Efficacy of antidepressants: similar but different

Received 16 February 2011; Accepted 15 March 2011; First published online 14 April 2011

Fountoulakis & Moller (2010) have re-analysed the Kirsch et al. (2008) meta-analysis of newer antidepressants and reached quite different conclusions to the original study. In particular they highlight that both venlafaxine and paroxetine exceed the NICE criteria for ‘clinical significance’ of 3 points on the Hamilton Rating Scale for Depression (HAMD) and that the Kirsch et al. analysis underestimates the effect size of newer antidepressants due to methodological reasons. They also go on to claim that placebo response in trials declines with increasing baseline severity while antidepressant response remains the same. They argue this means that while much of the placebo response is due to expectancy effect this is not true for antidepressants and that therefore the effects are not additive, undermining the validity of randomized controlled trials.

We have also performed a re-analysis of the Kirsch et al. data (Horder et al. 2010) and made broadly similar findings. In particular we found that the effect size of venlafaxine and paroxetine exceeded the NICE threshold (but see below for differences in methodology) and that, contrary to what Kirsch et al. claim, controlling for baseline severity did not strongly affect this superiority (although the differences between the drugs was not statistically significant).

However, our analysis disagrees in two important respects from that of Fountoulakis & Moller. First, we criticized Kirsch et al. for their idiosyncratic method of meta-analysis where they calculated mean antidepressant and mean placebo improvement scores separately, subtracting one from the other to estimate the overall effect size. This approach treats each trial as being two entirely separate experiments, ignoring trial-specific effects and assuming that placebo improvement is the same across all trials. Fountoulakis & Moller actually use a very similar method to re-analyse the data and are subject to the same criticisms. They claim that Kirsch et al. have made a mistake in their analysis when they find a pooled effect size in the antidepressant arms of 9.60 rather than 10.04–10.16, which then biases down the overall effect size to 1.80 rather than 2.18−2.68.

Looking at the data it is clear that this difference is due to Kirsch et al. using the standard error as an estimator of inverse variance while Fountoulakis & Moller weight their meta-analysis either by sample size only or by using the standard deviation as an estimator of inverse variance. This latter approach does not include the sample size information inherent in using the standard error. Since using the standard error estimate would be the more usual and correct approach it would be unfair to characterize this as a mistake on the part of Kirsch et al. certainly when compared to their other methodological errors. When we looked at the data using standard methods where the effect size for each study (antidepressant change score minus placebo change score) is pooled using a random-effects model weighted by the standard error as an estimate of inverse variance we found an overall effect size of 2.70 on the HAMD, using a fixed-effects model this was 2.40. There is an argument that when we are concerned with optimizing a point estimate of effect size (as with a threshold score), rather than maximizing precision, then fixed-effect weighting by sample size would be the preferred method and in this case the results are similar with an effect size of 2.65.

Second, we found that when we re-analysed the data the apparent decline in placebo response with increasing baseline severity and constant response in the antidepressant arm was actually an artefact of the use of the standardized mean difference (SMD) by Kirsch et al. It is not necessary to use the SMD to pool studies when they all use the same outcome measure (change scores on the HAMD in this case) and it can be actively misleading when effect size correlates with variance, as it does in this dataset. Looking at the HAMD effect sizes directly we found that there was in fact a linear relationship between baseline severity and effect size in the antidepressant arms and essentially no relationship in the placebo arms – effectively the opposite conclusion to that emphasized by both Kirsch et al. and Fountoulakis & Moller when they use the SMD data.

Acknowledgements

Thanks to Jamie Horder and Robert Waldmann for helpful discussions.
Statement of Interest

None.

References


Paul R. Matthews

*OBMH NHS Foundation Trust, West Oxfordshire CMHT, Nuffield Health Centre, Witney, UK*