

MYTH OR CURE?
THE PLACEBO ARGUMENT IN THE
CONTEMPORARY DEBATE ON THE
PHARMACOLOGICAL TREATMENT OF
DEPRESSION

Juha Rudanko

Syventävien opintojen opinnäytetyö

Tampereen yliopisto

Lääketieteen ja biotieteiden tiedekunta

Elokuu 2017

Tampereen yliopisto
Lääketieteen ja biotieteiden tiedekunta
Yleislääketiede

RUDANKO JUHA: MYTH OR CURE? THE PLACEBO ARGUMENT IN THE
CONTEMPORARY DEBATE ON THE PHARMACOLOGICAL TREATMENT OF
DEPRESSION

Kirjallinen työ, 58 s.
Ohjaaja: professori Raimo Puustinen

Elokuu 2017

Avainsanat: masennuslääkkeet, SSRI, hoitosuosituks

Masennuslääkkeet ovat keskeinen depression hoitomuoto yleislääketieteessä ja psykiatriassa. Psykiatri Joanna Moncrieff ja psykologi Irving Kirsch esittävät plaseboargumentin kyseenalaistaakseen niiden tehon.

Yleisen plaseboargumentin mukaan masennuslääkkeiden tehoero lumelääkkeeseen verrattuna on vaatimaton. Moncrieff ja Kirsch pyrkivät myös osoittamaan, että tilastollisesti merkitsevät erot masennus- ja lumelääkkeen välillä ovat kliinisesti merkityksettömiä. He myös pyrkivät osoittamaan, että lääkkeillä ei ole todellista farmakologista vaikutusta masennukseen. Tätä osaa argumentista kutsun "aktiivisen plasebon" argumentiksi.

Analysoin työssäni *Myth or cure? The placebo argument in the contemporary debate on the pharmacological treatment of depression* kriittisesti Kirschin ja Moncrieffin argumentin osa-alueet. Peilaan heidän argumentaatiotaan psykiatrian valtavirtaan, jota edustaa depression Käypä hoito -suositus.

Työn tavoite on selvittää, onnistuvatko Moncrieff ja Kirsch osoittamaan keskeinen masennuksen hoitomuoto tehottomaksi.

Yleinen plaseboargumentti pitää käyttämäni materiaalin perusteella paikkansa, mutta sekä kriitikkojen että Käypä hoito -suosituksen esittämiä väitteitä masennuslääkkeiden tehon kliinisestä merkityksestä on mahdotonta arvioida. Kriitikkojen "aktiivinen plasebo" -argumentti epäonnistuu, koska se on ristiriidassa heidän käyttämiensä tutkimusten kanssa.

Contents

Introduction	1
1. “Antidepressants are effective for all severities of depression”: the guidelines	6
1.1 Short-term treatment	6
1.2 Long-term treatment	18
2. Placebo: the Kirsch-Moncrieff argument	23
2.1 Moncrieff on antidepressants versus placebo	23
2.1.1 Short-term treatment	23
2.1.2 Severity	25
2.1.3 Long-term treatment	30
2.2 The placebo argument: Kirsch	33
3. “Active placebo”	37
3.1 “Active placebo”: Kirsch	37
3.2 Moncrieff on antidepressants versus “active placebo”	41
3.3 Moncrieff on antidepressants versus other drugs	42
4. Assessing the Kirsch-Moncrieff Argument	44
4.1 Short-term treatment	44
4.2 Severity	45
4.3 Long-term treatment	47
4.4 “Active placebo”	48
Conclusion	52
Bibliography	55

Introduction

Depression is one of the leading causes of morbidity both in Finland and worldwide.¹ In current treatment guidelines, pharmacological treatment is presented as central, especially in cases of moderate or severe depression.² The drugs of first choice for depression are the selective serotonin reuptake inhibitors (SSRIs). Other newer drugs, such as the serotonin-noradrenaline reuptake inhibitors (SNRIs), are also possibilities for first-line treatment of depression. The tricyclic antidepressants (TCAs), the first choice drug treatment before the advent of the SSRIs, are still widely in use, and are possibilities for patients whose depression does not respond to SSRI or SNRI treatment.

Psychiatrists claim that two thirds of depressed patients taking antidepressants experience a significant amelioration of symptoms, whilst one third of patients taking placebo experience the same effect.³ In the era of evidence-based medicine, the ultimate justification for the use of antidepressants in treating depressed patients is this apparent superiority compared to placebo in randomized controlled trials.

In recent years, critics have challenged the evidence on the effectiveness of antidepressants. A number of meta-analyses, which have shown little difference between active drugs and placebo, have been published.⁴ Some critics have questioned whether antidepressants have any specific pharmacological effect on depression at all. This essay explores these critiques of the evidence base of antidepressant treatment. I will focus on the work of two prominent critics: Irving Kirsch and Joanna Moncrieff.

Psychologist Irving Kirsch, currently Director of the Program in Placebo Studies and lecturer at Harvard Medical School, has co-authored a number of meta-analyses investigating drug and placebo responses in depression. He has also written a book intended for non-specialist readers, *The Emperor's New Drugs*, which outlines his thinking on antidepressants.⁵ Kirsch states that he

I would like to thank Professor Raimo Puustinen for patient supervision of this project, and insightful comments on numerous versions of the manuscript. I would also like to thank psychologist Samuli Kangaslampi for incisive comments on the manuscript and especially for clarifying some of the statistics in the studies discussed here.

¹ Kessler RC & Bromet EJ. The Epidemiology of Depression Across Cultures. *Annu Rev Public Health* 2013; 34:119-138. Lönnqvist J, Henriksson M, Marttunen M *et al.* *Psykiatria*. Helsinki: Kustannus Oy Duodecim 2011.

² Depression. Current Care Guidelines. Working group set up by the Finnish Medical Society Duodecim and the Finnish Psychiatric Association. Helsinki: The Finnish Medical Society Duodecim, 2016.

³ *ibid.*

⁴ Eg. Kirsch I, Moore TJ, Scoboria A *et al.* *The Emperor's New Drugs: An Analysis of Antidepressant Medication Data Submitted to the U.S. Food and Drug Administration*. Prevention & Treatment 2002. I was only able to obtain an online copy without the original page numbers, so the numbers given below refer only to pages within this single article.

⁵ Kirsch I. *The Emperor's New Drugs: Exploding the Antidepressant Myth*. London: The Bodley Head 2009.

came to be interested in the efficacy of antidepressants through his original interest in the placebo response in depression. He approaches the issues covered in this essay based on the meta-analyses he has co-authored. In Finland, the debate in the media over the efficacy of antidepressants has largely arisen from news coverage of Kirsch's meta-analyses.

Joanna Moncrieff is a practicing psychiatrist based in the UK. Her project is a critique of contemporary psychiatry as a whole, and is by no means limited to arguing that antidepressants are no better than placebo in treating depression.⁶ She questions the concept of mental illness in contemporary psychiatry, arguing that psychiatry is fundamentally political. Invoking Marx and Foucault, she defines her task as uncovering the interests that have been at work in establishing what she calls the "disease-centered" model of psychiatric disorder.⁷ She argues that psychiatry conceptualizes the disorders it diagnoses as "chemical imbalances" in the brain, and that psychiatry understands the pharmacological agents it employs as working by correcting these imbalances.⁸ Her central thesis is that this understanding of psychiatric disorder or its treatment is not based on strong evidence, but has become entrenched in psychiatry because it serves powerful interests.⁹

She argues that the disease-centered model should be replaced with what she calls the "drug-centered" one. This model acknowledges that psychiatric drugs do not have any specific effect on some underlying biological pathology, but they do have unspecific effects, such as sedation, that can be useful for relieving the distress of some depressed patients.¹⁰

Despite their differing foci, Kirsch and Moncrieff share their major arguments, and some of their evidence. Exploring the work of both critics provides a more complete picture of these arguments.

The central thrust of their critique is what I call the placebo argument. In short, it is the claim that antidepressants are not significantly more effective than placebo, that differences between active drugs and placebo in antidepressant trials are very small. In other words, it states that the clinical improvement seen in patients taking antidepressants is almost completely explained by a non-specific response that occurs with placebo treatment as well.

This is the general claim that the critics make. There are two more specific claims, which are aspects of the general argument. The first, which I call the clinical relevance argument, is that, even though there are statistically significant differences between active drug and placebo in antidepressant trials, these are so small as to be clinically meaningless. Even though the critics

⁶ Moncrieff J. *The Myth of the Chemical Cure: A Critique of Psychiatric Drug Treatment*. Houndmills: Palgrave Macmillan 2008.

⁷ *ibid.*, 13.

⁸ *ibid.*, 9-11.

⁹ *ibid.*, 13.

¹⁰ *ibid.*, 14-29.

emphasize this, they reference such limited material on this issue that an in-depth analysis is not possible here. I return to clinical relevance in the Conclusion.

The second claim attempts to explain the small differences between drug and placebo. This is the “active placebo” argument, which states that the statistical superiority of drug over placebo in antidepressant trials is not due to a true pharmacological effect, but rather to patients breaking blind, as the more frequent side effects of the active drug lead patients to realize that they are receiving active drug treatment, boosting their expectations of recovery.

The aim of this essay is to consider to what extent the critics hit the mark. Are they able to undermine the basis of the contemporary approach to treating depression, or are their arguments straw men, misrepresenting the way psychiatry conceptualizes these key treatments?

The conventional best practice of treating depression will be represented by the treatment guidelines (hereafter “guidelines”) published by the Finnish Medical Society Duodecim.¹¹ Doctors routinely refer to them to guide their clinical decision-making; they are also a major source for teaching medical students. The guidelines also explicitly refer to the debate on the efficacy of antidepressants. For the purposes of this study, they are sufficient representatives of mainstream psychiatry.

The guidelines cover many aspects of the diagnosis and treatment of depression. I only discuss the evidence summaries for short- and long-term antidepressant drug treatment for depression, authored by Professor of Psychiatry Erkki Isometsä, since this essay is focused on the debate on the effectiveness of such drugs. All references to the guidelines, unless otherwise specified, are to these evidence summaries appended to the full guideline.¹²

This essay is organized around the three major aspects of the critics’ placebo argument. The material in different sections will inevitably overlap, as the different aspects of the overall argument are closely connected. In the first section, I analyze arguments on the effectiveness of antidepressants compared to placebo. I first critically examine the evidence that the guidelines draw on to support the view that antidepressants are effective for short- and long-term treatment of depression. In the next section, I examine the evidence that Moncrieff uses to dispute such claims. I consider the evidence on three aspects of her argument: antidepressant versus placebo for short-term treatment, the effect of the initial severity of depression on drug/placebo differences, and drug

¹¹ Duodecim 2016.

¹² As far as I can tell, the evidence summaries are only available online, so it is not possible to be as accurate with references to them as I have sought to be with references to other materials. I will refer to them with “Duodecim 2016”. In other cases, I have tried to include page numbers for the exact locations of arguments and data, so the reader can easily check whether I have presented the material fairly. The evidence summaries are freely available online in Finnish. Acute treatment: <http://kaypahoito.fi/web/kh/suosituksset/suositus?id=nak04327&suositusid=hoi50023>. Long-term treatment: <http://kaypahoito.fi/web/kh/suosituksset/suositus?id=nak04347&suositusid=hoi50023> and <http://kaypahoito.fi/web/kh/suosituksset/suositus?id=nak04350&suositusid=hoi50023>

versus placebo for long-term treatment. In the final part of this section I discuss the evidence that Kirsch presents for his version of this aspect of the placebo argument. The focus is on Kirsch's own meta-analyses.

In the next section, I discuss the "active placebo" aspect of the overall argument. The critics present two "active placebo" claims. The first is that double-blind antidepressant trials are not truly blind, because the more frequent incidence of side effects in the drug treatment group enables patients to guess they are receiving active treatment, changing expectations of recovery. The second is that other pharmacological agents, which are not considered specific antidepressants, show superiority over placebo in clinical trials. The idea is that like antidepressants, these drugs act as "active placebos", producing side effects, which enable patients to guess which treatment group they have been assigned to.

Throughout, I focus my discussion primarily on the SSRIs, SNRIs, and other newer pharmacological agents widely used today. There is inevitably some overlap with older drugs, particularly TCAs, since many of the studies discussed include these older drugs. Joanna Moncrieff in particular bases much of her argument on older TCA trials. Likewise, there is some overlap with other forms of depression – psychotic depression and bipolar disorder are discussed in passing – but the focus is on major depressive disorder.

There are three important aspects of the critics' argument that it is not possible to discuss systematically in the main body of this essay – I will outline these issues in the Conclusion. The first is the question of clinical relevance. It is crucial to the critical argument, but there is very limited material for assessing it, and it is not systematically addressed by the guidelines. The second one is the question of how the use of antidepressants affects depression outcome in real-world settings. The third is the critics' conceptual argument: that depression is not a malfunction of the brain's monoamine system, and that therefore, the notion that depression can be cured by using drugs correcting an imbalance in that system is a myth.

Pharmacological treatment is central to contemporary psychiatry and to the treatment of psychiatric disorders in general practice. Furthermore, depression is common and disabling; effective treatments are essential. If the critics are right, the implications for both psychiatry and general practice are vast. Is the widespread prescribing of antidepressants justified? Why does clinical experience seem to confirm the effectiveness of antidepressants? Why do so many patients respond to placebo in antidepressant trials? How big should the drug/placebo difference in clinical trials be to justify the use of antidepressants? How should we help people suffering from depression? What does the effectiveness or lack thereof of antidepressants tell us about the etiology of depression? There are a plethora of issues for clinical trials, meta-analyses, and more conceptual research to address in the territory opened up by this debate.

In conclusion, I attempt to draw some implications from the evidence and arguments covered in this essay for the individual clinician facing the depressed patient, who is desperate for alleviation of their suffering. Is prescribing an antidepressant a reasonable way of attempting to ease their distress?

1. “Antidepressants are effective for all severities of depression”: the guidelines

The guidelines note that there has been critical debate on how clinically relevant drug/placebo differences are, but argue that the evidence is unequivocal: it shows that antidepressants are effective in all severities of depression.¹³ The guidelines present evidence for both acute-phase and long-term treatment. In this section, I critically discuss the way that the guidelines present the findings and conclusions of these meta-analyses.

The guidelines cite eleven meta-analyses in support of the conclusion that antidepressants are effective for acute-phase treatment. I omit two of these studies: the first one because it concerns reboxetine, which is no longer in use, and the second one because it is a study of dysthymia rather than major depressive disorder. The guidelines further discuss four meta-analyses in support of the conclusion that antidepressants are effective for preventing depression relapse over the long term; I turn to these in section 1.2.

1.1 Short-term treatment

The first meta-analysis cited by the guidelines, Turner *et al.* 2008, focuses on the effects of publication bias by comparing published and unpublished double-blind placebo-controlled trials for twelve antidepressants.¹⁴ The unpublished trials are drawn from studies submitted to the Food and Drug Administration (FDA) in the United States by industry for approval of the antidepressants. Even though drug companies are under no obligation to publish the results of all the trials they run, the FDA does require them to submit all the trials they have conducted in order to gain approval for a new drug. Included in the meta-analysis are 74 studies including 12 564 patients submitted to the FDA for the approval of the following antidepressants: bupropion, citalopram, duloxetine, escitalopram, fluoxetine, mirtazapine, nefazodone, paroxetine, sertraline, and venlafaxine.¹⁵ Of these, citalopram, escitalopram, fluoxetine, paroxetine and sertraline are SSRIs. Nefazodone has been withdrawn from the market.

The researchers categorized the studies based on whether the FDA considered the trial result positive or negative. Of the published studies, 94% were deemed positive by the FDA, whereas when the unpublished studies are included, only 51% of all trials show superiority over placebo.¹⁶ As the guidelines point out, the antidepressants fared worse when all the studies were included, but

¹³ Duodecim 2016.

¹⁴ Turner EH, Matthews AM, Linardatos E *et al.* Selective publication of antidepressant trials and its influence on apparent efficacy. *The New England Journal of Medicine* 2008; 358: 252-260. For brevity, I will only use the first author's name and year of publication when referring to a study in the main text.

¹⁵ *ibid.*, 258, Figure 3.

¹⁶ *ibid.*, 255.

they were still all superior to placebo.¹⁷ The overall Hedges's *g* effect size of all the antidepressants is 32 per cent higher when only published trials are included.¹⁸

The second meta-analysis cited by the guidelines is co-authored by Irving Kirsch, one of the leading critics of antidepressant use.¹⁹ This study also attempts to work around publication bias by utilizing trials submitted to the FDA. The researchers used the Freedom of Information Act to request the data submitted to the FDA for the approval of six antidepressants: fluoxetine, venlafaxine, nefazodone, paroxetine, sertraline and citalopram. In total, the meta-analysis included 35 clinical trials. The meta-analysis focused on investigating the effects that the initial severity of depression has on patients' response to antidepressants. Severity was rated on the Hamilton depression scale.

Overall, the study found that patients taking antidepressants improved by a mean of 9.60 points on the Hamilton scale, and patients taking placebo improved by 7.80 points. In other words, the mean difference between drug and placebo was 1.80.²⁰ The statistical analysis shows that the superiority of drug over placebo grows as the initial severity of depression rises. Partly this is because drug response improves as the initial Hamilton score rises, partly because placebo response diminishes as the initial Hamilton score grows.²¹

In all but one of the trials the patients were very severely depressed. Thus, the one study including moderately depressed patients was deemed an outlier, and the researchers reran the statistical analysis with this one study excluded.²² In this analysis, the relationship of drug efficacy to initial severity of depression is flat. In other words, drug efficacy remains constant even as the initial severity of depression rises (within the very severe range). In this analysis, antidepressants are superior to placebo. Response to placebo in this analysis decreases as initial severity of depression rises.²³ In summary, this meta-analysis shows that when the outlier study is excluded, drug efficacy is flat as initial depression becomes more severe, whereas placebo performs worse. It is not that drug response is better the more depressed the patient is but that they are less likely to respond to placebo.²⁴

The guidelines state that the study unequivocally shows the superiority of antidepressants over placebo. As we have seen, it is true that when the one outlier is excluded, the drugs outperform

¹⁷ *ibid.*, 259; Duodecim 2016.

¹⁸ Turner, Matthews & Linardatos *et al.* 2008, 258. Of the SSRIs, sertraline fares the worst: its effect size is 64% bigger when only published studies are included, dropping to 0.26 when unpublished studies are included. With paroxetine, the difference is the smallest: the effect size drops from 0.36 to 0.32.

¹⁹ Kirsch I, Deacon BJ, Huedo-Medina TB *et al.* Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med* 2008; 5: 260-268.

²⁰ *ibid.*, 263.

²¹ *ibid.*, 263.

²² *ibid.*, 263.

²³ *ibid.*, 265, Figure 3.

²⁴ *ibid.*, 266.

placebo. However, the guidelines also state that the difference between drug and placebo does not seem clinically significant in cases of mild depression, with a score of less than 20 points on the Hamilton scale cited as the criterion of mild depression. This is different from Kirsch *et al.*'s criterion: according to them, mild depression is indicated by a score of 8-13 on the Hamilton scale, and severe depression by a score of 19-22.²⁵ Thus a score of less than 20 could be categorized as severe or moderate depression by Kirsch *et al.*

Furthermore, as we have seen, there is only one study included in the meta-analysis where the mean initial severity of depression was less than 20 points on the Hamilton scale. In this fluoxetine trial, the baseline severity was 17 points, which would indicate *moderate* depression according to the criteria used by Kirsch *et al.* In other words, if we follow the criteria used by Kirsch and colleagues, no studies of mild depression were included in the meta-analysis. The guidelines claim that on the basis of this meta-analysis, the drug/placebo difference seems clinically insignificant in mild depression. It is difficult to see how this claim could be made, as it is not possible to say anything about mild depression on the basis of this study.

The guidelines also seem to imply that even though the drug/placebo difference may not be clinically relevant for mild depression, it is clinically relevant for more severe depression, though this is not explicitly stated. This completely ignores Kirsch *et al.*'s central conclusion, which is that the drug/placebo difference reaches clinical significance only in the case of the most severely depressed patients. As noted, the average drug/placebo difference was 1.8 points on the Hamilton scale. The authors point out that the National Institute for Clinical Excellence (NICE) in the UK has adopted a 3-point difference as the threshold of clinical relevance.²⁶ Excluding the outlier study, the researchers showed that for the most severely depressed patients, there is a difference between drug and placebo that is clinically relevant according to the NICE criteria.²⁷ Note that whilst the guidelines imply that the drug/placebo difference is only clinically *irrelevant* in the case of mild depression, Kirsch *et al.*'s conclusion is that the difference reaches clinical *relevance* only in the case of the most severely depressed patients.

Melander *et al.* 2008 is the third meta-analysis cited by the guidelines.²⁸ This study refers to Kirsch *et al.* 2008 and other studies which question the efficacy of antidepressants and poses the question: if antidepressants are not more effective than placebo in a clinically meaningful sense, why were they approved for sale by the authorities in the first place?²⁹

²⁵ *ibid.*, 266; Kirsch 2009, 31.

²⁶ Kirsch, Deacon & Huedo Medina *et al.* 2008, 266.

²⁷ *ibid.*, 266.

²⁸ Melander H, Salmonson T, Abadie E *et al.* A regulatory Apologia - A review of placebo-controlled studies in regulatory submissions of new-generation antidepressants. *European Neuropsychopharmacology* 2008; 18: 623-627.

²⁹ *ibid.*, 624.

The authors note that the meta-analyses, which question the clinical relevance of antidepressant/placebo differences, focus on average differences in Hamilton scores between drug and placebo groups. They argue that this is an inadequate standard. They maintain that such group-level differences in mean depression scores are useful for establishing the statistical significance of the drug effect, but that clinical relevance should be decided on the basis of the “percentage of patients achieving a clinically meaningful response”.³⁰ The threshold of this clinically meaningful response is then defined as at least a 50% reduction in Hamilton score.³¹ No justification for this figure is given, though the authors do point out that it is widely used in antidepressant studies, and that, since a Hamilton score of at least 18 is required for inclusion in most trials, a patient would need to experience at least a nine point reduction to be counted a responder.³² The authors also note that this should be compared to the average absolute difference of around two points “usually observed”.³³

The study included all the trials that had been submitted by industry for the approval of six SSRIs and two SNRIs for sale in Sweden. It included randomized placebo-controlled trials that lasted at least four weeks. In total, the study included 56 studies with 7374 patients. As the meta-analysis is based on all studies submitted to the regulator for approval, the influence of publication bias should be minimal.³⁴

The response rate for patients taking an active drug was 48%, and for patients taking placebo it was 32%.³⁵ All the individual antidepressants were found to be superior to placebo. The difference in response rates for individual antidepressants versus placebo ranged from 13.5% to 19.3%-units. No relationship between baseline severity of depression and response was found.³⁶

The guidelines mention the results noted above, but they also claim that the meta-analysis covered all severities of depression, and showed the effectiveness of antidepressants in mild depression, as well.³⁷ It is hard to understand this claim. The mean initial depression scores on the Hamilton scale ranged from 19.8 to 23.8, with an average score of 21.6. The range of individual scores varied from 17.6 to 28.4.³⁸ In other words, the least depressed patient included in the meta-analysis had a score of 17.6, and the most severely depressed patient had a score of 28.4. As noted above, the cut off point for mild depression that Kirsch uses is 13 and the one for moderate depression is 18. In the Hamilton questionnaire available on Duodecim’s Terveystietäminen portal, a

³⁰ *ibid.*, 624.

³¹ *ibid.*, 624.

³² *ibid.*, 626.

³³ *ibid.*, 626.

³⁴ *ibid.*, 626.

³⁵ *ibid.*, 626.

³⁶ *ibid.*, 625.

³⁷ Duodecim 2016.

³⁸ Melander, Salmonson & Abadie *et al.* 2008, 624, table 1.

database widely used by Finnish clinicians, the range for mild depression is 8-15 points, and severe depression is indicated by a score of over 15 points; this classification omits moderate depression. Regardless of the criteria used, it is clear that the Melander *et al.* 2008 meta-analysis does not include mildly depressed patients. Under the criteria Kirsch uses, the least depressed patient is moderately depressed, under the second criteria he or she is severely depressed. The guidelines' assertion that Melander *et al.* 2008 shows antidepressants to be effective in mild depression is incorrect.

The fourth meta-analysis cited in the guidelines looks at trials involving imipramine and paroxetine.³⁹ Imipramine is a tricyclic antidepressant that is not in use in Finland, and paroxetine is an SSRI. The aim of the study is similar to Kirsch *et al.* 2008 discussed above: to assess the effect that the initial severity of depression has on the efficacy of drug and placebo. The authors refer to Kirsch *et al.* 2008 and note that this meta-analysis mostly included very severely depressed patients.⁴⁰ Fournier *et al.* 2010 includes patients with lower scores on the Hamilton scale, including one trial of minor depressive disorder in the meta-analysis. One other important feature of the study design is that it excludes trials that include a placebo washout period. A placebo washout is used to exclude strong placebo responders before the trial starts. The authors argue that this potentially weakens response to placebo in the trial proper.⁴¹

Six trials were included.⁴² Overall, the Cohen's *d* effect size for patients with mild to moderate depression was 0.11, severe depression 0.17 and very severe depression 0.47. For these same groups, number-needed-to-treat values were 16, 11, and 4.⁴³ Using the NICE criteria according to which the minimum threshold of clinical relevance is an improvement of three points on the Hamilton scale, the authors point out that the drug/placebo difference reaches clinical relevance when the initial severity of depression is 25 or higher on the Hamilton scale. In other words, the difference only becomes clinically relevant in the very severe range of depression. The other NICE criterion for clinical relevance is an effect size of over 0.50. Using this criterion, the authors point out that the drug/placebo difference reaches clinical relevance only when the initial severity of depression is 27 or greater on the Hamilton scale.⁴⁴ The researchers reran the analysis whilst excluding the one study of minor depressive disorder; this did not change the threshold of clinical relevance.⁴⁵

³⁹ Fournier JC, DeRubeis RJ, Hollon SD *et al.* Antidepressant drug effects and depression severity: A patient-level meta-analysis. *JAMA* 2010; 303: 47-53.

⁴⁰ *ibid.*, 48.

⁴¹ *ibid.*, 48-49.

⁴² *ibid.*, 50.

⁴³ *ibid.*, 51.

⁴⁴ *ibid.*, 51.

⁴⁵ *ibid.*, 51.

Fournier *et al.* 2010 concludes that antidepressants provide clinically relevant benefits only for severely depressed patients.⁴⁶ The guidelines argue that this conclusion is not justified. They point out that a relatively small number of patients were included, that only six studies were included and that only two antidepressants were studied.⁴⁷ This is the strongest critical argument made, but it ignores the reason for the limited number of trials. Fournier *et al.* 2010 points out that the majority of antidepressant trials are conducted with patients who are severely depressed, even though they do not represent the majority of patients in clinical practice.⁴⁸

The guidelines also state that Fournier *et al.* 2010 includes patients whose depression is less severe than the depression covered by the guidelines.⁴⁹ It is difficult to understand this claim. The guidelines cover all severities of major depressive disorder, and explicitly state that antidepressants are superior to placebo even in mild depression. The claim that the patients in Fournier *et al.* 2010 are not depressed enough is also curious because it specifies that these less severely depressed patients are in the 8-18 range on the Hamilton scale. However, the minimum cut off point for patients in the one trial on minor depressive disorder included by Fournier *et al.* 2010 is ten points on the Hamilton scale.⁵⁰ All the other trials have a cut off point of 14 points or higher.⁵¹ Furthermore, as noted, Fournier and colleagues reran the statistical analysis with the one trial of minor depressive disorder excluded, which did not change the threshold of clinical relevance.

The next meta-analysis, Undurraga and Baldessarini 2012, includes 142 randomized placebo-controlled trials of 19 antidepressants approved for sale in the United States published between 1980-2011, with a total of 27 127 patients.⁵² 36.6% of the trials involved SSRIs, 26.8% tricyclic antidepressants, 9.9% atypical antidepressants, and 3.5% MAO inhibitors.⁵³ The outcome measured is response, which is defined as a specific reduction in the depression score on the different scales used in the individual studies. The vast majority of the trials use some variation of the Hamilton scale. On that scale, response is defined as a reduction of at least 50%.⁵⁴

The meta-analysis showed the superiority of drugs over placebo: the overall drug/placebo pooled rate ratio for response was 1.42.⁵⁵ The individual trials show consistent superiority of drug over placebo, as there is only one study in which there were more responders in the placebo group.⁵⁶

⁴⁶ *ibid.*, 52.

⁴⁷ Duodecim 2016.

⁴⁸ Fournier, DeRubeis, Hollon *et al.* 2010, 53-54.

⁴⁹ Duodecim 2016.

⁵⁰ The trial is Barrett *et al.* See Table 1 in the supplementary material in Fournier, DeRubeis & Hollon *et al.* 2010.

⁵¹ The cut off points for all the trials are given in Table 1 in *ibid.*

⁵² Undurraga J & Baldessarini RJ. Randomized, placebo-controlled trials of antidepressants for acute major depression: Thirty-year meta-analytic review. *Neuropsychopharmacology* 2012; 37: 851-864.

⁵³ *ibid.*, 853.

⁵⁴ *ibid.*, 852.

⁵⁵ *ibid.*, 853.

This meta-analysis provides evidence for the claim in the guidelines that antidepressants, including SSRIs, are superior to placebo in the treatment of depression. However, a fundamental problem with this meta-analysis is that it makes no attempt to account for publication bias: only published trials are included,⁵⁷ and over 97 per cent of them were funded by drug manufacturers.⁵⁸ Furthermore, the study does not differentiate outcomes based on the initial severity of depression. The authors note that their results are in line with other studies showing that drug/placebo differences are generally moderate.⁵⁹

The next meta-analysis cited by the guidelines is Gibbons *et al.* 2012.⁶⁰ It refers to both Turner *et al.* 2008 and Kirsch *et al.* 2008 discussed above in noting that recent studies have suggested that the superiority of antidepressants over placebo has been overstated.⁶¹ The authors argue that the sort of “vote counting” methodology used in Turner *et al.* 2008 in finding that the number of studies deemed positive by the FDA is markedly reduced when unpublished trials are included is inadequate for assessing clinical effectiveness.⁶² As an aside, it should be noted that Turner *et al.* 2008 makes no strong claims about the clinical effectiveness of antidepressants. That study is focused on the effect of publication bias, not on the question of whether drug/placebo differences are clinically relevant. The authors do note that including unpublished studies shows antidepressants to be less effective than a meta-analysis including only published trials, but that all the drugs were still superior to placebo, and they make no conclusions about clinical relevance.⁶³

Gibbons *et al.* 2012 also questions the methodology used by Kirsch and colleagues, arguing that patient level data is required to draw the conclusion made in that study. In other words, the authors maintain that using average study-level initial severity of depression instead of patient-level initial severity of depression is inadequate.⁶⁴

The study includes patient-level data of placebo-controlled trials of fluoxetine conducted by Eli Lilly and venlafaxine trials conducted by Wyeth. For fluoxetine, 12 trials comprising 2 635 patients were included.⁶⁵ Venlafaxine IR (instant release) and ER (extended release) were studied separately. For venlafaxine IR, 11 trials with 2 421 patients, and for venlafaxine ER, 10 trials with 2

⁵⁶ See Table 1 in *ibid.*, 854-856.

⁵⁷ *ibid.*, 852.

⁵⁸ *ibid.*, 853.

⁵⁹ *ibid.*, 858.

⁶⁰ Gibbons RD, Hur K, Brown CH *et al.* 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.

⁶¹ *ibid.*, 573.

⁶² *ibid.*, 573.

⁶³ Turner, Matthews & Linardatos *et al.* 2008, 259.

⁶⁴ Gibbons, Hur & Brown *et al.* 2012, 573.

⁶⁵ *ibid.*, 574.

461 participants were included.⁶⁶ The outcomes measured were change in the patient's Hamilton score, response, and remission. Response was defined as a 50% or greater reduction in Hamilton score, and remission as a score of less than 8. All outcomes were measured at six weeks.⁶⁷

For adult patients taking fluoxetine, the average reduction in their Hamilton score was 10.12 points and for those taking placebo, it was 7.52 points, yielding a drug/placebo difference of 2.6 points. The overall response rate for fluoxetine was 55.1% and for placebo 33.7%. The remission rate in the fluoxetine group was 45.8% and in the placebo group it was 30.2%.⁶⁸ For adult patients taking venlafaxine ER, average reduction in Hamilton score was 12.39 versus 10.21 for placebo, for a difference of 2.18 points. For IR, the figures are 14.32, 10.71, and 3.61, respectively.⁶⁹

No statistically significant effect of initial severity of depression on either absolute improvement or response was found for fluoxetine or either type of venlafaxine versus placebo.⁷⁰

The authors argue that even though average reductions in the drug and placebo groups do not differ markedly from each other, this is "an enormous difference" from a public health perspective.⁷¹ They note that the difference in response rates between the fluoxetine and placebo groups means that one additional patient responds for every five patients treated with fluoxetine.⁷² Likewise, the guidelines conclude that this study showed that active drugs were superior to placebo in mild, moderate and severe depression.

It should be noted that even though Gibbons and colleagues criticize Kirsch *et al.* 2008, and reach different conclusions, they reach those conclusions using different criteria. The NICE criteria for clinical relevance are not used. According to those criteria, the overall fluoxetine/placebo difference, 2.60 points on the Hamilton scale, is clinically irrelevant. Only venlafaxine IR reaches clinical relevance under these criteria, with an overall difference of 3.61 points.

It is difficult to assess the validity of this study, as no indication is given of exactly how depressed the patients included in the trials were. As noted in Fournier *et al.* 2010, most trials of antidepressants are conducted with patients who are quite significantly depressed. The range of initial depression scores is relevant because the study includes separate analyses of patients with lower and higher severities of depression. Furthermore, Gibbon *et al.* 2012 note that only 52.2% of patients had a Hamilton depression score in week six of their trials.⁷³ As far as I can tell, the

⁶⁶ *ibid.*, 574.

⁶⁷ *ibid.*, 575. The study also includes four fluoxetine trials with geriatric patients and another four trials of fluoxetine with adolescent patients. As my focus is on adult depression, I will ignore these trials. This is possible, as the meta-analyses were conducted separately for the different patient groups.

⁶⁸ *ibid.*, 576.

⁶⁹ *ibid.*, 576.

⁷⁰ *ibid.*, 576-577.

⁷¹ *ibid.*, 578.

⁷² *ibid.*, 578.

⁷³ *ibid.*, 575.

assessment of patients at six weeks is based on a Bayes estimate of available data on the patient, even if they had dropped out prior to six weeks. The authors do not clarify this further.⁷⁴

Taylor *et al.* 2014 is a meta-analysis of all published and unpublished short-term trials of the new antidepressant agomelatine. Agomelatine is a serotonin receptor (specifically 5HT_{2c}) antagonist and an agonist of melatonergic MT₁ and MT₂ receptors. Thus, its mechanism of action is quite distinct from other antidepressants.

The study included 20 trials with 7460 patients.⁷⁵ Twelve trials compared agomelatine to placebo and thirteen studies compared agomelatine to established antidepressants (escitalopram, fluoxetine, sertraline, paroxetine, and venlafaxine).⁷⁶ The risk of publication bias should be negligible, since the authors received data on all trials used for regulatory approval from the European Medicines Agency (EMA), and on all trials conducted by the manufacturer, Servier.⁷⁷ Trial participants were diagnosed with major depressive disorder according to DSM-IV criteria. All but two studies used the Hamilton scale for assessing severity of depression. At baseline, patients had a mean Hamilton score of 27.⁷⁸ The main outcome was change in mean depression score at the end of the trial. Response and remission, as defined by the original trials, were studied as secondary outcomes. The majority of the studies defined response as a 50% reduction in depression score and remission as a Hamilton score of 7 or less or a MADRS score of 12 or less.

Agomelatine was superior to placebo for the primary outcome, with a standardized mean difference of 0.24. Agomelatine was also superior to placebo for response: the relative risk was 1.25. No statistically significant difference between agomelatine and placebo was found for remission.⁷⁹ Patients on agomelatine were no more likely than patients on placebo to discontinue treatment due to adverse events or due to any reason.⁸⁰ There were no statistically significant differences between agomelatine and other antidepressants (considered as a group) for the primary or secondary outcomes.⁸¹ There was evidence of publication bias in the literature: agomelatine was not superior to placebo for response or remission in the unpublished studies. Published studies tended to favour agomelatine versus other antidepressants, whereas unpublished studies tended to favour established antidepressants over agomelatine.⁸²

⁷⁴ I would like to thank Samuli Kangaslampi for clarifying this for me.

⁷⁵ Taylor D, Sparshatt A, Varma S *et al.* Antidepressant efficacy of agomelatine: meta-analysis of published and unpublished studies. *BMJ* 2014; 348:g1888, 1. I was only able to access an electronic copy of this paper, and was unable to find the page numbers for the print version, so the page numbers given here refer only to pages within this single article.

⁷⁶ *ibid.*, 3.

⁷⁷ *ibid.*, 3.

⁷⁸ *ibid.*, 3.

⁷⁹ *ibid.*, 3.

⁸⁰ *ibid.*, 4.

⁸¹ *ibid.*, 3-4.

⁸² *ibid.*, 5.

The authors conclude that agomelatine is an effective antidepressant, but that its effect size when compared to placebo is small. Agomelatine did not show superiority over placebo for remission; the authors note that this could be because some of the trials did not report this outcome.⁸³ The authors make an indirect reference to the antidepressant debate by noting that agomelatine's effect size might cast doubt on its clinical relevance and lend credence to the critics' argument that the use of antidepressants should be limited.⁸⁴ They suggest that such a conclusion should be considered with three caveats. They note that placebo treatment has a significant effect size in depression,⁸⁵ presumably suggesting that this leads to smaller effect sizes for active drugs. They point out that effect sizes for active drugs are greater in relapse prevention studies. Furthermore, they argue that the small effect sizes for antidepressants are similar to effect sizes for drugs used in other fields of medicine, such as ACE inhibitors for prevention of cardiovascular events and thrombolysis for survival in acute stroke.⁸⁶ This discussion is germane to the question of clinical relevance: how great should the difference favoring active drugs be to justify their use in clinical practice? I touch on these questions in the Conclusion. The guidelines do not comment on the magnitude of agomelatine's effect size or its clinical relevance; they merely point out agomelatine's statistical superiority over placebo.⁸⁷

Pae *et al.* 2015 is a meta-analysis of short-term trials of vortioxetine for major depressive disorder. Vortioxetine is the newest SSRI; it received FDA approval in 2013. Twelve randomized controlled trials comparing vortioxetine to placebo or to another antidepressant (venlafaxine, duloxetine, or agomelatine) were included. Mean Hamilton or MADRS scores at baseline ranged from 28.5 to 34.1.⁸⁸

The primary outcome was change in depression score on the 24-item Hamilton scale or the MADRS scale. Secondary outcomes were response and remission. Response was defined as a 50% or greater reduction in initial HAMD/MADRS score and remission as 7 points or less on the HAMD or 10 points or less on the MADRS.⁸⁹

Intention-to-treat data was available for 4947 patients. Vortioxetine was superior to placebo, with a standardized mean difference of -0.217. Vortioxetine was superior to placebo for response and remission: odds ratios were 1.652 and 1.399, respectively. There was significant heterogeneity

⁸³ *ibid.*, 4.

⁸⁴ *ibid.*, 5.

⁸⁵ *ibid.*, 5.

⁸⁶ *ibid.*, 5.

⁸⁷ Duodecim 2016.

⁸⁸ Pae C-U, Wang S-M, Han C et al. Vortioxetine: a meta-analysis of 12 short-term, randomized, placebo-controlled clinical trials for the treatment of major depressive disorder. *J. Psychiatry Neurosci.* 2015; 40: 174-186, 177.

⁸⁹ *ibid.*, 175.

among trials, but no single trial was found to have a strong effect on the overall results.⁹⁰ For discontinuation for any reason, there was no statistically significant difference between vortioxetine and placebo, but the discontinuation rate due to adverse events was significantly higher in the vortioxetine than the placebo group. The discontinuation rate due to lack of efficacy was significantly higher in the placebo group.⁹¹

Intention-to-treat data was available for 2843 participants for vortioxetine versus other antidepressants (grouped together).⁹² There was no statistically significant difference for vortioxetine versus the other drugs for the primary outcome⁹³ or for response or remission.⁹⁴

The authors note that even though the statistical superiority of vortioxetine over placebo was demonstrated, the standardized mean difference of -0.22 is of questionable clinical relevance. They point out that the SMD translates to a two-point difference in MADRS score.⁹⁵ The authors refer to the debate on the minimal clinically important difference (MCID) in drug-placebo comparisons.⁹⁶ Using Duru and Fantino's criteria for the MCID – a difference of 2 points on the MADRS or a 10% difference in response rate – Pae and colleagues argue that “vortioxetine may meet the marginal standard criterion for an antidepressant to be considered effective for treating MDD.”⁹⁷ This is directly pertinent to the issue of clinical relevance, which I touch on in the Conclusion. The guidelines offer no comment on the magnitude of vortioxetine's effect size compared to placebo or on its clinical relevance.

Jakubovski *et al.* 2016 is a meta-analysis studying the relationship between SSRI dose and treatment response.⁹⁸ It is not primarily a study of SSRI efficacy versus placebo, but it is discussed here because the guidelines include it amongst the meta-analyses they must support of the overall argument that antidepressants are effective treatments for depression and because it is highly relevant to the critics' “active placebo” argument discussed later.

Forty studies with forty-nine active treatment arms with 10 039 patients were included. The drugs studied were fluoxetine, fluvoxamine, paroxetine, sertraline, citalopram and escitalopram. SSRI doses were converted to imipramine equivalents.⁹⁹ Jakubovski *et al.* found a statistically significant association between SSRI dose and efficacy in reducing the severity of depression, as

⁹⁰ *ibid.*, 178-179.

⁹¹ *ibid.*, 181.

⁹² *ibid.*, 179.

⁹³ *ibid.*, 180.

⁹⁴ *ibid.*, 181.

⁹⁵ *ibid.*, 183.

⁹⁶ *ibid.*, 182-183.

⁹⁷ *ibid.*, 183.

⁹⁸ Jakubovsky E, Varigonda AL, Freemantle N *et al.* Systematic Review and Meta-analysis: Dose-Response Relationship of Selective Serotonin Reuptake Inhibitors in Major Depressive Disorder. *Am. J. Psychiatry* 2016; 173: 174-183.

⁹⁹ *ibid.*, 175.

measured on the Hamilton or Montgomery-Åsberg scales.¹⁰⁰ The result was statistically significant both when SSRI dose was examined by dosing category and when it was examined as a continuous variable.¹⁰¹ The result remained significant when only trials using intention-to-treat data were included.¹⁰² There was also a statistically significant association of SSRI dose with treatment response. This finding remained significant when only intention-to-treat data was included.¹⁰³

The meta-analysis also investigated the relationship between dose and adverse events. There was a statistically significant association between increased SSRI dose and dropout because of unwanted effects.¹⁰⁴ However, there was a small statistically significant association between higher SSRI dose and a lower likelihood of dropout for any reason.¹⁰⁵

The authors conclude that the “meta-analysis demonstrated a significant association between higher SSRI doses and greater measured efficacy of SSRIs [...]”¹⁰⁶ and that this benefit comes at the cost of increased adverse events.¹⁰⁷ The guidelines echo this conclusion.¹⁰⁸ This is one way of interpreting the results. However, the results might also be compatible with the Kirsch-Moncrieff argument that antidepressant superiority over placebo is due to unblinding because of side effects.

To summarize: the guidelines assert that antidepressants are clearly superior to placebo in acute-phase treatment of depression, and that no methodologically valid meta-analysis questions their effectiveness. It is difficult to agree with this claim on the basis of the above review. No specific criteria are presented in the guidelines for assessing the methodological validity of meta-analyses. The guidelines do note that they try to focus on studies in which the effects of publication bias is minimal, and which are otherwise as independent as possible of the pharmaceutical industry.¹⁰⁹

However, Kirsch *et al.* 2008 is described as a good quality study, well suited to application in the Finnish context. This study is cited in support of the claim that antidepressants are effective in treating all severities of depression. As we have seen, the study itself reaches quite a different conclusion: that antidepressants are effective only for the most severely depressed patients, and that even for those patients, their apparent superiority over placebo derives mostly from the declining efficacy of placebo. It seems reasonable to conclude that at least this study is a methodologically

¹⁰⁰ *ibid.*, 177.

¹⁰¹ *ibid.*, 177.

¹⁰² *ibid.*, 177.

¹⁰³ *ibid.*, 177.

¹⁰⁴ *ibid.*, 178.

¹⁰⁵ *ibid.*, 177.

¹⁰⁶ *ibid.*, 178.

¹⁰⁷ *ibid.*, 178-179.

¹⁰⁸ Duodecim 2016.

¹⁰⁹ Duodecim 2016.

valid meta-analysis that puts into question the effectiveness of antidepressants, and thus contradicts the claim that no such meta-analysis exists.

It should also be noted that studies such as Mellander *et al.* 2008 and Gibbons *et al.* 2012 explicitly refer to the conclusions drawn by Kirsch and colleagues, and attempt to show that a different analysis of the data yields a different conclusion. If Kirsch *et al.* 2008 reached the conclusion that antidepressants are generally effective, as the guidelines imply, it would make no sense for these other studies to position themselves in relation to it, as there would be no need to defend the mainstream view of antidepressants against the study's conclusions.

1.2 Long-term treatment

The guidelines maintain that antidepressants are not only superior to placebo for acute treatment of depression, but that they are effective for the prevention of relapse. The guidelines argue that maintenance treatment is generally well tolerated, because patients have already demonstrated response to antidepressants. They note that it is common for patients to discontinue treatment, but claim that this is rarely due to unwanted drug effects.¹¹⁰ The guidelines base these arguments on four meta-analyses.

The first, Geddes *et al.* 2003 studied 31 discontinuation trials.¹¹¹ All the 4410 patients included had responded to active drug treatment in the acute phase of depression. Patients were randomly allocated to continue or discontinue active drug treatment and were followed up for at least one month.¹¹² Two trials were on noradrenaline reuptake inhibitors, four on MAO inhibitors, 15 on TCAs, 10 on SSRIs, and one on some other type of antidepressant. The trials were published between 1973-2001. There were various criteria used for relapse, ranging from specific scores on the Hamilton or the Montgomery-Åsberg scale to hospital admission or the indication to change drug treatment.¹¹³

Treatment with the active drug consistently reduced the risk of relapse. The average relapse rate in the placebo groups was 41%; in the drug groups, it was 19%. The pooled odds ratio for relapse was 0.30. In the drug groups, more patients (18%) withdrew from the trials than in the placebo groups (15%).¹¹⁴ Even though the studies were quite heterogeneous in terms of the drugs studied, duration of the maintenance treatment, and the definition of relapse, the results consistently favored the effectiveness of active drug treatment in the prevention of relapse.¹¹⁵

¹¹⁰ Duodecim 2016.

¹¹¹ Geddes JR, Carney SM, Davies C *et al.* Relapse prevention with antidepressant drug treatment in depressive disorders: a systematic review. *The Lancet* 2003; 361: 653-661.

¹¹² *ibid.*, 653-654.

¹¹³ *ibid.*, 656-657.

¹¹⁴ *ibid.*, 659.

¹¹⁵ *ibid.*, 660.

It should be noted that almost half (15) of the 31 studies were of TCAs, which are not the first line of drug treatment for depression today. A more fundamental problem with the trials was that all patients had been on active drug treatment prior to the trials. The authors note that this raises the possibility that some of the relapses in the placebo groups were due to withdrawal of the active drug rather than to an actual recurrence of depression. The authors point out that, if this is the case, the effectiveness of maintenance treatment with antidepressants might have been inflated.¹¹⁶ The guidelines ignore this potential weakness.

The second study, Hansen *et al.* 2008, is a review and meta-analysis of comparative and placebo controlled trials of newer antidepressants for the prevention of relapse and recurrence of depression.¹¹⁷ Four trials comparing antidepressants head-to-head and 23 trials comparing antidepressant with placebo were included.¹¹⁸ The antidepressants studied were all “second-generation”, i.e. representative of current clinical practice. Included were bupropion, citalopram, duloxetine, escitalopram, fluoxetine, fluvoxamine, mirtazapine, nefazodone, paroxetine, sertraline, trazodone, and venlafaxine. The study investigated two questions: are there differences in the effectiveness of antidepressants in maintaining remission for adults with major depressive disorder? What is the overall effect size for drug versus placebo, and does it persist over time?¹¹⁹

All patients had responded to acute treatment with an antidepressant prior to randomization to continue active or placebo treatment. The definition of this response varied. Some trials used a specific cut-off point on the Hamilton scale, the Beck Depression Inventory (BDI), the MADRS scale, or the CGI scale (such as <10 points on the Hamilton scale); some a reduction of initial depression score by a specific percentage (such as a 50% reduction in Hamilton score); and some studies combined these ratings with DSM diagnostic criteria.¹²⁰ The authors note that in most trials, loss of response or remission was defined by an increase in Hamilton or MADRS score above a specific cut-off point.¹²¹ However, no exact definition for relapse or remission is given for any of the trials, even though the exact criteria for inclusion are given for all of them.

The authors did not consider individual drugs versus placebo, but combined the different drug treatment groups into one, which was compared to the placebo group. The placebo-controlled trials showed superiority of active drug over placebo in relapse prevention. In the twelve trials with a follow-up period of less than one year, 22% of patients in the active drug group relapsed, whereas

¹¹⁶ *ibid.*, 660.

¹¹⁷ Hansen RA, Gaynes BN, Thieda P *et al.* Prevention of major depressive disorder relapse and recurrence with second-generation antidepressants: A systematic review and meta-analysis. *Psychiatr Serv* 2008; 59: 1121-1130.

¹¹⁸ *ibid.*, 1121.

¹¹⁹ *ibid.*, 1122.

¹²⁰ See Tables 1 and 2 in *ibid.*.

¹²¹ *ibid.*, 1123-1124.

42% of patients in the placebo group did so. In the 11 trials that were longer than 12 months, the relapse rate in the active treatment group was 26% and in the placebo group it was 48%. The meta-analysis of trials shorter than 12 months yielded a pooled relative risk for relapse of 0.54, translating into a number-needed-to-treat of 5 to prevent one additional relapse. In the longer trials, the overall relative risk of relapse was 0.56, and the NNT was 5.¹²² There were four trials comparing different active drugs head-to-head; were no significant differences between them.¹²³

Twenty-two trials provided sufficient data for analyzing adverse events; loss to follow-up in these trials was 7% in the active treatment group and 5% in the placebo group. Based on data from 18 trials, there was no statistically significant difference in loss to follow-up due to adverse events between active and placebo treatment. Overall, 50% of patients assigned to active treatment, and 68% of patients assigned to placebo were lost to follow-up.¹²⁴

The study shows clear superiority of active drugs over placebo, and the guidelines note the relative risk of 0.56.¹²⁵ However, the guidelines describe this as the relative risk in intention-to-treat analysis.¹²⁶ Hansen *et al.* 2008 mentions that intention-to-treat data was included for trials for which it was available.¹²⁷ However, their paper does not indicate how many trials included intention-to-treat data. The authors did assess the quality of each trial according to several criteria, one of which was the inclusion of intention-to-treat analysis.¹²⁸ Only two of the placebo-controlled trials were rated as good quality, the rest were categorized as fair quality.¹²⁹ It is impossible to tell which of these trials included intention-to-treat analysis on the basis of the article.

Intention-to-treat is crucial here, since so many patients in the trials were lost to follow-up. It would therefore be essential to know how the trials categorized patients that dropped out. Given that 50% of patients in the drug group and 68% in the placebo group were lost to follow-up, it is hard to draw any conclusions from the study's results. The guidelines ignore this.

The third study, Williams *et al.* 2009, is a meta-analysis of eleven antidepressant maintenance trials published between 1998-2007.¹³⁰ The study compared relapse rates among patients receiving active drug and placebo in long-term maintenance treatment. In nine of the trials, maintenance treatment was preceded first by 0-12 (mean = 7.5) weeks of acute treatment and second by 0-26 (mean = 17) weeks of continuation treatment with a specific antidepressant. Only

¹²² *ibid.*, 1125.

¹²³ *ibid.*, 1125.

¹²⁴ *ibid.*, 1126.

¹²⁵ Duodecim 2016.

¹²⁶ *ibid.*.

¹²⁷ Hansen, Gaynes, Thieda *et al.* 2008, 1123.

¹²⁸ *ibid.*, 1123.

¹²⁹ See Table 2 in *ibid.* for quality ratings of placebo-controlled trials.

¹³⁰ Williams N, Simpson AN, Simpson K *et al.* Relapse rates with long-term antidepressant drug therapy: a meta-analysis. *Hum. Psychopharmacol. Clin. Exp.* 2009; 24: 401-408.

patients who had been categorized as remitters or responders in the acute phase moved onto the continuation phase, and again, only those who were remitters/responders in continuation treatment were randomized into double-blind placebo controlled long-term (24-156 weeks, mean = 60) maintenance treatment.¹³¹ The baseline Hamilton score for admission to the acute phase ranged from 14 to 18.¹³² For both the acute and continuation phases, remission was defined as a Hamilton score of eight or lower and response as a score of twelve or less plus a 50% reduction from the baseline score.¹³³

In one trial, 10 sessions of electroconvulsive therapy (ECT) were used for acute treatment. Another trial included patients who did not meet DSM depression criteria at baseline and who then received two months of placebo treatment before being randomized to maintenance treatment with placebo or sertraline.¹³⁴ In almost all the trials, the active drug was an SSRI or SNRI in widespread clinical use today; the drugs were fluoxetine, citalopram, sertraline, venlafaxine, escitalopram, and duloxetine. One trial used a TCA, nortriptyline, and another used a nortriptyline-lithium combination.¹³⁵ The study that used nortriptyline-lithium was the only one that included psychotic patients.¹³⁶

The risk of relapse was 23% for drug and 51% for placebo at one year, 34% for drug and 62% for placebo at two years, and 45% for drug and 7% for placebo at three years.¹³⁷ In all the individual studies, relapse rates were higher for placebo than for active drug.¹³⁸ The number of previous episodes or the baseline Hamilton score in the acute phase did not affect relapse rates.¹³⁹

The final meta-analysis, Glue *et al.* 2010, studied 54 double-blind placebo-controlled relapse prevention trials.¹⁴⁰ SNRIs were studied in 21 of the trials and SSRIs in another 21 trials. Two trials studied noradrenaline reuptake inhibitors, five MAOIs, and eight trials investigated other types of antidepressants (gepirone, mianserine, bupropion).¹⁴¹ In a majority of studies, patients were diagnosed with major depressive disorder. In nine studies, other depressive diagnoses such as bipolar disorder and dysthymia were included. 31 studies used various DSM criteria; the rest used a

¹³¹ *ibid.*, 403.

¹³² *ibid.*, 402.

¹³³ *ibid.*, 403.

¹³⁴ *ibid.*, 404.

¹³⁵ *ibid.*, 405.

¹³⁶ *ibid.*, 404.

¹³⁷ *ibid.*, 404. Samuli Kangaslampi points out that there has to be an error in the figure for the placebo group at three years in the original article, since the *cumulative* risk of relapse cannot fall.

¹³⁸ *ibid.*, 405.

¹³⁹ *ibid.*, 405.

¹⁴⁰ Glue P, Donovan MR, Kolluri S *et al.* Meta-analysis of relapse prevention antidepressant trials in depressive disorders. Australian and New Zealand Journal of Psychiatry 2010; 44: 697-705.

¹⁴¹ *ibid.*, 699-700.

variety of other diagnostic criteria, such as Research Diagnostic criteria. Various criteria for defining response and relapse were used.¹⁴²

Odds ratios for relapse were determined based on drug class, patient age, diagnostic system and depression subtype, treatment duration, and year of publication. All classes of active drug were found to reduce the risk of relapse. The result was statistically significant. There was statistically significant heterogeneity between drug classes; the exclusion of MAOI trials eliminated the statistical significance of this heterogeneity.¹⁴³

The authors note that for patients who experienced response to active medication in the acute phase of treatment, the odds of relapse were reduced by two thirds by continued antidepressant use.¹⁴⁴ The authors point out the same alternative explanation for such results that Moncrieff and Kirsch do, *i.e.* that higher relapse rates in the placebo group might be due to drug withdrawal rather than a true relapse of depression. They argue that were this the case, active drugs would fare better in trials where patients randomized to placebo were abruptly moved from active drug to placebo than in those trials where the active drug patients were taking in the acute phase was gradually down-titrated before they started receiving only placebo. Data on drug titration was available in 28 studies, and in these, patients were abruptly switched from drug to placebo in two trials, whereas in the other trials active drug was down-titrated for a period of 1-16 weeks. The authors found no “obvious increase in patients relapsing early in the post-randomization phase” when comparing abrupt and gradual withdrawal.¹⁴⁵

The authors also point out the possibility of publication bias, but maintain that “diligent searches were made of pharmaceutical company and non-commercial clinical trial databases” in order to discover all relevant trials.¹⁴⁶ The authors do not indicate what kind of access they had to pharmaceutical company databases, and they did not obtain full data from the FDA, so some publication bias may remain.

¹⁴² *ibid.*, 699.

¹⁴³ *ibid.*, 700.

¹⁴⁴ *ibid.*, 700.

¹⁴⁵ *ibid.*, 702.

¹⁴⁶ *ibid.*, 702.

2: Placebo: the Kirsch-Moncrieff argument

In this section I summarize the argument presented by Kirsch and Moncrieff that antidepressants are not significantly superior to placebo. Their arguments move from the observation of the moderate effectiveness of antidepressants to attempting to explain this relative lack of efficacy; this further elucidation of the argument is presented in section 3. A critical examination of the validity of these arguments is given in section 4.

2.1 Moncrieff on antidepressants versus placebo

Moncrieff presents the bulk of her version of the placebo argument in chapter nine of *The Myth of the Chemical Cure*, in which she asks whether such a thing as an antidepressant exists. She notes that textbooks state the superiority of antidepressants over placebo to be around 20-40%.¹⁴⁷ She points out that most antidepressant trials show slight superiority for drug over placebo, but that some trials show placebo to be equal or even superior to drug.¹⁴⁸ As is typical of Moncrieff, she relies heavily on early antidepressant trials.

There are four elements of the drug/placebo difference that Moncrieff discusses: drug versus placebo for short-term treatment, the effect of the severity of depression on drug/placebo differences, drug versus placebo for long-term treatment, and drug versus placebo for the long-term outcome of depression. I discuss each of these lines of argument below. In the chapter of *The Myth of the Chemical Cure* in which she presents the bulk of her argument on these issues, she also discusses the question of the efficacy of other active drugs not considered antidepressants for the treatment of depression, but I will consider this discussion in the section on the “active placebo” argument.

2.1.1 Short-term treatment

Moncrieff refers to a review of trials of imipramine versus placebo published in 1975, in which a majority of trials was found to show no superiority of this older tricyclic antidepressant over placebo.¹⁴⁹ She points to both older and more recent reviews of trials which suggest the effectiveness of antidepressants, tricyclics in particular, is not impressive. She discusses the Medical Research Council trial of imipramine, phenelzine, electroconvulsive therapy and placebo conducted in the UK. This early study, published in 1965, showed little difference between

¹⁴⁷ Moncrieff 2008, 139.

¹⁴⁸ *ibid.*, 139.

¹⁴⁹ *ibid.*, 139.

phenelzine, an MAO inhibitor, when compared to placebo. Likewise, the tricyclic antidepressant imipramine did not fare much better than placebo.¹⁵⁰

Another study she refers to, published by the National Institutes of Mental Health in 1970, compared imipramine, chlorpromazine and placebo. Chlorpromazine is an early dopamine antagonist, and is used as an antipsychotic medication, but is not indicated for depression. Moncrieff notes several weaknesses in how the data in this study is presented, and that the results were “essentially negative”.¹⁵¹

Moncrieff maintains that newer meta-analyses show lower superiority for drug over placebo.¹⁵² Her choice of studies here is curious. Storosum *et al.* 2001, the newest study she refers to, evaluates the effectiveness of TCAs versus placebo, even though it was published in the SSRI era.¹⁵³ Bech *et al.* 2000, the second one she refers to here, is a meta-analysis of a single drug, fluoxetine.¹⁵⁴ The results favor fluoxetine. The third study Moncrieff refers to here, Khan, Warner & Brown 2000, is a meta-analysis examining suicide risk and symptom reduction in the active treatment and placebo arms of antidepressant trials.¹⁵⁵ It concludes that receiving placebo in an antidepressant clinical trial poses no additional suicide risk for the depressed patient. The authors also note that patients receiving placebo experience significant symptom reduction, but not to the same degree as those receiving the active drug.¹⁵⁶

Moncrieff then argues that categorical figures are misleading and that Kirsch’s 2002 paper demonstrates this.¹⁵⁷ Kirsch *et al.* 2002 used the Freedom of Information Act to request trial data submitted to the FDA for six SSRIs: fluoxetine, paroxetine, sertraline, venlafaxine, nefazodone, and

¹⁵⁰ *ibid.*, 140-141.

¹⁵¹ *ibid.*, 141.

¹⁵² *ibid.*, 139.

¹⁵³ Storosum JG, Elferink AJ, van Zwieten BJ *et al.* Short-term efficacy of tricyclic antidepressants revisited: a meta-analytic study. *Eur Neuropsychopharmacol* 2001; 11: 173-180. I was unable to obtain a copy of this paper, but based on the abstract, it found a modest but significant effect for TCAs for response. I assume that this is the 2001 Storosum paper Moncrieff is referring to, even though it does not appear in her bibliography, which only includes one article co-authored by Storosum in 2001, and that study concerns suicide risk.

¹⁵⁴ Bech P, Cialdella P, Haugh MC *et al.* 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.

¹⁵⁵ Khan A, Warner HA, Brown WA. Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials: an analysis of the Food and Drug Administration database. *Arch. Gen. Psychiatry* 2000; 57: 311-317. The authors situate their paper in the debate over the use of placebo controls in antidepressant trials, noting that some commentators have argued that the use of placebo controls is unethical, since effective antidepressant medications are available. The basis of such an argument is that patients in the placebo arm of antidepressant trials are exposed to unnecessary morbidity or even mortality. Khan, Warner & Brown maintain that this argument is not based on evidence. They also note that although several meta-analyses show antidepressants to be superior to placebo, different researchers draw different conclusions about the magnitude of this superiority and its clinical relevance.

¹⁵⁶ *ibid.*, 316.

¹⁵⁷ Kirsch, Moore & Scoboria *et al.* 2002.

citalopram. All the trials used the Hamilton depression scale to evaluate patients. 38 trials for which last observation carried forward data was available were included in the meta-analysis.

Moncrieff points out that though the analysis shows an 18 % superiority for drug over placebo in categorical terms, this only means a 1.7 point difference on the Hamilton scale.¹⁵⁸ She goes on to assert that this difference is “too small to be meaningful” by “all assessments of clinical validity”.¹⁵⁹ Moncrieff notes that the difference is so small that it could easily be explained by non-specific drug effects or “enhanced placebo effects”.¹⁶⁰

The next study that Moncrieff draws on in support of her argument is the systematic review of antidepressant research conducted by NICE in the UK.¹⁶¹ Moncrieff notes that the review did find a statistically significant difference between drug and placebo, but concluded that this difference was so small that its clinical relevance was questionable. However, instead of reaching the conclusion that drug/placebo differences are clinically irrelevant, Moncrieff points out that NICE instead analyzed the data in terms of differences in response and remission rates, which enabled them to conclude that drug/placebo differences are clinically significant, even though the same data had led them to question that significance.¹⁶²

Unfortunately, I was unable to obtain a copy of the full 2004 NICE guideline, despite repeated online searches and ordering a copy from the British Library. The version I was able to obtain from the British Library only included the guidance, not the review of the evidence on which the guidance is based. It is not possible to properly explore the critics’ arguments relating to NICE’s materials.

2.1.2 Severity

Moncrieff moves onto considering the question of whether the efficacy of antidepressants is dependent on the severity of depression. This is an important question, since antidepressants might be ineffective for milder forms of depression, but have a more significant effect in more severe forms of the disorder. This notion is echoed in the guidelines, which state that treatment with antidepressants is especially important in severe depression.¹⁶³

¹⁵⁸ Moncrieff 2008, 140.

¹⁵⁹ *ibid.*, 140.

¹⁶⁰ *ibid.*, 140.

¹⁶¹ *ibid.*, 140.

¹⁶² *ibid.*, 140.

¹⁶³ Duodecim 2016.

NICE has gone as far as recommending against the use of antidepressants for mild depression, and recommending them only for moderate and severe depression.¹⁶⁴ The 2004 NICE guideline states: “Randomised controlled trial (RCT) evidence indicates that for many patients there is little clinically important difference between antidepressants and placebo, and the placebo response is greatest in mild depression.”¹⁶⁵ In the case of moderate or severe depression, NICE is more careful in its assessment than the Finnish guidelines, stating that “[...]There is more evidence for the effectiveness of antidepressant medication than in milder depression.”¹⁶⁶ Noting that medication is as effective as psychological interventions, and is cheaper, NICE recommends that the clinician offer antidepressants to moderately depressed patients.¹⁶⁷

Moncrieff argues against the notion that antidepressants are effective in more severe depression. Again, she refers to some older studies to support her argument. First, a 1976 review found that supposed predictors of antidepressant response were not based on much evidence.¹⁶⁸ That study concluded that the relationship of depression severity and antidepressant response is unclear.¹⁶⁹ The second study, a bit more recent, that she refers to found that the evidence on whether so-called endogenous depression - which is quite severe - responds better or worse to antidepressants than other types of depression was mixed.¹⁷⁰

She moves on to discuss meta-analyses published in the 1990s and 2000s. She notes that Angst *et al.* 1993 found antidepressants to show greater superiority over placebo for patients with more severe depression, but argues that these effects were not strong, and that they did not reach statistical significance for the most part.¹⁷¹

Angst *et al.* 1993 is a meta-analysis of 38 trials of moclobemide versus placebo or tricyclics.¹⁷² The authors present subscales of the Hamilton scale: psychic, somatic, retardation, and agitation.¹⁷³ They investigated drug versus placebo response on these subscales and the effect of baseline severity of depression on response. Response was defined as a reduction in Hamilton score of 50% or more.¹⁷⁴ A score of less than 23 was defined as low, 23-27 as medium and 28 or greater

¹⁶⁴ National Institute for Clinical Excellence. Depression: Management of depression in primary and secondary care. Clinical Guideline 23 2004; 1-63. I use this earlier NICE guideline, since it is the one the critics refer to.

¹⁶⁵ *ibid.*, 17.

¹⁶⁶ *ibid.*, 18.

¹⁶⁷ *ibid.*, 19.

¹⁶⁸ Moncrieff 2008, 142.

¹⁶⁹ *ibid.*, 142.

¹⁷⁰ *ibid.*, 142.

¹⁷¹ Angst J, Scheidegger P, Stabl M. Efficacy of moclobemide in different patient groups. Results of new subscales of the Hamilton Depression Rating Scale. Clin. Neuropharmacol. 1993; 16: S55-S62.

¹⁷² *ibid.*, 55.

¹⁷³ *ibid.*, 56.

¹⁷⁴ For brevity, I ignore the Global Assessment of Efficacy scores and the analysis of type of depression (agitated vs. retarded).

as high on the total Hamilton scale. In the placebo group, statistical significance was reached only on the retardation subscale, with 34% of patients with a low initial HAMD score and 16% with a high score achieved response.¹⁷⁵ For moclobemide, more severely depressed patients tended to respond better, with the results statistically significant on the somatic and agitation subscales.¹⁷⁶ More severely depressed patients also tended to respond better to the TCAs imipramine and clomipramine, with statistically significant results on the psychic, somatic, retardation, and agitation subscales for the former, and only on the agitation subscale for the latter.¹⁷⁷

More pronounced effects of severity on antidepressant versus placebo differences were found in Khan *et al.* 2002, but Moncrieff points out that full data was provided only for “investigational” antidepressants.¹⁷⁸ Khan *et al.* 2002 examined 45 trials conducted for the approval of the following antidepressants: fluoxetine, sertraline, paroxetine, venlafaxine, nefazodone, mirtazapine, and bupropion. The trial data was extracted from the FDA database. Trial outcomes were categorized into superior, supportive, and equivocal based on drug versus placebo response.¹⁷⁹ A superior outcome here means that the active drug was clearly statistically superior over placebo, equivocal that there was no difference, and supportive something in between.¹⁸⁰ The trials were divided into three groups based on mean initial severity of depression, assessed on the Hamilton scale. In the least severe group, the mean Hamilton score was 24 or less, in the next group, between 24 and 28, and in the most severe group, 28 or more.¹⁸¹

48% of trials were found to show superiority of the investigational antidepressant over placebo, and 64% of trials showed superiority of an established antidepressant over placebo. For the group of trials with the lowest initial mean Hamilton score, only one in ten trials showed statistical superiority over placebo for the investigational antidepressant. In the second group, such superiority was found in 49% of trials, and in the group of trials with the highest mean Hamilton score, this figure was 71.4%.¹⁸²

There was a statistically significant relationship between the mean initial Hamilton score and change in the score in the investigational antidepressant arms of these trials. No such statistically significant relationships were found for established antidepressants or placebo.¹⁸³ The authors summarize their findings by noting that “the higher the mean initial HAM-D score, the

¹⁷⁵ *ibid.*, 56.

¹⁷⁶ *ibid.*, 57.

¹⁷⁷ *ibid.*, 58.

¹⁷⁸ Khan A, Leventhal RM, Khan SR *et al.* 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.

¹⁷⁹ *ibid.*, 41.

¹⁸⁰ *ibid.*, 42.

¹⁸¹ *ibid.*, 42.

¹⁸² *ibid.*, 43.

¹⁸³ *ibid.*, 43.

greater the change with antidepressants [...].”¹⁸⁴ The mean change in Hamilton score in the investigational antidepressant arms of the trials was 9.8 in the group of trials with the lowest initial severity, 10.5 in the middle severity group, and 12.6 in the group with the highest severity.

Moncrieff then discusses an unpublished version of the 2008 Kirsch *et al.* 2008 meta-analysis discussed above. She points to the results I have already discussed: that patients on the active drug responded more to the treatment as the initial severity of depression rose, whereas the response to placebo diminished. She emphasizes that the drug/placebo difference was only four points for the most severely depressed patients – a difference she considers “of doubtful clinical relevance”.¹⁸⁵ She goes on to argue that the difference is so small that it could be explained by “active placebo effects”. She also suggests that the response pattern in the group of patients receiving active drug could be explained by regression to the mean, *i.e.* by the patients who are initially the most depressed tending to improve the most.¹⁸⁶

Moncrieff notes that two other relatively recent meta-analyses – Walsh *et al.* 2002 and NICE 2004 – showed no consistent relationship of the initial severity of depression with drug response.

Walsh *et al.* 2002 investigated how placebo response in antidepressant trials has changed over the years.¹⁸⁷ Much like Khan, Warner & Brown 2000, the authors situate their study in the context of the debate on whether it is ethical to include a placebo group in antidepressant trials, given that supposedly effective medication is available.¹⁸⁸ The study included seventy-five antidepressant trials conducted between 1981-2000. Both SSRI and TCA trials were included.¹⁸⁹ As usual, response was defined as a 50% reduction in Hamilton score.¹⁹⁰ The study’s main objective was to ascertain whether placebo and drug response correlates with year of publication, *i.e.* whether these responses have changed over the years.¹⁹¹

Walsh *et al.* 2002 found that the proportion of patients responding to placebo increased significantly with year of publication.¹⁹² The same was true for patients responding to active drug, though the latter result was not statistically significant, if certain covariates were included.¹⁹³ The authors note that “[t]here were no statistically significant associations between the proportion of

¹⁸⁴ *ibid.*, 43.

¹⁸⁵ Moncrieff 2008, 143.

¹⁸⁶ *ibid.*, 143. Samuli Kangaslampi points out that this argument is made questionable by the fact that such an effect is not seen in the placebo group.

¹⁸⁷ Walsh TB, Seidman SN, Sysko R *et al.* Placebo Response in Studies of Major Depression: Variable, Substantial, and Growing. *JAMA* 2002; 14: 1840-1847.

¹⁸⁸ *ibid.*, 1840-1841.

¹⁸⁹ *ibid.*, 1842-1843.

¹⁹⁰ *ibid.*, 1842.

¹⁹¹ *ibid.*, 1841.

¹⁹² *ibid.*, 1842.

¹⁹³ *ibid.*, 1843.

patients taking placebo who responded and pro-rated average initial HRSD score.”¹⁹⁴ In the analysis of the active drug groups, average initial Hamilton scores were only analyzed as a covariate of TCA response. The relationship of this covariate with drug response was found to be statistically insignificant.

Moncrieff then discusses a number of individual trials which show a stronger antidepressant effect in more severe depression, but which do so on the basis of *post hoc* analysis, *i.e.* by looking for statistically significant effects in the data without predetermining which outcomes will be considered. This mode of inquiry can easily yield positive results because of random effects.¹⁹⁵

Moncrieff then points out that it “has long been believed that antidepressants are relatively ineffective in severe depression accompanied by psychotic delusions.”¹⁹⁶ Arguing that it is not the presence of psychotic symptoms but rather the general ineffectiveness of antidepressants that explains the apparent lack of significant antidepressant effects in psychotic depression, she refers to a study, which found that the severity of depression rather than the presence of delusions predicted a more modest response to antidepressants.¹⁹⁷ This study, Kocsis *et al.* 1990, investigated response to imipramine and amitriptyline, both TCAs, in different severities of depression and in non-psychotic versus psychotic depression.¹⁹⁸

The study situates itself in the context of discussion on the optimal treatment in psychotic depression: on whether the modest response to TCAs often observed in psychotically depressed patients is caused by the psychosis or by the severity of their depression.¹⁹⁹ The study included 132 depressed patients, who were categorized into non-psychotic severely depressed, non-psychotic moderately depressed and psychotically depressed. Patients who fulfilled criteria for schizoaffective disorder were excluded, but it should be noted that patients with both unipolar depression and bipolar illness were included in both the non-psychotic and psychotic groups. Severity of depression was rated on the Hamilton scale both before and after the four-week treatment period, and patients were divided into severely and moderately depressed by using the median initial Hamilton score in the nonpsychotic group as the cut-off point. Patients were categorized as good, intermediate or poor responders based on an algorithm combining Hamilton, SADS-C global assessment scale, clinical global improvement, and clinical global severity ratings.²⁰⁰

¹⁹⁴ *ibid.*, 1842.

¹⁹⁵ Moncrieff 2008, 143.

¹⁹⁶ *ibid.*, 144.

¹⁹⁷ *ibid.*, 144.

¹⁹⁸ Kocsis JH, Croughan JL, Katz MM *et al.* Response to treatment with antidepressants of patients with severe or moderate nonpsychotic depression and of patients with psychotic depression. *Am. J. Psychiatry* 1990; 147: 621-624.

¹⁹⁹ *ibid.*, 621.

²⁰⁰ *ibid.*, 622.

67% of moderately depressed patients were found to be good responders, whilst 39% and 32% of the severely and psychotically depressed, respectively, were categorized as good responders. Difference in response among the three groups was statistically significant. Difference in response was not statistically significant when comparing the severe and psychotic groups, but was significant when comparing the moderately depressed group with the combined severe and psychotic groups.²⁰¹ The authors conclude that the results suggest that severely depressed patients might respond less favorably to antidepressant treatment than moderately depressed ones.²⁰²

Moncrieff argues that another study of bipolar illness strengthens her case. This study, Sachs *et al.* 2007, found that patients in the depressive phase of bipolar disorder, who were taking mood stabilizer drugs, did not respond better to drug than to placebo.²⁰³ The study was a randomized, double blind trial comparing the use of paroxetine or bupropion with placebo as adjunct therapy for depression in type I and II bipolar disorder. It included 366 patients who were randomized to active or placebo treatment for 26 weeks. The primary outcome was euthymia for eight weeks.²⁰⁴ All of the patients were taking mood stabilizer drugs: lithium, valproate, carbamazepine or other drugs approved by the FDA for preventing mania.²⁰⁵

Supplementing mood stabilizer treatment with paroxetine or bupropion was not superior to supplementation with placebo. 27.3% of patients in the placebo group experienced euthymia for eight weeks, whereas 23.5% of patients in the active drug groups did so. This result was not statistically significant ($p=0.40$).²⁰⁶ Noting that depressions in bipolar illness tend to be more severe than other types of depression, Moncrieff argues that this study provides further support for her claim that antidepressants are no better than placebo even in the most severe forms of depression.²⁰⁷

2.1.3 Long-term treatment

Moncrieff notes that continuation of antidepressant treatment for several months after an acute episode has resolved is conventionally recommended.²⁰⁸ She argues that even though several studies have shown that patients are at greater risk of relapse if they stop taking antidepressants, the discontinuation design of these studies undermines their results. She refers to several studies which demonstrate that antidepressant discontinuation results in adverse events.²⁰⁹

²⁰¹ *ibid.*, 622-623.

²⁰² *ibid.*, 624.

²⁰³ Sachs GS, Nierenberg AA, Calabrese JR *et al.* Effectiveness of adjunctive antidepressant treatment for bipolar depression. *N. Engl. J. Med.* 2007; 356: 1711-1722.

²⁰⁴ *ibid.*, 1711.

²⁰⁵ *ibid.*, 1712.

²⁰⁶ *ibid.*, 1719.

²⁰⁷ Moncrieff 2008, 144.

²⁰⁸ *ibid.*, 148.

²⁰⁹ *ibid.*, 149.

Dilsaver *et al.* 1987 reviews withdrawal symptoms of patients whose TCA treatment has been discontinued.²¹⁰ The authors outline four different types of withdrawal “syndrome”: gastrointestinal and general somatic distress possibly accompanied by anxiety, sleep disruption, movement disorders, and hypomania and mania.²¹¹ They note that it is possible to erroneously interpret withdrawal symptoms as relapse.²¹² Hindmarch *et al.* 2000 investigated the effect of discontinuation of fluoxetine, sertraline, paroxetine and citalopram on cognitive and psychomotor performance, social functioning, and depressive symptoms.²¹³ Significant effects were not found for the discontinuation of fluoxetine, sertraline, or citalopram – the discontinuation of paroxetine did produce such effects on the measures adopted.²¹⁴ The authors note that depressive symptoms reappeared so rapidly in patients discontinuing paroxetine that it is more likely that they were experiencing antidepressant discontinuation syndrome rather than recurrence of depression.²¹⁵ They also point out that such a withdrawal syndrome should not be considered common to all SSRIs, since it was notable only for paroxetine.²¹⁶

Moncrieff argues that withdrawal symptoms are not distinguished from true relapse in discontinuation studies, which means that some of the relapses observed in placebo groups might be due to discontinuation symptoms. She also speculates that such symptoms could unblind patients, i.e. they realize they have been withdrawn to placebo. Given that generally only patients who have already responded to drug treatment are included in continuation/maintenance studies, she argues that such unblinding may cause patients to think they will do worse on placebo, causing them to have a worse outcome.²¹⁷

Moncrieff includes only one review of discontinuation studies as evidence for her argument in this section. Viguera *et al.* 1998 examined 27 trials with 3037 patients.²¹⁸ In each trial, one group of patients continued on antidepressant treatment and another discontinued it. In 25 trials, those discontinuing treatment were withdrawn to placebo. In 17 trials, the active drug was a TCA, in five, it was a MAOI, and in another five, it was an SSRI.²¹⁹ The primary outcome was relapse, defined variously by clinical assessment, some psychiatric rating scale, Research Diagnostic Criteria or

²¹⁰ Dilsaver SC, Greden JF, Snider RM. Antidepressant withdrawal syndromes: phenomenology and pathophysiology. *Int. Clin. Psychopharmacol.* 1987; 2: 1-19.

²¹¹ *ibid.*, 2.

²¹² *ibid.*, 11. The study does refer to the frequency of withdrawal symptoms, but it is not primarily a study of their incidence.

²¹³ Hindmarch I, Kimber S, Cockle SM. Abrupt and brief discontinuation of antidepressant treatment: effects on cognitive function and psychomotor performance. *Int. Clin. Psychopharmacol.* 2000; 15: 305-318.

²¹⁴ *ibid.*, 313.

²¹⁵ *ibid.*, 316.

²¹⁶ *ibid.*, 316-317.

²¹⁷ Moncrieff 2008, 149.

²¹⁸ Viguera AC, Baldessarini RJ, Friedberg J. *et al.* Discontinuing antidepressant treatment in major depression. *Harv. Rev. Psychiatry* 1998; 5: 293-306.

²¹⁹ *ibid.*, 296.

hospital readmission.²²⁰ Patients had been stabilized for an average of 23.1 weeks (range 0-192) before discontinuation, and were followed up for an average of 16.6 months (range 6-60).

Relapse rates for patients withdrawn from antidepressant treatment were 6.24 per cent per month (SD \pm 5.34) and 1.85 per cent per month (SD \pm 1.51) for patients continuing active drug treatment.²²¹ There was a 3.37-fold difference in the time to 50% risk of relapse favoring active drug treatment.²²² The ratio of relapse rates in the placebo versus the drug group fell over time. The ratio of placebo versus drug relapse rates was 3.69 at two months, 2.21 at 12 months, and 1.34 at five years.²²³ Relapse rate was not associated with length of stabilization on antidepressant.²²⁴ In fact, as Moncrieff notes, relapse rates were higher for those who had been stabilized on an antidepressant for longer, but these results were not statistically significant.²²⁵ There were no significant differences in relapse rates between those studies that used an abrupt discontinuation of active drug treatment and those that used gradual withdrawal. In fact, the relapse rates were higher in the studies that used gradual withdrawal, but the result was not statistically significant.²²⁶

The authors argue that there are two possible ways of interpreting their results. The first is the conventional one: that the clear superiority of long-term antidepressant treatment over withdrawal implies that such drugs have an important prophylactic effect in recurrent depression. The second is that “[...] drug discontinuation itself may represent a clinically significant stressor that may temporarily increase relapse risk.”²²⁷ The authors point out that this hypothesis is consistent with the finding that relapse risk off versus on drug treatment decreased over time.²²⁸

Moncrieff argues that “People who believe that their recovery is attributable to antidepressant drugs are likely to feel anxious and vulnerable if those drugs are withdrawn.”²²⁹ She maintains that psychological effects are probably as significant, if not more so, than the pharmacological stress hypothesized in Viguera *et al.* 1998. Moncrieff further asserts that “psychological explanations would fit with the longer time to relapse and might explain how people who have been stable for longer periods are more at risk, since they are likely to be more psychologically dependent on the drugs.”²³⁰

²²⁰ *ibid.*, 295.

²²¹ *ibid.*, 296.

²²² *ibid.*, 297.

²²³ *ibid.*, 298.

²²⁴ *ibid.*, 298.

²²⁵ *ibid.*, 298; Moncrieff 2008, 149-150.

²²⁶ Viguera, Baldessarini & Friedberg *et al.* 1998, 299.

²²⁷ *ibid.*, 302.

²²⁸ *ibid.*, 302.

²²⁹ Moncrieff 2008, 150.

²³⁰ *ibid.*, 150.

2.2 The placebo argument: Kirsch

Kirsch's version of the placebo argument in *The Emperor's New Drugs* relies heavily on meta-analyses he has co-authored. The most recent one, Kirsch *et al.* 2008, has already been discussed in section 2.1; I will discuss only the two other studies here.

The first one is the 1998 study by Kirsch and Sapirstein.²³¹ The aim of this meta-analysis was to study response to placebo in antidepressant trials.²³² Kirsch & Sapirstein distinguish placebo "response" from the "placebo effect." They define the placebo response as the "change that occurs following administration of a placebo."²³³ They note that this response could also come about without the administration of a placebo – it could occur because of spontaneous remission, a change in the patient's life circumstances, or other such factors not related to the patient taking the placebo pill. The authors define the placebo effect as "the difference between the placebo response and changes that occur without the administration of a placebo."²³⁴ In other words, the "placebo effect" is the effect that the administration of a placebo has, when other causes for a patient's improvement have been eliminated.

The meta-analysis investigated two questions: How big is the placebo *response* compared to the response to active drug in antidepressant trials? How big is the "placebo *effect*" in such trials? The first question was answered based on a meta-analysis of randomized controlled drug trials.²³⁵ The second question was answered on the basis of comparing placebo response in such trials with the response of people assigned to the waiting-list group in psychotherapy trials. The authors note that in drug trials, a no-treatment group is usually not included, but in psychotherapy trials, there is often a group of patients randomized to remain on the waiting list.²³⁶ In other words, in psychotherapy trials there is a group of patients who do not receive any treatment. In order to tease out the placebo "effect" it is necessary to compare patients receiving placebo with patients receiving no treatment.

Randomized placebo-controlled trials of patients between the ages of 18-75 with a primary diagnosis of depression in the acute phase of treatment that included sufficient data to calculate effect sizes were included in the first part of this meta-analysis.²³⁷ 19 trials with a total of 2 318 patients were considered to meet the criteria. The range of medications studied was considerable: amitriptyline, amylobarbitone, fluoxetine, imipramine, paroxetine, isocarboxazid, trazodone,

²³¹ Kirsch I & Sapirstein G. Listening to Prozac but hearing placebo: A meta-analysis of antidepressant medication. *Prevention & Treatment* 1998; 5. As with the 2002 Kirsch & Moore paper, I was only able to obtain an electronic copy without the original page numbers.

²³² *ibid.*, 2.

²³³ *ibid.*, 8.

²³⁴ *ibid.*, 8. See also Kirsch 2009, 11.

²³⁵ Kirsch & Sapirstein 1998, 3.

²³⁶ *ibid.*, 8.

²³⁷ For the calculation of effect sizes, see *ibid.*, 3-4.

lithium, liothyronine, adinazolam, amoxapine, phenelzine, venlafaxine, maprotiline, tranylcypromine, and bupropion.²³⁸ Only fluoxetine and paroxetine are SSRIs.

Mean effect sizes were 1.55 SDs for medication response and 1.16 for placebo response, giving a mean effect for medication of 0.39 SDs. Kirsch & Sapirstein note that 75% of the response to medication is accounted for by placebo response, leaving 25% of the drug response as the “true” drug effect.²³⁹ In other words, they write, “[...] for a typical patient, 75% of the benefit obtained from the active drug would also have obtained from an inactive placebo.”²⁴⁰

The authors considered the possibility that the type of medication used might have an impact on drug and placebo response rates.²⁴¹ They categorized the medications into TCAs/tetracyclics, SSRIs, other antidepressants, and other medications. The last category includes amylobarbitone, lithium, liothyronine, and adinazolam, which were studied in the drug trials, but which are not conventionally considered antidepressants.²⁴² The placebo response as a proportion of active drug response was 0.74 for SSRIs and 0.76 for the other categories.²⁴³ Kirsch notes that the type of active medication used had virtually no effect on the superiority of drug over placebo.²⁴⁴

For the second part of the meta-analysis, 19 psychotherapy trials with no-treatment or waiting list groups were included.²⁴⁵ There were no statistically significant differences in the severity of depression between the drug and psychotherapy trials.²⁴⁶ The mean effect size for no-treatment/waiting list response was 0.37.²⁴⁷ Thus, the effect size for placebo “effect” was 0.79 (effect size for placebo response (1.16) minus effect size for no treatment). The authors conclude that half of the response to active drug is due to the placebo “effect”, one quarter is attributable to nonspecific factors, and only one quarter can be attributed to the administration of the active drug.²⁴⁸

The authors proceed to arguing that the superiority of drug over placebo might be caused not by the drugs having true pharmacological effects that alleviate depression, but because they act as “active placebos”, causing more side effects than placebo pills. The argument is based on the observation that there is an exceptionally high correlation between drug and placebo responses, and

²³⁸ *ibid.*, 3.

²³⁹ *ibid.*, 4.

²⁴⁰ *ibid.*, 4.

²⁴¹ See the discussion in *ibid.*, 6.

²⁴² *ibid.*, 6-7.

²⁴³ *ibid.*, 7.

²⁴⁴ Kirsch 2009, 12.

²⁴⁵ Kirsch & Sapirstein 1998, 9-10.

²⁴⁶ *ibid.*, 11.

²⁴⁷ *ibid.*, 9.

²⁴⁸ *ibid.*, 11. See also Kirsch 2009, 11.

that medications not conventionally considered antidepressants were found as effective as antidepressants.²⁴⁹

The second meta-analysis that Kirsch bases his argument on, Kirsch & Moore 2002, uses data submitted to the FDA for approval of fluoxetine, paroxetine, sertraline, venlafaxine, nefazodone, and citalopram.²⁵⁰ Publication bias does not affect this data, since all trials submitted to the FDA were obtained using the Freedom of Information Act.²⁵¹ In nine trials, there was missing data.²⁵²

All trials assessed depression severity on the Hamilton scale, and with the exception of one trial on mild depression, patients in all trials were either moderately or severely depressed (range of mean Hamilton scores 21.0 – 29.7).²⁵³

Weighted mean improvement scores were recorded for all trials, which had full data: these were the trials on fluoxetine, venlafaxine, and nefazodone.²⁵⁴ For these medications, the mean difference in improvement was 1.80 points on the Hamilton scale. Placebo response was 82 per cent of the drug response. The superiority of the active drugs over placebo was statistically significant.²⁵⁵

There were twelve fixed-dose studies included, whilst in most of the trials, dose was determined individually within a predetermined range. To assess the possible effect of dosage, last observation carried forward data was compared for patients receiving the lowest and highest prescribed doses in the fixed-dose studies. Mean improvement on the lowest dose was 9.57 points, and 9.97 points on the highest dose; this difference was not statistically significant.²⁵⁶

Kirsch & Moore note that the mean drug/placebo difference, 1.80 on the Hamilton scale, is of questionable clinical relevance. However, they point out that mean improvement in the placebo group was equal to or superior to improvement in the drug group in only four trials, and placebo was not significantly superior to drug in any of the trials. They concede that it is possible that there is a small but real drug effect, but go on to present the “active placebo” argument as another possible explanation.²⁵⁷ Kirsch & Moore refer to the same Rabkin *et al.* 1986 study that Kirsch references in *The Emperor’s New Drugs* as evidence that the ability of doctors and patients to guess whether the patient has been assigned to the drug or placebo arm of a double-blind trial exceeds

²⁴⁹ Kirsch & Sapirstein 1998, 12.

²⁵⁰ Kirsch & Moore 2002, 2. See also the discussion in Kirsch 2009, 27-30.

²⁵¹ Kirsch & Moore 2002, 3.

²⁵² *ibid.*, 6.

²⁵³ *ibid.*, 3.

²⁵⁴ *ibid.*, 6.

²⁵⁵ *ibid.*, 6.

²⁵⁶ *ibid.*, 6.

²⁵⁷ *ibid.*, 7.

chance.²⁵⁸ They also propose the more frequent side effects caused by drug rather than placebo administration as an explanation for this ability to guess,²⁵⁹ but ignore the fact that the Rabkin *et al.* 1986 data lends no weight to this thesis, as I will discuss in section 4.

They also discuss the possibility that drug and placebo effects are not additive. An assumption of placebo-controlled trials is that the drug response is the sum of the true pharmacological effect of the drug and the placebo response. In this case, the difference between drug and placebo response represents the “true” drug effect. If, however, these effects are not additive in this way, the drugs might be more effective than they seem, because the methodology of the trials is inadequate. Kirsch & Moore note that placebo alcohol and placebo caffeine are not fully additive in their effects when compared to the real thing. They propose a “balanced placebo design” for drug trials to sort out these effects; a full discussion is beyond the scope of this essay.²⁶⁰

²⁵⁸ *ibid.*, 7; Rabkin JG, Markowitz JS, Stewart J *et al.* How blind is blind? Assessment of patient and doctor medication guesses in a placebo-controlled trial of imipramine and phenelzine. *Psychiatry Research* 1986; 19: 75-86.

²⁵⁹ Kirsch & Moore 2002, 7.

²⁶⁰ *ibid.*, 8-9.

3. “Active placebo”

Both Kirsch and Moncrieff argue that the modest superiority of antidepressants over placebo is not due to a true pharmacological effect, but rather to patients breaking blind because of side effects of the active drug. This is the “active placebo” argument. Another aspect of the argument is the observation that other drugs, such as antipsychotic agents, benzodiazepines, and mood stabilizers, that are not conventionally considered antidepressants, have been found to be superior to placebo in treating depression. Since these drugs are supposedly not antidepressants, the critics argue, their superiority over placebo in treating depression stems from their side effects. The implication that the critics draw from this is that it demonstrates that conventional antidepressants, such as SSRIs, have no specific effect on depression, either. The superiority they demonstrate over placebo is due to the same reason as the superiority of other drugs over placebo in treating depression: the unwanted side effects they produce.

Another aspect of this argument that Moncrieff in particular emphasizes is the idea that even though antidepressants have no specific pharmacological effect on depression, they have unspecific effects, such as sedation, which in certain situations can be useful in alleviating the symptoms from which the depressed patient suffers.

3.1 “Active placebo”: Kirsch

Kirsch argues that in double-blind trials, patients’ expectations of getting better are tempered by their knowledge that it is possible that they have been given a placebo. Breaking blind, according to Kirsch, leads to a change of expectations. If patients know they are taking placebo, they are less likely to think they are getting better, and if they know they are on the active drug, this enhances their expectation of recovery. This expectation, Kirsch maintains, affects recovery.²⁶¹

Kirsch writes that the only reason he can think of why such side effects of antidepressant medication as insomnia, diarrhea or nausea might be correlated with the alleviation of depression is that they lead the patient to break blind.²⁶² Kirsch also writes: “the correlation between side effects and improvement is so strong as to be almost perfect.”²⁶³

Kirsch refers to the 2008 Sneed *et al.* study as evidence for his position.²⁶⁴ Sneed *et al.* 2008 compared the response of patients in trials comparing different antidepressants with each other with the response of patients in placebo-controlled trials. The authors refer to Krell *et al.* 2004 in

²⁶¹ Kirsch 2009, 16.

²⁶² *ibid.*, 18.

²⁶³ *ibid.*, 18.

²⁶⁴ Sneed JR, Rutherford BR, Rindskopf D *et al.* Design makes a difference: a meta-analysis of antidepressant response rates in placebo-controlled versus comparator trials in late-life depression. *American Journal of Geriatric Psychiatry* 2008; 16: 65-73.

pointing out that patients who have higher expectations of responding to treatment tend to improve better than patients whose expectations are more modest.²⁶⁵ They point out that patients in comparator trials know they are receiving a medication that is supposed to be effective. In placebo-controlled trials, patients can only hope to receive such medication, and the doubt that they may end up receiving dummy pills potentially dampens their response to antidepressants.²⁶⁶

The Sneed *et al.* 2008 meta-analysis included 16 trials studying the treatment of acute-phase unipolar major depression in patients 60 years or older. Seven trials were placebo-controlled and nine were comparator trials.²⁶⁷ The drugs studied were escitalopram, fluoxetine, paroxetine, citalopram, venlafaxine, sertraline, nortriptyline, doxepin, imipramine, clomipramine, trimipramine, mirtazapine and bupropion.²⁶⁸ The primary outcome was response, defined in most trials as a 50% reduction in Hamilton score, but in two trials as 50% reduction in MADRS score and in three trials as a CGI score of one or two.²⁶⁹ In comparator trials, the odds of being a responder were 1.82 times higher than in placebo-controlled trials. They estimated the response rate as 60% in comparator trials and 46% in placebo-controlled trials.²⁷⁰

Kirsch argues that this difference in response rates stems from the patients in the drug comparator trials knowing that they were definitely receiving an active drug, thus enhancing their expectations of recovery.²⁷¹ Sneed *et al.* 2008 agrees that this is a possible interpretation of the results. The authors refer to Schweizer *et al.* 1998 and note that when patients suffering from such ailments as chronic pain, anxiety and Parkinson's disease are unaware that they are being given treatment, they do not respond as well as when they know treatment is being administered.²⁷² They also point out that it is possible that the observed differences stem from study doctors having different expectations in different types of trials, and raise the possibility that patients in placebo-controlled trials might be more severely depressed than patients in comparator trials, which might affect response rates.²⁷³

Kirsch further points out that there is a correlation between experiencing side effects from SSRIs and experiencing improvement. He refers to Greenberg *et al.* 1994 as evidence.²⁷⁴ This is a meta-analysis of all placebo-controlled fluoxetine trials published through 1992. Thirteen trials

²⁶⁵ *ibid.*, 66.

²⁶⁶ *ibid.*, 66.

²⁶⁷ *ibid.*, 67.

²⁶⁸ *ibid.*, 68.

²⁶⁹ *ibid.*, 68.

²⁷⁰ *ibid.*, 70.

²⁷¹ Kirsch 2009, 17.

²⁷² Sneed, Rutherford & Rindskopf *et al.* 2008, 70.

²⁷³ *ibid.*, 71.

²⁷⁴ Greenberg RP, Bornstein RF, Zborowski MJ et al. A meta-analysis of fluoxetine outcome in the treatment of depression. *Journal of Nervous and Mental Disease* 1994; 182: 547-551.

were included.²⁷⁵ The overall Cohen's *d* effect size for fluoxetine was .40, which the authors consider moderate.²⁷⁶ They point out that the average patient taking fluoxetine was better off than two thirds of patients on placebo, noting that the results were quite reliable: it would take more than 200 unpublished trials with zero effect size to render the results statistically nonsignificant.²⁷⁷

The authors did not find significant correlations between effect size and patients' mean age, percentage of women, or fluoxetine dosage.²⁷⁸ However, effect size was significantly correlated with the proportion of patients reporting side effects.²⁷⁹ This was true both for clinician and patient reported outcomes. It should be noted, though, that only half of the studies provided data on side effects.²⁸⁰ The authors speculate on the same conclusion that Kirsch does: that the correlation between side effects and effect size is due to unblinding. Greenberg *et al.* also argue that trials with "active placebo" comparators might lend weight to the unblinding hypothesis.²⁸¹

Kirsch argues that one would expect the correlation to go in the opposite direction: that the unwanted side-effects of SSRIs, such as sexual dysfunction, insomnia and gastrointestinal distress would cause the patient to be more depressed, not less. He maintains that the reason that patients experiencing side effects show superior drug responses to those not experiencing side effects must be that those effects lead the patient to realize they are in the active treatment group.²⁸²

Kirsch argues that the 2009 paper by Barbui, Cipriani and Kirsch provides further evidence for his position. The aim of the study was to isolate the true pharmacological effect from possible "active placebo" effects. The study is an analysis of all published and unpublished trials conducted by GlaxoSmithKline on paroxetine. According to Kirsch 2009, it concludes that adjusting for drug/placebo differences in side effects eliminates the statistical significance of drug/placebo differences in efficacy.²⁸³ I was unable to find the Cipriani, Barbui & Kirsch paper on PubMed in January 2017. Kirsch lists it as unpublished in 2009. Presumably it is still unpublished, making it impossible to assess its relevance.

There is also some evidence that patients and doctors in supposedly double-blind trials are able to guess which treatment the patient is receiving, thus breaking blind. Kirsch points to Rabkin *et al.* 1986, which investigated this question.²⁸⁴ The study investigated patients' and doctors' guesses on assignment to treatment group in a randomized double-blind trial comparing

²⁷⁵ *ibid.*, 548.

²⁷⁶ *ibid.*, 549.

²⁷⁷ *ibid.*, 549.

²⁷⁸ *ibid.*, 549. It should be noted that data on mean dosage was unavailable for most of the trials, so the meta-analysis utilized the maximum dosage allowed to assess the effects of fluoxetine dosage.

²⁷⁹ *ibid.*, 549.

²⁸⁰ *ibid.*, 550.

²⁸¹ *ibid.*, 550.

²⁸² Kirsch 2009, 18.

²⁸³ Kirsch 2009, 19.

²⁸⁴ Rabkin, Markowitz & Stewart *et al.* 1986.

imipramine, phenelzine and placebo. 137 patients were assigned to either one of the drug arms or the placebo arm of the trial after a ten-day placebo washout.²⁸⁵ The patients included met Research Diagnostic Criteria for major, minor or intermittent depression. After six weeks of treatment, doctors and patients were asked to guess what treatment the patient had been assigned to. The patients' response to this question was categorized as either active drug or placebo, whereas the doctors' guesses for the specific active drugs were recorded separately.

The study found that 79% of patients were able to correctly identify whether they were in the drug or the placebo arm of the study. Doctors could do this in 87% of cases.²⁸⁶ Patients receiving the active drug were more likely to guess their group correctly. 87% of patients on imipramine, 96% of patients on phenelzine, but only 59% of patients in the placebo group did so.²⁸⁷ Kirsch points out that the likelihood of the patients randomly guessing correctly is exceedingly small.²⁸⁸

Kirsch goes on to point out studies comparing antidepressants with “active placebos”, i.e. drugs, which produce side effects but should have no pharmacological effect on depression. The aim of using such drugs as “placebo” is to prevent patients from breaking blind.²⁸⁹ Kirsch refers to Moncrieff *et al.* 2005 here in discussing nine antidepressant trials that used atropine as an “active placebo”. This study is a Cochrane Review based on virtually the same material as the 1998 Moncrieff, Wessely & Hardy paper. There is slight variation in the results²⁹⁰, but for my purposes, the discussion I present on the 1998 paper in section 4.2 suffices.

A further aspect of the active placebo argument is that there is some evidence that active drugs, which are not antidepressants, are superior to placebo in the treatment of depression. In the 1998 Kirsch & Sapirstein study, four active drugs that are not antidepressants are included. These are amylobarbitone, lithium, liothyronine, and adinazolam.²⁹¹ Amylobarbitone is a barbiturate, an older class of sedative. Lithium is used as a mood stabilizer in bipolar illness. Liothyronine is a synthetic thyroid hormone used for hypothyroidism, and adinazolam is a benzodiazepine, a newer class of anxiolytic.²⁹² The drug/placebo difference in effect size for these non-antidepressants when

²⁸⁵ *ibid.*, 77.

²⁸⁶ *ibid.*, 78.

²⁸⁷ *ibid.*, 79.

²⁸⁸ Kirsch 2009, 15-16.

²⁸⁹ *ibid.*, 19.

²⁹⁰ Cf. Moncrieff, Wessely & Hardy 1998, 230 and Moncrieff, Wessely & Hardy 2004, 7-8. Moncrieff J, Wessely S, Hardy R. Meta-analysis of trials comparing antidepressants with active placebos. *Br. J. Psychiatry* 1998; 172: 227-231. Moncrieff J, Wessely S & Hardy R. Active placebos versus antidepressants for depression (Review). *Cochrane Database of Systematic Reviews* 2004; 1: 1-31.

²⁹¹ Kirsch & Sapirstein 1998, 7.

²⁹² Kirsch 2009, 13.

given for depression was 0.76, whereas the effect size for the SSRIs, the drugs of choice for depression today, was 0.74.²⁹³

The paper suggests that it is possible that these drugs have an indirect effect on depression, but in this case, it would be expected that they would not be as effective as antidepressants supposed to have a more specific effect. In fact, there is no difference in superiority over placebo between antidepressants and these other active drugs. The paper also suggests that perhaps these four drugs should be classified as antidepressants, but then the definition would need to be based merely on superiority over placebo in clinical trials rather than the underlying pharmacological mechanism.²⁹⁴

However, the explanation that Kirsch and Sapirstein embrace in this paper is the same one Kirsch presents in *The Emperor's New Drugs*: that the non-antidepressant active drugs are superior to placebo in the treatment of depression because they cause more side effects than placebo. *I.e.*, the antidepressants and the other active drugs are more effective because patients realize they are receiving active treatment as a consequence of side effects.²⁹⁵

3.2 Moncrieff on antidepressants versus “active placebo”

One line of argument that Moncrieff pursues is that antidepressants seem less effective when they are compared to “active placebos” than when the comparator is an inert placebo. By definition a placebo is inert, but what Moncrieff means by “active placebo” is a substance that is not supposed to have any pharmacological effect on the disorder being studied, but which is sufficiently active pharmacologically that it produces some of the unwanted effects of the active drug being studied. Such “active placebos” should reduce the chance of unblinding.

In chapter nine of *The Myth of the Chemical Cure*, Moncrieff discusses a 1998 meta-analysis of trials using an “active placebo” that she co-authored.²⁹⁶ The Moncrieff, Wessely & Hardy study included nine trials comparing imipramine, amitriptyline, and nortriptyline, all TCAs, with atropine. The trials were published between 1961-1984, with only one trial published in the 1980s.²⁹⁷ A variety of outcomes for calculating effect size were used.²⁹⁸ The authors note that two of the trials demonstrated a statistically significant superiority of drug over atropine.²⁹⁹ In most trials, the results favored the antidepressant, but most results were not statistically significant.³⁰⁰

²⁹³ Kirsch & Sapirstein 1998, 7.

²⁹⁴ *ibid.*, 7.

²⁹⁵ *ibid.*, 7-8.

²⁹⁶ Moncrieff 2008, 147.

²⁹⁷ Moncrieff, Wessely & Hardy 1998, 228.

²⁹⁸ *ibid.*, 228.

²⁹⁹ *ibid.*, 229.

³⁰⁰ *ibid.*, 230.

Moncrieff points out that in trials with inpatients, no statistically significant differences between drug and “active placebo” were found, and “only small differences” were observed in all but one of the trials involving outpatients.³⁰¹ She notes that results still favored the antidepressants, and that the quality of the trials has rightly been questioned, but maintains that even the use of atropine as “active placebo” was not sufficient to maintain the integrity of the blind.³⁰²

In two trials, there was an attempt to assess the integrity of the blind by asking assessors to guess to which treatment arm patients had been randomized. The authors state that “[...] the guesses were more accurate than would be predicted by chance,” even though they note that the effect was not statistically significant.³⁰³ The authors note that in the Weintraub & Aronson 1963 trial, raters assessed patients who they guessed to be in the drug arm to be more improved. Moncrieff argues in *The Myth of the Chemical Cure* that this suggests that assessments of patients may have been influenced by treatment expectations.³⁰⁴ Moncrieff and colleagues also point out that in the Hollister *et al.* 1964 trial, more side effects were reported in the drug group despite the use of atropine as an “active placebo”, which means that “residual unblinding effects may have occurred”.³⁰⁵ Moncrieff argues that this is evidence of the possibility that the use of atropine was insufficient for maintaining the integrity of the blind.³⁰⁶

The authors refer to other meta-analyses and point out that their results show more modest efficacy for antidepressants than other studies.³⁰⁷ This suggests that the use of an “active placebo” yields more modest results for antidepressants.

3.3 Moncrieff on antidepressants versus other drugs

Moncrieff’s placebo argument is not limited to pointing out the limited superiority of antidepressants over placebo. An important aspect of it is her claim that drugs other than those considered specifically antidepressants have been found to be superior to placebo in the treatment of depression in clinical trials. Moncrieff notes that in the 1960s, when benzodiazepines were new, many trials found them to be equivalent or superior to antidepressants for depression, but in the 1970s most trials found antidepressants superior. She points out that when the new benzodiazepine alprazolam arrived on the market in the 1980s, trials found it effective for treating depression.³⁰⁸

³⁰¹ Moncrieff 2008, 147.

³⁰² *ibid.*, 147.

³⁰³ Moncrieff, Wessely & Hardy 1998, 230.

³⁰⁴ Moncrieff 2008, 149.

³⁰⁵ Moncrieff, Wessely & Hardy 1998, 230.

³⁰⁶ Moncrieff 2008, 147.

³⁰⁷ Moncrieff, Wessely & Hardy 1998, 230.

³⁰⁸ Moncrieff 2008, 144-146.

In *The Myth of the Chemical Cure*, she presents a table summarizing the results of seventeen trials or reviews of various drugs versus either an established antidepressant or placebo for treating depression. These studies were published between 1955-2006, most of them having been published before 1990. The drugs studied range from benzodiazepines and antipsychotics to amphetamine-derivatives and opioids.³⁰⁹

Davies & Shepherd 1955 found the antihypertensive/antipsychotic drug reserpine superior to placebo for treating depression. The 1982 Robertson & Trimble review of antipsychotics for depression found antipsychotics superior to placebo and in most comparisons equivalent with antidepressants. Blashki *et al.* 1971 found amylobarbitone broadly as effective as amitriptyline. Schatzberg & Cole 1978 reviewed 20 trials comparing benzodiazepines with antidepressants; in half benzodiazepines were either equal to antidepressants or superior to placebo. Imlah 1985, Feighner *et al.* 1983, Rickels *et al.* 1987, and Weissman *et al.* 1992 found alprazolam superior to placebo and at least equal to imipramine or amitriptyline.³¹⁰

Rickels *et al.* 1970 found pemoline, a stimulant now withdrawn from the market, and methylphenidate, a stimulant used for treating ADHD, superior to placebo. Hare *et al.* 1964 found a combination of dexamphetamine and amylobarbitone equal in efficacy as imipramine. Robinson *et al.* 1990 found buspirone, indicated for treating anxiety, superior to placebo for treating depression. Fabre 1990 found a trend favoring buspirone over placebo. Emrich *et al.* found buprenorphine, a strong opioid, superior to placebo for depression. Philipp *et al.* 1999, Szegedi *et al.* 2005, and Kasper *et al.* 2006 found the medicinal herb St. John's wort equivalent to imipramine and paroxetine and superior to placebo.³¹¹

³⁰⁹ *ibid.*, 145-146.

³¹⁰ *ibid.*, 145.

³¹¹ *ibid.*, 145-146.

4. Assessing the Kirsch-Moncrieff argument

In the previous two sections, I have attempted to present the major evidence that the Kirsch-Moncrieff arguments are based on. The aim here is to ask the central question of this essay: how does the evidence and argumentation stand up to scrutiny?

4.1 Short-term treatment

The fundamental problem with Moncrieff's argument about the lack of efficacy of antidepressants for short-term treatment is the limited amount of relevant evidence. To be sure, Moncrieff amasses an impressive number of trials in support of her argument, but a close reading of the relevant section of her book shows that there are only two studies that really matter here: Kirsch *et al.* 2002 and NICE 2004.

The value of the 1965 Medical Research Council study for Moncrieff's argument is questionable, as it concerns early antidepressants no longer in clinical use. Likewise, the 1975 National Institutes of Mental Health study has virtually no relevance for our discussion of the claim that antidepressants in clinical use today are no better than placebo. Again, the 1976 Bielski & Friedel review concerns tricyclic antidepressants, so it is of limited value in assessing Moncrieff's argument as it relates to the SSRIs and other modern antidepressants. The same goes for the Joyce & Paykel review, published in 1989. All these studies may have value in understanding how the notion of an antidepressant was constructed historically. To be fair to Moncrieff, this is partly what she is doing, and understanding this construction is part of her overall project.

The other studies referred to in this section of *The Myth of the Chemical Cure* illustrate her point that older meta-analyses show antidepressants to be more effective than newer ones. I have discussed some of the newer ones. Bech *et al.* 2000, Khan, Warner & Brown 2000, and Storosum *et al.* 2001 show modest superiority of drug over placebo. Moncrieff does not dispute this.

Thus, in this section, Moncrieff is really left with only Kirsch 2002 and NICE 2004 to make the argument that antidepressants are so ineffective that the drug effect is virtually irrelevant. Moncrieff asserts that the 1.7-point superiority of drug over placebo in the Kirsch *et al.* 2002 study is clinically meaningless by "all assessments of clinical validity."³¹² She gives no justification for this claim, and gives no indication of which assessments of clinical validity she is referring to, though, to be sure, the difference of 1.7 points would be clinically irrelevant under the NICE criteria. Unfortunately, I am unable to give any analysis of the other crucial study, the NICE 2004 review.

³¹² *ibid.*, 140.

I have been critical of Moncrieff in this section, but I think this is the part of the argument that stands up to scrutiny, though Moncrieff is not very convincing in making her case. At least on the basis of the material covered in this essay, it seems clear that the effectiveness of antidepressants is modest. In virtually all the trials covered, the superiority of drug over placebo amounts to only a few points on a particular rating scale. I am fully aware that I am not using any rigorous criteria in claiming this, but rather making a general observation. Furthermore, whether this moderate superiority, however defined, makes a clinically relevant difference, is another question, which I will briefly return to in the Conclusion.

4.2 Severity

As we have seen, Moncrieff moves from criticizing the evidence on acute phase treatment to arguing against the commonly held assumption that antidepressants are more effective in treating more severe forms of depression. This part of her argument is undermined by her disingenuous use of the evidence.

Again, part of the problem here is Moncrieff's use of old studies. A couple of the studies she refers to – Bielski & Friedel 1976 and Joyce & Paykel 1989 – do not really even demonstrate the lack of efficacy of antidepressants in more severe depression, they merely point out that the evidence is unclear. The Angst *et al.* 1993 study that Moncrieff notes found weak and mostly not statistically significant effects for greater drug efficacy in more severe depression, studied only one antidepressant, moclobemide versus placebo and TCAs.³¹³

Moncrieff muddles the waters even further with her rhetoric on the 2002 Khan *et al.* study. It seems that she considers a problem the fact that full data was available only for so-called investigational antidepressants. She fails to mention that these antidepressants are some of the most common in clinical use today – and much more relevant than older TCAs such as imipramine that she relies on to make much of her argument. To be sure, it is a weakness of Khan *et al.* 2002 that the authors do not name the “established” antidepressants; I assume that they mean the same older antidepressants (imipramine, amitriptyline, and trazodone) investigated in Khan *et al.* 2000.³¹⁴

In any case, these “investigational” antidepressants are the most clinically relevant today. Moncrieff notes that the relationship of initial severity of depression with drug response appeared to be weaker for the established antidepressants,³¹⁵ but this is almost irrelevant.

Moncrieff's use of the Walsh *et al.* 2002 study is likewise disingenuous. Referring to this study, she writes: “... one recent meta-analysis found no effect of initial severity on treatment

³¹³ Angst, Scheidegger & Stabl 1993.

³¹⁴ This assumption is based on the reference in Khan, Leventhal & Khan *et al.* 2002, 41 to Khan, Warner & Brown 2000. See Khan, Warner & Brown 2000, 312.

³¹⁵ Moncrieff 2008, 143.

response.”³¹⁶ What she fails to mention is that the study did not primarily investigate the question that she is discussing. Walsh *et al.* 2002 is not a study of how the initial severity of depression affects drug response or drug/placebo differences, it is a study concerned with how placebo responses have changed over the years.³¹⁷ The effect of the initial severity of depression on SSRI response was not even investigated. Moncrieff’s enlistment of this study in support of her argument that antidepressants have limited efficacy in severe depression is dubious.

Moncrieff’s arguments about psychotic depression and bipolar illness are also curious. She utilizes the Kocsis *et al.* 1990 paper in arguing that the general ineffectiveness of antidepressants rather than the presence of psychotic symptoms explains the apparent lack of efficacy of antidepressants in treating psychotic depression. This should be taken with a number of caveats. First, like most of the studies that Moncrieff draws on, this study examines tricyclic antidepressants, not SSRIs or other more recent drugs most widely in use today. Second, the number of patients included was small: there were only 25 patients in the psychotic group.³¹⁸

Furthermore, Kocsis *et al.* 1990 included both patients diagnosed with unipolar depression and bipolar disorder.³¹⁹ Given that these are two distinct diagnostic categories, it is possible that there are differences between unipolar depression and a depressive episode in bipolar illness, which confound the results. Furthermore, it should be noted that, even if Kocsis *et al.* 1990 could be construed as lending some weight to Moncrieff’s particular argument about the relationship between depression severity and drug response, it does not seem to support her more general argument about the effectiveness of antidepressants, since 67% of moderately depressed patients responded well to active drug treatment.³²⁰

Moncrieff’s use of the Sachs *et al.* 2007 study to make her argument does not stand up to scrutiny, either. Again, Moncrieff discusses a study whose primary objective was to answer a different question than the one necessary to make her point. The study investigated the use of adjunct antidepressant treatment in the depressive phase of bipolar illness rather than the effect of the severity of depression on drug/placebo differences.³²¹ And again, Moncrieff bases her argument on an assumption, and does not provide the actual figures on severity in this trial. The baseline data is available³²², but since the study investigated a different outcome, the authors have not analyzed the potential effect of depression severity on drug/placebo differences in outcome.

³¹⁶ *ibid.*, 143.

³¹⁷ Walsh, Seidman & Sysko *et al.* 2002, 1840.

³¹⁸ Kocsis, Croughan, Katz *et al.* 1990, 623, Table 1.

³¹⁹ *ibid.*, 623.

³²⁰ *ibid.*, 622.

³²¹ Sachs, Nierenberg & Calabrese *et al.* 2007, 1711.

³²² *ibid.*, 1717.

In Sachs *et al.* 2007, the average baseline severity of depression on the MADRS scale is 24.5 in the drug group and 24.0 in the placebo group.³²³ The maximum result on the MADRS scale is 60 points. According to Jääskeläinen & Miettunen 2011, a score of 24 would indicate mild depression and a score of 25 would indicate moderate depression.³²⁴ Thus, the figures on severity in Sachs *et al.* 2007 are not even compatible with Moncrieff's claim that bipolar depression is worse than unipolar depression. I by no means wish to dispute that in general, depression is more severe for those suffering from bipolar disorder, but merely to point out the problematic way that Moncrieff develops her argument.

As I have already noted, it is also possible that there is something different about depression in bipolar disorder, which causes it to react differently to medication than unipolar depression. It is quite questionable to draw conclusions about the efficacy of treatments by assuming that diagnosis makes no difference. Furthermore, all participants in Sachs *et al.* 2007 were on mood stabilizers. Such drugs are used for unipolar depression only in special cases. It is hard to justify Moncrieff's enlistment of Sachs *et al.* 2007 as evidence for her argument.

4.3 Long-term treatment

We have seen that Moncrieff wishes to undermine the evidence on maintenance treatment with antidepressants, as well. Discontinuation of antidepressant treatment can cause withdrawal effects.³²⁵ The 2004 NICE guideline notes that "All patients prescribed antidepressants should be informed that, although the drugs are not associated with tolerance and craving, discontinuation/withdrawal symptoms may occur on stopping, missing doses or, occasionally, on reducing the dose of the drug."³²⁶

Moncrieff argues that withdrawal effects undermine the evidence base on relapse prevention studies, since it is not possible to distinguish them from actual recurrence of depression. Whilst it is completely possible that this is true, Moncrieff is not able to prove her point. She refers to only Viguera *et al.* 1998 in making this argument. As we have seen, the study showed clear superiority for drug over placebo. The results are completely compatible with Moncrieff's thesis, but as the authors point out, they are also compatible with the conventional view that antidepressants have a real prophylactic effect.³²⁷

³²³ *ibid.*, 1717, Table 2.

³²⁴ Jääskeläinen E & Miettunen J 2011. Psykiatriset arviointiasteikot kliinisessä työssä. Lääketieteellinen Aikakauskirja Duodecim 2011; 127:1719-1725, supplementary online material.

³²⁵ For a good summary of evidence on withdrawal symptoms, see Wilson E & Lader M 2015. A review of the management of antidepressant discontinuation symptoms. *Therapeutic Advances in Psychopharmacology* 2015; 5: 357-368, 357-359.

³²⁶ NICE 2004, 25-26.

³²⁷ Viguera, Baldessarini & Friedberg 1998, 302.

Moncrieff argues that people who attribute their recovery to drugs will feel “anxious and vulnerable” when the drugs are withdrawn, but she provides no evidence for this claim.³²⁸ Whatever its factual basis, it should be noted that if applied to the Viguera *et al.* 1998 meta-analysis, it assumes that patients know they have been withdrawn. In 25 of the trials, patients were withdrawn to placebo. Thus, Moncrieff assumes that they have been unblinded, presumably due to withdrawal symptoms, since this is the only way they would know that they were now receiving placebo instead of the active drug they had been taking until then. However, Viguera *et al.* 1998 did not study such symptoms.

Moncrieff also maintains that psychological effects are probably more important for explaining the increased risk of relapse after withdrawal than the pharmacological stress hypothesized by Viguera *et al.* 1998.³²⁹ She gives no evidence for this position, nor does she give a definition of psychological dependence, presumably distinct from the physiological dependence resulting in withdrawal symptoms. But she also seems to assume that it is an established fact that people who have been stabilized on antidepressants for longer are at greater risk of relapse when treatment is withdrawn. To be sure, the relapse rates in such patients were higher in Viguera *et al.* 1998, but this result was not statistically significant, and Moncrieff provides no further evidence for her assumption.

4.4 “Active placebo”

There are three aspects of the “active placebo” argument I have discussed: the correlation between side effects and response, the effectiveness of drugs other than “antidepressants” in treating depression, and the evidence from antidepressant trials utilizing “active placebos”. How do they stand up to scrutiny?

Kirsch maintains that he cannot think of any other reason beside unblinding that would explain the correlation between such unwanted side-effects of antidepressant use as gastrointestinal distress and insomnia and alleviation of depression.³³⁰ This is disingenuous. There is another possible explanation for this correlation that is noted in Greenberg *et al.* 1994, one of the studies that Kirsch utilizes in making his case. The patients who are predisposed to respond to the therapeutic effects of the drugs might be similarly physiologically predisposed to experience side effects.³³¹ Kirsch ignores this possibility, even though it is as compatible with the data as the unblinding hypothesis is.

³²⁸ Moncrieff 2008, 150.

³²⁹ *ibid.*, 150; Viguera, Baldessarini & Friedberg 1998, 302. To be precise: “pharmacological stress” is Moncrieff’s term. Viguera, Baldessarini & Friedberg merely refer to “drug continuation itself”.

³³⁰ Kirsch 2009, 18.

³³¹ Greenberg, Bornstein & Zborowski *et al.* 1994, 550.

Kirsch also makes a strong claim about the correlation between response to medication and side effects being almost perfect.³³² There is evidence of such a correlation, but Kirsch goes too far in stating that the correlation is almost perfect. Since the implications are so significant, such a claim should be backed up with a wealth of evidence. However, Kirsch gives only a single reference for his claim, the same Greenberg *et al.* 1994 study of one antidepressant, fluoxetine.³³³

The Rabkin *et al.* 1986 study is crucial to the “active placebo” argument, as it demonstrated that the ability of doctors and patients to guess whether patients have been assigned to drug or placebo treatment in a double-blind trial exceeds chance. This evidence is important for assessing such trials in general, but Kirsch ignores the full implications of the study for the particular argument that he is making.

Rabkin *et al.* 1986 not only investigated the ability to guess, but also some factors that correlated with this ability. They found that for patients, assignment to active drug treatment and treatment response were associated with greater accuracy in guessing.³³⁴ Doctors guessed more accurately for patients that did not respond to placebo than for placebo responders, and for those that responded to imipramine than those who did not.³³⁵

Crucially for Kirsch’s argument, Rabkin *et al.* 1986 also investigated the association between side effects and guessing accuracy. Data on these was available for only 57% of patients. Events were considered side effects only if they had not been reported at the start of the trial.³³⁶ The authors did not find any evidence that side effects had an impact on patients’ guesses of their assignment to active drug or placebo.³³⁷ Kirsch ignores this. The data on side effects in this trial is limited, but this is the one trial that actually investigated the crucial question for the “active placebo” argument: whether the premise that more frequent side effects lead patients to break blind is correct. And the data does not support the Kirsch-Moncrieff argument that more frequent side effects in active drug groups enable patients to guess they are receiving an active drug, thus inflating their expectations.³³⁸

³³² Kirsch 2009, 18.

³³³ *ibid.*, 18.

³³⁴ Rabkin, Markowitz & Stewart *et al.* 1986, 79.

³³⁵ *ibid.*, 81.

³³⁶ *ibid.*, 80.

³³⁷ *ibid.*, 81.

³³⁸ In fact, the results were exactly the opposite to what would be expected on the basis of the Kirsch-Moncrieff hypothesis, but it should be stressed that they were not statistically significant, so we can only draw the conclusion that the results in no way support the critics’ argument. Patients who were taking an active drug, but who guessed they were taking placebo, reported *more* side effects than those who were correct in guessing they were taking an active drug. Likewise, patients taking placebo who guessed incorrectly did not experience significantly more side effects than those who correctly guessed they were in the placebo arm of the trial. *ibid.*, 81.

The second aspect of the “active placebo” argument that I discussed was the evidence from using “active” rather than inert placebos as comparators in antidepressant trials. The idea here was that the use of “placebos” that produce side effects should protect blinding, and therefore trials utilizing such a design would yield more modest results for antidepressant efficacy.

The one study that Moncrieff presents as evidence for this argument, the 1998 Moncrieff, Wessely & Hardy meta-analysis, is problematic for a number of reasons. Firstly, the trials are old. The newest one was published in 1984, the others between 1961-1975. All trials are of TCAs.³³⁹ As Moncrieff herself points out, the trials are of questionable quality.³⁴⁰

Most importantly, it is dubious whether one can draw implications in support of the Kirsch-Moncrieff “active placebo” argument from this study. To really make the argument, one should be able to demonstrate that the premise of the argument is correct. The premise is that experiencing side effects leads people to guess they are in the drug group. Both Moncrieff and Kirsch seem to treat this premise as obviously true, but it is not. I have already shown how the data from the one trial that explicitly addressed this issue, Rabkin *et al.* 1986, in no way lends credence to this premise.

The Moncrieff, Wessely & Hardy 1998 meta-analysis does not provide any evidence for this premise. Moncrieff makes much of the fact that in one trial, more side effects were reported in the drug group.³⁴¹ However, there is no data on whether there was any correlation between side effects and guessing; guessing treatment group was not studied. Furthermore, Moncrieff fails to mention in *The Myth of the Chemical Cure* that even though there were two trials that did assess guessing treatment group, and the guesses, according to her, exceeded chance³⁴², this result was not statistically significant.³⁴³ Not reaching statistical significance here means, by definition, that the guesses did not exceed chance.³⁴⁴ Patients’ guesses were not even assessed.

Both Kirsch and Moncrieff attempt to bolster their “active placebo” argument by maintaining that drugs other than conventional antidepressants are as effective in treating depression as SSRIs, SNRIs, and TCAs. It is not possible to delve into the pharmacology of these drugs here, or to survey the literature for evidence of their effectiveness in depression, but I want to offer a couple of tentative criticisms of their conclusions.

As regards lithium, it does not seem surprising that it was superior to placebo, given that it is indicated for bipolar illness, and that it is used as an adjunct for treatment of treatment-resistant depression. Perhaps lithium’s mood-stabilizing effect caused an amelioration of depressive

³³⁹ Moncrieff, Wessely & Hardy 1998, 228.

³⁴⁰ Moncrieff 2008, 147.

³⁴¹ Moncrieff 2008, 147-148.

³⁴² *ibid.*, 147.

³⁴³ Moncrieff, Wessely & Hardy 1998, 230.

³⁴⁴ I would like to thank Samuli Kangaslampi for making this point.

symptoms. Whether lithium should be called an antidepressant or merely a “mood stabilizer” is another question; I cannot consider this here. The point is merely to point out that there are other possible, perhaps even reasonable explanations of the data besides the Kirsch-Moncrieff hypothesis.

Kirsch’s argument here is based on the assumption that drugs such as the SSRIs are considered antidepressants because mainstream psychiatry maintains that they have a specific effect on the underlying pathophysiology of depression – they affect serotonin levels in the synapse – and that by this criterion, lithium is not an antidepressant, because it does not affect serotonin.

It is also not surprising that barbiturates and benzodiazepines would be superior to placebo for the treatment of depression. They might produce pleasant effects in patients, which might be experienced as an alleviation of depression, even if they had no specific effect on the monoamine system. Moncrieff’s reference to a trial showing alprazolam superior to placebo is particularly striking. Given how addictive alprazolam is known to be, it would be quite surprising if it did not produce pleasant effects which patients found more effective in alleviating depression than placebo. The same can be said for buprenorphine, another drug that Moncrieff refers to in this context.

The concept of an antidepressant is intimately linked with how depression is defined. My discussion here is necessarily a mere sketch, since a deeper conceptual analysis of how depression should be understood is beyond the scope of this essay.

Conclusion

On the basis of the evidence discussed in this essay, the statistical superiority of antidepressants over placebo is not in doubt. In all the meta-analyses cited here, and in almost all individual trials, active drugs fare better than placebo. It would be difficult to disagree with the guidelines' claim that the effect size of antidepressants is consistent.

However, it is also difficult to dispute the critics' general placebo argument that the superiority of drug over placebo is modest. In the evidence cited here, patients on drug treatment tend to improve by a few more points on the various scales compared to those taking placebo.

This makes the issue of clinical relevance crucial. The critics argue that though the drugs show statistical superiority, this is of doubtful relevance clinically. They refer to NICE criteria on clinical relevance in making this point. Unfortunately, I was unable to access the full NICE guideline, so it is not possible to assess how NICE came to define a particular cut off point for relevance. The critics themselves do not present any evidence on why relevance should be defined in a particular way. It is curious that since both Moncrieff and Kirsch repeatedly refer to the NICE criteria, they give no indication of why they consider the criteria so authoritative. We are left with anecdotal argument and pure assertion.

The guidelines also refer to clinical relevance. They argue that antidepressants have clinically relevant effectiveness even in mild depression. However, we have seen that the guidelines make dubious claims about mild depression based on meta-analyses that do not study mild depression. Furthermore, the guidelines do not directly comment on the clinical relevance of antidepressants in more severe depression. They merely assert that antidepressant treatment becomes more important as the severity of depression grows. The guidelines do not give any definition of clinical relevance. On this crucial issue, we are left in the lurch by both sides of the debate.

We have seen how the critics attempt to explain the modest superiority of drug over placebo with the "active placebo" argument. Both Kirsch and Moncrieff maintain that the greater incidence of side effects in active drug treatment groups unblinds patients, and thus changes their expectations of recovery. However, we have seen that the critics are unable to prove this argument. Especially for Kirsch, the Rabkin *et al.* 1986 study is central to the argument. We have seen that this study lends no weight to the unblinding hypothesis. On the basis of the material the critics present, it is neither possible to prove or disprove the hypothesis. Another hypothesis – that patients who are predisposed to benefit from drug treatment are also predisposed to experience unwanted effects – is at least as compatible with the evidence as the Kirsch-Moncrieff unblinding hypothesis.

Whereas Kirsch focuses on acute treatment, Moncrieff also argues that antidepressants are ineffective for long-term maintenance treatment. The guidelines maintain that antidepressants are

clearly superior to placebo for preventing relapse. On this issue, it is not possible to come to a definitive conclusion based on the evidence cited here. It is possible that the superiority of drug over placebo seen in the maintenance studies cited by the guidelines is a true drug effect. However, Moncrieff's argument that greater rate of relapse for patients taking placebo is due to withdrawal effects is also compatible with the data.

Moncrieff argues that even though antidepressants are widely prescribed today, outcomes for depressed patients remain poor.³⁴⁵ A fundamental problem with the literature reviewed here is that it is almost completely focused on depression scores on a particular rating scale, *i.e.* on what symptoms patients report, or on assessments made by doctors, rather than on how depressed individuals do in real life. In their discussion of antidepressant effectiveness, the guidelines do not include a single study on how antidepressant treatment affects any real-world outcome, such as ability to maintain or resume employment. To be sure, in the section giving guidance on suicidal behavior, they do include one reference to a study showing a correlation between the rise in antidepressant prescriptions and a fall in the suicide rate.³⁴⁶

Moncrieff critically discusses some of the evidence on suicide in the *Myth of the Chemical Cure*.³⁴⁷ She also points out some studies suggesting that episodes of depression last longer for patients taking antidepressants, and that people taking medication are less likely to return to work.³⁴⁸ Assessing these arguments is left for further study, but I would suggest that such outcomes, rather than an arbitrary cut-off point on a rating scale assessing symptoms, would be central for a meaningful definition of clinical relevance.

The individual clinician facing the individual patient suffering from depression has to make the decision whether to prescribe. As the guidelines point out, placebo is a tool used in clinical trials, not a treatment option in clinical practice.³⁴⁹ To be sure, there are a variety of other treatment options, such as psychotherapy, which are not covered in this essay. But considering only the simplified choice of whether to prescribe an antidepressant, I believe that the evidence covered here overall supports the use of drug treatment, if the patient is motivated to try it. Patient motivation is, of course, important in virtually any kind of medical treatment, but perhaps here it is more important still, since the weighing of benefits and risks is perhaps trickier than with some common treatments in other fields of medicine. Average response to drug, while better than average placebo response, is modest. The drugs have unwanted effects, and patients may experience withdrawal effects from prolonged use. It is also possible, though by no means proven, that there are subgroups

³⁴⁵ Moncrieff 2008, 150.

³⁴⁶ Korkeila J, Salminen JK, Hiekkanen H *et al.* Use of antidepressants and suicide rate in Finland: an ecological study. *J Clin Psychiatry* 2007; 68:505-511.

³⁴⁷ Moncrieff 2008, 151-152.

³⁴⁸ *ibid.*, 151.

³⁴⁹ Duodecim 2016.

of patients who are both predisposed to respond better to drugs and experience more unwanted effects.

We have seen that there is a consistent statistical superiority of drugs over placebo. This represents the average, so the individual may respond better – or, of course, worse. Given all of this, an antidepressant can still make a difference for the individual patient, even if most of their response is due to non-specific effects.

Both Kirsch and Moncrieff employ the language of myth in making their case. Moncrieff maintains that psychiatry is beset by the “myth of the chemical cure” – the myth that psychiatric disorders are primarily biochemical imbalances to be corrected by applying chemicals. Kirsch called his book *The Emperor’s New Drugs*, implying that underneath all the pomp and circumstance of drug marketing, the emperor has no clothes – the drugs do not do much. Both seem to imply that a myth means an untruth.

A myth is not true in the strictest sense, and may be used for malicious purposes, but a myth can also embody something essential. It can perhaps distill an important truth into a more memorable or understandable form. In the case of psychiatry, there is little doubt that the “myth of the chemical cure” has been misused in marketing antidepressants. But perhaps it contains a kernel of truth – not that psychiatric disorders are reducible to imbalances in brain chemicals, but that biochemical treatments do have a role in helping people afflicted by such maladies.

Bibliography

- Angst J, Scheidegger P, Stabl M. Efficacy of moclobemide in different patient groups. Results of new subscales of the Hamilton Depression Rating Scale. *Clin. Neuropharmacol.* 1993; 16: S55-S62.
- Bech P, Cialdella P, Haugh MC *et al.* 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.
- Bielski RJ & Friedel RO. Prediction of tricyclic antidepressant response: a critical review. *Arch. Gen. Psychiatry* 1976; 33: 1479-1489.
- Depression. Current Care Guidelines. Working group set up by the Finnish Medical Society Duodecim and the Finnish Psychiatric Association. Helsinki: The Finnish Medical Society Duodecim, 2016.
- Dilsaver SC, Greden JF, Snider RM. Antidepressant withdrawal syndromes: phenomenology and pathophysiology. *Int. Clin. Psychopharmacol.* 1987; 2: 1-19.
- Elkin I, Shea MT, Watkins JT *et al.* National Institute of Mental Health Treatment of Depression Collaborative Research Program. General Effectiveness of Treatments. *Arch. Gen. Psychiatry* 1989; 46: 971-982.
- Fournier JC, DeRubeis RJ, Hollon SD *et al.* Antidepressant drug effects and depression severity: A patient-level meta-analysis. *JAMA* 2010; 303: 47-53.
- Geddes JR, Carney SM, Davies C *et al.* Relapse prevention with antidepressant drug treatment in depressive disorders: a systematic review. *The Lancet* 2003; 361: 653-661.
- Gibbons RD, Hur K, Brown CH *et al.* 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.
- Glue P, Donovan MR, Kolluri S *et al.* Meta-analysis of relapse prevention antidepressant trials in depressive disorders. *Australian and New Zealand Journal of Psychiatry* 2010; 44: 697-705.
- Greenberg RP, Bornstein RF, Zborowski MJ *et al.* A meta-analysis of fluoxetine outcome in the treatment of depression. *Journal of Nervous and Mental Disease* 1994; 182: 547-551.
- Hansen RA, Gaynes BN, Thieda P *et al.* Prevention of major depressive disorder relapse and recurrence with second-generation antidepressants: A systematic review and meta-analysis. *Psychiatr. Serv.* 2008; 59: 1121-1130.
- Hindmarch I, Kimber S, Cockle SM. Abrupt and brief discontinuation of antidepressant treatment: effects on cognitive function and psychomotor performance. *Int. Clin. Psychopharmacol.* 2000; 15: 305-318.
- Jakubovsky E, Varigonda AL, Freemantle N *et al.* Systematic Review and Meta-analysis: Dose-Response Relationship of Selective Serotonin Reuptake Inhibitors in Major Depressive Disorder. *Am. J. Psychiatry* 2016; 173: 174-183.

- Joyce PR & Paykel ES. Predictors of drug response in depression. *Arch. Gen. Psychiatry* 1989; 46: 89-99.
- Jääskeläinen E & Miettunen J 2011. Psykiatriset arviointiasteikot kliinisessä työssä. *Lääketieteellinen Aikakauskirja Duodecim* 2011; 127:1719-1725.
- Kessler RC & Bromet EJ. The Epidemiology of Depression Across Cultures. *Annu. Rev. Public Health* 2013; 34:119-138.
- Khan A, Leventhal RM, Khan SR *et al.* 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.
- Khan A; Warner HA, Brown WA. Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials: an analysis of the Food and Drug Administration database. *Arch. Gen. Psychiatry* 2000; 57: 311-317.
- Kirsch I. *The Emperor's New Drugs: Exploding the Antidepressant Myth.* London: The Bodley Head 2009.
- Kirsch I, Deacon BJ, Huedo-Medina TB *et al.* Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med* 2008; 5: 260-268.
- Kirsch I, Moore TJ, Scoboria A *et al.* The Emperor's New Drugs: An analysis of antidepressant medication data submitted to the U.S. Food and Drug Administration. *Prevention & Treatment* 2002.
- Kirsch I & Sapirstein G. Listening to Prozac but hearing placebo: A meta-analysis of antidepressant medication. *Prevention & Treatment* 1998.
- Kocsis JH, Croughan JL, Katz MM *et al.* Response to treatment with antidepressants of patients with severe or moderate nonpsychotic depression and of patients with psychotic depression. *Am. J. Psychiatry* 1990; 147: 621-624.
- Korkeila J, Salminen JK, Hiekkanen H *et al.* Use of antidepressants and suicide rate in Finland: an ecological study. *J Clin Psychiatry* 2007; 68:505-511.
- Lönnqvist J, Henriksson M, Marttunen M *et al.* *Psykiatria.* Helsinki: Kustannus Oy Duodecim 2011.
- Malt UF, Robak OH, Madsbu HP *et al.* The Norwegian naturalistic treatment study of depression in general practice (NORDEP)-I: randomised double blind study. *Br. Med. J.* 1999; 318: 1180-1184.
- Medical Research Council. Clinical trial of the treatment of depressive illness. *Br. Med. J.* 1965; 1: 881-886.
- Melander H, Salmonson T, Abadie E *et al.* A regulatory Apologia - A review of placebo-controlled studies in regulatory submissions of new-generation antidepressants. *European Neuropsychopharmacology* 2008; 18: 623-627.
- Moncrieff J. *The Myth of the Chemical Cure: A Critique of Psychiatric Drug Treatment.* Houndmills: Palgrave Macmillan 2008.

Moncrieff J, Wessely S, Hardy R. Meta-analysis of trials comparing antidepressants with active placebos. *Br. J. Psychiatry* 1998; 172: 227-231.

Moncrieff J, Wessely S & Hardy R. Active placebos versus antidepressants for depression (Review). *Cochrane Database of Systematic Reviews* 2004; 1: 1-31.

National Institute for Clinical Excellence. Depression: Management of Depression in Primary and Secondary Care. Clinical Practice Guideline Number 23. National Institute for Clinical Excellence 2004.

Pae C-U, Wang S-M, Han C *et al.* Vortioxetine: a meta-analysis of 12 short-term, randomized, placebo-controlled clinical trials for the treatment of major depressive disorder. *J. Psychiatry Neurosci.* 2015; 40: 174-186.

Paykel ES, Hollyman JA, Freeling P *et al.* Predictors of therapeutic benefit from amitriptyline in mild depression: a general practice placebo-controlled trial. *J. Affect. Disord.* 1988; 14: 83-95.

Phillipp M, Kohlen R, Hiller KO. Hypericum extract versus imipramine or placebo in patients with moderate depression: randomised multicentre study of treatment for eight weeks. *Br. Med. J.* 1999; 319: 1534-1538.

Rabkin JG, Markowitz JS, Stewart J *et al.* How blind is blind? Assessment of patient and doctor medication guesses in a placebo-controlled trial of imipramine and phenelzine. *Psychiatry Research* 1986; 19: 75-86.

Raskin A, Schulterbrandt JG, Reatig N *et al.* Differential response to chlorpromazine, imipramine, and placebo. A study of subgroups of hospitalized depressed patients. *Arch. Gen. Psychiatry* 1970; 23: 164-173.

Rogers SC & Clay PM. A statistical review of controlled trials of imipramine and placebo in the treatment of depressive illnesses. *Br. J. Psychiatry* 1975; 127: 599-603.

Sachs GS, Nierenberg AA, Calabrese JR *et al.* Effectiveness of adjunctive antidepressant treatment for bipolar depression. *N. Engl. J. Med.* 2007; 356: 1711-1722.

Sneed JR, Rutherford BR, Rindskopf D *et al.* Design makes a difference: a meta-analysis of antidepressant response rates in placebo-controlled versus comparator trials in late-life depression. *American Journal of Geriatric Psychiatry* 2008; 16: 65-73.

Smith A, Traganza E, Harrison G. Studies on the effectiveness of antidepressant drugs. *Psychopharmacol. Bull.* 1969; Suppl-53.

Storosum JG, Elferink AJ, van Zwieten BJ *et al.* Short-term efficacy of tricyclic antidepressants revisited: a meta-analytic study. *Eur Neuropsychopharmacol* 2001; 11: 173-180.

Storosum JG, van Zwieten BJ, van den BW *et al.* Suicide risk in placebo-controlled studies of major depression. *Am. J. Psychiatry* 2001; 158: 1271-1275.

Taylor D, Sparshatt A, Varma S *et al.* Antidepressant efficacy of agomelatine: meta-analysis of published and unpublished studies. *BMJ* 2014; 348:g1888.

Turner EH, Matthews AM, Linardatos E et al. Selective publication of antidepressant trials and its influence on apparent efficacy. *The New England Journal of Medicine* 2008; 358: 252-260.

Undurraga J & Baldessarini RJ. Randomized, placebo-controlled trials of antidepressants for acute major depression: Thirty-year meta-analytic review. *Neuropsychopharmacology* 2012; 37: 851-864.

Viguera AC, Baldessarini RJ, Friedberg J. et al. Discontinuing antidepressant treatment in major depression. *Harv. Rev. Psychiatry* 1998; 5: 293-306.

Walsh TB, Seidman SN, Sysko R et al. Placebo Response in Studies of Major Depression: Variable, Substantial, and Growing. *JAMA* 2002; 14: 1840-1847.

Williams N, Simpson AN, Simpson K et al. Relapse rates with long-term antidepressant drug therapy: a meta-analysis. *Hum. Psychopharmacol. Clin. Exp.* 2009; 24: 401-408.

Wilson E & Lader Malcolm 2015. A review of the management of antidepressant discontinuation symptoms. *Therapeutic Advances in Psychopharmacology* 2015; 5: 357-368.