

What Are the Comparative Benefits and Harms of Augmentation Treatments in Major Depression?

Richard C. Shelton, MD

The majority of depressed patients do not experience sufficient response to their initial antidepressant medication.¹ Augmentation strategies, particularly the use of atypical antipsychotics to augment selective serotonin reuptake inhibitors (SSRIs), have proliferated over the last 15 years, and 2 atypical antipsychotics, aripiprazole and quetiapine, have received US Food and Drug Administration approval for adjunctive therapy in treatment-resistant depression.² Alternatively, data in treatment-resistant depression exist for most of the atypicals and for other treatments, including lithium, bupropion, methylphenidate, pindolol, and buspirone. However, which is the most effective and safe? This question is difficult to answer without comparing across studies, since there are few head-to-head trials.

The traditional approach to summarizing and evaluating data across trials is meta-analysis, which is a statistical method for comparing results from different studies. Aggregating data in this way allows the assessment of the relative size of effect in a way that increases statistical power far beyond that found in an individual study. Meta-analysis also can help resolve discrepancies when studies disagree in their outcomes and conclusions. Meta-analysis not only increases the precision of estimates of effect but also tests for publication bias—that is, that only positive studies may have been previously published. The simplest approach involves identifying a measure, such as the Hamilton Depression Rating Scale (HDRS),³ that is in common across multiple trials. The data then can be aggregated across trials and compared against a common comparator (eg, another antidepressant), all comparators (if more than 1 treatment is used for comparison), or placebo.

There are major limitations with meta-analysis. For example, LeLorier et al⁴ compared earlier meta-analyses of small trials with the results of subsequent large-scale studies and found that 35% of the meta-analyses were not supported by the subsequent large-scale studies. One reason for the discrepant findings is that there are certain strict preconditions to meta-analysis that are sometimes violated.⁵ There are many steps that are needed to harmonize data across multiple sets. Considerations include the population

studied, the study design itself, differential sample sizes, outcome measures, rates of attrition, the length of the study, and other factors.⁵ These sources of heterogeneity can reduce the precision of the comparative effects. However, the principal limitation is that standard meta-analyses are typically intended to assess 1 intervention (eg, a specific antipsychotic) or a set of interventions compared to a single specific outcome—usually a placebo (or equivalent) control. While this approach does allow an estimate of the relative effect against a control condition, it does not give a direct estimate of comparative effectiveness. If there are several studies comparing a treatment against a particular comparator (or type of comparator), meta-analysis may be able to address the comparative effectiveness and harms of 2 treatments. However, in most cases, such data are not available. Therefore, traditional meta-analyses usually cannot answer a comparative effectiveness question across multiple treatments.

Network meta-analysis (sometimes called multiple treatment comparison meta-analysis or mixed treatment meta-analysis⁶) is intended to solve the latter problem in that it can summarize and compare data across trials with multiple treatments for a given indication when there is limited head-to-head data.⁵ The network meta-analysis begins with a geometrical arrangement of individual treatments (nodes) followed by lines connecting any nodes that have actually been compared directly. For a clinical trial in which A is compared against an active comparator (B) and placebo (C), both A and B would be connected with C and with each other. The relative effectiveness of these 3 nodes can then be contrasted, and the lines are then given statistical weighting. If, then, treatment A is compared against treatment D and placebo (C), then a relative weighting of A versus D can be calculated, but the relative effect of B and D can also be estimated since both were compared with A and C. Individual treatments can then be added to the matrix, generating relative effects for many different treatments. The “strength” or validity of the inferred comparative effectiveness is greater if a given node has had multiple comparisons in several different trials. Those with fewer connections have to be interpreted with greater caution than those with many.⁶ Appropriate use of network meta-analysis must still consider factors such as heterogeneity of design or population, which could be used as preselection variables for the studies included. However, the weight of individual sources of heterogeneity (or uncertainty) can be determined within the network via sensitivity analysis,⁷ which can be factored into a comparison. However, the

Submitted: October 14, 2014; accepted October 15, 2014.

Corresponding author: Richard C. Shelton, MD, Department of Psychiatry and Behavioral Neurobiology, The University of Alabama at Birmingham, SC 1026, 1720 2nd Ave South, Birmingham, AL 35294 (rshelton@uab.edu).

J Clin Psychiatry 2015;76(4):e531–e533 (doi:10.4088/JCP.14com09586).

© Copyright 2015 Physicians Postgraduate Press, Inc.

problem remains that the more heterogeneous the studies included in a network analysis are, the more uncertain are the conclusions.

The publication “Comparative Efficacy, Acceptability, and Tolerability of Augmentation Agents in Treatment-Resistant Depression: Systematic Review and Network Meta-Analysis” by Zhou et al⁸ in this issue of the *Journal* reports the results of a network meta-analysis of antidepressant augmentation trials in treatment-resistant depression involving 11 augmenting agents: aripiprazole, bupropion, buspirone, lamotrigine, lithium, methylphenidate, olanzapine, pindolol, quetiapine, risperidone, and thyroid hormone. A total of 48 studies involving 6,654 participants were initially evaluated. A 6-week time point in a study was used for comparison purposes if the 6-week data were available. If not, then the closest time point to 6 weeks was used. Different studies used either the HDRS or the Montgomery-Asberg Depression Rating Scale (MADRS)⁹ as the primary measure of outcome. Since no single scale was used across all studies, the primary outcome was the proportion of participants achieving *response*, defined as a 50% reduction in the relevant scale (HDRS or MADRS), a widely accepted definition.¹⁰ A secondary outcome was *remission*, defined as an HDRS score of ≤ 7 or a MADRS score of ≤ 10 ⁹; if these 2 scales were not employed, then a comparable definition was used. Other outcomes tested included acceptability outcome, which was discontinuation for any cause, and a tolerability outcome, which was discontinuation due to side effects.

The analysis involved 2 steps. The first was used in instances in which there were head-to-head comparisons. The differences were estimated using either a fixed or random effects model (depending on the level of heterogeneity of the studies), and outcomes were expressed as odds ratios (ORs) and 95% confidence intervals (CIs). Following that, a Bayesian random effects network analysis was conducted. The Bayesian model allows a test of the association of 2 variables conditional on the associations of a third. In our earlier model of A (a treatment) compared to C (placebo) and D (a second treatment) compared with C, the relationship between A and C can be assessed as conditional on the effect of D on C. In the Bayesian network meta-analysis, the effects are expressed as ORs and 95% credible intervals (CrIs), which are conceptually similar to confidence intervals.¹¹ The outcomes then are ranked based on ORs as the (first, second, third, etc) best regimens. Odds ratios generated by standard fixed or random effects models (ORs and CIs) can also be compared to the estimates generated by the Bayesian network analysis (ORs and CrIs) as a check on the model. This step was followed by a sensitivity analysis, which also tested the results when certain characteristics of studies were excluded, eg, dosage adequacy, treatment duration, blinded design, and other variables. This helps to test the relative impact of various sources of heterogeneity between studies.

The results are somewhat surprising. Of the various augmentation approaches tested, only quetiapine

(OR = 1.92), aripiprazole (OR = 1.85), thyroid hormone (OR = 1.84), and lithium (OR = 1.56) were significantly more effective than placebo based on response rates. For remission, thyroid hormone (OR = 2.94), risperidone (OR = 2.17), quetiapine (OR = 2.08), buspirone (OR = 1.86), aripiprazole (OR = 1.83), and olanzapine (OR = 1.79) were superior to placebo, although it should be noted that remission rates are typically low in treatment-resistant depression studies. There were no differences from placebo in overall acceptability (all-cause discontinuation), but there were important differences in tolerability compared with placebo: quetiapine (OR = 3.85), olanzapine (OR = 3.36), aripiprazole (OR = 2.51), and lithium (OR = 2.30) were significantly less well tolerated, and quetiapine was significantly less well tolerated than thyroid hormone. The sensitivity analysis, which corrected for sources of heterogeneity in studies, found stronger effects for aripiprazole and quetiapine than thyroid hormone and lithium.

What, then, can we make of the data? The strength of network meta-analysis relative to the results of individual studies is both statistical power and the ability to compare across multiple interventions when good head-to-head comparisons are not available. The principal weaknesses of the approach are consistent with the problems associated with meta-analysis in general. A meta-analysis is more reliable if the studies included are both large and similar in design and execution. The more heterogeneity in the studies, the less reliable are the estimated outcomes. The authors made a good effort to test for some of these effects by doing a sensitivity analysis in which studies that lacked certain characteristics were left out. However, the variation in the included studies introduces a good deal of uncertainty in the results.

The odds ratios generated in the analysis are relative to placebo. However, a major source of variability in outcomes in depression trials is placebo response (or remission). The studies analyzed had highly variable placebo response rates, which would then be expected to affect the odds ratios generated. To take examples from the augmentation trials with olanzapine, the placebo response rate ranged from 10%¹² to 50%.¹³ High placebo response in some of the clinical trials could certainly affect the data and conclusions. In the case of olanzapine, for example, high placebo effect in some trials would affect the OR in the meta-analysis (1.40).

The choice of 2 atypical antipsychotic augmentation medications, quetiapine and aripiprazole, as “winners” in the meta-analysis appears to ignore the rates of discontinuation due to side effects. For quetiapine and aripiprazole, the ORs for response were 1.92 and 1.85, respectively, while ORs for discontinuation were 3.85 and 2.51, respectively. Further, these were rates of intolerability in only short-term trials and do not take into consideration the risks associated with longer-term use, such as weight gain, metabolic syndrome, and tardive dyskinesia. In fairness, the authors do interpret their data with caution about the adverse effects of these medications. The beneficial effects of thyroid hormone (for response, OR = 1.84), by contrast, were comparable to both of

the atypical antipsychotics, and it was much more tolerable (OR = 1.36). The effects of thyroid hormone diminished in the sensitivity analysis because of heterogeneity. This can be attributable, in part, to the lack of well-designed larger-scale trials, which is due to the fact that no pharmaceutical company supported the studies. Thyroid hormone augmentation fared reasonably well in step 3 of the STAR*D study (although the study was not placebo controlled) with a 24.3% response rate in combination with several types of antidepressants (32.4% with the SSRI citalopram), which was better than other step 3 options.¹

The results of meta-analyses should always be interpreted with caution. Rather than necessarily informing clinical practice, perhaps the strongest use of network meta-analysis results is to inform future larger-scale trials. In the current funding and policy climate, these types of studies are very unlikely to happen. The typical current approach is to use large pragmatic effectiveness studies (including data from treatment networks). While there is great value in these studies, they seldom are able to distinguish the relative benefits of one treatment against another (although they are somewhat better at comparing harms). In the absence of larger controlled studies, meta-analysis may have to substitute.

Author affiliations: Department of Psychiatry and Behavioral Neurobiology, The University of Alabama, Birmingham.

Potential conflicts of interest: Dr Shelton has been a consultant to Bristol-Myers Squibb, Cerecor, Clintara, Cyberonics, Forest, Janssen, Medtronic, MSI Methylation Sciences, Naurex, Pamlab, Pfizer, Ridge Diagnostics, Shire Pic, and Takeda; and has received grant/research support from Alkermes, Assurex Health, Avanir, Cerecor, Elan, Forest, Janssen, Naurex, Novartis, Otsuka, Pamlab, and Takeda.

Funding/support: This commentary was supported by Pamlab, Inc.

Role of the sponsor: The sponsor had no role in the writing or review of the commentary.

REFERENCES

1. Rush AJ, Trivedi MH, Wisniewski SR, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry*. 2006;163(11):1905–1917.
2. Shelton RC, Osuntokun O, Heinloth AN, et al. Therapeutic options for treatment-resistant depression. *CNS Drugs*. 2010;24(2):131–161.
3. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960;23(1):56–62.
4. LeLorier J, Grégoire G, Benhaddad A, et al. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med*. 1997;337(8):536–542.
5. Walker E, Hernandez AV, Kattan MW. Meta-analysis: its strengths and limitations. *Cleve Clin J Med*. 2008;75(6):431–439.
6. Mills EJ, Thorlund K, Ioannidis JP. Demystifying trial networks and network meta-analysis. *BMJ*. 2013;346:f2914.
7. Kleijnen JPC. Sensitivity Analysis and Related Analyses: a Survey of Statistical Techniques. International Symposium Theory and applications of Sensitivity Analysis of Model Output in computer simulation. Belgirate, Italy; September 25–27, 1995.
8. Zhou X, Ravindran AV, Qin B, et al. Comparative efficacy, acceptability, and tolerability of augmentation agents in treatment-resistant depression: systematic review and network meta-analysis. *J Clin Psychiatry*. 2015;76(4):e487–e498.
9. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. 1979;134(4):382–389.
10. Frank E, Prien RF, Jarrett RB, et al. Conceptualization and rationale for consensus definitions of terms in major depressive disorder: remission, recovery, relapse, and recurrence. *Arch Gen Psychiatry*. 1991;48(9):851–855.
11. Chen M-H, Shao Q-M. Monte Carlo Estimation of Bayesian Credible and HPD Intervals. *J Comput Graph Statist*. 1999;8(1):69–92.
12. Shelton RC, Tollefson GD, Tohen M, et al. A novel augmentation strategy for treating resistant major depression. *Am J Psychiatry*. 2001;158(1):131–134.
13. Corya SA, Williamson D, Sanger TM, et al. A randomized, double-blind comparison of olanzapine/fluoxetine combination, olanzapine, fluoxetine, and venlafaxine in treatment-resistant depression. *Depress Anxiety*. 2006;23(6):364–372.