



# Statistically Significant Antidepressant-Placebo Differences on Subjective Symptom-Rating Scales Do Not Prove That the Drugs Work: Effect Size and Method Bias Matter!

Michael P. Hengartner<sup>1\*</sup> and Martin Plöderl<sup>2</sup>

<sup>1</sup> Department of Applied Psychology, Zurich University of Applied Sciences, Zurich, Switzerland, <sup>2</sup> Department for Crisis Intervention and Suicide Prevention and Department for Clinical Psychology, University Clinic for Psychiatry, Psychotherapy, and Psychosomatics, Paracelsus Medical University, Salzburg, Austria

**Keywords:** antidepressant, meta-analysis, efficacy, effectiveness, effect size, clinical significance, method bias

## OPEN ACCESS

### Edited by:

Stefan Borgwardt,  
Universität Basel, Switzerland

### Reviewed by:

Bertus F. Jeronimus,  
University of Groningen, Netherlands  
Stefan Weinmann,  
Vivantes Klinikum, Germany  
Glen Spielmans,  
Metropolitan State University,  
United States

### \*Correspondence:

Michael P. Hengartner  
michaelpascal.hengartner@zhaw.ch

### Specialty section:

This article was submitted to  
Public Mental Health,  
a section of the journal  
Frontiers in Psychiatry

**Received:** 13 March 2018

**Accepted:** 01 October 2018

**Published:** 17 October 2018

### Citation:

Hengartner MP and Plöderl M (2018)  
Statistically Significant  
Antidepressant-Placebo Differences  
on Subjective Symptom-Rating Scales  
Do Not Prove That the Drugs Work:  
Effect Size and Method Bias Matter!  
*Front. Psychiatry* 9:517.  
doi: 10.3389/fpsy.2018.00517

Following the publication of a recent meta-analysis by Cipriani et al. (1), various opinion leaders and news reports claimed that the effectiveness of antidepressants has been definitely proven (2). E.g., Dr. Pariente, spokesperson for the Royal College of Psychiatrists, stated that this study “finally puts to bed the controversy on antidepressants, clearly showing that these drugs do work in lifting mood and helping most people with depression” (<https://www.theguardian.com/science/2018/feb/21/the-drugs-do-work-antidepressants-are-effective-study-shows>). We surely would embrace drug treatments that effectively help most people with depression, but based on work that has contested the validity of mostly industry-sponsored antidepressant trials (3–6) we remain skeptical about antidepressants’ clinical benefits. The most recent meta-analysis indeed concludes that antidepressants are more effective than placebo but also acknowledges that risk of bias was substantial and that the mean effect size of  $d = 0.3$  was modest (1). Unfortunately, no clarification is given what this effect size means and whether it can be expected to be clinically significant in real-world routine practice. In this opinion paper we therefore ponder over how the reported effect size of  $d = 0.3$  relates to clinical significance and how method bias undermines its validity, in order that the public, clinicians, and patients can judge for themselves whether antidepressants clearly work in most people with depression.

## STATISTICAL VS. CLINICAL SIGNIFICANCE

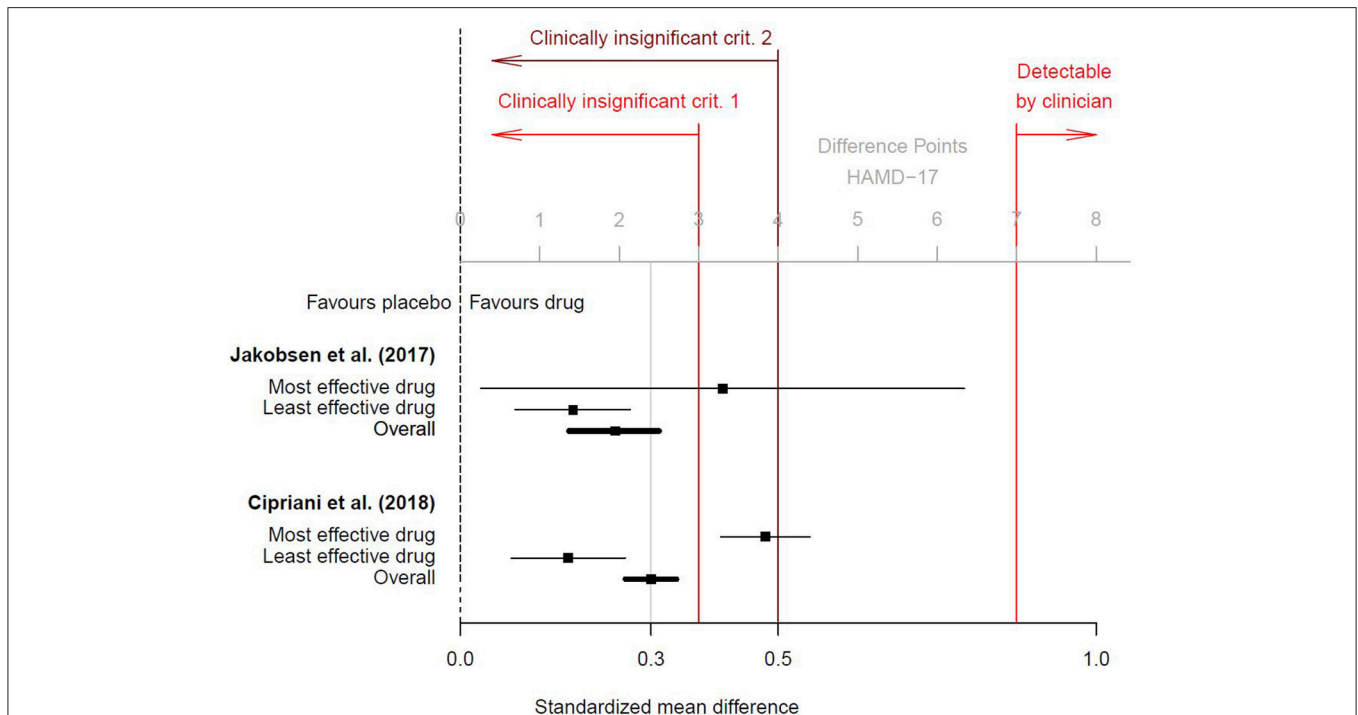
Based on statistically significant drug-placebo differences, authors commonly conclude that antidepressants are effective regardless of the clinical significance of effect sizes. Cipriani et al. (7) even complained that there was “an undue focus on the binary and polarizing question of clinical significance” (p. 462). However, statisticians repeatedly cautioned that statistical significance does not imply practical relevance (8–10). A statistically significant result neither proves that the null hypothesis is false nor that the alternative hypothesis is true (8, 9, 11). Interpreting a statistically significant drug-placebo difference as evidence that drugs work is therefore a logical fallacy (12). The null hypothesis is always false, as a true null-association between natural variables (i.e.,  $d = 0.0$ ) is nearly impossible due to residual confounding and correlational noise (8, 9). The American Statistical Association (10) formally states that “A  $p$ -value, or statistical significance, does not measure the size of an effect or the importance of a result” and they further emphasize that

“Any effect, no matter how tiny, can produce a small  $p$ -value if the sample size or measurement precision is high enough ...” (p. 132). With a total sample size of  $n = 116,477$  as in the most recent meta-analysis (1), it is therefore not surprising that any given drug-placebo difference, however small it may be, reaches statistical significance. Thus, since statistical significance does not imply clinical significance (10, 12, 13), readers need to consider what the reported mean effect of  $d = 0.3$  practically means.

As shown in **Figure 1**, this effect size corresponds to approximately 2 points on the Hamilton Rating-Scale for Depression 17-item version (HAMD-17; range 0–52 points), but per convention a difference  $<3$  points or an effect size  $d < 0.5$  (corresponding to  $<4$  HAMD-17 points) are considered clinically irrelevant (14, 15). Research suggests that drug-placebo differences  $<3$  points are undetectable by clinicians and that at least 7 HAMD-17 points are necessary for a clinician to detect a minimal improvement in a patient’s clinical presentation (16). As a result, the average treatment effect of  $d = 0.3$  must be considered undetectable and therefore clinically insignificant in real-world routine practice. Interestingly, a previous meta-analysis by Jakobsen et al. (14) found comparable effect sizes, but the authors defined clinical significance a-priori and therefore questioned the real-world benefits of antidepressants. The effect sizes reported by Cipriani et al. (1) and Jakobsen et al. (14) are plotted in **Figure 1**.

Here we report Cohen’s  $d$  effect sizes for the sake of completeness and because they are often reported in meta-analyses. However, we emphasize that cut-offs such as  $d = 0.2$  (“small” effect size) or  $d = 0.5$  (“medium” effect size) are arbitrary and should be interpreted with caution (17). Cohen’s  $d$  is calculated as the mean HAMD-17 difference between treatment groups divided by their pooled standard deviation. When samples are homogeneous and inter-individual variability is low, then the standard deviation is small. All things being equal, the smaller the standard deviation, the larger Cohen’s  $d$ . E.g., a group difference of 2 HAMD-17 points will yield an effect size of  $d = 0.4$  when the pooled standard deviation is 5 ( $2/5 = 0.4$ ), but only an effect size of  $d = 0.2$  when the pooled standard deviation is 10 ( $2/10 = 0.2$ ). The clinical significance of Cohen’s  $d$  further depends on the outcome. A  $d = 0.3$  referring to mortality necessarily has more practical relevance than  $d = 0.3$  based on subjective (and often transient) symptom ratings.

When based on approximately normally-distributed interval scales,  $d = 0.3$  indicates that, first, the outcome of antidepressants and placebo overlap by 88%, second, that only 62% of participants in the antidepressant group score above the mean of the placebo group and, conversely, 38% score below the mean (referred to as Cohen’s  $U_3$ ), and, third, that if you pick a person at random from the antidepressant group, he/she will have a minor chance of 58% to have the better outcome than a person picked at random



**FIGURE 1** | Clinical significance of antidepressants, based on the results of Cipriani et al. (1); additional online information (p. 150) and of Jakobsen et al. (14). Black squares are the standardized mean differences  $d$  (drug vs. placebo) for the most and least effective drug and for the overall effect. Horizontal lines are the related 95% confidence intervals. Two conventions for clinical insignificance were used. Criterion 1 was a difference of  $<3$  points on the HAMD-17 scale (corresponding to  $d < 0.4$ ), and criterion 2 was  $d < 0.5$ . Only differences of at least 7 points on the HAMD-17 scale were found to be detectable by clinicians (16). To transform standardized mean differences into mean point-differences on the HAMD-17 (or vice versa), we assumed a pooled standard deviation of  $SD = 8.0$ , as suggested by Moncrieff and Kirsch (16) and which conforms to data provided in the online appendix by Cipriani et al. (1).

from the placebo group (probability of 50% indicates no benefit at all) (17). Finally, assuming a placebo response rate of 35–40% in moderate-to-severe depression (18), based on the Furukawa formula (19), the number needed to treat (NNT) is approximately 9 [see also (20), who calculated a NNT of 8–10 based on the results reported by (1)]. This indicates that, relative to placebo, 9 patients need to undergo antidepressant pharmacotherapy for 1 patient to benefit. In consequence, 8 of 9 patients would equally benefit from an inert placebo pill without risk to eventually suffer from adverse pharmacologic effects (14, 21) and debilitating withdrawal symptoms upon discontinuation of drug treatment (22, 23). A brief synopsis of these findings is that antidepressants might work in a small minority of patients who do not benefit from placebo [see also (24)], but for the vast majority an inert placebo pill that conveys no health risks would work just as well.

## ADDRESSING COMMON OBJECTIONS

A frequently cited paper by Leucht et al. (25) claims that the effect of antidepressants is comparable to that of other medications in general medicine, but note that several general medicine drugs have effect sizes  $d > 0.8$ , whereas the effect size of antidepressants is  $d = 0.3$ . Moreover, the general medicine drugs with small effect sizes reported in Leucht et al. (25) were mostly based on objective, severe clinical outcomes such as mortality or cardiovascular events (i.e., “hard” outcomes). Efficacy of antidepressants, in contrast, is exclusively based on subjective symptom ratings (i.e., “soft” outcomes). To provide a fair comparison of the efficacy of antidepressants and general medicine drugs, researchers should base the effect size of antidepressants likewise on a severe clinical outcome such as for instance (fatal) suicide attempts. In that case the effect size of antidepressants would be close to zero and favoring placebo (26–30). This compares very unfavorably to most general medicine drugs.

Another unsubstantiated objection is that the efficacy of antidepressants is poor due to inadequate psychometric properties of the HAMD-17 [e.g., its poor content validity (31)]. We do not intend to defend the validity of the HAMD-17, but instead we want to stress that when the efficacy of antidepressants relies on other outcome measures, effect sizes are not higher. First, when efficacy is based on patient self-reports such as the Beck Depression Inventory (BDI), mean effect sizes are even smaller (i.e.,  $d < 0.3$ ) than those based on the HAMD-17 (32, 33). Second, a meta-analysis of all escitalopram trials sponsored by Forest and Lundbeck, which applied the Montgomery-Asberg Depression Rating Scale (MADRS), produced a mean effect size of  $d = 0.32$  (24). Third, there is no evidence from clinical trials that antidepressants work when efficacy is based on severe clinical outcomes such as suicide attempts (26–30). I.e., the HAMD-17 is not accountable for antidepressants’ poor efficacy.

A third objection is that critics of antidepressants unjustifiably promote psychotherapy although talk therapy is no better than pharmacotherapy. In response to these concerns we would like to state that we have also written about the limitations and biases in psychotherapy research (34). We further agree that in the

short-term (i.e., acute treatment), the outcome of psychotherapy and pharmacotherapy is comparable (35). Cuijpers and Cristea (36), two prominent psychotherapy researchers, proposed that enhanced placebo effects could explain the short-term outcome of both pharmacotherapy and psychotherapy. Nevertheless, in the long-term, psychotherapy conveys less physical health risks and its effect on depression (i.e., sustained remission and relapse prevention) appears to be superior to pharmacotherapy according to several meta-analyses of direct comparisons (37–39).

## THE EFFICACY OF ANTIDEPRESSANTS IS OVERESTIMATED

The average treatment effect detailed above, albeit minor, yet is most likely an overestimation due to various systematic biases that inflate the apparent efficacy of antidepressants, including, in particular, unblinding of outcome assessors (3, 36, 40). Treatment effects in antidepressant trials are commonly rated by clinicians who can identify with high accuracy which patients receive the active drug and which inert placebo based on the reporting, or a suspicious lack thereof, of recognizable side effects such as nausea or dry mouth (36, 41). Several lines of evidence suggest that drug-placebo differences might be inflated when efficacy estimates are based on subjective symptom rating-scales such as the HAMD-17.

First, it has consistently been shown that treatment effects are larger when the outcome is rated by unblinded assessors, thus efficacy estimates are inflated due to assessors’ treatment expectancies (42–44). Second, when active placebos that mimic common antidepressant side effects are applied instead of inert placebos, the estimated treatment effects are substantially smaller because assessors are more effectively blinded (45). Third, antidepressants’ efficacy has been shown to be substantially smaller when estimates are based on patients’ self-reported depression symptoms instead of observer-ratings (32, 33), suggesting that patients do not perceive the same benefit as (unblinded) clinicians attribute to the drugs. Fourth, with respect to dropouts due to any reason, which is regarded as an objective measure of real-world effectiveness (46), antidepressants are, on average, not superior to placebo (1, 47). Finally, fifth, evidence for assessor bias was also shown in the most recent meta-analysis, where antidepressants were judged more efficacious when they were novel as compared to when they were older (1). Since a drug does not lose its pharmacologic effect simply because it has been on the market for a few years, this is evidence for a systematic overestimation of novel drugs due to clinicians’ treatment expectancies.

Given that the mean drug-placebo difference is only about 2 HAMD-17 points, even a minor bias in symptom-ratings could fully account for antidepressants’ treatment effect. Indeed, taking the observer bias into account, Gotzsche (48) calculated that the effect of antidepressants, relative to placebo, is virtually zero (OR = 1.02). Note that there are many more systematic biases than unblinding of outcome assessors that we did not consider here. These include, for instance, the selective inclusion

of participants (patients who are known to preferably respond to the experimental drug are included in the trials, while none-responders and patients who experienced bothersome side effects prior to the actual trial are excluded), patient expectancy bias (patients believe that the drugs work, thus producing an enhanced placebo response which takes effect as soon as a patient realizes that he/she receives the active drug), inadequate management of missing data (the common procedure of “last observation carried forward” produces inflated efficacy estimates), and outcome reporting bias (quite often only results for the most convenient outcome are reported and interpreted) (3, 49, 50).

## CONCLUSIONS

Contrary to the predominant interpretation we contend that antidepressants do not work in most patients, given that only 1 of 9 people benefit, whereas the remaining 8 are unnecessarily put at risk of adverse drug effects. To be clear, antidepressants can have strong mental and physical effects in some patients that may be considered helpful for some time (51), but there is no

evidence that the drugs can cure depression (3, 40, 48). Insomnia, fatigue, loss of appetite, psychomotor agitation, and suicidal acts are recognized depression symptoms (52), but newer-generation antidepressants may cause precisely these symptoms (14, 29, 46, 53). This is not what we would expect from drugs that effectively treat depression. Moreover, emerging evidence from well-controlled long-term pharmacoepidemiologic studies suggests that antidepressants may increase this risk of serious medical conditions (21, 54, 55), including dementia (56), stroke (57), obesity (58), and all-cause mortality (57, 59, 60). Antidepressants may have clinically meaningful short-term benefits in a small minority of patients, but the most recent meta-analytic evidence does not indicate that they work in the majority of patients. A careful re-evaluation of risks and benefits is therefore needed before the controversy about the utility of antidepressants can be put to bed.

## AUTHOR CONTRIBUTIONS

MPH drafted the manuscript. MP contributed significantly in writing and critical revision.

## REFERENCES

- Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet* (2018) 391:1357–66. doi: 10.1016/S0140-6736(17)32802-7
- Adlington K. Pop a million happy pills? Antidepressants, nuance, and the media. *BMJ* (2018) 360:k1069. doi: 10.1136/bmj.k1069
- Hengartner MP. Methodological flaws, conflicts of interest, and scientific fallacies: Implications for the evaluation of antidepressants' efficacy and harm. *Front Psychiatry* (2017) 8:275. doi: 10.3389/fpsy.2017.00275
- Ebrahim S, Bance S, Athale A, Malachowski C, Ioannidis JP. Meta-analyses with industry involvement are massively published and report no caveats for antidepressants. *J Clin Epidemiol.* (2016) 70:155–63. doi: 10.1016/j.jclinepi.2015.08.021
- Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence based medicine—selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ* (2003) 326:1171–3. doi: 10.1136/bmj.326.7400.1171
- Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med.* (2008) 358:252–60. doi: 10.1056/NEJMsa065779
- Cipriani A, Salanti G, Furukawa TA, Egger M, Leucht S, Ruhe HG, et al. Antidepressants might work for people with major depression: where do we go from here? *Lancet Psychiat.* (2018) 5:461–3. doi: 10.1016/S2215-0366(18)30133-0
- Cohen J. The earth is round ( $P < 0.05$ ). *Am Psychol* (1994) 49:997–1003. doi: 10.1037/0003-066x.50.12.1103
- Szucs D, Ioannidis JPA. When null hypothesis significance testing is unsuitable for research: a reassessment. *Front Hum Neurosci.* (2017) 11:390. doi: 10.3389/fnhum.2017.00390
- Wasserstein RL, Lazar NA. The ASA's statement on p-values: context, process, and purpose. *Am Stat.* (2016) 70:129–33. doi: 10.1080/00031305.2016.1154108
- Wagenmakers EJ. A practical solution to the pervasive problems of p values. *Psychon B Rev.* (2007) 14:779–804. doi: 10.3758/Bf03194105
- Hengartner MP. What is the threshold for a clinically minimally important drug effect? *BMJ Evid Based Med.* (2018). doi: 10.1136/bmjebm-2018-111056
- Kirk RE. Practical significance: a concept whose time has come. *Educ Psychol Meas.* (1996) 56:746–59.
- Jakobsen JC, Katakam KK, Schou A, Hellmuth SG, Stallknecht SE, Leth-Moller K, et al. Selective serotonin reuptake inhibitors versus placebo in patients with major depressive disorder. A systematic review with meta-analysis and Trial Sequential Analysis. *BMC Psychiatry* (2017) 17:58. doi: 10.1186/s12888-016-1173-2
- Moncrieff J, Kirsch I. Efficacy of antidepressants in adults. *BMJ* (2005) 331:155–7. doi: 10.1136/bmj.331.7509.155
- Moncrieff J, Kirsch I. Empirically derived criteria cast doubt on the clinical significance of antidepressant-placebo differences. *Contemp Clin Trials* (2015) 43:60–2. doi: 10.1016/j.cct.2015.05.005
- Durlak JA. How to select, calculate, and interpret effect sizes. *J Pediatr Psychol.* (2009) 34:917–28. doi: 10.1093/jpepsy/jsp004
- Furukawa TA, Cipriani A, Atkinson LZ, Leucht S, Ogawa Y, Takeshima N, et al. Placebo response rates in antidepressant trials: a systematic review of published and unpublished double-blind randomised controlled studies. *Lancet Psychiat.* (2016) 3:1059–66. doi: 10.1016/S2215-0366(16)30307-8
- Furukawa TA, Leucht S. How to obtain NNT from Cohen's d: comparison of two methods. *PLoS ONE* (2011) 6:e19070. doi: 10.1371/journal.pone.0019070
- McCormack J, Korownyk C. Effectiveness of antidepressants. *BMJ* (2018) 360:k1073. doi: 10.1136/bmj.k1073
- Carvalho AF, Sharma MS, Brunoni AR, Vieta E, Fava GA. The safety, tolerability and risks associated with the use of newer generation antidepressant drugs: a critical review of the literature. *Psychother Psychosom.* (2016) 85:270–88. doi: 10.1159/000447034
- Fava GA, Gatti A, Belaise C, Guidi J, Offidani E. Withdrawal symptoms after selective serotonin reuptake inhibitor discontinuation: a systematic review. *Psychother Psychosom.* (2015) 84:72–81. doi: 10.1159/000370338
- Fava GA, Benasi G, Lucente M, Offidani E, Cosci F, Guidi J. Withdrawal symptoms after serotonin-noradrenaline reuptake inhibitor discontinuation: systematic review. *Psychother Psychosom.* (2018) 87:195–203. doi: 10.1159/000491524
- Thase ME, Larsen KG, Kennedy SH. Assessing the ‘true’ effect of active antidepressant therapy v. placebo in major depressive disorder: use of a mixture model. *Br J Psychiatry* (2011) 199:501–7. doi: 10.1192/bjp.bp.111.093336

25. Leucht S, Hierl S, Kissling W, Dold M, Davis JM. Putting the efficacy of psychiatric and general medicine medication into perspective: review of meta-analyses. *Br J Psychiatry* (2012) 200:97–106. doi: 10.1192/bjp.bp.111.096594
26. Baldessarini RJ, Lau WK, Sim J, Sum MY, Sim K. Suicidal risks in reports of long-term controlled trials of antidepressants for major depressive disorder II. *Int J Neuropsychopharmacol.* (2017) 20:281–4. doi: 10.1093/ijnp/pyw092
27. Braun C, Bschor T, Franklin J, Baethge C. Suicides and suicide attempts during long-term treatment with antidepressants: a meta-analysis of 29 placebo-controlled studies including 6,934 patients with major depressive disorder. *Psychother Psychosom.* (2016) 85:171–9. doi: 10.1159/000442293
28. Healy D, Whitaker C. Antidepressants and suicide: risk-benefit conundrums. *J Psychiatry Neurosci.* (2003) 28:331–7.
29. Fergusson D, Doucette S, Glass KC, Shapiro S, Healy D, Hebert P, Hutton B. *BMJ* (2005) 330:396. doi: 10.1136/bmj.330.7488.396
30. Stone M, Laughren T, Jones ML, Levenson M, Holland PC, Hughes A, et al. Risk of suicidality in clinical trials of antidepressants in adults: analysis of proprietary data submitted to US Food and Drug Administration. *BMJ* (2009) 339:b2880. doi: 10.1136/bmj.b2880
31. Bagby RM, Ryder AG, Schuller DR, Marshall MB. The Hamilton Depression Rating Scale: has the gold standard become a lead weight? *Am J Psychiatry* (2004) 161:2163–77. doi: 10.1176/appi.ajp.161.12.2163
32. Greenberg RP, Bornstein RF, Greenberg MD, Fisher S. A meta-analysis of antidepressant outcome under “blinder” conditions. *J Consult Clin Psychol.* (1992) 60:664–9.
33. Spielmans GI, Gerwig K. The efficacy of antidepressants on overall well-being and self-reported depression symptom severity in youth: a meta-analysis. *Psychother Psychosom.* (2014) 83:158–64. doi: 10.1159/000356191
34. Hengartner MP. Raising awareness for the replication crisis in clinical psychology by focusing on inconsistencies in psychotherapy research: how much can we rely on published findings from efficacy trials? *Front Psychol* (2018) 9:256. doi: 10.3389/fpsyg.2018.00256
35. Cuijpers P, Sijbrandij M, Koole SL, Andersson G, Beekman AT, Reynolds CF, III. The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: a meta-analysis of direct comparisons. *World Psychiatry* (2013) 12:137–48. doi: 10.1002/wps.20038.
36. Cuijpers P, Cristea IA. What if a placebo effect explained all the activity of depression treatments? *World Psychiatry* (2015) 14:310–1. doi: 10.1002/wps.20249
37. Biesheuvel-Liefveld KE, Kok GD, Bockting CL, Cuijpers P, Hollon SD, van Marwijk HW, et al. Effectiveness of psychological interventions in preventing recurrence of depressive disorder: meta-analysis and meta-regression. *J Affect Disord.* (2015) 174:400–10. doi: 10.1016/j.jad.2014.12.016
38. De Maat S, Dekker J, Schoevers R, De Jonghe F. Relative efficacy of psychotherapy and pharmacotherapy in the treatment of depression: a meta-analysis. *Psychother Res.* (2006) 16:566–78. doi: 10.1080/10503300600756402
39. Spielmans GI, Berman MI, Usitalo AN. Psychotherapy versus second-generation antidepressants in the treatment of depression: a meta-analysis. *J Nerv Ment Dis.* (2011) 199:142–9. doi: 10.1097/NMD.0b013e31820caefb
40. Moncrieff J. Are antidepressants as effective as claimed? No, they are not effective at all. *Can J Psychiatry* (2007) 52:96–7. doi: 10.1177/070674370705200204
41. Even C, Siobud-Dorocant E, Dardennes RM. Critical approach to antidepressant trials. Blindness protection is necessary, feasible and measurable. *Br J Psychiatry* (2000) 177:47–51.
42. Hrobjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Hilden J, Boutron I, et al. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. *BMJ* (2012) 344:e1119. doi: 10.1136/bmj.e1119
43. Khan A, Faucett J, Lichtenberg P, Kirsch I, Brown WA. A systematic review of comparative efficacy of treatments and controls for depression. *PLoS ONE* (2012) 7:e41778. doi: 10.1371/journal.pone.0041778
44. Hrobjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Hilden J, Boutron I, et al. Observer bias in randomized clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors. *CMAJ* (2013) 185:E201–11. doi: 10.1503/cmaj.120744
45. Moncrieff J, Wessely S, Hardy R. Active placebos versus antidepressants for depression. *Cochrane Database Syst Rev.* (2004) 1:CD003012. doi: 10.1002/14651858.CD003012.pub2.
46. Barbui C, Furukawa TA, Cipriani A. Effectiveness of paroxetine in the treatment of acute major depression in adults: a systematic re-examination of published and unpublished data from randomized trials. *CMAJ* (2008) 178:296–305. doi: 10.1503/cmaj.070693
47. Arroll B, Elley CR, Fishman T, Goodyear-Smith FA, Kenealy T, Blashki G, et al. Antidepressants versus placebo for depression in primary care. *Cochrane Database Syst Rev.* (2009) 3:CD007954. doi: 10.1002/14651858.CD007954.
48. Gotzsche PC. Why I think antidepressants cause more harm than good. *Lancet Psychiat.* (2014) 1:104–6. doi: 10.1016/S2215-0366(14)70280-9
49. Wang SM, Han C, Lee SJ, Jun TY, Patkar AA, Masand PS, et al. Efficacy of antidepressants: bias in randomized clinical trials and related issues. *Expert Rev Clin Pharmacol.* (2018) 11:15–25. doi: 10.1080/17512433.2017.1377070
50. Antonuccio DO, Danton WG, DeNelsky GY, Greenberg RP, Gordon JS. Raising questions about antidepressants. *Psychother Psychosom.* (1999) 68:3–14. doi: 10.1159/000012304.
51. Moncrieff J, Cohen D. Do antidepressants cure or create abnormal brain states? *PLoS Med* (2006) 3:e240. doi: 10.1371/journal.pmed.0030240
52. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders DSM-5.* Washington, DC: American Psychiatric Association (2013).
53. Breggin PR. Suicidality, violence and mania caused by selective serotonin reuptake inhibitors (SSRIs): a review and analysis. *Int J Risk Saf Med.* (2004) 16:31–49.
54. Andrews PW, Thomson JA Jr, Amstadter A, Neale MC. Primum non nocere: an evolutionary analysis of whether antidepressants do more harm than good. *Front Psychol.* (2012) 3:117. doi: 10.3389/fpsyg.2012.00117
55. Moret C, Isaac M, Briley M. Problems associated with long-term treatment with selective serotonin reuptake inhibitors. *J Psychopharmacol.* (2009) 23:967–74. doi: 10.1177/0269881108093582
56. Richardson K, Fox C, Maidment I, Steel N, Loke YK, Arthur A, et al. Anticholinergic drugs and risk of dementia: case-control study. *BMJ* (2018) 361:k1315. doi: 10.1136/bmj.k1315
57. Smoller JW, Allison M, Cochrane BB, Curb JD, Perlis RH, Robinson JG, et al. Antidepressant use and risk of incident cardiovascular morbidity and mortality among postmenopausal women in the Women’s Health Initiative study. *Arch Int Med.* (2009) 169:2128–39. doi: 10.1001/archinternmed.2009.436
58. Gafoor R, Booth HP, Gulliford MC. Antidepressant utilisation and incidence of weight gain during 10 years’ follow-up: population based cohort study. *BMJ* (2018) 361:k1951. doi: 10.1136/bmj.k1951.
59. Coupland C, Dhiman P, Morris R, Arthur A, Barton G, Hippisley-Cox J. Antidepressant use and risk of adverse outcomes in older people: population based cohort study. *BMJ* (2011) 343:d4551. doi: 10.1136/bmj.d4551
60. Maslej MM, Bolker BM, Russell MJ, Eaton K, Durisko Z, Hollon SD, et al. The mortality and myocardial effects of antidepressants are moderated by preexisting cardiovascular disease: a meta-analysis. *Psychother Psychosom.* (2017). 86:268–82. doi: 10.1159/000477940

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Hengartner and Plöderl. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.