

ARTICLE COVERSHEET
LWW_CONDENSED(7.75X10.75)
SERVER-BASED

Article : JCP50124

Creator : dpc_lww

Date : Monday November 23rd 2015

Time : 23:08:54

Number of Pages (including this page) : 6

Antidepressants

Conundrums and Complexities of Efficacy Studies

Ronald W. Pies, MD

AQ1AQ2

RECENT CONCERNS

It is by now a truism in the lay press that “antidepressants are no more effective than sugar pills.”¹ Although the best available evidence does not support this conclusion,^{2–4} public skepticism regarding the efficacy of antidepressants is understandable, given several recent developments. Furthermore, concerns over possible adverse effects of antidepressants, including increased risk of “suicidality” in younger populations, have continued to darken the public’s perception of these agents.

Recently, these controversies came to a head after a reanalysis⁵ of SmithKline Beecham’s randomized, double-blind Study 329, comparing paroxetine and imipramine with placebo in adolescents with unipolar major depression. The reanalysis, which used previously confidential documents, concluded that—contrary to the original findings—“...neither paroxetine nor high dose imipramine showed efficacy for major depression in adolescents, and there was an increase in harms with both drugs.”⁵ The “harms” included clinically significant increases in suicidal ideation and behavior. In an editorial accompanying the BMJ paper, Dr David Henry commented, “It’s not clear whether it was deliberate or accidental, but [the original report] wrongly gave the impression that an antidepressant drug was effective and safe in children and adolescents.”⁶

Furthermore, the issue of “ghost writing” of antidepressant studies and/or heavy involvement of pharmaceutical industry authors has received renewed scrutiny. One recent review concluded that “There is a massive production of meta-analyses of antidepressants for depression authored by or linked to the industry, and they almost never report any caveats about antidepressants in their abstracts.”⁷ In addition, the issue of publication bias in the antidepressant literature remains a thorny problem. Thus, Turner et al⁸ compared 74 Food and Drug Administration–registered antidepressant trials submitted for regulatory approval (for 12 antidepressants involving 12,564 patients) with the published literature. They found evidence of the “file drawer effect,” that is, publication bias in favor of positive studies. In the backdrop of these findings are some influential meta-analyses suggesting that the efficacy of antidepressants for major depression is exaggerated and/or limited to severe depression.^{9,10}

Taken in toto, these revelations can only deepen the public’s concern—if not cynicism—regarding both the validity of antidepressant studies and possible harms from these agents. Yet, amidst the flurry of media reports on Study 329, the many nuances of antidepressant research were often obscured, and the “bigger picture” was generally missed. For despite the apparent failure of Study 329, the preponderance of research still supports modest-to-moderate efficacy—and overall safety—for several commonly used antidepressants.^{2–4,11,12}

BROADER CONCERNS AND COMPLEXITIES

That said, there are still many theoretical and practical problems in the way antidepressant studies are carried out and interpreted; moreover, questions remain as to the degree of major depression (mild, moderate, or severe) for which antidepressants are effective. Underlying these issues are even broader concerns, such as the marked heterogeneity in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-V) construct of “major depressive disorder” (MDD),¹³ the dissimilarity between research subjects in antidepressant registration trials and “real-world” patients,¹⁴ problems and limitations inherent in all meta-analyses,¹⁵ and rising rates of the placebo response in studies of depression conducted in the past few decades.¹⁶

To complicate matters further: the outcome of antidepressant studies may depend critically on what “level” of data is assessed. For example, based on the Hamilton Depression Scale (HAM-D), Gibbons et al¹¹ (2012) have pointed to the difference between average study-level initial severity and antidepressant response, on the one hand, and patient-level data, on the other. These researchers argue that “...relatively small overall mean differences can translate into relatively large patient-level differences in clinically interpretable and meaningful endpoints such as response and remission.”

Thus, Gibbons et al¹¹ recently examined the short-term efficacy of antidepressants for treating major depression in youth, adults, and geriatric populations. The authors carried out a reanalysis of all

From the SUNY Upstate Medical University, Syracuse, NY; and Tufts University School of Medicine, Boston, MA.

Reprints: Ronald W. Pies, MD, Department of Psychiatry, SUNY Upstate Medical University, 750 E Adams St, Syracuse, New York 13210

(e-mail: piesr@upstate.edu; ronpies@massmed.org).

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

ISSN: 0271-0749

DOI: 10.1097/JCP.0000000000000455

intent-to-treat person-level longitudinal data during the first 6 weeks of treatment of MDD. Data were derived from both published and unpublished studies, conducted by the manufacturers of fluoxetine and venlafaxine. These included 20 randomized, placebo-controlled trials of fluoxetine (12 adult, 4 geriatric, and 4 youth) and 21 adult trials of venlafaxine. Complete longitudinal patient records were obtained, allowing the authors to examine associations between treatment response and baseline severity measured at the patient level. The study found that patients in all age and drug groups had significantly greater improvement relative to placebo controls, and baseline severity of depression did not affect symptom reduction. The authors concluded that "...The results do not support previous findings that antidepressants show little benefit except for severe depression. The antidepressants fluoxetine and venlafaxine are efficacious for major depression, in all age groups although more so in youth and adults compared with geriatric patients. Baseline severity was not significantly related to degree of treatment advantage over placebo."¹¹

In a companion article, Gibbons et al¹² examined suicidal thoughts and behavior in the same set of randomized, placebo-controlled studies of fluoxetine and venlafaxine and again used longitudinal patient-level data. The suicide items from the Children's Depression Rating Scale-Revised and the HAM-D, as well as adverse event reports of suicide attempts and suicide during active treatment, were analyzed in 9185 patients. The study found that "Fluoxetine and venlafaxine decreased suicidal thoughts and behavior for adult and geriatric patients. This protective effect is mediated by decreases in depressive symptoms with treatment. For youths, no significant effects of treatment on suicidal thoughts and behavior were found, although depression responded to treatment. No evidence of increased suicide risk was observed in youths receiving active medication."¹²

The use of longitudinal patient-level data is a genuine strength of the Gibbons et al studies. However, both studies by Gibbons et al came in for withering criticism in subsequent letters to the editor,^{17,18} followed by vigorous rejoinders by Gibbons et al. A complete discussion of these critiques is beyond the scope of this editorial, but a few points are worth noting. For example, inclusion of the LYAQ fluoxetine trial in the Gibbons et al data—which looked at subjects with comorbid attention-deficit/hyperactivity disorder and depression—may have distorted the aggregate analysis because 19% of LYAQ subjects did not have depression.¹⁸ There is also controversy over the use of rating scales (such as the Children's Depression Rating Scale-Revised) as a measure of suicidality, as opposed to spontaneous reports of suicidal thoughts or behaviors.

OTHER STUDIES SUPPORTING ANTIDEPRESSANT EFFICACY AND SAFETY

While acknowledging shortcomings in the 2 studies by Gibbons et al, it is important to note that the results of these studies are largely consistent with several other recent analyses. For example, Vöhringer and Ghaemi³ conducted their own reanalysis of the US Food and Drug Administration database MDD studies specifically analyzed by Kirsch et al. The reanalysis corrected for a statistical "floor effect" so that relative (instead of absolute) effect size differences were calculated; that is, drug-placebo differences were adjusted for baseline severity of illness. This led to an increase in nonstandardized effect size from 0.32 (as per Kirsch et al) to 0.40. Contrary to Kirsch et al (2008)—who found antidepressants effective only in the most severely depressed patients—the Vöhringer and Ghaemi³ reanalysis found that "...antidepressants are effective in acute depressive episodes that are moderate to severe..." although not in mild depressive episodes.

In partial contrast, Stewart et al¹⁹ analyzed 6 placebo-controlled antidepressant studies of patients with nonsevere MDD (HAM-D score < 23) and found that "mild-moderate MDD can benefit from antidepressants," with the number needed to treat in the range of 3 to 8 (number needed to treat < 10 is considered clinically significant). It seems fair to conclude that the effectiveness of antidepressants for "mild" cases of major depression is still unclear—but there is little doubt that antidepressants are effective acutely in moderate-to-severe major depression.

Recently, Thorlund et al²⁰ conducted a meta-analysis of selective serotonin reuptake inhibitors and serotonin-norepinephrine reuptake inhibitors in adults 60 years and older, using data from 15 randomized, controlled trials. With respect to achieving a partial response, the authors found "...clear evidence of the effectiveness of sertraline, paroxetine, and duloxetine" in this study population.

Furthermore, although the issue of "suicidality" in younger populations treated with antidepressants remains controversial, 2 recent analyses suggest a "neutral" effect of selective serotonin reuptake inhibitors on suicide risk in children and adolescents.^{21,22} Thus, Ghaemi²¹ concluded that "SRIs increase suicide risk in 1% of children, and lead to completed suicide in about 1 in 500, which is the same as their prevention rate. Their overall effect is probably neutral when benefits are weighed against harms." Similarly, Carroll²² found "...evidence of equipoise between the therapeutic outcome of preventing suicide and any potential drug-related provocation of suicide among adolescents treated for MDD with fluoxetine."

Notably, a recent consensus conference on antidepressant safety highlighted the marked ambiguity in the term "suicidality"—which can mean anything from suicidal ideation to a completed suicide. The consensus authors pointedly observed that "...no deaths from suicide were reported in any of the 24 pediatric [antidepressant] trials involving 4,582 patients."²³ Indeed, suicidal ideation per se is a poor predictor of completed suicide,²⁴ which remains a poorly studied phenomenon; for example, randomized controlled studies routinely exclude subjects at high risk for suicide. Finally, it is important to note that, in adult populations, there are no controlled studies demonstrating increased rates of completed suicides associated with newer antidepressants.¹²

ARE WE HAMSTRUNG BY THE HAM-D?

The HAM-D (or HDRS) is so widely used in antidepressant research; it has become nearly synonymous with measures of antidepressant efficacy. (It has sometimes been said that most antidepressant studies are illustrations of how strongly antidepressants affect the HAM-D, more so than how well they treat depression.) However, the HAM-D itself is subject to variability, depending on the level of experience of the rater: for example, poorly trained raters tend to produce results that diminish the effect of the antidepressant.²⁵ Furthermore, as Bagby et al²⁶ (2004) have observed that "...many [HAM-D] scale items are poor contributors to the measurement of depression severity; others have poor inter-rater and retest reliability."

Indeed, much may depend on which version of the HAM-D (HAM-D17, HAM-D21, HAM-D24, etc) or which HAM-D item cluster is used. In a key study using the HAM-D6, Bech⁴ (2010) reported on a "reallocation" of HAM-D items, focusing on the 6 items measuring severity of clinical depression—depressed mood, guilt, work and interests, tiredness, anxiety, and psychomotor retardation. For second-generation antidepressants in placebo-controlled trials, application of the HAM-D6 resulted in clinically significant effect sizes of 0.40 or greater.⁴

That said, the HAM-D is not necessarily the last word in measuring antidepressant response. Indeed, it is rare to find studies of

antidepressant efficacy that examine “quality of life” (QOL) in study subjects, although arguably, it is precisely this factor that matters most to our patients. Two notable exceptions to the “HAMD world view” are the studies by Skevington and Wright²⁷ (2001) and Berlim et al²⁸ (2007). The Skevington and Wright study examined general practice patients with moderate depression ($n = 106$) by DSM-IV criteria. Subjects completed the 100-item World Health Organization Quality of Life Assessment (WHOQOL) and the Beck Depression Inventory, before the start of antidepressant treatment and 6 weeks afterward. Depression decreased significantly for 2 months, with 74% reported feeling better. The 100-item World Health Organization Quality of Life Assessment scores increased in 24 of the 25 facets, “...demonstrating that QOL improves significantly in the 8 weeks following the start of antidepressant treatment...” and that “...antidepressants significantly and comprehensively improve QOL” in this sample.

In a Brazilian study by Berlim et al,²⁸ 73 patients presenting with a severe episode of major depression were assessed by the WHOQOL BREF (26 items derived from the 100-item WHOQOL) and the Beck Depression Inventory at the start of antidepressant treatment and again after a mean of 12 weeks. The depressed patients' QOL scores significantly improved in all the assessed domains (ie, physical health, psychological, social relations, environmental, and overall QOL) during the study period. Moreover, there was significant improvement in depressive symptoms between test and retest (effect sizes ranged from 0.49 to 1.08; ie, medium-to-large effects).

Although these are relatively small, uncontrolled studies, they suggest that QOL may be enhanced by antidepressant treatment—perhaps a more important clinical measure than HAMD scores.

OTHER PROBLEMS WITH ANTIDEPRESSANT STUDIES

As noted previously,¹³ the DSM-IV/V construct of MDD is extraordinarily elastic and heterogeneous. Moreover, DSM-V field trials have revealed “questionable” reliability ($\kappa = 0.20$ – 0.39) for the diagnosis of MDD.²⁹ Given that a patient with MDD may have had the requisite symptoms for as little as 2 weeks, or as long as 2 years; be able to work or not be able; and so forth, it is almost inevitable that the DSM criteria will capture a wide range of MDD types and severity. Most randomized, controlled studies of MDD do not attempt to distinguish between MDD symptoms present for, say, 2 versus 12 months. Nor do most studies look specifically at subjects with MDD who meet criteria for the specifier, “with melancholic features”—a more serious form of MDD that is poorly responsive to placebo, compared with nonmelancholic MDD.³⁰ Indeed, some have argued that “melancholia” (a related construct) ought to be considered a distinct mood disorder.³¹ It seems axiomatic that the presence or absence of melancholic features will result in discrepant findings, with respect to drug versus placebo differences.

Finally, we have yet to solve the puzzle of increasing placebo response rates in more recent studies of major depression. (Contrary to much misrepresentation in the lay press, the placebo condition is much more than “a sugar pill” and usually includes 8 or more hours of weekly, supportive contact with professional staff.³²) As 1 review noted, “Since the response to placebo is variable, often substantial, and increasing, it is not surprising that in many randomized controlled trials the response associated with placebo is similar to that associated with an established antidepressant.”¹⁶ One possible factor in this trend is the changing demographics of research subjects assigned to the placebo condition. Bridge et al³³ (2009) hypothesized that, in some multisite studies of MDD, subjects with less severe depressive illness may have been recruited—which

would likely inflate placebo response rates and diminish drug-placebo differences.

CONCLUSIONS

The preponderance of data from randomized, controlled studies suggests that modern antidepressants are modestly effective and generally safe, in the acute treatment of major depression. Notwithstanding these upbeat conclusions, many problems continue to complicate the research literature, including but not limited to publication bias, rising rates of the placebo response, marked heterogeneity in the construct of “MDD,” and the exclusion of “real-world” patients (eg, those with comorbidities or high suicide risk) from randomized, controlled studies.

Going forward, we need to increase “transparency” in publication, by, for example, making raw study data available to journal reviewers.³⁴ We need to reexamine our entry criteria for antidepressant trials so that study subjects more closely resemble those we treat in clinical practice. We need to refine our recruitment methods so that highly placebo-responsive subjects are not included in randomized controlled studies. We must also address the marked heterogeneity in the DSM-V construct of MDD, via more fine-grained examination of MDD subgroups, for example, patients with chronic and melancholic major depressive episodes. Furthermore, with respect to antidepressant efficacy, we must move beyond the HAMD toward broader assessments of “QOL.” Finally, we need to intensify our outreach and education efforts so that the important role of antidepressant treatment is better appreciated by the general public.

ACKNOWLEDGMENTS

This editorial grew out of discussions with Drs David Osser, Nassir Ghaemi, Peter Kramer, Bernard Carroll, Donald Klein, and other colleagues. The opinions expressed here, however, are solely those of the author, who would also like to thank the peer reviewers for their useful suggestions.

AUTHOR DISCLOSURE INFORMATION

The author declares no conflicts of interest.

REFERENCES

1. Begley S. Why antidepressants are no better than placebos. Accessed at: <http://www.newsweek.com/why-antidepressants-are-no-better-placebos-71111>. Accessed January 28, 2010.
2. Pies R. Are antidepressants effective in the acute and long-term treatment of depression? Sic et Non. *Innov Clin Neurosci*. 2012;9:31–40.
3. Vöhringer PA, Ghaemi SN. Solving the antidepressant efficacy question: effect sizes in major depressive disorder. *Clin Ther*. 2011;33:B49–B61.
4. Bech P. Is the antidepressive effect of second-generation antidepressants a myth? *Psychol Med*. 2010;40:181–186.
5. Le Noury J, Nardo JM, Healy D, et al. Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. *BMJ*. 2015;351:h4320.
6. Henry D, Fitzpatrick T. Liberating the data from clinical trials. *BMJ*. 2015;351:h4601.
7. Ebrahim S, Bance S, Athale A, et al. Meta-analyses with industry involvement are massively published and report no caveats for antidepressants. *J Clin Epidemiol*. 2015. pii: S0895-4356:00429–1.
8. Turner EH, Matthews AM, Linardatos E, et al. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med*. 2008;358:252–260.
9. Kirsch I, Deacon BJ, Huedo-Medina TB, et al. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med*. 2008;5:e45.

AQ3

10. Fournier JC, DeRubeis RJ, Hollon SD, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA*. 2010;303:47–53.
11. Gibbons RD, Hur K, Brown CH, et al. Benefits from antidepressants: synthesis of 6-week patient-level outcomes from double-blind placebo-controlled randomized trials of fluoxetine and venlafaxine. *Arch Gen Psychiatry*. 2012;69:572–579.
12. Gibbons RD, Brown CH, Hur K, et al. Suicidal thoughts and behavior with antidepressant treatment: reanalysis of the randomized placebo-controlled studies of fluoxetine and venlafaxine. *Arch Gen Psychiatry*. 2012;69:580–587.
13. Goldberg D. The heterogeneity of “major depression”. *World Psychiatry*. 2011;10:226–228.
14. Preskorn SH, Macaluso M, Trivedi M. How commonly used inclusion and exclusion criteria in antidepressant registration trials affect study enrollment. *J Psychiatr Pract*. 2015;21:267–274.
15. Feinstein AR. Meta-analysis: statistical alchemy for the 21st century. *J Clin Epidemiol*. 1995;48:71–79.
16. Walsh BT, Seidman SN, Sysko R, et al. Placebo response in studies of major depression: variable, substantial, and growing. *JAMA*. 2002;287:1840–1847.
17. Spielmans GI, Jureidini J, Healy D, et al. Inappropriate data and measures lead to questionable conclusions. *JAMA Psychiatry*. 2013;70:121–122.
18. Carroll BJ. Suicide risk and efficacy of antidepressant drugs. *JAMA Psychiatry*. 2013;70:123–124.
19. Stewart JA, Deliyannides DA, Hellerstein DJ, et al. Can people with nonsevere major depression benefit from antidepressant medication? *J Clin Psychiatry*. 2012;73:518–525.
20. Thorlund K, Druyts E, Wu P, et al. Comparative efficacy and safety of selective serotonin reuptake inhibitors and serotonin-norepinephrine reuptake inhibitors in older adults: a network meta-analysis. *J Am Geriatr Soc*. 2015;63:1002–1009.
21. Ghaemi SN. Classic Study of the Month: The FDA analysis of antidepressants and suicide. *The Psychiatry Letter*. 2015. Accessed at: www.psychiatryletter.org.
22. Carroll BJ. Adolescents with depression. *JAMA*. 2004;292:2578.
23. Meyer RE, Salzman C, Youngstrom EA, et al. Suicidality and risk of suicide—definition, drug safety concerns, and a necessary target for drug development: a consensus statement. *J Clin Psychiatry*. 2010;71:e1–e21.
24. de Leon J, Baca-García E, Blasco-Fontecilla H. From the serotonin model of suicide to a mental pain model of suicide. *Psychother Psychosom*. 2015;84:323–329.
25. Kobak KA, Kane JM, Thase ME, et al. Why do clinical trials fail? The problem of measurement error in clinical trials: time to test new paradigms? *J Clin Psychopharmacol*. 2007;27:1–5.
26. Bagby RM, Ryder AG, Schuller DR, et al. The Hamilton Depression Rating Scale: has the gold standard become a lead weight? *Am J Psychiatry*. 2004;161:2163–2177.
27. Skevington SM, Wright A. Changes in the quality of life of patients receiving antidepressant medication in primary care: validation of the WHOQOL-100. *Br J Psychiatry*. 2001;178:261–267.
28. Berlim MT, Pargendler J, Brenner J, et al. Significant improvement in the quality of life of Brazilian depressed outpatients 12 weeks following the start of antidepressants. *Psychiatry Res*. 2007;153:253–259.
29. Regier DA, Narrow WE, Clarke DE, et al. DSM-5 field trials in the United States and Canada, part II: test-retest reliability of selected categorical diagnoses. *Am J Psychiatry*. 2013;170:59–70.
30. Brown W. Treatment response in melancholia. *Acta Psychiatr Scand Suppl*. 2007;125–129.
31. Parker G, Fink M, Shorter E, et al. Issues for DSM-5: whither melancholia? The case for its classification as a distinct mood disorder. *Am J Psychiatry*. 2010;167:745–747.
32. Preskorn S. A dangerous idea. *J Prac Psych Behav Health*. 1996;2:231–234.
33. Bridge JA, Birmaher B, Iyengar S, et al. Placebo response in randomized controlled trials of antidepressants for pediatric major depressive disorder. *Am J Psychiatry*. 2009;166:42–49.
34. Klein DF, Glick ID. Conflict of interest, journal review, and publication policy. *Neuropsychopharmacology*. 2008;33:3023–3026.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

AQ1 = The first heading ("INTRODUCTION AND RECENT CONCERNS") was changed to "RECENT CONCERNS." Please check whether this change is appropriate.

AQ2 = All headings were tagged as first level. Please check.

AQ3 = Please update details for Ref. 7.

END OF AUTHOR QUERIES