpers P, van Straten A, Schuurmans J, et al. (2010b) Psychotherapy for chronic major depression and dysthymia: a meta-analysis. *Clin Psychol Rev* 30: 51–62.

Fountoulakis KN, Kaprinis SG and St Kaprinis G (2004) Geriatric depression: a challenge for the 21st century. *Aging Clin Exp Res* 16: 173–175.

Fountoulakis KN and Möller HJ (2010) Efficacy of antidepressants: a re-analysis and re-interpretation of the Kirsch data. *Int J Neuropsychopharmacol* 1–8.

Fountoulakis KN, O'Hara R, Iacovides A, et al. (2003) Unipolar late-onset depression: A comprehensive review. *Ann Gen Hosp Psychiatry* 2: 11.

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

Gelfand S, Ullmann LP and Krasner L (1963) The placebo response: an experimental approach. *J Nerv Ment Dis* 136: 379–387.

Ghaemi SN (2008) Why antidepressants are not antidepressants: STEP-BD, STAR*D, and the return of neurotic depression. *Bipolar Disord* 10: 957–968.

Horder J, Matthews P and Waldmann R (2010) Placebo, Prozac and PLoS: significant lessons for psychopharmacology. *J Psychopharmacol* 25: 1277–1288.

Kaptchuk TJ, Kelley JM, Conboy LA, et al. (2008) Components of placebo effect: randomised controlled trial in patients with irritable bowel syndrome. *BMJ* 336: 999–1003.

Kirsch I (2000) Are drug and placebo effects in depression additive? *Biol Psychiatry* 47: 733–735.

Kirsch I (2004) Conditioning, expectancy, and the placebo effect: comment on Stewart-Williams and Podd (2004). *Psychol Bull* 130: 341–343; discussion 344-345.

Kirsch I (2005) Placebo psychotherapy: synonym or oxymoron? *J Clin Psychol* 61: 791–803.

Kirsch I (2008a) Antidepressant drugs 'work', but they are not clinically effective. *Br J Hosp Med (Lond)* 69: 359.

Kirsch I (2008b) Challenging received wisdom: antidepressants and the placebo effect. *Mcgill J Med* 11: 219–222.

Kirsch I (2009a) Antidepressants and the placebo response. *Epidemiol Psichiatr Soc* 18: 318–322.

Kirsch I (2009b) *The Emperor's New Drugs: Exploding the Antidepressant Myth*. London: The Bodley Head.

Kirsch I, Deacon BJ, Huedo-Medina TB, et al. (2008) Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med* 5: e45.

Kirsch I and Johnson BT (2008) Moving beyond depression: How full is the glass? *BMJ* 336: 629–630.

Kirsch I and Moncrieff J (2007) Clinical trials and the response rate illusion. *Contemp Clin Trials* 28: 348–351.

Konarski JZ, Kennedy SH, Segal ZV, et al. (2009) Predictors of non-response to cognitive behavioural therapy or venlafaxine using glucose metabolism in major depressive disorder. *J Psychiatry Neurosci* 34: 175–180.

Martin SD, Martin E, Rai SS, et al. (2001) Brain blood flow changes in depressed patients treated with interpersonal psychotherapy or venlafaxine hydrochloride: preliminary findings. *Arch Gen Psychiatry* 58: 641–648.

Mayberg HS, Silva JA, Brannan SK, et al. (2002) The functional neuroanatomy of the placebo effect. *Am J Psychiatry* 159: 728–737.

McAllister-Williams RH (2008) Do antidepressants work? A commentary on "Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration" by Kirsch et al. *Evid Based Ment Health* 11: 66–68.

Möller H (2009a) Antidepressants: controversies about their efficacy in depression, their effect on suicidality and their place in a complex psychiatric treatment approach. *World J Biol Psychiatry* 10: 180–195.

Möller H and Maier W (2010a) Evidence-based medicine in psychopharmacotherapy: possibilities, problems and limitations. *Eur Arch Psychiatry Clin Neurosci* 260: 25–39.

Möller HJ (2008) Isn't the efficacy of antidepressants clinically relevant? A critical comment on the results of the metaanalysis by Kirsch et al. 2008. *Eur Arch Psychiatry Clin Neurosci* 258: 451–455.

Möller HJ (2009b) Antidepressants: controversies about their efficacy in depression, their effect on suicidality and their place in a complex psychiatric treatment approach. *World J Biol Psychiatry* 10: 180–195.

Möller HJ (2009c) Is evidence sufficient for evidence-based medicine? *Eur Arch Psychiatry Clin Neurosci* 259(Suppl. 2): S167–S172.

Möller HJ and Broich K (2010) Principle standards and problems regarding proof of efficacy in clinical psychopharmacology. *Eur Arch Psychiatry Clin Neurosci* 260: 3–16.

Möller HJ and Maier W (2010b) Evidence-based medicine in psychopharmacotherapy: possibilities, problems and limitations. *Eur Arch Psychiatry Clin Neurosci* 260: 25–39.

Moncrieff J and Kirsch I (2005) Efficacy of antidepressants in adults. *BMJ* 331: 155–157.

Moncrieff J, Wessely S and Hardy R (2004) Active placebos versus antidepressants for depression. *Cochrane Database Syst Rev* CD003012.

Posternak MA, Zimmerman M, Keitner GI, et al. (2002) A reevaluation of the exclusion criteria used in antidepressant efficacy trials. *Am J Psychiatry* 159: 191–200.

Rush AJ (2007) STAR*D: what have we learned? *Am J Psychiatry* 164: 201–204.

Trivedi MH and Rush H (1994) Does a placebo run-in or a placebo treatment cell affect the efficacy of antidepressant medications? *Neuropsychopharmacology* 11: 33–43.

Turner EH, Matthews AM, Linardatos E, et al. (2008) Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 358: 252–260.

Waring DR (2008) The antidepressant debate and the balanced placebo trial design: an ethical analysis. *Int J Law Psychiatry* 31: 453–462.

Whalley B, Hyland ME and Kirsch I (2008) Consistency of the placebo effect. *J Psychosom Res* 64: 537–541.