Acta Neuropsychiatrica

cambridge.org/neu

Scientific debate instead of beef; challenging misleading arguments about the efficacy of antidepressants

Michael P. Hengartner

Department of Applied Psychology, Zurich University of Applied Sciences (ZHAW), PO Box 707, CH-8037 Zurich, Switzerland

Commentary

Cite this article: Hengartner MP. (2019) Scientific debate instead of beef; challenging misleading arguments about the efficacy of antidepressants. *Acta Neuropsychiatrica* **31**:235–236. doi: 10.1017/neu.2019.23

Received: 27 March 2019
Revised: 1 May 2019
Accepted: 1 May 2019
First published online: 4 June 2019

Key words:
antidepressants; bias; efficacy; selective reporting; SSRI

Author for correspondence: Michael P. Hengartner,
Email: michaelpascal.hengartner@zhaw.ch

In a recent commentary with the polemic title 'Antidepressants; what's the beef?', Goodwin and Nutt argued that the benefit-risk ratio of antidepressants had been questioned inappropriately (Goodwin & Nutt, 2019). Personally I think it is a great achievement that our medical system can offer pharmacological treatments to people who suffer from serious clinical depression, and like Goodwin and Nutt I accept that antidepressants may be useful in some patients (Hengartner & Plöderl, 2018). Nevertheless, and this is where my position deviates from Goodwin and Nutt, I am also concerned about the overestimation of efficacy and the minimisation of harm (Hengartner, 2017). There are many misrepresentations in the commentary by Goodwin and Nutt, all of which systematically inflate the apparent benefits of antidepressants, and in this letter, I will discuss five of them.

First misleading claim: '[European regulators] found around an average 16% greater response rate following active treatment than placebo for newer antidepressants (which included SSRIs)' (Goodwin & Nutt, 2019).

Here Goodwin and Nutt ignore that binary response rates derived from continuous symptom scales (commonly defined as ≥50% symptom reduction) are an inappropriate and occasionally even deceptive construct (Hengartner, 2017; Senn, 2018). Most people would probably be surprised to learn that even if the response rate for antidepressants is 56% and that for placebo is 40% (hence 16% difference), it still could be that, on average, antidepressants are no better than placebo. A simple example follows: imagine that all participants in a trial have a baseline depression score of 20, so that a change ≥10 points from baseline to end of treatment is considered response and change <10 points non-response. Further assume that there were 100 patients in the drug and placebo arm each. In the antidepressant group, 56 people improved by exactly 10 points, 24 had 9 points, and the remaining 20 had 8 points. In the placebo group, 40 people improved by 10 points and the remaining 60 people had 9 points. So what's the result? Response rate would be 56% for antidepressants and 40% for placebo, but mean change score would be 9.4 for both antidepressants and placebo. That is, mean improvement would not differ between drug and placebo arm! This example illustrates precisely why response rates may erroneously suggest drug efficacy even when a true benefit is lacking. Moreover, a critical reader might also wonder why Goodwin and Nutt quote the more favourable 16% response rate difference reported by Melander *et al.* in a relatively small ($n = 7374$) meta-analysis from 2008 rather than the 10% difference reported in a much larger ($n = 27\,422$) and more recent SSRI meta-analysis by Jakobsen *et al.* (2017). Selective reporting is the only explanation that comes to my mind.

Second misleading claim: 'The number needed to treat (NNT) in studies with a mean drug-placebo difference on the HDRS scale of around 3 is between 5 and 7 and this effect size compares reasonably well with most drugs used in medicine' (Goodwin & Nutt, 2019).

Here Goodwin and Nutt selectively report efficacy from an arbitrary subgroup of 'true benefiters' delineated post hoc in five escitalopram trials, and they ignore that the mean drug–placebo difference across all participants in SSRI trials is about two points, not three points (Hengartner & Plöderl, 2018). Thus, the NNT is not between 5 and 7, but rather between 8 and 10 (Hengartner & Plöderl, 2018). Moreover, post hoc analyses as those carried out by Thase *et al.* (quoted by Goodwin and Nutt) that allegedly delineate a subgroup of 'true benefiters' are by and large a statistical artefact due to random outcome variation and arbitrary subgroup selection (Senn, 2018). When a specific patient shows improvements that are considerably larger than the average drug effect, then this symptom change is usually not brought about by the drug but rather by other factors, for example, the patient may have fell in love or he/she just had few good days due to random symptom fluctuations (Senn, 2018). Moreover, seeing a bimodal distribution (i.e. benefiters vs. non-benefiters) in the six graphs provided by Thase *et al.* requires a lot of imagination; in three graphs (b, c, and d) the distribution is obviously not bimodal. Another issue is whether a NNT between 5 and 7 compares 'reasonably well' with most drugs used in general medicine. Apart from this NNT being overestimated (NNT is between 8 and 10 across all antidepressant trials), the efficacy of most drugs used in general

medicine is assessed based on hard outcomes such as mortality or cardiovascular events. Comparing a NNT for partial and often transient symptom reduction derived from a subjective rating scale to a NNT for objective rates of mortality or cardiovascular events, is an inappropriate comparison (Hengartner & Plöderl, 2018).

Third misleading claim: 'A meta-analysis of the effect of SSRIs on HRSD items in regulatory trials, showed that depressed mood itself was the most sensitive. The effect size for the whole scale was 0.27, while that for mood per se was 0.4' (Goodwin & Nutt, 2019).

Several aspects of this analysis by Hieronymus *et al.* (quoted by Goodwin and Nutt) are problematic. First of all, by 'most sensitive' Goodwin and Nutt apparently mean that the depressed mood item displays the largest effect size, that is, relatively high responsiveness. However, whether this particular item is a sensitive outcome measure remains speculative, since it has never been validated as an outcome measure. Thus far, we do not know whether it has good criterion validity with respect to quality of life, subjective distress, or general level of functioning. Imagine, for instance, that most patients who show marked improvements in the depressed mood item also become agitated, irritable, and insomniac, then its criterion validity as a sensitive outcome measure would be questionable. Moreover, a single depression item necessarily has poor content validity, because depression is a complex, multi-dimensional disorder. Depressed mood is just one aspect of major depression, and there is no reason to assume that it is more important than for instance feelings of worthlessness or loss of interests and pleasure. Finally, the approach chosen by Hieronymus *et al.* to establish efficacy is misleading, as it merely compared d effect size estimates for single items to the full scale. Given that the depressed mood item is an ordinal variable with five levels, d is an inappropriate effect size estimate, as its calculation requires at least interval scale.

Fourth misleading claim: 'Analysis of the long-term efficacy of antidepressants shows that in terms of protecting patients against a subsequent relapse to depression these medicines have an NNT of less than 3, which is a remarkable efficacy for any form of treatment' (Goodwin & Nutt, 2019).

Here Goodwin and Nutt refer to findings from discontinuation trials, where select groups of patients who responded particularly well to antidepressants and who remained mostly symptom-free for some time are either abruptly switched to placebo or maintained on active drug. These trials cannot provide information about the long-term efficacy of antidepressants, as they merely indicate that stopping antidepressants abruptly in patients who improved on them can cause considerable health problems that may qualify as depression relapse. Since in most cases, relapses occur shortly after stopping the drug (i.e. 'preventive' effects are detectable during the first 1–3 months only, thereafter the survival curves for placebo and active drug run parallel), it is very likely that many relapses were actually withdrawal reactions (Hengartner, 2017). So, instead of stating that antidepressants are 'protecting patients against a subsequent relapse to depression', the more accurate interpretation would be that abruptly stopping antidepressants can cause severe withdrawal reactions (note that withdrawal symptoms such as depressed mood, insomnia, agitation, anxiety, and gastrointestinal complaints are rated as depression symptoms according to the HRSD). In addition, the stated 'NNT of less than 3' is evidently false. In the meta-analysis of Geddes *et al.* (quoted by Goodwin and Nutt), the relapse rate on placebo versus SSRI was 37% versus 15%, so the absolute risk reduction is 22%, which produces a NNT of 4.5.

Fifth misleading claim: 'For no SSRI is the drop-out rate statistically higher than placebo, which is what the literal interpretation of doing more harm than good would require' (Goodwin & Nutt, 2019).

I consider this is a flawed argument as I detail below, but for now, I will accept it at face value. A conversion of this argument would then be that doing more good than harm requires that the drop-out rate for SSRI is consistently lower than placebo, which it is not (Hengartner & Plöderl, 2018). Therefore, and according to Goodwin and Nutt's very own line of reasoning, it follows that antidepressants do no more good than harm. But why do I consider this a flawed argument? Say, for instance, 20% of antidepressant users in a trial drop out due to severe adverse events and 20% of placebo users drop out due to inefficacy. If there are no further drop outs, then the drop-out rate would be 20% each (so no difference between treatment arms), and according to Goodwin and Nutt, this means that the drug does no more harm than good. Critical readers will easily understand that this argumentation is problematic.

Goodwin and Nutt (2019) state at the end of their commentary that the glass is half full but not empty. In the spirit of finding some common ground, I accept that the glass is not empty (i.e. antidepressants may be useful in some patients), but according to the literature, the glass is certainly not half full. In fact, the glass is rather full by one-ninth, which corresponds to the average NNT of 9 for newer antidepressants (Hengartner & Plöderl, 2018). Therefore, researchers should avoid inflating the apparent benefits of these drugs and instead acknowledge that they may also cause various adverse events, including rare but serious ones (Hengartner, 2017; Jakobsen *et al.*, 2017). This is the critical information that should be given to patients who consider starting an antidepressant, since a balanced benefit-risk evaluation is required to provide fully informed consent.

Author ORCIDs. Michael Hengartner, 0000-0002-2956-2969

Author contributions. MPH was the sole author and did all the writing.

Financial support. No financial support was received for this work.

Statement of interest. I declare no conflicts of interests.

References

Goodwin GM and Nutt D (2019) Antidepressants; what's the beef? *Acta Neuropsychiatrica* **31**, 59–60.

Hengartner MP (2017) Methodological flaws, conflicts of interest, and scientific fallacies: Implications for the evaluation of antidepressants' efficacy and harm. *Front Psychiatry* **8**, 275.

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.

Jakobsen JC, Katakam KK, Schou A, Hellmuth SG, Stallknecht SE, Leth-Moller K, Iversen M, Banke MB, Petersen IJ, Klingenberg SL, Krogh J, Ebert SE, Timm A, Lindschou J and Gluud C (2017) Selective serotonin reuptake inhibitors versus placebo in patients with major depressive disorder. A systematic review with meta-analysis and Trial Sequential Analysis. *BMC Psychiatry* **17**, 58.

Senn S (2018) Statistical pitfalls of personalized medicine. *Nature* **563**, 619–621.

https://doi.org/10.1017/neu.2019.23 Published online by Cambridge University Press