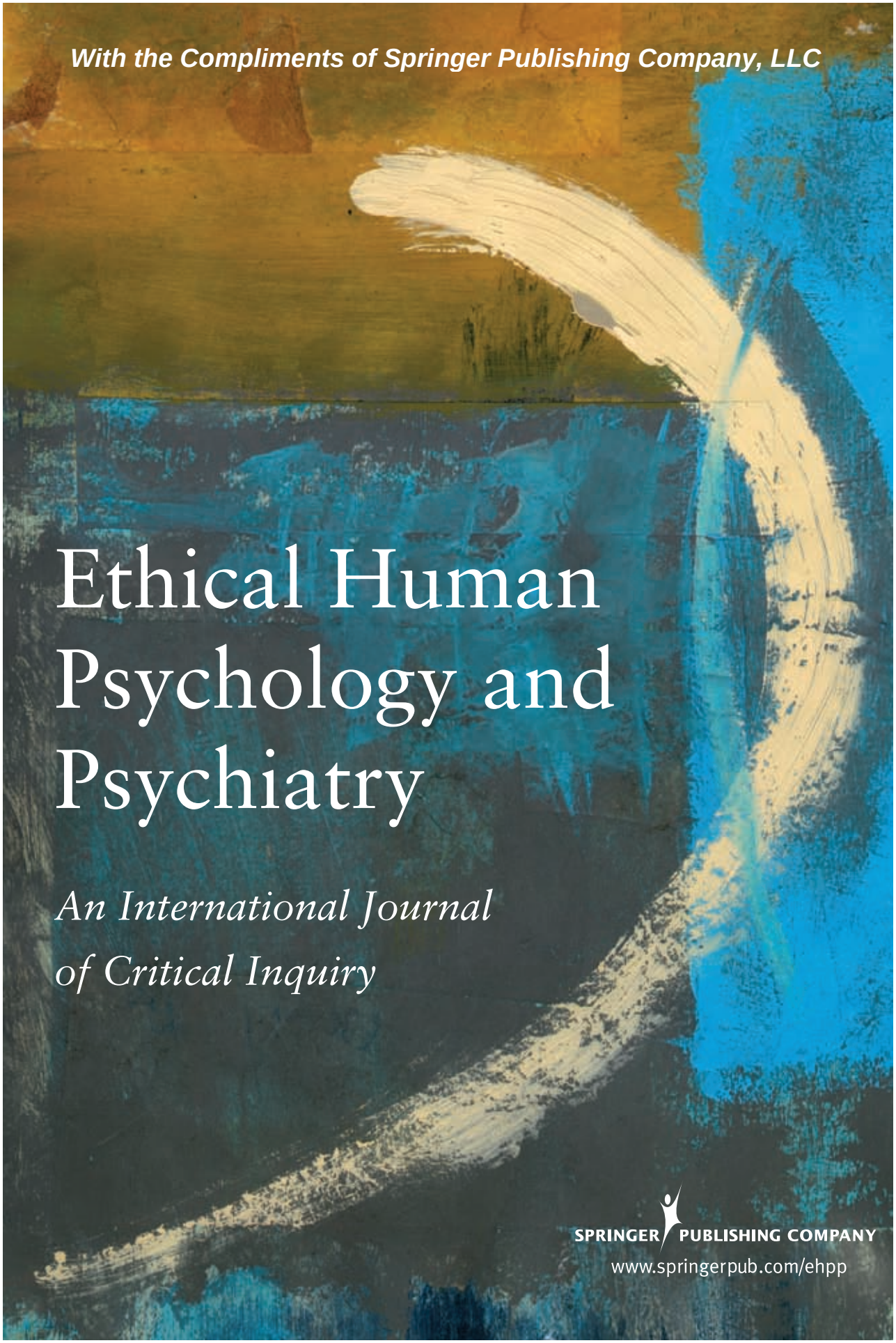

Ethical Human Psychology and Psychiatry

An International Journal of Critical Inquiry

SPRINGER PUBLISHING COMPANY
www.springerpub.com/ehpp

Ethical Human Psychology and Psychiatry, Volume 13, Number 2, 2011

Research Critique

Reanalyzing a Randomized Controlled Trial of Combination Antidepressant Treatment With Mirtazapine: Confidence Intervals Suggest Substantial Uncertainty

Jeffrey R. Lacasse, PhD

Arizona State University

This article subjects a randomized controlled trial (RCT) published in the *American Journal of Psychiatry* to a methodological and statistical critique, including a reanalysis of the effect size statistics presented. The published trial tested the use of combination antidepressants (mirtazapine coprescribed with either bupropion, venlafaxine, or fluoxetine) at treatment initiation as compared with fluoxetine monotherapy. The authors report that combination therapy was effective, with a number-needed-to-treat (NNT) statistic of 3–5, a strong effect size. Scrutiny of the methodology and clinical trial registration shows that 4 of 6 preregistered outcomes were statistically nonsignificant, 1 outcome was not reported, and 1 unregistered outcome was published. The well-critiqued Hamilton Depression Inventory was the only positive outcome measure. Calculating confidence intervals for the reported NNT demonstrates substantial uncertainty (95% CI for NNT = 2.3–18.0). In an era of evidence-based psychiatric practice, there is insufficient evidence to recommend combination therapy at initiation of treatment.

Keywords: evidence-based practice; mirtazapine; antidepressant; statistical reform

The limitations of antidepressant monotherapy in the treatment of major depression are well established (Leventhal & Antonuccio, 2009). No more than 30% of depressed patients will have full remission of their depression in response to a single antidepressant (Pigott, 2011; Trivedi et al., 2006). Given these acknowledged limitations, there is a need for treatments that have demonstrated evidence of superiority over antidepressant monotherapy.

In a recent issue of the *American Journal of Psychiatry*, Blier and colleagues (2010) reported a randomized controlled trial (RCT) sponsored by Organon Pharmaceuticals, the manufacturer of mirtazapine. The investigators claim that combining mirtazapine with fluoxetine, venlafaxine, or bupropion is more effective in the treatment of major depression than fluoxetine monotherapy + placebo. The clinical implication is that prescribers should consider prescribing two antidepressants (one of them mirtazapine) at treatment initiation. This is a reasonable conclusion given the findings as presented, in that combination treatment "yielded a number needed to treat of 3 to 5 over fluoxetine monotherapy,

DOI: 10.1891/1559-4343.13.2.149

which is similar to the advantage of clozapine over conventional antipsychotics" (Blier et al., 2010, p. 286). An editorial in the same issue states, while noting limitations, "The results are striking, and the message is encouraging" (Rush, 2010, p. 241). As reported, these are robust findings, and given the potential ramifications of these findings for clinical practice,[1] they warrant detailed consideration.

METHODOLOGICAL CRITIQUE

Antidepressant trials have been frequently critiqued for methodological problems (Moncrieff, 2001), and this trial is no exception. The dosing of fluoxetine monotherapy was not optimized, and most of the patients had depression with melancholic features, both of which may have disadvantaged fluoxetine monotherapy (Rush, 2010). Unsurprisingly, this type of confound favors mirtazapine, the drug manufactured by the sponsor of the trial (see Smith, 2005).

The basic design of this trial has been questioned. A letter to the editor published in response to this study (El-Mallakh, Kaur, & Lippman, 2010) argues that because the trial lacked a mirtazapine + placebo group, the authors' conclusions are invalid. This argument is critically important—if it is correct, other analyses may be beside the point.

This trial was preregistered, as is required of all contemporary trials (De Angelis et al., 2005). Trial registration consists of the investigators preregistering their outcome variables, so that the selective reporting so common in psychiatric trials (e.g., Turner, Matthews, Linardatos, Tell, & Rosenthal, 2008) can be prevented or at least identified. Comparing the trial registration record (i.e., "ISRCTN44468346—Assessment of Augmentation Strategies to Optimize the Therapeutic Response to Mirtazapine in Major Depression," 2005) with the published article facilitates the identification of "spin," which is often found in an industry-sponsored trials (Lexchin, 2011).

Of the seven registered outcome measures, four were statistically nonsignificant: (a) the Montgomery-Asberg Depression Rating Scale (MADRS); (b) the percentage of responders per MADRS score; (c) the Clinical Global Impressions (CGI) Severity Scale; (d) and the Improvement Scale (see Table 1). The Symptom Checklist-90-R (SCL-90-R) was preregistered as a secondary outcome, but the results are not reported. It is interesting that this heterogenous group of depression rating scales did not find combination therapy to be superior to fluoxetine monotherapy. The only positive findings from preregistered variables resulted from use of the Hamilton Depression Rating Scale (HAM-D). At the conclusion of the 6-week trial, HAM-D scores were roughly 4.5 points lower in the combination therapy group as compared to fluoxetine monotherapy, and there were more remitted patients in the combination therapy group.

The positive results for combination therapy found on the HAM-D should be considered in light of the purpose and existing critiques (e.g., Jacobs & Cohen, 2010) of the HAM-D. The HAM-D was developed to measure the impact of antidepressants (Healy, 1999) and heavily weights sleep and mood. A resolution of significant appetite and sleep problems can result in an improvement in as many as 10 points. A review of the adverse effects in the published trial (See Figure 2 in Blier et al., 2010) shows that combination therapy caused much higher rates of increased appetite and sedation than fluoxetine monotherapy. Some proportion of the reported improvement in HAM-D scores was likely the result of these "adverse" effects, unrelated to the mood component of clinical depression. This

TABLE 1. Comparison of Trial Registration to Published Results of Trial ID# ISCRTN44468346

Variable	Registration Status	Reported	Implication
Total HAM-D	Registered	~4.5 points lower in combination therapy	Primary evidence of efficacy
Total MADRS	Registered	*ns*	Does not support combination therapy
CGI improvement	Registered	*ns*	Does not support combination therapy
CGI severity	Registered	*ns*	Does not support combination therapy
% Responders	Registered	*ns*	Does not support combination therapy
% Remitters	Registered as HAM-D ≤8	Reported as HAM-D ≤7	Small discrepancy; either a change in endpoint or typographical error
SCL-90-R	Registered	Not reported	Registered secondary outcome unreported
6-month prolongation trial	Mentioned in registration but no outcomes specified	Withdrawal of combination meds leads to relapse	Unregistered outcome published in support of combination therapy

Note. *ns* = nonsignificant difference; HAM-D = Hamilton Depression Rating Scale; MADRS = Montgomery-Asberg Depression Rating Scale; CGI = Clinical Global Impressions; SCL-90-R = Symptom Checklist 90-R.

is particularly of note given the fact that other rating scales failed to find combination therapy effective.

Finally, the adverse effect profile of mirtazapine combination therapy could have compromised the blind (see Cohen, 2005). Patients on the mirtazapine combinations reported sedation at much higher rates than patients who were taking fluoxetine monotherapy. It is very likely that raters were able to discern who was taking combination therapy, which could have led to inflated perceived efficacy due to confirmatory bias (Moncrieff, 2008; Nickerson, 1998). Although the researchers could have tested the degree to which the blind was preserved, they did not do so.

STATISTICAL CRITIQUE

The authors make a strong argument in favor of combination mirtazapine therapy by claiming an NNT statistic of 3 to 5 for remission of depression. An NNT of 3 would mean that for every three patients who are treated with combination antidepressants as compared to those patients who are receiving fluoxetine monotherapy, one additional patient experiences remission of depression. This is a low number that indicates considerable

TABLE 2. Comparison of Point Estimates to Confidence Intervals

Medication Compared With Monotherapy	NNT	95% CI (Newcombe-Wilson Hybrid Score)
As claimed by authors:		
All combinations	3–5	Not provided
As calculated:		
All combinations	3.7*	(2.3, 18.0)
Fluoxetine + mirtazapine	3.7*	(2.0, 106.6)
Venlafaxine + mirtazapine	3.1*	(1.9, 15.5)
Bupropion + mirtazapine	4.7	*ns***

Note. CI = confidence interval; NNT = number needed to treat; *ns* = nonsignificant difference.
* $p < .05$. ***ns* result, confidence interval includes both harm and benefit.

efficacy; to compare, the advantage of venlafaxine over selective serotonin reuptake inhibitor (SSRI) monotherapy has been reported as an NNT of 17, with the 95% confidence interval running from 12 to 26 (Nemeroff et al., 2008).

Confidence intervals are the well-accepted standard for characterizing uncertainty for effect size statistics such as NNT (Fidler, Thomason, Cumming, Finch, & Leeman, 2004; Ziliak & McCloskey, 2008). Providing only a point estimate implies a false sense of certainty; therefore confidence intervals should be reported whenever an NNT statistic is given (Altman, 1998). In fact, many journals require confidence intervals for effect size statistics, likely because there is increasing recognition that omitting them is poor scientific practice (Ziliak & McCloskey, 2008). To illustrate the point, omitting confidence intervals is basically equivalent to a pollster reporting that a political candidate has 51% of the vote, but not reporting the margin of error: The information being withheld is necessary to understand the real-world significance of the results.

However, although the article was published in an elite psychiatric journal, no confidence intervals are reported, and so I calculated them manually (Newcombe, 1998) from the information provided in the article. The results suggest substantial uncertainty (see Table 2); the confidence intervals for each combination are wide, perhaps due to the small size of the subgroups. The NNT for fluoxetine + mirtazapine is just as likely to be 2.0 as 106.6, and the NNT for venlafaxine + mirtazapine ranges from 1.9 to 15.5. The NNT for bupropion + mirtazapine is not statistically significant, with the confidence interval ranging from harm to benefit, presenting a clinically important signal that such treatment could actually harm rather than help. If all mirtazapine combinations are combined, which is likely inappropriate, then the confidence interval ranges from 2.3 to 18.0. Thus, reporting NNT statistics using confidence intervals makes it clear that the effect size is much less certain than that reported by the authors.

DISCUSSION

Based on the critiques discussed in this article, there is seemingly insufficient evidence to justify combination antidepressant therapy with mirtazapine at the initiation of treatment.

Although the authors claim an NNT of 3 to 5, additional analysis suggests substantial uncertainty. The publication of this article possibly points to a failure of the peer-review process. Peer reviewers and/or journal editors could have insisted that confidence intervals for effect sizes be provided. However, although this issue was raised years ago (Simon, 2005), the statistical reporting standards of the *American Journal of Psychiatry* currently lag behind many social science journals. This may lend support to those who claim that psychiatric publishing lacks scientific standards (e.g., McLaren, 2009) or that many journal articles are propagandistic (Gambrill, 2010, 2011). Psychiatric journals can increase their credibility by insisting on rigorous scientific standards, including comprehensive reporting of statistical information.

NOTE

1. A much shorter version of this article was submitted to the *American Journal of Psychiatry* in April of 2010. It was rejected due to space constraints.

REFERENCES

Altman, D. G. (1998). Confidence intervals for the number needed to treat. *British Medical Journal*, *317*(7168), 1309–1312.

Blier, P., Ward, H. E., Tremblay, P., Laberge, L., Hébert, C., & Bergeron, R. (2010). Combination of antidepressant medications from treatment initiation for major depressive disorder: A double-blind randomized study. *American Journal of Psychiatry*, *167*(3), 281–288. doi:10.1176/appi.ajp.2009.09020186

Cohen, D. (2005). Clinical psychopharmacology trials: "Gold standard" or fool's gold? In S. Kirk (Ed.), *Mental disorders in the social environment: Critical perspectives* (pp. 347–367). New York: Columbia University Press.

De Angelis, C. D., Drazen, J. M., Frizelle, F. A., Haug, C., Hoey, J., Horton, R., et al. (2005). Is this clinical trial fully registered? A statement from the International Committee of Medical Journal Editors. *Annals of Internal Medicine*, *143*(2), 146–148.

El-Mallakh, R. S., Kaur, G., & Lippman, S. (2010). Placebo group needed for interpretation of combination trial. *American Journal of Psychiatry*, *167*(8), 996–997. doi:10.1176/appi.ajp.2010.10030453

Fidler, F., Thomason, N., Cumming, G., Finch, S., & Leeman, J. (2004). Editors can lead researchers to confidence intervals, but can't make them think: Statistical reform lessons from medicine. *Psychological Science*, *15*(2), 119–126. doi:10.1111/j.0963-7214.2004.01502008.x

Gambrill, E. (2010). Evidence-informed practice: Antidote to propaganda in the helping professions? *Research on Social Work Practice*, *20*(3), 302–320.

Gambrill, E., & Reiman, A. (2011). A propaganda index for reviewing problem framing in articles and manuscripts: An exploratory study. *PLoS One*, *6*(5), e19516. doi:10.1371/journal.pone.0019516

Healy, D. (1999). *The antidepressant era*. Cambridge, MA: Harvard University Press.

ISRCTN44468346—Assessment of augmentation strategies to optimize the therapeutic response to mirtazapine in major depression. (2005). Retrieved March 10, 2011, from http://www.controlled-trials.com/ISRCTN44468346

Jacobs, D. H., & Cohen, D. (2010). The make-believe world of antidepressant randomized controlled trials—An afterword to Cohen and Jacobs (2010). *Journal of Mind and Behavior*, *31*(1, 2), 23–36.

Leventhal, A. M., & Antonuccio, D. O. (2009). On chemical imbalances, antidepressants, and the diagnosis of depression. *Ethical Human Psychology and Psychiatry*, *11*(3), 199–214.
Lexchin, J. (2011). Those who have the gold make the evidence: How the pharmaceutical industry biases the outcomes of clinical trials of medications. *Science and Engineering Ethics*. doi:10.1007/s11948-011-9265-3
McLaren, N. (2009). Science and the psychiatric publishing industry. *Ethical Human Psychology and Psychiatry*, *11*(1), 29–36.
Moncrieff, J. (2001). Are antidepressants overrated? A review of methodological problems in antidepressant trials. *The Journal of Nervous and Mental Disease*, *189*(5), 288–295.
Moncrieff, J. (2008). *The myth of the chemical cure: A critique of psychiatric drug treatment*. New York: Palgrave.
Nemeroff, C. B., Entsuah, R., Benattia, I., Demitrack, M., Sloan, D. M., & Thase, M. E. (2008). Comprehensive analysis of remission (COMPARE) with venlafaxine versus SSRIs. *Biological Psychiatry*, *63*(4), 424–434. doi:10.1016/j.biopsych.2007.06.027
Newcombe, R. G. (1998). Interval estimation for the difference between independent proportions: Comparison of eleven methods. *Statistics in Medicine*, *17*(8), 873–890.
Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220.
Pigott, H. E. (2011). STAR*D: A tale and trail of bias. *Ethical Human Psychology and Psychiatry*, *13*(1), 6–28.
Rush, A. J. (2010). Combining antidepressant medications: A good idea? *American Journal of Psychiatry*, *167*(3), 241–243. doi:10.1176/appi.ajp.2009.09121768
Simon, S. D. (2005, December 15). *Confidence intervals are needed to evaluate clinical importance*. Unpublished letter to the American Journal of Psychiatry. Retrieved from http://www.webcitation.org/6021IguIR
Smith, R. (2005). Medical journals are an extension of the marketing arm of pharmaceutical companies. *PLoS Medicine*, *2*(5), e138. doi:10.1371/journal.pmed.0020138
Trivedi, M. H., Rush, A. J., Wisniewski, S. R., Nierenberg, A. A., Warden, D., Ritz, L., et al. (2006). Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: Implications for clinical practice. *American Journal of Psychiatry*, *163*(1), 28–40. doi:10.1176/appi.ajp.163.1.28
Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A., & Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *The New England Journal of Medicine*, *358*(3), 252–260. doi:10.1056/NEJMsa065779
Ziliak, S. T., & McCloskey, D. N. (2008). *The cult of statistical significance: How the standard error costs us jobs, justice, and lives*. Ann Arbor, MI: University of Michigan Press.

Jeffrey R. Lacasse is an assistant professor of the School of Social Work at Arizona State University in Phoenix. He is a board member of the International Society for Ethical Psychology and Psychiatry, and currently acts as secretary of Healthy Skepticism, an international nonprofit organization working to improve health by reducing harm from misleading drug promotion.

Correspondence regarding this article should be directed to Jeffrey R. Lacasse, PhD, Arizona State University, School of Social Work, 411 N. Central Avenue, Suite 871, Phoenix, AZ 85004-0689. E-mail: jeffrey.lacasse@asu.edu