Efficacy of new generation antidepressants: differences seem illusory

Del Re, A C; Spielmans, Glen I; Flückiger, Christoph; Wampold, Bruce E

Abstract: BACKGROUND: Recently, Cipriani and colleagues examined the relative efficacy of 12 new-generation antidepressants on major depression using network meta-analytic methods. They found that some of these medications outperformed others in patient response to treatment. However, several methodological criticisms have been raised about network meta-analysis and Cipriani’s analysis in particular which creates the concern that the stated superiority of some antidepressants relative to others may be unwarranted. MATERIALS AND METHODS: A Monte Carlo simulation was conducted which involved replicating Cipriani’s network meta-analysis under the null hypothesis (i.e., no true differences between antidepressants). The following simulation strategy was implemented: (1) 1000 simulations were generated under the null hypothesis (i.e., under the assumption that there were no differences among the 12 antidepressants), (2) each of the 1000 simulations were network meta-analyzed, and (3) the total number of false positive results from the network meta-analyses were calculated. FINDINGS: Greater than 7 times out of 10, the network meta-analysis resulted in one or more comparisons that indicated the superiority of at least one antidepressant when no such true differences among them existed. INTERPRETATION: Based on our simulation study, the results indicated that under identical conditions to those of the 117 RCTs with 236 treatment arms contained in Cipriani et al.’s meta-analysis, one or more false claims about the relative efficacy of antidepressants will be made over 70% of the time. As others have shown as well, there is little evidence in these trials that any antidepressant is more effective than another. The tendency of network meta-analyses to generate false positive results should be considered when conducting multiple comparison analyses.

DOI: https://doi.org/10.1371/journal.pone.0063509

Posted at the Zurich Open Repository and Archive, University of Zurich
ZORA URL: https://doi.org/10.5167/uzh-86434
Journal Article
Published Version

Originally published at:
Del Re, A C; Spielmans, Glen I; Flückiger, Christoph; Wampold, Bruce E (2013). Efficacy of new generation antidepressants: differences seem illusory. PLoS ONE, 8(6):e63509.
DOI: https://doi.org/10.1371/journal.pone.0063509
Efficacy of New Generation Antidepressants: Differences Seem Illusory

A. C. Del Re1,2*, Glen I. Spielmans3, Christoph Flückiger4, Bruce E. Wampold5,6

1 VA Palo Alto Health Care System, Palo Alto, California, United States of America, 2 Stanford University Medical School, Stanford, California, United States of America, 3 Metropolitan State University, St. Paul, Minnesota, United States of America, 4 University of Zurich, Zurich, Switzerland, 5 University of Wisconsin-Madison, Madison, Wisconsin, United States of America, 6 Research Institute, Modum Bad Psychiatric Center, Ykersund, Norway

Abstract

Background: Recently, Cipriani and colleagues examined the relative efficacy of 12 new-generation antidepressants on major depression using network meta-analytic methods. They found that some of these medications outperformed others in patient response to treatment. However, several methodological criticisms have been raised about network meta-analysis and Cipriani’s analysis in particular which creates the concern that the stated superiority of some antidepressants relative to others may be unwarranted.

Materials and Methods: A Monte Carlo simulation was conducted which involved replicating Cipriani’s network meta-analysis under the null hypothesis (i.e., no true differences between antidepressants). The following simulation strategy was implemented: (1) 1000 simulations were generated under the null hypothesis (i.e., under the assumption that there were no differences among the 12 antidepressants), (2) each of the 1000 simulations were network meta-analyzed, and (3) the total number of false positive results from the network meta-analyses were calculated.

Findings: Greater than 7 times out of 10, the network meta-analysis resulted in one or more comparisons that indicated the superiority of at least one antidepressant when no such true differences among them existed.

Interpretation: Based on our simulation study, the results indicated that under identical conditions to those of the 117 RCTs with 236 treatment arms contained in Cipriani et al.’s meta-analysis, one or more false claims about the relative efficacy of antidepressants will be made over 70% of the time. As others have shown as well, there is little evidence in these trials that any antidepressant is more effective than another. The tendency of network meta-analyses to generate false positive results should be considered when conducting multiple comparison analyses.

Findings:

Interpretation:

Citation: Del Re AC, Spielmans GJ, Flückiger C, Wampold BE (2013) Efficacy of New Generation Antidepressants: Differences Seem Illusory. PLoS ONE 8(6): e63509. doi:10.1371/journal.pone.0063509

Editor: Christian Gluud, Copenhagen University Hospital, Denmark

Received January 11, 2013; Accepted March 20, 2013; Published June 3, 2013

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: Dr. AD is supported by a postdoctoral affiliation with the Center for Health Care Evaluation at the VA Palo Alto Health Care System. Dr. CF is supported by the University of Zurich. Dr. GS is supported by Metropolitan State University. Dr. BW is supported by the University of Wisconsin – Madison and Modum Bad Psychiatric Center. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: acdelre@gmail.com

Introduction

Recently, Cipriani and colleagues [1] examined the relative efficacy of 12 new-generation antidepressants on major depression. They applied a random-effects meta-analytic model that used a Bayesian approach [2] (often referred to as network meta-analysis) to examine 117 randomized controlled trials (RCTs) and concluded, “Mirtazapine, escitalopram, venlafaxine, and sertraline were significantly more efficacious than duloxetine (estimated odds ratios [OR] 1.39, 1.33, 1.30 and 1.27, respectively), fluoxetine (1.37, 1.32, 1.28, and 1.25, respectively), fluvoxamine (1.41, 1.35, 1.30, and 1.27, respectively), paroxetine (1.35, 1.30, 1.27, and 1.22, respectively), and reboxetine (2.03, 1.95, 1.89, and 1.85, respectively) [and that] reboxetine was significantly less efficacious than all the other antidepressants tested” (pgs. 746). If these results are reliable, this meta-analysis would have important implications for clinical practice.

Determining the relative efficacy of competing treatments for a particular disorder is of critical importance in evidence-based medicine for improving the quality of care and reducing costs. Cipriani et al.’s efforts to use a sophisticated method to determine which antidepressants are more effective than others is commendable. Nevertheless, a number of concerns have been raised with regard to network meta-analysis [3–6]. For example, Trinquart, Abbé, and Ravaud [3] showed that reporting bias had a particularly pernicious effect on the results of network meta-analyses because the bias extended to treatments for which there was no bias; in this way the reporting bias of a particular treatment affected the ranking of all treatments, regardless of whether there was bias for the other treatments. Further, in antidepressant research, selective publication of results from placebo-controlled trials has been well-documented [7]. Across the medical literature in general, including head-to-head antidepressant trials, reporting bias occurs frequently [8–11]. Indeed, the reported superiority of escitalopram over citalopram reported in Cipriani et al. was
Differences in Antidepressant Efficacy

Materials and Methods

A Monte Carlo simulation was conducted that involved replicating Cipriani’s network meta-analysis under the null hypothesis (i.e., no true differences between antidepressants). An overview of the simulation strategy is as follows: (1) 1000 simulations were generated under the null hypothesis of no differences among the 12 antidepressants, (2) each of the 1000 simulations were network meta-analyzed, (3) the total number of false positive results from the network meta-analyses were calculated.

One-thousand simulated datasets were generated that were equivalent to Cipriani’s dataset in all aspects (viz., sample size, antidepressants, and treatment comparisons). The response for each participant in each study was randomly generated under a null hypothesis of no differences in which the probability of a positive response was equal to the response rate for the total sample reported in Cipriani et al. The studies in Cipriani’s meta-analysis included 25,928 participants, of whom 0.57 had a positive response to treatment. Consequently, in each of the simulations, the probability that an individual participant, regardless of arm in the study, would have a positive response was set at 0.57 prior to running the simulations. A 1 was assigned for a positive response, and 0 otherwise, such that \( P(\text{response} = 1) = 0.57 \). The response of all participants in the 117 trials were randomly repeated 1000 times under the null hypothesis that the probability of a response was not conditioned on the antidepressant administered.

The simulated data were then analyzed with network meta-analysis using the ‘netmeta’ network meta-analysis package in the R statistical software program [13]. Odds Ratios \( (OR) \) were calculated from the resulting logistic point estimates, along with the log odds of the standard errors. Before conducting the simulation study, the software was validated by analyzing the results of the 117 RCTs and we found that the results were in accord with Cipriani et al. at a level of precision of \( \pm 0.01 \). For each of the 1000 simulations, the OR as well the standard error for each of the \( k(k−1)/2 = 66 \) comparisons between pairs of the \( k = 12 \) antidepressants were used to determine the statistical significance of the pairwise comparisons: A false positive occurred when the 95% confidence interval for the OR did not include 1, indicating that one antidepressant was considered to be superior to the comparison.

Results

For each of the 1000 replications, let \( x \) be the number of false positives (i.e., falsely concluding that antidepressant A was superior to antidepressant B based on \( \sigma = 0.05 \)) obtained for that replication. The values of \( x \) range from 0 (for that replication, there were no statistically significant differences) to 66 (all pairwise comparisons for the replication yielded statistical significance). A frequency \( f(x) \) for each value of \( x \) was constructed to indicate the number of replications that produced that value of \( x \). The probability distribution \( X \) was determined by computing \( f(x)/1000 \) for each value of \( x \). That is, \( P(X = x) = f(x)/1000 \), which is the probability of obtaining \( x \) number of significant comparisons. If network meta-analysis maintains the proper family-wise error rates, the probability of obtaining one or more false positives would be less than 0.05 (i.e., \( P(X \geq 1) < 0.05 \)). The probability distribution of \( X \) is shown in Figure 1, truncated at \( X = 18 \). For this distribution \( P(X \geq 1) = 0.72 \); thus, in more than 7 times out of 10, the network meta-analysis resulted in one or more comparisons that indicated the superiority of at least one antidepressant when no such true differences among them existed.

The expectation of the probability distribution of \( X \) was equal to 2.68 (i.e., \( E(X) = \sum_{x=0}^{66} x \cdot P(X = x) = 2.68 \)). On average, it is expected that 2.68 significant differences will be detected by chance. Interestingly, if the 66 comparisons were conducted independently as a binomially distributed random variable with parameters \( n = 66 \) comparisons and probability of false rejection of \( p = 0.05 \), the expectation would be \( np = 3.30 \). Consequently, network meta-analysis offers some protection, in that the expected number of false positives was reduced from 3.30 to 2.68. However, the relative reduction is not comforting as there remains a probability of 0.72 of detecting one or more differences by chance.

Discussion

Based on the results of their network meta-analysis, Cipriani et al. [1] concluded, “Our findings might help to choose among new-generation antidepressants” (p. 753). Conclusions about the
relative efficacy of medical antidepressants should only be made if the null hypothesis of no treatment differences is rejected, conventionally based on an error rate of 0.05. Based on our simulation study, the results indicated that under identical conditions to those of the 117 RCTs with 236 treatment arms contained in Cipriani et al.’s meta-analysis, one or more false claims about efficacy will be made over 70% of the time. As others [5], [6], [12] have shown, the pattern of results in the 117 RCTs are consistent with the null hypothesis that the 12 antidepressants are equally efficacious, so the findings of those analyses and our simulation suggest that Cipriani et al.’s conclusions should not be relied upon to make clinical decisions.

The results of the simulation study are specific to the parameters of the RCTs in the Cipriani meta-analysis in terms of number of trials, number of participants per trial, and response rate. Consequently, network meta-analyses may perform differently under different scenarios. However, it is clear that the use of network meta-analysis to examine all k(k−1)/2 pairwise comparisons of k treatments with conventional alpha levels is problematic. The implications of conducting multiple statistical tests on the family-wise error rates are well known [14].

If one is interested in the comparison of two particular treatments, say Treatment A and Treatment B, and there are insufficient number of trials comparing A and B, then the use of a method that increases the power of the comparison by incorporating information from indirect paths, in the way that network meta-analysis does, may be warranted. In that way, the method is used to test a focused hypothesis with a powerful meta-analytic strategy [15–17]. One of the advantages of Bayesian MTC methods over conventional meta-analysis is that one can examine and take into consideration the likelihood of the alternative hypothesis (i.e., estimate the probability that a treatment should be preferred). For example, conducting a rank probability test using MTC methods will estimate the probability of each treatment to be superior to other treatments (i.e., the probability of treatment A ranking first, ranking 2nd, etc and the probability of treatment B ranking first, and so on).

If one wanted to extend the method to compare one treatment, say treatment A, to various other treatments, say B, C, and D, then one could compare A to the effects of B, C, and D pooled, although the details of how this would work in the context of network meta-analysis has not yet been derived. In the antidepressant trials reviewed in Cipriani et al., reboxetine appears to be inferior to other antidepressants, a result that has been meta-analytically examined and confirmed [9]. If this is the conjecture of interest, then the analysis should be focused on the relative efficacy of reboxetine versus other antidepressants. Indeed, we examined the 8 trials in Cipriani et al. that directly compared reboxetine to one of the other 12 antidepressants and conducted a standard meta-analysis. A restricted maximum likelihood univar-

![Figure 1. Probability of Significant Contrasts from Simulation (66 Contrasts).](image-url)
Conjecture.
not needed to reach this conclusion and emphasizing the
p = .03), suggesting in this case that utilizing the indirect paths was
was inferior to the other antidepressants (OR = 0.81, [0.68, 0.98],
iate meta-analysis [18–20] of these 8 trials reveals that reboxetine
similar to the antidepressants versus
unacceptably high. Such analyses do not address whether the set of
when applying this method, as the probability of false positive is
than others. We have shown in one case that caution must be used
The purpose of network meta-analysis is to identify which
antidepressants from a set of antidepressants are more effective
than others. We have shown in one case that caution must be used when
applying this method, as the probability of false positive is
unacceptably high. Such analyses do not address whether the set of
antidepressants is effective, for which would be established by
conventional meta-analyses of the antidepressants versus placebo controls.

Conclusion

The results of this simulation indicated that under identical
conditions to those of Cipriani et al.’s meta-analysis, one or more
false claims about efficacy will be made 71% of the time. As other
research has shown, the pattern of results in the 117 RCTs is
consistent with the null hypothesis that the 12 antidepressants are
equally efficacious. The findings of those analyses and our
simulation suggest that Cipriani et al.’s conclusions should not be
relied upon to make clinical decisions, although the inferiority of
reboxetine was confirmed by a meta-analysis of direct
comparisons.

Supporting Information

Appendix S1  R code to Replicate Output of MTC Meta-

Analysis.

Acknowledgments

Disclaimer

The views expressed here are those of the authors and do not necessarily
represent those of the Department of Veterans Affairs.

Author Contributions

Conceived and designed the experiments: AD CF GS BW. Performed the
experiments: AD. Analyzed the data: AD. Contributed reagents/
materials/analysis tools: AD CF GS BW. Wrote the paper: AD CF GS BW.

References

1. Cipriani A, Furukawa TA, Salanti G, Geddes JR, Higgins JP, et al. (2009)
Comparative efficacy and acceptability of 12 new-generation antidepressants: A
multiple-treatments meta-analysis. The Lancet 373: 746–758.
2. Ades AE, Sculpher M, Sutton A, Abrams K, Cooper N, et al. (2006) Bayesian
Methods for Evidence Synthesis in Cost-Effectiveness Analysis. Pharmacoecon-
omics 24: 1–19.
3. Trinquart L, Abbé A, Ravaud P (2012) Impact of Reporting Bias in Network
Meta-Analysis of Antidepressant Placebo-Controlled Trials. PLoS ONE 7: 1–8.
4. Trinquart L, Chatellier G, Ravaud P (2012) Adjustment for reporting bias in
network meta-analysis of antidepressant trials. BMC Med Res Methodol 12: 1–15.
5. Wampold BE, Serlin RC (2013) Meta-analytic methods to test relative efficacy.
Quality and Quantity. DOI: 10.1007/s11135-012-9800-6. In press.
6. Imberger G, Vejlby AD, Hansen SB, Moller AM, Wetterslev J (2011) Statistical
Multiplicity in Systematic Reviews of Anaesthesia Interventions: A Quantifica-
tion and Comparison between Cochrane and Non-Cochrane Reviews. PLoS
ONE 6(12): e28422.
7. Howard GS, Maxwell SE, Fleming KJ (2000)The proof of the pudding: An
illustration of the relative strengths of null hypothesis, meta-analysis, and
Bayesian analysis. Psychol Methods 5: 315–332.
8. Gartlehner G, Hansen RA, Morgan LC, Thaler K, Lux L, et al. (2011) Comparative
effects and harms of second-generation antidepressants for treating major
depressive disorder: An updated meta-analysis. Ann Intern Med 155: 772–785.
9. van Valkenhorst G, Kuijer J (2013) gemtc: GeMTC network meta-analysis.
R package version 0.1–3.
10. Ioannidis JP (2009) Ranking antidepressants. Lancet 373: 1739–60.
11. Svensson S, Mansfield PR (2004) Escitalopram: Superior to citalopram or a
chiral chimera? Psychopharmacology 73: 10–16.
12. Eyding D, Lelgemann M, Grouven U, Harter M, Kromp M, et al. (2010)
meta-analysis of published and unpublished placebo and selective serotonin
reuptake inhibitor controlled trials. BMJ 341: c4737.
13. van Valkenhorst G, Kuijer J (2013) gemtc: GeMTC network meta-analysis.
R package version 0.1–3.
14. Rosnow RL, Rosenthal R (1988) Focused tests of significance and effect size
estimation in counseling psychology. J Couns Psychol35: 203–208.
15. Imberger G, Vejlby AD, Hansen SB, Moller AM, Wetterslev J (2011) Statistical
Multiplicity in Systematic Reviews of Anaesthesia Interventions: A Quantifica-
tion and Comparison between Cochrane and Non-Cochrane Reviews. PLoS
ONE 6(12): e28422.
16. Howard GS, Maxwell SE, Fleming KJ (2000)The proof of the pudding: An
illustration of the relative strengths of null hypothesis, meta-analysis, and
Bayesian analysis. Psychol Methods 5: 315–332.
17. Wampold BE, Davis B, Good RH III (1990) Hypothesis validity of clinical
research. J Consult Clin Psychol 58: 360–367.
18. Viechtbauer W (2010) Conducting meta-analyses in R with the metafor
package. J Stat Software 36: 1–48.
19. Del Re AC, Hoyt WT (2012) MAd: meta-analysis with mean differences.
R package version 0.8.
20. Del Re AC (2013) RcmdrPlugin. MA: Graphical User Interface for Meta-
Analysis in R. R package version 0.0–1.
21. Levin JR, Serlin RC, Seaman MA (1994) A controlled, powerful multiple-
comparison strategy for several situations. Psych Bull; 115: 153–159.