Solving the Antidepressant Efficacy Question: Effect Sizes in Major Depressive Disorder

Paul A. Vöhringer, MD[1,2]; and S. Nassir Ghaemi, MD, MPH[3]

[1]Facultad de Medicina, Hospital Clinico, Universidad de Chile, Santiago, Chile; [2]Mood Disorders Program, Department of Psychiatry, Tufts Medical Center, Boston, Massachusetts; and [3]Tufts University School of Medicine, Boston, Massachusetts

ABSTRACT

Background: Numerous reviews and meta-analyses of the antidepressant literature in major depressive disorders (MDD), both acute and maintenance, have been published, some claiming that antidepressants are mostly ineffective and others that they are mostly effective, in either acute or maintenance treatment.

Objective: The aims of this study were to review and critique the latest and most notable antidepressant MDD studies and to conduct our own reanalysis of the US Food and Drug Administration database studies specifically analyzed by Kirsch et al.

Methods: We gathered effect estimates of each MDD study. In our reanalysis of the acute depression studies, we corrected analyses for a statistical floor effect so that relative (instead of absolute) effect size differences were calculated. We also critiqued a recent meta-analysis of the maintenance treatment literature.

Results: Our reanalysis showed that antidepressant benefit is seen not only in severe depression but also in moderate depression and confirmed a lack of benefit for antidepressants over placebo in mild depression. Relative antidepressant versus placebo benefit increased linearly from 5% in mild depression to 12% in moderate depression to 16% in severe depression. The claim that antidepressants are completely ineffective, or even harmful, in maintenance treatment studies involves unawareness of the enriched design effect, which, in that analysis, was used to analyze placebo efficacy. The same problem exists for the standard interpretation of those studies, although they do not prove antidepressant efficacy either, since they are biased in favor of antidepressants.

Conclusions: In sum, we conclude that antidepressants are effective in acute depressive episodes that are moderate to severe but are not effective in mild depression. Except for the mildest depressive episodes, correction for the statistical floor effect proves that antidepressants are effective acutely. These considerations only apply to acute depression, however. For maintenance, the long-term efficacy of antidepressants is unproven, but the data do not support the conclusion that they are harmful. (*Clin Ther.* 2011;33:B49–B61)

Key words: antidepressants, effect size, meta-analysis, randomized clinical trials.

INTRODUCTION

Much controversy has surrounded recent meta-analyses and randomized clinical trials (RCTs) of antidepressant efficacy in major depressive disorder (MDD), including in the nonscientific media. In this review, we use the concept of effect sizes to make clinical and scientific sense of what has become a cultural debate.

Examined here are the most prominent RCTs or meta-analyses of RCTs published in the last 5 years for both acute and maintenance efficacy of antidepressants in MDD. A summary of the review of these studies is provided in **Table I.**

In acute depression RCTs, some reviews involve reanalysis of the US Food and Drug Administration (FDA) database of RCTs conducted by pharmaceutical companies. The major nonpharmacuetical industry study is the National Institute of Mental Health (NIMH)–sponsored Sequenced Alternatives for Treatment-Resistant Depression (STAR*D) project.[1] The pharmaceutical trials have been analyzed and reanalyzed by different authors, with the most media attention being given to the analysis by Kirsch et al.[2] Other published analyses are also important.[3]

Accepted for publication November 9, 2011.
doi:10.1016/j.clinthera.2011.11.019
0149-2918/$ - see front matter

Table I. Summary of analysis of reviews of antidepressant efficacy in RCTs of MDD.

Study	N	Trials Reviewed	Effect Sizes (95% CI)	Comments
Rush et al[1] STAR*D RCT	3671	1	67% acute remission, 26% maintenance remission	No pbo group. Good acute efficacy is shown, but maintenance efficacy is about one half less than acute efficacy.
Kocsis et al[23] and Kornstein et al[5] Maintenance RCT of venlafaxine vs pbo	First maintenance study (year 0) n = 1096 Second maintenance study (year 1) n = 114	2	92% 2-year efficacy reported; this reflects 11% of original sample	"Super-enrichment" design. Second maintenance study sample was only ~10% of the initial sample
Turner et al[7] MA of FDA database of RCTs	12,564	74	0.37 (0.33, 0.41) for published studies vs 0.15 (0.08, 0.22) for unpublished studies. ES of 0.31(0.27, 0.35) when all studies are combined.	31% of studies were unpublished, accounting for 27.5% of the sample
Kirsch et al[2] MA of FDA database	5133	35	Overall standardized ES was 0.61. Absolute ES HDRS of 9.6 drug and 7.8 pbo.	NICE criterion for clinical significance was absolute ES of 3 HDRS points or standardized ES of d = 0.5 for AD-pbo difference. Overall nonstandardized effect size of 0.32 increases to 0.40 when corrected for baseline severity (authors do not discuss)
Horder et al[8] Reanalysis of Kirsch et al[2]	5133	35	Absolute HDRS difference between AD and pbo = 2.70 (including negative unpublished studies)	Reanalysis was based on (1) random effects rather than fixed effects model as in Kirsch et al and (2) pooling ES differences study by study rather than summing all studies and then ES difference. These changes produce a much larger ES near the NICE threshold.
Davis et al[14] Narrative summary of MAs and RCTs	Not reported	Not reported	Mean acute difference between AD and pbo = 23.6% Mean maintenance difference between AD and pbo = 36%	Uncritical about bias toward ADs in maintenance studies using the enriched design
Fountoulakis and Möller[13] Reanalysis of Kirsch et al[2]	5133	35	Mean AD ES was 10.05, not 9.60, as in Kirsch et al. AD-pbo difference was 2.18, not 1.80 as in Kirsch et al. Venlafaxine and paroxetine absolute HDRS ES were 3.12 and 3.22, respectively, exceeding NICE threshold. Nefazodone and fluoxetine did not.	Reanalysis was based on weighting the mean difference by sample size.
Andrews et al[6] MA of maintenance RCTs	3454	46	Risk difference AD-pbo for relapse = 0.20, meaning 20% increased rate of relapse with AD than with pbo.	MA used an "enriched" design in favor of the pbo arm.
Briscoe and El-Mallakh[22] Reanalysis of maintenance RCTs	449	5	5 RCTs examined for AD efficacy after 6 mo. Four of 5 studies showed no benefit with AD over pbo.	Only analysis to correct for enriched design, which is biased in favor of ADs. Removes relapses due to AD withdrawal.
Vöhringer and Ghaemi (present study) Reanalysis of Kirsch et al[2] MA to correct for statistical floor effect	5133	35	Relative effect size for mild depression was 5% (HDRS < 24), 12% for moderate (24 < HDRS > 28), and 16% for severe depression (HDRS > 28).	NICE criterion is met by 11.5% relative difference between AD and pbo. This analysis disproves the claim by Kirsch et al that only severe depression has clinically meaningful ES. Moderate depression also met NICE criterion.

AD = antidepressant; CI = confidence interval; ES = effect size; FDA = US Food and Drug Administration; HDRS = Hamilton Depression Rating Scale; MA = meta-analysis; MDD = major depressive disorder; NICE = National Institute for Health and Clinical Excellence (UK); pbo = placebo; RCT = randomized clinical trial.