

## Short report

## Initial depression severity and response to antidepressants v. placebo: patient-level data analysis from 34 randomised controlled trials

Jonathan Rabinowitz, Nomi Werbeloff, Francine S. Mandel, François Menard, Lauren Marangell and Shitij Kapur

**Summary**

Several often-cited meta-analyses have reported that the efficacy of antidepressant medications depends on the severity of depression. They found that drug–placebo differences increased as a function of initial severity, which was attributed to decreased responsiveness to placebo among patients with severe depression rather than to increased responsiveness to medication. We retested this using patient-level data and also undertaking a meta-analysis of trial-level data from 34 randomised placebo controlled trials ( $n=10737$ ) from the NEWMEDS registry. Although our trial-level data support previous findings, patient-level data did not show any significant effect of initial depression severity on drug v. placebo difference.

**Declaration of interest**

J.R. has received research support, and/or consultancy fees and/or travel support from Amgen, Janssen, JNJ, Eli Lilly, Pfizer, BiolineRx, F. Hoffmann-La Roche, Piere-Fabre, Kamada, Spatz Medical, Ethypharm and Avraham Pharmaceuticals and is on the scientific advisory board of MedAvante. F.M. is an employee of H. Lundbeck A/S. F.M. is an employee of Pfizer. L.M. was an employee of Eli Lilly and Company when this work was done. S.K. has received grant support from GSK and has served as a consultant and/or speaker for AstraZeneca, BiolineRx, BMS-Otsuka, Eli Lilly, Janssen (J&J), Lundbeck, NeuroSearch, Pfizer, F. Hoffmann-La Roche, Servier and Solvay Wyeth.

**Copyright and usage**

© The Royal College of Psychiatrists 2016.

One of the most controversial studies on the treatment of depression,<sup>1</sup> a meta-analysis conducted by Kirsch *et al*<sup>2</sup> cited 1500 times, found that the efficacy of antidepressant treatment is attributable to decreased responsiveness to placebo among patients who were severely depressed rather than to increased responsiveness to medication. That analysis included data from 35 published and unpublished studies on fluoxetine, venlafaxine, nefazodone and paroxetine conducted between 1987 and 1999. A more recent analysis of the same data-set did not find that initial severity determined drug–placebo differences.<sup>3</sup> Khan *et al* examined the association of baseline severity and outcome in 45 phase II and III antidepressant clinical trials.<sup>4</sup> They found that in the active treatment group in trials that included patients with severe depression, more severe depression at baseline was associated with more symptom reduction, yet in the placebo group with less reduction. Fournier *et al*,<sup>5</sup> using patient-level data from six studies, found a significant interaction between baseline severity and treatment, confirming Kirsch *et al*'s conclusion. However, it was only due in small degree to decreased responsiveness to placebo. We retested the hypothesis that the relationship between initial severity and antidepressant efficacy is attributable to decreased responsiveness to placebo among patients with severe depression rather than to increased responsiveness to medication. We conducted both patient-level and meta-analysis of trial-level data.

**Method**

We used patient- and trial-level data from 34 randomised placebo-controlled trials ( $n=10737$ ) (1987–2007) of citalopram, duloxetine, escitalopram, quetiapine and sertraline from the NEWMEDS registry<sup>6</sup> (see online supplement DS1 for a list of studies). This included all acute placebo-controlled trials of major depressive disorder in adult populations sponsored or owned by Pfizer (12 studies; active:  $n=2455$ , placebo:  $n=888$ ), Lilly (11 studies; active:  $n=2425$ , placebo:  $n=1134$ ), AstraZeneca (4 studies; active:  $n=1021$ , placebo:  $n=524$ ) and Lundbeck (7 studies; active:  $n=1509$ , placebo  $n=781$ ) on these five compounds. In three

studies the Hamilton Rating Scale for Depression (HRSD)<sup>7</sup> was estimated based on the Montgomery–Åsberg Depression Rating Scale (MADRS)<sup>8</sup> using equipercentile linking, which gives an equivalent score of one measure on the other measure. It was done using data from 16 studies that included both measures.

Analysis of covariance of change from baseline on the HRSD, using last observation carried forward, was examined testing for baseline, a dichotomous variable of placebo v. active treatment and their interaction. A significant interaction of placebo v. active treatment and baseline score would support the hypothesis. To further test the hypothesis, linear and quadratic regression equations of baseline severity and change from baseline to end-point were run separately for placebo and active treatment. A larger regression coefficient for placebo than active treatment would support the hypothesis. Analysis was repeated for citalopram and sertraline only – the two drugs in common with Kirsch *et al*,<sup>2</sup> and also repeated by study, to examine possible study-level differences. Mixed-models analysis was used to test the data under the assumption that data were missing at random.

As an alternative test, we compared effects for those patients with low ( $<22$ ), medium (22–25) and high (above 25) HRSD baseline scores. A significant interaction of baseline group and placebo v. active treatment and greater effects within placebo than active treatment groups would support the hypothesis. In addition, we repeated the above analysis using trial-level data to see whether differences in results obtained might be the result of differences in methodology. Trial-level data were weighted by adjusted inverse variance as done by Kirsch *et al*.<sup>2</sup>

**Results**

Patient-level results did not support our hypothesis. The interaction of placebo v. active treatment and baseline severity was not significant ( $F=1.19$ ,  $P=0.28$ ;  $B=0.045$  (95% CI  $-0.035$  to  $0.125$ ),  $\beta=0.059$ ). Linear and quadratic regression results were the same for both models and were almost the same for active treatment and for placebo ( $R^2=0.06$ , s.e.=8.1;  $R^2=0.04$ ,

**Table 1** Placebo–active differences in Hamilton Rating Scale for Depression (HRSD) scores and drop-out rates by group

	HRSD score, mean (s.d.)		Drug–placebo difference, <sup>a</sup> mean (95% CI), s.e.	Drop-out, %
	Baseline	Change from baseline		
Treatment group			–2.05 (–2.38 to –1.72) 0.17	
Placebo ( <i>n</i> = 3258)	23.0 (4.1)	–8.8 (8.1)		35.6
Active ( <i>n</i> = 7323)	23.1 (4.2)	–10.8 (8.4)		35.0
Baseline severity				
Low (less than 22)			–2.04 (–2.50 to –1.58) 0.24	
Placebo ( <i>n</i> = 1328)	19.3 (2.7)	–7.1 (7.2)		34.8
Active ( <i>n</i> = 3046)	19.4 (2.8)	–9.1 (7.4)		34.5
Medium (22–25)			–1.82 (–2.40 to –1.24) 0.30	
Placebo ( <i>n</i> = 1102)	23.8 (0.8)	–9.2 (8.0)		31.9
Active ( <i>n</i> = 2345)	23.8 (0.8)	–11.0 (8.1)		33.3
High (above 25)			–2.41 (–3.17 to –1.64) 0.39	
Placebo ( <i>n</i> = 828)	28.0 (2.4)	–10.7 (9.1)		40.0
Active ( <i>n</i> = 1932)	28.1 (2.6)	–13.1 (9.4)		35.6

a. Largest pairwise difference high –2.41 v. medium –1.82,  $t=1.16$ , d.f. = 6205,  $P=0.25$ .

s.e. = 7.9, respectively) showing a small increase in change from baseline as a function of baseline severity in both the active and placebo groups (see online Fig. DS1). The above analysis repeated by study found that the interaction was not significant in 25 trials ( $P > 0.18$  and  $P < 0.93$ ), trend level in 3/34 trials ( $P = 0.06$ , 0.08, 0.09) and significant in 6/34 trials. The additional mixed-models analysis found that the interaction of placebo v. active treatment and baseline severity was significant in 3/34 studies. For the citalopram and sertraline studies, two compounds overlapping with Kirsch *et al*'s study, there was no significant interaction of placebo v. active treatment and baseline severity ( $F=0.29$ ,  $P=0.59$ ). The linear and quadratic regression results were the same for both regression models and were almost the same for active treatment and for placebo ( $R^2=0.03$ , s.e. = 8.4;  $R^2=0.02$ , s.e. = 8.1, respectively). Results did not differ using random- and fixed-effects models. The difference in active v. placebo change from baseline between the low, medium and high severity groups was not statistically significant ( $P=0.25$ , Table 1).

To test whether the differences between these results and those of Kirsch *et al* are the result of differences between a meta-analysis that is limited by having only aggregate-level data on a study and an analysis of patient-level data, analysis was repeated on aggregate-level data. Unlike the patient-level analysis, when examining baseline severity as a continuous variable there was a significant interaction between placebo v. active treatment and baseline severity ( $F=9.27$ ,  $P=0.002$ ). Drug and placebo efficacy increased as initial severity increased, with baseline severity explaining more of the variance in the placebo group than in the active treatment group. The linear and quadratic regression results for active treatment were  $R^2=0.26$ ,  $R^2=0.28$  and for placebo  $R^2=0.32$ ,  $R^2=0.40$  (online Fig. DS2). The difference in active v. placebo change from baseline between the lower, medium and high severity groups was not statistically significant ( $P=0.99$ ).

## Discussion

Baseline severity was not associated with a more pronounced change from baseline in the active- v. placebo-treated patients when using patient-level data, but was evident to some extent when using aggregate trial-level data. The patient-level analysis does not support the findings of the previous meta-analyses<sup>2,4,5</sup> that antidepressants act at the same magnitude irrespective of initial severity while placebo changes as a function of baseline severity. Patient-level data are more sensitive than trial-level data in measuring the effects in question as they allow for adjusting each patient's change score by their baseline value and other patient-level characteristics.

The difference between our results and Kirsch *et al*'s<sup>2</sup> appear to be the result of differences in methodology – meta-analysis v. patient-level analysis. This is supported by our finding that when examining the same drugs as Kirsch *et al* using patient-level analysis, we did not find the effect that their study did. However, we note that for the most part our studies did not overlap with those included in the work of Kirsch *et al*. Caution is advised when examining positive relationships between baseline severity and symptom improvement as these may be the result of regression to the mean.

**Jonathan Rabinowitz**, PhD, **Nomi Werbeloff**, PhD, School of Social Work, Ramat Gan, Bar Ilan University, Israel; **Francoise S. Mandel**, PhD, Pfizer Development Operations, New York, USA; **François Menard**, MD, Lundbeck Research Department, Paris, France; **Lauren Marangell**, MD, Department of Psychiatry, University of Texas Health Science Center, Houston, Texas, USA; **Shitij Kapur**, MD, Institute of Psychiatry, Psychology and Neuroscience, Kings College London, London, UK

**Correspondence:** Jonathan Rabinowitz, Bar Ilan University, Ramat Gan, Israel 52900. Email: Jonathan.Rabinowitz@biu.ac.il

Received 19 Apr 2015, final revision 23 Dec 2015, accepted 27 Feb 2016

## Funding

The research leading to these results has received support from the Innovative Medicine Initiative Joint Undertaking under grant agreement number: 115008 of which resources are composed of a European Federation of Pharmaceutical Industries and Associations (EFPIA) in-kind contribution and financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013). The funding sources were not involved in the collection, analysis, and interpretation of data; in the writing of the report; and nor in the decision to submit the paper for publication.

## References

- Jarrett C. The 10 most controversial psychology studies ever published. *Res Dig: British Psychological Society* 2014; 19 September (<http://digest.bps.org.uk/2014/09/the-10-most-controversial-psychology.html>).
- Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med* 2008; **5**: e45.
- Fountoulakis KN, Veroniki AA, Siamouli M, Moller HJ. No role for initial severity on the efficacy of antidepressants: results of a multi-meta-analysis. *Ann Gen Psychiatry* 2013; **12**: 26.
- Khan A, Leventhal RM, Khan SR, Brown WA. Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. *J Clin Psychopharmacol* 2002; **22**: 40–5.
- Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, Shelton RC, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA* 2010; **303**: 47–53.
- Institute of Medicine. Models of data sharing: THE NEWMEDS CONSORTIUM. *Sharing Clinical Research Data: Workshop Summary*: 37–38. The National Academies Press, 2013 (<http://www.nap.edu/read/18267/chapter/5#39>).
- Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960; **23**: 56–62.
- Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979; **134**: 382–9.

