BMJ Open

# What are the chances for personalized treatment with antidepressants? Detection of patient by treatment interaction with a variance ratio meta-analysis



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2019-034816 |
| Article Type: | Original research |
| Date Submitted by the Author: | 08-Oct-2019 |
| Complete List of Authors: | Plöderl, Martin; Christian Doppler Clinic, Paracelsus Medical University, Dept. of Clinica Psychology and Dept. for Crisis Intervention and Suicide Prevention, University Clinic for Psychiatry, Psychotherapy, and Psychosomatics<br>Hengartner, Michael; Zurich University of Applied Sciences/ZHAW, Section for Clinical Psychology and Health Psychology |
| Keywords: | Depression & mood disorders < PSYCHIATRY, STATISTICS & RESEARCH METHODS, PSYCHIATRY |
| | |

# *What are the chances for personalized treatment with antidepressants? Detection of patient by treatment interaction with a variance ratio meta-analysis*

Martin Plöderl[1]*, Michael P. Hengartner[2]

[1] Department of Crisis Intervention and Suicide Prevention/Christian Doppler Clinic, Paracelsus Medical University, Salzburg, Austria.

[2] Department of Applied Psychology, Zurich University of Applied Sciences, Zurich, Switzerland.



**Short Title:** Personalized antidepressants treatment

**Word count:** 2932

***Corresponding Author**

Martin Plöderl PhD

Dept. of Crisis Intervention and Suicide Prevention

Christian Doppler Klinik,

Paracelsus Medical University,

Ignaz Harrer-Straße 79

A-5020 Salzburg

AUSTRIA

Phone: +43-5-7255-56679

Fax: +43-5-7255-4344

E-mail: M.Ploederl@salk.at

# ABSTRACT

## Objectives

To investigate if there is notable patient by treatment interaction in the pharmacological treatment of depression, a necessary but largely unexplored prerequisite of personalized antidepressant treatment.

## Design

Meta-analytic variance-comparison of treatment-outcome between drug arms and placebo arms of clinical trials, based on the assumption that patient by treatment interaction should lead to larger variances in drug arms than placebo arms. To put the results into context, we run simple simulations, assuming different definitions and rates of those who respond especially well to antidepressants.

## Data sources

163 randomized, placebo-controlled trials (51,137 patients) with complete results for pre-post differences, selected from a recently published systematic review.

## Analysis

Variance ratios (VR) and coefficients of variance ratios (CVR) of individual trials were meta-analytically combined. The analysis was repeated for classes of antidepressants and specific antidepressants.

## Results

Variance ratios (VR = 1.01, CI 0.99-1.02) and coefficient of variance ratios (CVR = 0.82, CI 0.80-0.84) of the antidepressant-treatment arms were comparable or smaller than in placebo-arms. Similar results were observed for classes of antidepressants and for specific antidepressants. Our simulation

analysis confirmed that equal variance ratios can only be obtained if they are not more than a few patients who respond slightly above average.

## Conclusions

The lack of increased of treatment-outcome variance in the antidepressants versus placebo groups in RCT indicates that no or only very small subgroups of patients respond particularly well to antidepressants. Thus, the scope for personalized treatment with antidepressants seems to be limited.

**Keywords:** antidepressants, personalized medicine, treatment heterogeneity, variance ratio, meta-analysis

## STRENGTHS AND LIMITATIONS OF THIS STUDY

- For the first time, the amount of patient by treatment interaction (treatment heterogeneity), a necessary prerequisite for personalized medicine, is estimated for the pharmacological treatment of depression with antidepressants.
- The data-base is from a systematic review of published and unpublished studies and is one of the largest so far, resulting in precise estimations of the main outcomes.
- The study results are important to inform further attempts in personalized medicine in psychiatry.
- As with all clinical trials, it remains an open question if our results can be replicated in real-world settings, for example among psychiatric inpatients with very severe depression.

**SUMMARY BOX**

**What is already known on this topic?**

- There is massive research effort to find subgroups of patients who respond especially well to antidepressant treatment (personalized medicine).
- Personalized antidepressant treatment rests on the assumption that treatment heterogeneity exists, i.e., that there is notable patient by treatment interaction.
- The existence of notable treatment heterogeneity has been questioned in other fields but has not been explored for antidepressant treatment.

**What this study adds**

- The results from our study suggest that no or only a very small subgroups of patients respond particularly well to antidepressants.
- The success for research into personalized treatment with antidepressants seems to be limited.
- The average drug-placebo difference appears to be the most accurate prediction of treatment response for individual patients

## INTRODUCTION

Personalized or precision medicine, i.e., applying medical interventions only to those patients known to benefit especially well to the intervention (henceforth termed "benefiters"), is important to increase benefits from treatment and to decrease harms. For example, if a drug with severe side effects is very effective in some patients with a specific genotype, it is crucial to know about these benefiters, because for all other patients, the risk-benefit ratio would be unfavorable. Similarly, if a drug is found to have only modest efficacy across all patients but notable side-effects, than it would be important to know if there are patients (e.g., those defined via a specific biomarker) who are benefiters. The latter example corresponds with the pharmacological treatment of major depression, because the average efficacy of antidepressants (AD) is modest, corresponding to, on average, only about 2 points difference on the Hamilton Depression Rating Scale (HDRS) between AD groups and placebo in randomized controlled short-term trials (RCT).[1-5] Put differently, according to our most recent estimate, there is about 88% overlap in distribution of depression scores between antidepressants and placebo at the end of acute treatment.[2]

Despite substantial research efforts, no predictors of treatment success with AD were found that were robust and reliable enough for use in clinical practice.[6-9] Thus, much of the variance of the treatment outcome remains unexplained thus far. Sources of outcome variation include variation between treatment arms (indicating that group means differ due to efficacy of treatments, e.g., AD versus placebo), variation between patients (indicating that the outcome differs from patient to patient, independent of the treatment received), variation within patients (indicating that the outcome for the same patient differs over time due to random symptom fluctuations), and patient-by-treatment interactions (indicating that treatment effects vary from patient to patient).[10] The quest for personalized medicine in the case of AD assumes that specific patients benefit more from AD than others, i.e., assumes that there is a patient-by-treatment interaction. Ideally, this can also be explained by a plausible causal mechanism, e.g. inter-individual differences in monoamine function.

Investing research efforts in personalized medicine only makes sense if there truly is a patient-by-drug interaction that explains some variance in the treatment outcome. Although the field is mostly enthusiastic about personalized medicine or precision psychiatry,[11] experts from various fields now start to dampen expectations and caution that personalized/precision medicine may fall short of expectations.[12-14]

Thus, we must remain mindful that there might be no notable subgroup of true AD benefiters and that the modest average treatment effect is the best we can hope for.[15] We further need to acknowledge that RCT are inherently limited to demonstrate patient-by-treatment interactions.[10] To identify patient-by-treatment interactions, repeated period cross-over trials are necessary, but these are hardly feasible with common AD due to delayed onset of therapeutic effect and relatively high rates of spontaneous remission. The most common trial design is the simple parallel-group trial, where patients are randomized to either AD or placebo. However, these trials can only identify mean differences between treatment arms (i.e., efficacy), whereas variation between patients, within patients, as well as patient-by-treatment interactions are part of the error term. Nevertheless, if patient-by-treatment interaction effects are present, then the variance in the treatment outcome should be increased in the drug group relative to the placebo group, because no comparable drug-by-patient interaction is present in the placebo group.[10 16 17] Thus, results from RCT can inform indirectly if there might be subgroups of benefiters.

The goal of this meta-analysis was to examine whether the outcome-variances between AD and placebo differ, in order to gauge the potential of personalized/precision psychiatry for treatment with AD.

# METHODS

## Data

Our analysis was based on short-term RCT of AD for patients with unipolar major depression, reported in the most recent systematic review.[18] The authors of this comprehensive study made the data available in a public repository https://data.mendeley.com/datasets/83rthbp8ys/2. This included 522 trials (with 21 different AD), of which 253 trials were suitable for further analysis, i.e., contained information about the outcome and also included a placebo-arm. In case trials had multiple treatment arms with different dosages of AD, these arms were aggregated. In case of trials compared different AD, the data of these arms were aggregated to only have one value for the drugs in these trials, similar as in a previous publication.[16] Additionally, we recorded different AD by their class (SNRI, SSRI, atypical AD, and tricyclic AD). For 168 (66%) of studies, the pre-post mean reduction of depression scores (M) and the related standard deviations (SD) were available, and only the analysis for these studies are reported here. Analysis for the 85 studies (34%) where only the mean value and standard deviation of the post-treatment depression scores were available are reported in an online supplement (https://osf.io/98kex/files/). Several additional variables were created for sensitivity analysis (see below).

## Patient and public involvement statement

This research was done without patient involvement. Patients were not invited to comment on the study design and were not consulted to develop patient relevant outcomes or interpret the results. Patients were not invited to contribute to the writing or editing of this document for readability or accuracy.

**Statistical analysis**

Meta-analysis

We calculated the variance ratio (VR) for each RCT and aggregated them by means of a random effect meta-analysis according to the procedure suggested by Winkelbeiner et al.;[16] see https://osf.io/qarvs/files/, using the metafor package in R. Because the pre-post-differences were significantly associated with their SD, we repeated the analysis using the coefficient of the variance ratio (CVR). This removes the effect of expected changes in the standard deviation due to changes in the mean.[19] A VR of 1.00 suggest equal variance of AD and placebo. If the VR exceeds 1, then the variance of the AD group is larger than in the placebo group. If the CVR exceeds one, than the increase of variance with increasing pre-post-differences is stronger in the AD than in the placebo arms. Sensitivity analysis included meta-regression models with the assessment instruments, year of publication, type of publication (published vs. unpublished), sample size, and drop-out rates.

Simulation analysis

To bring our results into context, we also ran simulation analysis with different definitions and probabilities for benefiters. We based these simulations on an efficacy of 2 points mean-difference between AD and placebo groups, as reported in a meta-analysis on the same dataset for trials using the HDRS-17 instrument.[5] We assumed a standard deviation of SD=8 and a mean difference between pre-and post-depression-scores of 11 points in the AD group (based on rounded means of these values observed in our data-base). We used different cut-offs to define patients as benefiters, ranging from 5 to 10 points superior treatment outcome with the HDRS-17, and a proportion of 5 to 50% in the AD group vs. 0% in the placebo group. To simulate placebo groups, we sampled from a normal distribution with the above parameters for the placebo group (M = 9, SD = 8, and 5,000,000 samples). We used a similar sampling procedure for the AD-group, but created benefiters by adding necessary HDRS responder points to an assumed fraction of the sample, and adding as much points

to the rest of the sample to end up with the overall efficacy of 2 HDRS points.

The R-code and data of this publication is available online (https://osf.io/98kex/files/).

## RESULTS

### Meta-analysis

Across all AD, the variance ratio was almost perfectly 1.00 with a narrow confidence interval (VR = 1.01, 95%CI = 0.99 – 1.02). This means that the variances in the AD and placebo groups are identical (Table 1). Similar findings were found for all classes of AD (Table 1) and for each individual drug (online supplement https://osf.io/98kex/files/). There was no significant sign of heterogeneity in the meta-analyses, except for SSRIs. A closer inspection revealed that this resulted from a single outlier (see footnote in Table 1).

----- Insert Table 1 around here -----

For the coefficient of variance ratio, results indicated that the increase of variance with increasing pre-post differences is less strong in AD than in placebo arms (CVR = 0.82, 95%CI = 0.80-0.84). Comparable results were found for all classes of AD and individual drugs (Table 1). The heterogeneity was statistically significant in nearly all meta-analyses of the CVR.

In the sensitivity analysis, the meta-regression models could not detect statistically significant effects for year of publication, type of publication, measurement-instruments, and sample size (see https://osf.io/98kex/files/). However, there was a small but statistically significant effect of the drop-out rates in drug-arms and the VRs. With increasing drop-out rates, the VRs slightly increased ($p < .04$).

**Simulation analysis**

As shown in Figure 1, most benefiter assumptions lead to VRs much different from those we observed in our study.

----- Insert Figure 1 around here -----

However, for liberal definitions of benefiters and low rates of these benefiters, the VRs can indeed be small. For example, if there are 10% benefiters, as defined with 6 HDRS points difference to average placebo response, the VR is within the confidence interval of our main finding (0.99-1.02).

## DISCUSSION

We found nearly identical treatment outcome variances for AD arms compared to placebo in RCT for the acute treatment of major depression in a large database, as indicated by variance ratios almost perfectly being VR = 1. The simplest explanation for the finding of similar variances is that there are constant treatment effects and no treatment heterogeneity, i.e. no patient-by-treatment interaction effects and no specific subgroups of patients who respond particularly well to the treatment.[20] Alternatively, such a subgroup of benefiters would be very small (≤10%) and the threshold to classify someone as benefiter would be low (≤6 HDRS points difference to average placebo response). To put this in context, according to anchor-based linkage studies, at least 6 points on the HDRS are necessary for a global impression of "minimally improved".[21 22] Consequently, the search for meaningful predictors of relative treatment response (compared to placebo) will probably fail or will at least be very difficult due to the small subgroup of weak benefiters. Therefore, the mean effect size estimate from parallel-group RCT remains the best guess for predicting treatment outcome for an individual patient. Furthermore, the results for the coefficient of variance indicated that the increase of variance associated with increasing larger pre-post-differences was stronger in the

placebo than the AD groups. There is no immediately plausible explanation for this finding, given that baseline severity does not predict differential treatment effects.[23-25]

Our findings are in line with Senn,[12] who argued that exploratory post-hoc delineation of putative benefiters, such as the "true benefiters" suggested by Thase *et al*.,[26] are simply statistical artefacts due to random symptom fluctuations and measurement error (see also Hengartner[15]). Our findings also replicate the findings for antipsychotics in the acute treatment of schizophrenia,[16] and several treatments in a review of various medical interventions.[17] Together, these studies indeed suggest that the promises of precision medicine may remain elusive and that the scope for personalized medicine might be smaller than previously hoped for.[12 13] Given the high expectations placed in precision medicine, such findings will probably cause disbelief and reluctance in many advocates of this enthusiastic movement. In anticipation of such critique, we would like to address two objections that are likely to be submitted in response to this paper.

First, as recently stressed by biostatistics professor Dr. Frank Harrell, to assume that there is treatment heterogeneity (i.e. significant patient-by-treatment interaction) when the average treatment effect is close to zero (which is the case with AD), would imply that there must be a large subgroup of patients where the treatment causes significant harm.[27] Although it has been suggested that AD may worsen the long-term outcome of depression in some patients,[28-31] there is no evidence that they may do harm in a large subgroup of patients in short-term trials. In the absence of consistent biologically-informed patient-specific treatment effects, our best treatment estimate for AD thus remains the average drug effect relative to placebo.[27]

Second, and closely related to the above argument, even after decades of massive research efforts there is no evidence of robust neurobiological and genetic predictors of differential treatment response in depression.[6-9 32] Biostatistics professor Dr. Stephen Senn once stated: „Unless patient by treatment interaction exists, it is pointless looking for gene by treatment interactions".[33] Thus, calling

for more genetic research into differential treatment effects clearly conflicts with the current literature and will most likely fail to yield the hoped-for results.

We acknowledge the following major limitation: as Cortés *et al*.[17] describe in their paper, equal variances are no definite proof for a lack of patient-by-treatment interactions. They hypothetically describe a situation that leads to equal variances in the treatment and in the control condition and with patient-by-treatment interactions, but this situation is highly unlikely.

In conclusion, the results of our meta-analysis suggest that there is no or at best a very small patient-by-treatment interaction. The lack of increased outcome-variance in the AD vs. placebo groups in parallel-group RCT indicates that no specific subgroup of patients may respond particularly well to AD. Thus, with the AD currently available, the scope for personalized AD treatments is probably limited and it is unlikely that precision psychiatry will succeed in finding clinical or biological predictors of differential treatment response that would account for a therapeutic effect that goes beyond a minimal clinical improvement.

# STATEMENTS

## Acknowledgements

We thank Cipriani et al.[18] and Winkelbeiner et al.[16] for making the data and the R-code publicly available.

## Contributors

MP and MPH were responsible for the conception and design of the study, interpretation of data, and drafting and revising the manuscript. MP performed the meta-analysis and simulation analysis. All authors had had full access to all the data and take responsibility for the integrity of the data and the accuracy of the data analysis.

## Ethical approval

No ethical approval is required since our study is a secondary analysis.

## Transparency and data-sharing statement

The lead author affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted. The data and statistical code are available online.

## Role of funding source

This research received no specific grant from any funding agency, commercial or not-for-profit sectors. The authors declare that they have no financial conflicts of interest.

**REFERENCES**

1. Hengartner MP. What is the threshold for a clinical minimally important drug effect? *BMJ Evid Based Med* 2018 doi: 10.1136/bmjebm-2018-111056

2. Hengartner MP, Plöderl M. Statistically significant antidepressant-placebo differences on subjective symptom-rating scales do not prove that the drugs work: Effect size and method bias matter! *Front Psychiatry* 2018;9:517. doi: 10.3389/fpsyt.2018.00517

3. Jakobsen JC, Katakam KK, Schou A, et al. Selective serotonin reuptake inhibitors versus placebo in patients with major depressive disorder. A systematic review with meta-analysis and Trial Sequential Analysis. *BMC Psychiatry* 2017;17(1):58. doi: 10.1186/s12888-016-1173-2

4. Moncrieff J, Kirsch I. Efficacy of antidepressants in adults. *BMJ* 2005;331(7509):155-7. doi: 10.1136/bmj.331.7509.155

5. Munkholm K, Paludan-Muller AS, Boesen K. Considering the methodological limitations in the evidence base of antidepressants for depression: a reanalysis of a network meta-analysis. *BMJ Open* 2019;9(6):e024886. doi: 10.1136/bmjopen-2018-024886

6. Perlman K, Benrimoh D, Israel S, et al. A systematic meta-review of predictors of antidepressant treatment outcome in major depressive disorder. *J Affect Disord* 2019;243:503-15. doi: 10.1016/j.jad.2018.09.067 [published Online First: 2018/10/05]

7. Simon GE, Perlis RH. Personalized medicine for depression: can we match patients with treatments? *Am J Psychiatry* 2010;167(12):1445-55. doi: 10.1176/appi.ajp.2010.09111680 [published Online First: 2010/09/17]

8. Widge AS, Bilge MT, Montana R, et al. Electroencephalographic Biomarkers for Treatment Response Prediction in Major Depressive Illness: A Meta-Analysis. *Am J Psychiatry* 2019;176(1):44-56. doi: 10.1176/appi.ajp.2018.17121358 [published Online First: 2018/10/04]

9. Tansey KE, Guipponi M, Perroud N, et al. Genetic predictors of response to serotonergic and noradrenergic antidepressants in major depressive disorder: a genome-wide analysis of individual-level data and a meta-analysis. *PLoS Med* 2012;9(10):e1001326. doi: 10.1371/journal.pmed.1001326

10. Senn S. Mastering variation: variance components and personalised medicine. *Stat Med* 2016;35(7):966-77. doi: 10.1002/sim.6739 [published Online First: 2015/09/30]

11. Fernandes BS, Williams LM, Steiner J, et al. The new field of 'precision psychiatry'. *BMC Med* 2017;15 doi: 10.1186/s12916-017-0849-x

12. Senn S. Statistical pitfalls of personalized medicine. *Nature* 2018;563(7733):619-21. doi: 10.1038/d41586-018-07535-2

13. Joyner MJ, Paneth N. Promises, promises, and precision medicine. *J Clin Invest* 2019;129(3):946-48. doi: 10.1172/JCI126119

14. Prasad V, Fojo T, Brada M. Precision oncology: origins, optimism, and potential. *Lancet Oncol* 2016;17(2):e81-e86. doi: 10.1016/S1470-2045(15)00620-8

15. Hengartner MP. Scientific debate instead of beef; challenging misleading arguments about the efficacy of antidepressants. *Acta Neuropsychiatr* 2019;31(4):235-36. doi: 10.1017/neu.2019.23

16. Winkelbeiner S, Leucht S, Kane JM, et al. Evaluation of Differences in Individual Treatment Response in Schizophrenia Spectrum Disorders: A Meta-analysis. *JAMA Psychiatry* 2019 doi: 10.1001/jamapsychiatry.2019.1530

17. Cortés J, González JA, Medina MN, et al. Does evidence support the high expectations placed in precision medicine? A bibliographic review [version 5; peer review: 1 approved, 2 approved with reservations, 3 not approved]. *F1000Research* 2019;7:30. doi: https://doi.org/10.12688/f1000research.13490.5

18. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet* 2018;391(10128):1357-66. doi: 10.1016/S0140-6736(17)32802-7

19. Nakagawa S, Poulin R, Mengersen K, et al. Meta-analysis of variation: ecological and evolutionary applications and beyond. *Methods in Ecology and Evolution* 2015;6(2):143-52. doi: 10.1111/2041-210x.12309

20. Cortes J, Gonzalez JA, Medina MN, et al. Does evidence support the high expectations placed in precision medicine? A bibliographic review. *F1000Res* 2018;7:30. doi: 10.12688/f1000research.13490.5

21. Furukawa TA, Akechi T, Azuma H, et al. Evidence-based guidelines for interpretation of the Hamilton Rating Scale for Depression. *J Clin Psychopharmacol* 2007;27(5):531-4. doi: 10.1097/JCP.0b013e31814f30b1

22. Leucht S, Fennema H, Engel R, et al. What does the HAMD mean? *J Affect Disord* 2013;148(2-3):243-8. doi: 10.1016/j.jad.2012.12.001

23. Rabinowitz J, Werbeloff N, Mandel FS, et al. Initial depression severity and response to antidepressants v. placebo: patient-level data analysis from 34 randomised controlled trials. *Br J Psychiatry* 2016;209(5):427-28. doi: 10.1192/bjp.bp.115.173906

24. Furukawa TA, Maruo K, Noma H, et al. Initial severity of major depression and efficacy of new generation antidepressants: individual participant data meta-analysis. *Acta Psychiatr Scand* 2018;137(6):450-58. doi: 10.1111/acps.12886

25. Hieronymus F, Lisinski A, Nilsson S, et al. Influence of baseline severity on the effects of SSRIs in depression: an item-based, patient-level post-hoc analysis. *Lancet Psychiat* 2019 doi: 10.1016/S2215-0366(19)30216-0

26. Thase ME, Larsen KG, Kennedy SH. Assessing the 'true' effect of active antidepressant therapy v. placebo in major depressive disorder: use of a mixture model. *Br J Psychiatry* 2011;199(6):501-7. doi: 10.1192/bjp.bp.111.093336

27. Harrell FE, Jr. Viewpoints on heterogeneity of treatment effect and precision medicine. Statistical Thinking, 2019.

28. Fava GA, Offidani E. The mechanisms of tolerance in antidepressant action. *Prog Neuropsychopharmacol Biol Psychiatry* 2011;35(7):1593-602. doi: 10.1016/j.pnpbp.2010.07.026

29. El-Mallakh RS, Gao Y, Jeannie Roberts R. Tardive dysphoria: the role of long term antidepressant use in-inducing chronic depression. *Med Hypotheses* 2011;76(6):769-73. doi: 10.1016/j.mehy.2011.01.020

30. Andrews PW, Kornstein SG, Halberstadt LJ, et al. Blue again: perturbational effects of antidepressants suggest monoaminergic homeostasis in major depression. *Front Psychol* 2011;2:159. doi: 10.3389/fpsyg.2011.00159

31. Hengartner MP, Angst J, Rossler W. Antidepressant Use Prospectively Relates to a Poorer Long-Term Outcome of Depression: Results from a Prospective Community Cohort Study over 30 Years. *Psychother Psychosom* 2018;87(3):181-83. doi: 10.1159/000488802

32. Garcia-Gonzalez J, Tansey KE, Hauser J, et al. Pharmacogenetics of antidepressant response: A polygenic approach. *Prog Neuropsychopharmacol Biol Psychiatry* 2017;75:128-34. doi: 10.1016/j.pnpbp.2017.01.011

33. Senn S. Individual response to treatment: is it a valid assumption? *BMJ* 2004;329(7472):966-68. doi: DOI 10.1136/bmj.329.7472.966

Table 1. Meta-Analytic results for variance ratios and coefficients of variance ratio

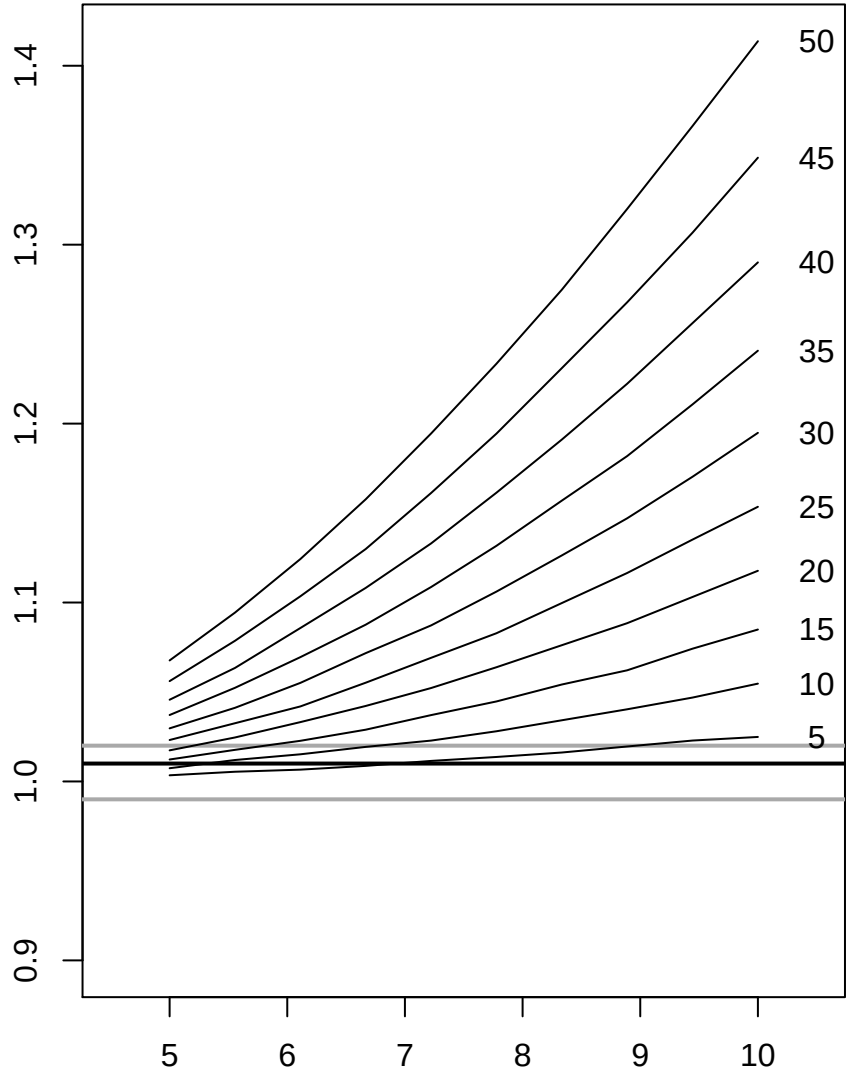
| | n | | | Variance Ratio (VR) Meta-Analysis | | | r (M, SD) | | Coefficient of Variance Ratio (CVR) Meta-Analysis | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Trials | AD | Placebo | VR (95% CI) | p | Q(df) | AD | Placebo | CVR (95% CI) | P | Q (df) |
| AD all | 168 | 32517 | 18620 | 1.01 (0.99-1.02) | .25 | 116.81 (167) | .56** | .53** | 0.82 (0.80-0.84) | .00 | 242.87 (167)** |
| SSRI | 89 | 13146 | 9024 | 1.01 (0.99-1.04) | .32 | 124.94 (88)**[a] | .56** | .45** | 0.83 (0.80-0.86) | .00 | 137.29 (88)** |
| SNRI | 52 | 9837 | 7100 | 1.01 (0.99-1.03) | .48 | 9.36 (51) | .55** | .57** | 0.81 (0.78-0.84) | .00 | 69.91 (51)** |
| Atypical | 54 | 8780 | 6293 | 1.00 (0.97-1.02) | .69 | 17.62 (53) | .71** | .74** | 0.83 (0.79-0.86) | .00 | 61.53(53) |
| Tricyclics | 11 | 754 | 771 | 1.04 (0.95-1.14) | .40 | 15.80 (10) | -.34 | .64* | 0.65 (0.58-0.72) | .00 | 6.96 (10) |

Notes:

p (M, SD) is the Pearson correlation of mean pre-post differences in depression (M) and the respective standard deviation (SD). Q(df) is the heterogeneity index for the meta-analysis. * $p < .05$, ** $p < .01$, [a]This result was caused by an outlier (Study Dube 2010, NCT00420004), and after removing this study, the heterogeneity index was Q(df=87)=54.54, p = 0.99.

**Legend for Figure 1.**

Figure 1. Results from simulation analyses (hypothetical variance ratios) for different definitions of benefiters (x-axis) and different percentages of benefiters (individual lines with the percentages on the right side of the line). The horizontal thick line is the result from our meta-analysis (VR = 1.01), the horizontal gray lines correspond with upper and lower limits of the confidence interval of the meta-analysis (0.99-1.02).



1Supplemental Table - Results for trials where only end of treatment depression scores were available

| | n | | | Variance Ratio Meta Analysis | | | r (M, SD) | | Coefficient of Variance Ratio Meta-Analysis | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Trials | AD | Placebo | VR (95% CI) | p | Q(df) | AD | Placebo | CVR (95% CI) | P | Q (df) |
| AD all | 84 | 10879 | 7346 | 0.98 (0.96-1.00) | .07 | 52.53 (83) | .36** | .54** | 1.15 (1.11-1.18) | .00 | 92.67 (83) |
| SSRI | 40 | 5424 | 3763 | 0.99(0.96-1.03) | .71 | 28.60 (39) | .63** | .37* | 1.13 (1.08-1,17) | .00 | 35.23 (39)** |
| SNRI | 13 | 1440 | 1288 | 0.99 (0.94-1.05) | .73 | 4.50 (12) | -.49 | .62* | 1.15 (1.07-1.24) | .00 | 9.82 (12)** |
| Atypical | 31 | 4245 | 3092 | 0.97 (0.94-1.00) | .07 | 12.63 (30) | .43* | .14 | 1.12 (1.07-1.17) | .00 | 34.37 (30) |
| Tricyclics | 13 | 670 | 643 | 0.93 (0.86-1.01) | .07 | 9.18(10) | .09 | .72** | 1.37 (1.20-1.56) | .00 | 15.64 (12) |

2Notes:

3p (M, SD) is the pearson correlation of mean pre-post differences in depression (M) and the respective standard deviation (SD). Q(df) ist the heterogenity index 4for the meta-analysis. * p < .05, ** p < .01.

Table 1. Meta-Analytic results for variance ratios and coefficients of variance ratios, individual AD, only studies reporting pre-post-differences of depression

| | n | | | Variance Ratio (VR) Meta-Analysis | | | r (M, SD) | | Coefficient of Variance Ratio (CVR) Meta-Analysis | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Trials | AD | Placebo | VR (95% CI) | p | Q(df) | AD | Placebo | CVR (95% CI) | P | Q (df) |
| Agomelatine | 0 | - | - | - | - | - | - | - | - | - | - |
| Amitriptyline | 11 | 754 | 771 | 1.04 (0.95-1.14) | .40 | 15.80 (10) | -34 | .64* | 0.65(0.58-0.72) | .00 | 6.96 (10) |
| Bupropion | 13 | 2095 | 1887 | 0.99 (0.94-1.03) | .99 | 2.57 (12) | .89** | .87** | 0.88 (0.82-0.94) | .00 | 7.70 (12) |
| Citalopram | 11 | 1819 | 1335 | 1.02 (0.97-1.07) | .50 | 4.35 (10) | .54** | .71* | 0.85 (0.79-0.92) | .00 | 4.83 (10) |
| Clomipramine | 0 | - | - | - | - | - | - | - | - | - | - |
| Desvenlafaxine | 7 | 2124 | 1241 | 1.00 (0.95-1.05) | .91 | 1.17 (6) | -.05 | -.03 | 0.83 (0.77-0.89) | .00 | 1.33 (6) |
| Duloxetine | 20 | 2953 | 2371 | 1.01 (0.97-1.05) | .66 | 1.81 (19) | .61** | .73** | 0.76 (0.71-0.81) | .00 | 25.63 (19) |
| Escitalopram | 16 | 2749 | 2114 | 1.06 (0.95-1.19) | .28 | 74.54 (15)**[a] | .65** | .05 | 0.90 (0.79-1.02) | .11 | 55.82 (15)** |
| Fluoxetine | 18 | 2765 | 1991 | 1.00 (0.96-1.05) | .95 | 6.96 (17) | .60** | .71** | 0.85 (0.78-0.92) | .00 | 20.58(17) |
| Fluvoxamine | 0 | - | - | - | - | - | - | - | - | - | - |
| Levomilnacipran | 5 | 1566 | 1032 | 1.03 (0.97-1.09) | .31 | 2.91 (4) | .20 | .45 | 0.83 (0.74-0.92) | .00 | 6.02 (4) |
| Milnacipran | 0 | - | - | - | - | - | - | - | - | - | - |
| Mirtazapine | 10 | 624 | 466 | 1.00 (0.91-1.09) | .96 | 8.74 (9) | .52 | .35 | 0.72 (0.63-0.83) | .00 | 10.91 (9) |
| Nefazodone | 9 | 767 | 528 | 1.01 (0.97-1.09) | .90 | 0.34 (8) | .27 | -.29 | 0.80 (0.71-0.91) | .00 | 5.09 (8) |
| Paroxetine | 37 | 4233 | 3333 | 1.00 (0.97-1.04) | .89 | 11.27 (36) | .28 | .26 | 0.82 (0.78-0.86) | .00 | 41.51 (36) |
| Reboxetine | 6 | 1035 | 944 | 0.99 (0.93-1.05) | .69 | 0.45 (5) | .51 | .66 | 0.92 (0.79-1.07) | .28 | 12.10 (5)* |
| Sertraline | 15 | 1534 | 1374 | 0.98 (0.90-1.06) | .57 | 27.03 (14) | .75** | .77** | 0.77 (0.71-0.83) | .00 | 11.68 (14) |
| Trazodone | 0 | - | - | - | - | - | - | - | - | - | - |

| | n | | | Variance Ratio (VR) Meta-Analysis | | | r (M, SD) | | Coefficient of Variance Ratio (CVR) Meta-Analysis | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Venlafaxine | 15 | 2159 | 1672 | 1.01 (0.96-1.06) | .74 | 2.00 (14) | .42 | .22 | 0.83 (0.75-0.85) | .00 | 11.57 (14) |
| Vilazodone | 5 | 1360 | 1064 | 0.99 (0.93-1.05) | .72 | 1.07 (4) | .55 | .28 | 0.81 (0.73-0.90) | .00 | 4.96 (4) |
| Vortioxetine | 14 | 3562 | 2098 | 1.00 (0.96-1.04) | .99 | 0.41 (13) | .32 | .79** | 0.81 (0.76-0.88) | .00 | 23.04 (13)* |

Notes:

r (M, SD) is the Pearson correlation coefficient of mean pre-post differences in depression (M) and the respective standard deviation (SD). Q(df) is the heterogeneity index of the meta-analysis. In case of AD without trials that reported pre-post differences, the results for the post-values of depression scores also resulted in VR ≈ 1.00 (Agomelatine: VR = 0.96, CI 0.92-1.01; Fluvoxamine VR = 0.99, CI 0.88-12; Trazodone: VR = 0.96, CI 0.88-1.06. For Clomipramine and Milnacipran, no placebo-controlled trials were available.

* p < .05, ** p < .01

[a]This result was caused by an outlier (Study Dube 2010, NCT00420004), and after removing this study, the heterogeneity index was Q(df=14)=7.36, p = 0.92.

BMJ Open

# What are the chances for personalized treatment with antidepressants? Detection of patient by treatment interaction with a variance ratio meta-analysis



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2019-034816.R1 |
| Article Type: | Original research |
| Date Submitted by the Author: | 27-Nov-2019 |
| Complete List of Authors: | Plöderl, Martin; Christian Doppler Clinic, Paracelsus Medical University, Dept. of Clinica Psychology and Dept. for Crisis Intervention and Suicide Prevention, University Clinic for Psychiatry, Psychotherapy, and Psychosomatics<br>Hengartner, Michael; Zurich University of Applied Sciences/ZHAW, Section for Clinical Psychology and Health Psychology |
| Primary Subject Heading: | Patient-centred medicine |
| Secondary Subject Heading: | Mental health |
| Keywords: | Depression & mood disorders < PSYCHIATRY, STATISTICS & RESEARCH METHODS, PSYCHIATRY |
| | |

# *What are the chances for personalized treatment with antidepressants? Detection of patient by treatment interaction with a variance ratio meta-analysis*

Martin Plöderl[1]*, Michael P. Hengartner[2]

[1] Department of Crisis Intervention and Suicide Prevention/Christian Doppler Clinic, Paracelsus Medical University, Salzburg, Austria.

[2] Department of Applied Psychology, Zurich University of Applied Sciences, Zurich, Switzerland.



**Short Title:** Personalized antidepressants treatment

**Word count:** 3075

***Corresponding Author**

Martin Plöderl PhD

Dept. of Crisis Intervention and Suicide Prevention

Christian Doppler Klinik,

Paracelsus Medical University,

Ignaz Harrer-Straße 79

A-5020 Salzburg

AUSTRIA

Phone: +43-5-7255-56679

Fax: +43-5-7255-4344

E-mail: M.Ploederl@salk.at

# ABSTRACT

## Objectives

To investigate if the treatment effect of antidepressants in patients with depression substantially varies in each patient (patient by treatment interaction or treatment heterogeneity), a necessary but largely unexplored prerequisite of personalized antidepressant treatment.

## Design

Meta-analytic variance-comparison of treatment-outcome between drug arms and placebo arms of clinical trials, based on the assumption that patient by treatment interaction should lead to larger variances in drug arms than placebo arms. To put the results into context, we run simple simulations, assuming different definitions and rates of those who respond especially well to antidepressants.

## Data sources

163 randomized, placebo-controlled trials (51,396 patients) with complete results for pre-post differences, selected from a recently published systematic review.

## Analysis

Variance ratios (VR) and coefficients of variance ratios (CVR) of individual trials were meta-analytically combined. The analysis was repeated for classes of antidepressants and specific antidepressants.

## Results

Variance ratios (VR = 1.01, CI 0.99-1.02) and coefficient of variance ratios (CVR = 0.82, CI 0.80-0.84) of the antidepressant-treatment arms were comparable or smaller than in placebo-arms. Similar results were observed for classes of antidepressants and for specific antidepressants. Our simulation

analysis confirmed that equal variance ratios can only be obtained if they are not more than a few patients who respond slightly above average.

## Conclusions

The lack of increased treatment-outcome variance in the antidepressants versus placebo groups in RCT indicates that no or only very small subgroups of patients respond particularly well to antidepressants. Thus, the scope for personalized treatment with antidepressants seems to be limited.

**Keywords:** antidepressants, personalized medicine, treatment heterogeneity, variance ratio, meta-analysis

## STRENGTHS AND LIMITATIONS OF THIS STUDY

- For the first time, the amount of patient by treatment interaction (treatment heterogeneity), a necessary prerequisite for personalized medicine, is estimated for the pharmacological treatment of depression with antidepressants.
- The data-base is from a systematic review of published and unpublished studies and is one of the largest so far, resulting in precise estimations of the main outcomes.
- The study results are important to inform further attempts in personalized (precision) medicine in psychiatry.
- As with all clinical trials, it remains an open question if our results can be replicated in real-world settings, for example among psychiatric inpatients with very severe depression.

## INTRODUCTION

Personalized or precision medicine, i.e., applying medical interventions only to those patients known to benefit especially well to the intervention (henceforth termed "benefiters"), is important to increase benefits from treatment and to decrease harms. For example, if a drug with severe side effects is very effective in some patients with a specific genotype, it is crucial to know about these benefiters, because for all other patients, the risk-benefit ratio would be unfavorable. Similarly, if a drug is found to have only modest efficacy across all patients but notable side-effects, than it would be important to know if there are patients (e.g., those defined via a specific biomarker) who are benefiters. The latter example corresponds with the pharmacological treatment of major depression, because the average efficacy of antidepressants (AD) is modest, corresponding to, on average, only about 2 points difference on the Hamilton Depression Rating Scale (HDRS) between AD groups and placebo in randomized controlled short-term trials (RCT).[1-5] Put differently, according to our most recent estimate, there is about 88% overlap in distribution of depression scores between antidepressants and placebo at the end of acute treatment.[2]

Despite substantial research efforts, no predictors of treatment success with AD were found that were robust and reliable enough for use in clinical practice.[6-9] Thus, much of the variance of the treatment outcome remains unexplained thus far. Sources of outcome variation include variation between treatment arms (indicating that group means differ due to efficacy of treatments, e.g., AD versus placebo), variation between patients (indicating that the outcome differs from patient to patient, independent of the treatment received), variation within patients (indicating that the outcome for the same patient differs over time due to random symptom fluctuations), and patient-by-treatment interactions (indicating that treatment effects vary from patient to patient).[10] The quest for precision psychiatry, in this case personalized AD treatment, assumes that specific patients benefit more from AD than others, i.e., assumes that there is a patient-by-treatment interaction. Ideally, this can also be explained by a plausible causal mechanism, e.g. inter-individual differences in

monoamine function. Investing research efforts in personalized medicine only makes sense if there truly is a patient-by-drug interaction that explains some variance in the treatment outcome. Although the field is mostly enthusiastic about personalized medicine or precision psychiatry,[11] experts from various fields now start to dampen expectations and caution that personalized/precision medicine may fall short of expectations.[12-14]

Thus, we must remain mindful that there might be no notable subgroup of true AD benefiters and that the modest average treatment effect is the best we can hope for.[15] We further need to acknowledge that RCT are inherently limited to demonstrating patient-by-treatment interactions.[10] To identify patient-by-treatment interactions, repeated period cross-over trials are necessary, but these are hardly feasible with common AD due to delayed onset of therapeutic effect and relatively high rates of spontaneous remission. The most common trial design is the simple parallel-group trial, where patients are randomized to either AD or placebo. However, these trials can only identify mean differences between treatment arms (i.e., efficacy), whereas variation between patients, within patients, as well as patient-by-treatment interactions are part of the error term. Nevertheless, if patient-by-treatment interaction effects are present, then the variance in the treatment outcome should be increased in the drug group relative to the placebo group, because no comparable drug-by-patient interaction is present in the placebo group.[10 16 17] Thus, results from RCT can inform indirectly if there might be subgroups of benefiters.

The goal of this meta-analysis was to examine whether the outcome-variances between AD and placebo differ, in order to gauge the potential of personalized/precision psychiatry for treatment with AD.

## METHODS

### Data

Our analysis was based on short-term RCT of AD for patients with unipolar major depression, reported in the most recent systematic review.[18] The authors of this comprehensive study made the data available in a public repository https://data.mendeley.com/datasets/83rthbp8ys/2. This included 522 trials (with 21 different AD), of which 254 trials were suitable for further analysis, i.e., contained information about the outcome and also included a placebo-arm. Where trials had multiple treatment arms with different dosages of AD, these arms were aggregated. Where trials compared different AD, the data of these arms were aggregated to only have one value for the drugs in these trials, similar as in a previous publication.[16] Additionally, we recorded different AD by their class (SNRI, SSRI, atypical AD, and tricyclic AD). For 169 (67%) of studies, the pre-post mean reduction of depression scores (M) and the related standard deviations (SD) were available, and only the analysis for these studies are reported here. Analysis for the 85 studies (33%) where only the mean value and standard deviation of the post-treatment depression scores were available are reported in the online supplementary table 1 (https://osf.io/98kex/files/). Several additional variables were created for sensitivity analysis (see below).

### Patient and public involvement statement

This research was done without patient involvement. Patients were not invited to comment on the study design and were not consulted to develop patient relevant outcomes or interpret the results. Patients were not invited to contribute to the writing or editing of this document for readability or accuracy.

**Statistical analysis**

Meta-analysis

We calculated the variance ratio (VR) for each RCT and aggregated them by means of a random effect meta-analysis according to the procedure suggested by Winkelbeiner et al.;[16] see https://osf.io/qarvs/files/, using the metafor package in R. Because the pre-post-differences were significantly associated with their SD, we repeated the analysis using the coefficient of the variance ratio (CVR). This removes the effect of expected changes in the standard deviation due to changes in the mean.[19] A VR of 1.00 suggest equal variance of AD and placebo. If the VR exceeds 1, then the variance of the AD group is larger than in the placebo group. If the CVR exceeds one, than the increase of variance with increasing pre-post-differences is stronger in the AD than in the placebo arms. Sensitivity analysis included meta-regression models with the assessment instruments, year of publication, type of publication (published vs. unpublished), sample size, and drop-out rates.

Simulation analysis

To bring our results into context, we also ran simulation analysis with different definitions and probabilities for benefiters. We based these simulations on an efficacy of 2 points mean-difference between AD and placebo groups, as reported in a meta-analysis on the same dataset for trials using the HDRS-17 instrument.[5] We assumed a standard deviation of SD=8 and a mean difference between pre-and post-depression-scores of 11 points in the AD group (based on rounded means of these values observed in our data-set). We used different cut-offs to define patients as benefiters, ranging from 5 to 10 points superior treatment outcome with the HDRS-17, and a proportion of 5 to 50% in the AD group vs. 0% in the placebo group. To simulate placebo groups, we sampled from a normal distribution with the above parameters for the placebo group (M = 9, SD = 8, and 5,000,000

samples). We used a similar sampling procedure for the AD-group, but created benefiters by adding necessary HDRS responder points to an assumed fraction of the sample, and adding as many points to the rest of the sample to end up with the overall efficacy of 2 HDRS points.

The R-code and data of this publication is available online (https://osf.io/98kex/files/).

## RESULTS

### Meta-analysis

Across all AD, the variance ratio was almost perfectly 1.00 with a narrow confidence interval (VR = 1.01, 95%CI = 0.99 – 1.02). This means that the variances in the AD and placebo groups are identical (Table 1). Similar findings were found for all classes of AD (Table 1) and for each individual drug (online supplementary table2, https://osf.io/98kex/files/). There was no significant sign of heterogeneity in the meta-analyses, except for SSRIs. A closer inspection revealed that this resulted from a single outlier (see footnote in Table 1).

----- Insert Table 1 around here -----

For the coefficient of variance ratio, results indicated that the increase of variance with increasing pre-post differences is less strong in AD than in placebo arms (CVR = 0.82, 95%CI = 0.80-0.84). Comparable results were found for all classes of AD and individual drugs (Table 1). The heterogeneity was statistically significant in nearly all meta-analyses of the CVR.

In the sensitivity analysis, the meta-regression models could not detect statistically significant effects for year of publication, type of publication, measurement-instruments, and sample size (see https://osf.io/98kex/files/). However, there was a small but statistically significant effect of the drop-out rates in drug-arms and the VRs. With increasing drop-out rates, the VRs slightly increased ($p < .04$).

**Simulation analysis**

As shown in Figure 1, most benefiter assumptions lead to VRs much different from those we observed in our study.

----- Insert Figure 1 around here -----

However, for liberal definitions of benefiters and low rates of these benefiters, the VRs can indeed be small. For example, if there are 10% benefiters, as defined with 6 HDRS points difference to average placebo response, the VR is within the confidence interval of our main finding (0.99-1.02).

## DISCUSSION

We found nearly identical treatment outcome variances for AD arms compared to placebo in RCT for the acute treatment of major depression in a large database, as indicated by variance ratios almost perfectly being VR = 1. The simplest explanation for the finding of similar variances is that there are constant treatment effects and no treatment heterogeneity, i.e. no patient-by-treatment interaction effects and no specific subgroups of patients who respond particularly well to the treatment.[20] Alternatively, such a subgroup of benefiters would be very small (≤10%) and the threshold to classify someone as benefiter would be low (≤6 HDRS points difference to average placebo response). To put this in context, according to anchor-based linkage studies, at least 6 points on the HDRS are necessary for a global impression of "minimally improved".[21 22] Consequently, the search for meaningful predictors of relative treatment response (compared to placebo) will probably fail or will at least be very difficult due to the small subgroup of weak benefiters. Therefore, the mean effect size estimate from parallel-group RCT remains the best guess for predicting treatment outcome for an individual patient. Furthermore, the results for the coefficient of variance indicated that the

increase of variance associated with increasing larger pre-post-differences was stronger in the placebo than the AD groups. There is no immediately plausible explanation for this finding, given that baseline severity does not predict differential treatment effects.[23-25]

Our findings are in line with Senn,[12] who argued that exploratory post-hoc delineation of putative benefiters, such as the "true benefiters" suggested by Thase *et al.*,[26] are simply statistical artefacts due to random symptom fluctuations and measurement error (see also Hengartner[15]). Our findings also replicate the findings for antipsychotics in the acute treatment of schizophrenia,[16] and several treatments in a review of various medical interventions.[17] Together, these studies indeed suggest that the promises of precision medicine may remain elusive and that the scope for personalized medicine might be smaller than previously hoped for.[12 13] Given the high expectations placed in biomarker-based precision medicine, such findings will probably cause disbelief and reluctance in many advocates of this enthusiastic movement. In anticipation of such critique, we would like to address two objections that are likely to be submitted in response to this paper.

First, as recently stressed by biostatistics professor Dr. Frank Harrell, to assume that there is treatment heterogeneity (i.e. significant patient-by-treatment interaction) when the average treatment effect is close to zero (which is the case with AD), would imply that there must be a large subgroup of patients where the treatment causes significant harm.[27] Although it has been suggested that AD may worsen the long-term outcome of depression in some patients,[28-31] there is no evidence that they may do harm in a large subgroup of patients in short-term trials. In the absence of consistent biologically-informed patient-specific treatment effects, our best treatment estimate for AD thus remains the average drug effect relative to placebo.[27]

Second, and closely related to the above argument, even after decades of massive research efforts there is no evidence of robust neurobiological and genetic predictors of differential treatment response in depression.[6-9 32] Biostatistics professor Dr. Stephen Senn once stated: „Unless patient by

treatment interaction exists, it is pointless looking for gene by treatment interactions".[33] Thus, calling for more genetic and neurobiological research into differential treatment effects clearly conflicts with the current literature and will most likely fail to yield the hoped-for results.

We acknowledge the following major limitation: as Cortés *et al*.[17] describe in their paper, equal variances are no definite proof for a lack of patient-by-treatment interactions. They hypothetically describe a situation that leads to equal variances in the treatment and in the control condition and with patient-by-treatment interactions, but this situation is highly unlikely. Furthermore, VRs of 1 are, theoretically, also possible with a small fraction of "super-responders" and a specific response for all others, as highlighted in a vivid Twitter-Discussion of our paper (https://twitter.com/Martin_Ploederl/status/1188006207497363457). It can indeed be debated what assumption is more plausible: a constant treatment effect, a hypothetical fraction of super-responders, or other highly specific premises. However, a small fraction of super-responders would obviously lead to non-normal distributions with notable peaks at very low-levels of depression scores. This was not observed so far, to our knowledge,[26] but could be further investigated with patient level data. Moreover, VRs would increase for a wide range of scenarios with varying fractions of benefiters and varying definitions of "benefiters."

Another potential problem may be, as one reviewer pointed out, that the VRs did not vary much across trials, as indicated by the Q statistics, and also by the low $I^2$ statistics. However, the main results remained the same for different estimators of heterogeneity, unweighted results, or with the Knapp and Hartung adjustment (see online supplementary table 3). Furthermore, by manually increasing the value of the heterogeneity, results remained comparable. Presumably, the low between-trial heterogeneity was caused by insufficient randomization of trials, or due to narrow and selective inclusion criteria for trial participants.[34]

In conclusion, the results of our meta-analysis suggest that there is no or at best a very small patient-by-treatment interaction. The lack of increased outcome-variance in the AD vs. placebo groups in parallel-group RCT indicates that no specific subgroup of patients may respond particularly well to AD. Thus, with the AD currently available, the scope for personalized AD treatments is probably limited and it is unlikely that precision psychiatry will succeed in finding clinical or biological predictors of differential treatment response that would account for a therapeutic effect that goes beyond a minimal clinical improvement.

# STATEMENTS

## Acknowledgements

We thank Cipriani et al.[18] and Winkelbeiner et al.[16] for making the data and the R-code publicly available.

## Contributors

MP and MPH were responsible for the conception and design of the study, interpretation of data, and drafting and revising the manuscript. MP performed the meta-analysis and simulation analysis. All authors had had full access to all the data and take responsibility for the integrity of the data and the accuracy of the data analysis.

## Ethical approval

No ethical approval is required since our study is a secondary analysis.

## Transparency and data-sharing statement

The lead author affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted. The data, statistical code, and supplementary tables are available online (https://osf.io/98kex/files/).

## Role of funding source

This research received no specific grant from any funding agency, commercial or not-for-profit sectors. The authors do not have any financial conflicts of interest. MP works in a public psychiatric hospital, where the treatment with antidepressants is a common practice, potentially causing conflicts of interest. For this reason, the research was done in his leisure time.

**REFERENCES**


1. Hengartner MP. What is the threshold for a clinical minimally important drug effect? *BMJ Evid Based Med* 2018 doi: 10.1136/bmjebm-2018-111056 [published Online First: 2018/08/30]
2. Hengartner MP, Plöderl M. Statistically significant antidepressant-placebo differences on subjective symptom-rating scales do not prove that the drugs work: Effect size and method bias matter! *Frontiers in Psychiatry* 2018;9:517. doi: 10.3389/fpsyt.2018.00517
3. Jakobsen JC, Katakam KK, Schou A, et al. Selective serotonin reuptake inhibitors versus placebo in patients with major depressive disorder. A systematic review with meta-analysis and Trial Sequential Analysis. *BMC Psychiatry* 2017;17(1):58. doi: 10.1186/s12888-016-1173-2 [published Online First: 2017/02/10]
4. Moncrieff J, Kirsch I. Efficacy of antidepressants in adults. *Bmj* 2005;331(7509):155-7. doi: 10.1136/bmj.331.7509.155 [published Online First: 2005/07/16]
5. Munkholm K, Paludan-Muller AS, Boesen K. Considering the methodological limitations in the evidence base of antidepressants for depression: a reanalysis of a network meta-analysis. *BMJ Open* 2019;9(6):e024886. doi: 10.1136/bmjopen-2018-024886
6. Perlman K, Benrimoh D, Israel S, et al. A systematic meta-review of predictors of antidepressant treatment outcome in major depressive disorder. *J Affect Disord* 2019;243:503-15. doi: 10.1016/j.jad.2018.09.067 [published Online First: 2018/10/05]
7. Simon GE, Perlis RH. Personalized medicine for depression: can we match patients with treatments? *Am J Psychiatry* 2010;167(12):1445-55. doi: 10.1176/appi.ajp.2010.09111680 [published Online First: 2010/09/17]
8. Widge AS, Bilge MT, Montana R, et al. Electroencephalographic Biomarkers for Treatment Response Prediction in Major Depressive Illness: A Meta-Analysis. *Am J Psychiatry* 2019;176(1):44-56. doi: 10.1176/appi.ajp.2018.17121358 [published Online First: 2018/10/04]
9. Tansey KE, Guipponi M, Perroud N, et al. Genetic predictors of response to serotonergic and noradrenergic antidepressants in major depressive disorder: a genome-wide analysis of individual-level data and a meta-analysis. *PLoS medicine* 2012;9(10):e1001326. doi: 10.1371/journal.pmed.1001326
10. Senn S. Mastering variation: variance components and personalised medicine. *Stat Med* 2016;35(7):966-77. doi: 10.1002/sim.6739 [published Online First: 2015/09/30]
11. Fernandes BS, Williams LM, Steiner J, et al. The new field of 'precision psychiatry'. *BMC medicine* 2017;15 doi: ARTN 80

10.1186/s12916-017-0849-x
12. Senn S. Statistical pitfalls of personalized medicine. *Nature* 2018;563(7733):619-21. doi: 10.1038/d41586-018-07535-2 [published Online First: 2018/11/30]
13. Joyner MJ, Paneth N. Promises, promises, and precision medicine. *J Clin Invest* 2019;129(3):946-48. doi: 10.1172/JCI126119
14. Prasad V, Fojo T, Brada M. Precision oncology: origins, optimism, and potential. *Lancet Oncol* 2016;17(2):e81-e86. doi: 10.1016/S1470-2045(15)00620-8
15. Hengartner MP. Scientific debate instead of beef; challenging misleading arguments about the efficacy of antidepressants. *Acta Neuropsychiatr* 2019;31(4):235-36. doi: 10.1017/neu.2019.23
16. Winkelbeiner S, Leucht S, Kane JM, et al. Evaluation of Differences in Individual Treatment Response in Schizophrenia Spectrum Disorders: A Meta-analysis. *JAMA psychiatry* 2019 doi: 10.1001/jamapsychiatry.2019.1530 [published Online First: 2019/06/04]
17. Cortés J, González JA, Medina MN, et al. Does evidence support the high expectations placed in precision medicine? A bibliographic review [version 5; peer review: 1 approved, 2 approved

with reservations, 3 not approved]. *F1000Research* 2019;7:30. doi: https://doi.org/10.12688/f1000research.13490.5
18. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet* 2018;391(10128):1357-66. doi: 10.1016/S0140-6736(17)32802-7 [published Online First: 2018/02/27]
19. Nakagawa S, Poulin R, Mengersen K, et al. Meta-analysis of variation: ecological and evolutionary applications and beyond. *Methods Ecol Evol* 2015;6(2):143-52. doi: 10.1111/2041-210x.12309
20. Cortes J, Gonzalez JA, Medina MN, et al. Does evidence support the high expectations placed in precision medicine? A bibliographic review. *F1000Res* 2018;7:30. doi: 10.12688/f1000research.13490.5
21. Furukawa TA, Akechi T, Azuma H, et al. Evidence-based guidelines for interpretation of the Hamilton Rating Scale for Depression. *J Clin Psychopharmacol* 2007;27(5):531-4. doi: 10.1097/JCP.0b013e31814f30b1
22. Leucht S, Fennema H, Engel R, et al. What does the HAMD mean? *J Affect Disord* 2013;148(2-3):243-8. doi: 10.1016/j.jad.2012.12.001
23. Rabinowitz J, Werbeloff N, Mandel FS, et al. Initial depression severity and response to antidepressants v. placebo: patient-level data analysis from 34 randomised controlled trials. *Br J Psychiatry* 2016;209(5):427-28. doi: 10.1192/bjp.bp.115.173906
24. Furukawa TA, Maruo K, Noma H, et al. Initial severity of major depression and efficacy of new generation antidepressants: individual participant data meta-analysis. *Acta Psychiatr Scand* 2018;137(6):450-58. doi: 10.1111/acps.12886
25. Hieronymus F, Lisinski A, Nilsson S, et al. Influence of baseline severity on the effects of SSRIs in depression: an item-based, patient-level post-hoc analysis. *Lancet Psychiat* 2019 doi: 10.1016/S2215-0366(19)30216-0
26. Thase ME, Larsen KG, Kennedy SH. Assessing the 'true' effect of active antidepressant therapy v. placebo in major depressive disorder: use of a mixture model. *Br J Psychiatry* 2011;199(6):501-7. doi: 10.1192/bjp.bp.111.093336 [published Online First: 2011/12/02]
27. Harrell FE, Jr. Viewpoints on heterogeneity of treatment effect and precision medicine. Statistical Thinking, 2019.
28. Fava GA, Offidani E. The mechanisms of tolerance in antidepressant action. *Prog Neuropsychopharmacol Biol Psychiatry* 2011;35(7):1593-602. doi: 10.1016/j.pnpbp.2010.07.026
29. El-Mallakh RS, Gao Y, Jeannie Roberts R. Tardive dysphoria: the role of long term antidepressant use in-inducing chronic depression. *Med Hypotheses* 2011;76(6):769-73. doi: 10.1016/j.mehy.2011.01.020
30. Andrews PW, Kornstein SG, Halberstadt LJ, et al. Blue again: perturbational effects of antidepressants suggest monoaminergic homeostasis in major depression. *Frontiers in psychology* 2011;2:159. doi: 10.3389/fpsyg.2011.00159
31. Hengartner MP, Angst J, Rossler W. Antidepressant Use Prospectively Relates to a Poorer Long-Term Outcome of Depression: Results from a Prospective Community Cohort Study over 30 Years. *Psychother Psychosom* 2018;87(3):181-83. doi: 10.1159/000488802 [published Online First: 2018/04/24]
32. Garcia-Gonzalez J, Tansey KE, Hauser J, et al. Pharmacogenetics of antidepressant response: A polygenic approach. *Prog Neuropsychopharmacol Biol Psychiatry* 2017;75:128-34. doi: 10.1016/j.pnpbp.2017.01.011
33. Senn S. Individual response to treatment: is it a valid assumption? *Bmj* 2004;329(7472):966-68. doi: DOI 10.1136/bmj.329.7472.966

34. Zimmerman M, Clark HL, Multach MD, et al. Have Treatment Studies of Depression Become Even Less Generalizable? A Review of the Inclusion and Exclusion Criteria Used in Placebo-Controlled Antidepressant Efficacy Trials Published During the Past 20 Years. *Mayo Clin Proc* 2015;90(9):1180-86. doi: 10.1016/j.mayocp.2015.06.016

Table 1. Meta-Analytic results for variance ratios and coefficients of variance ratio

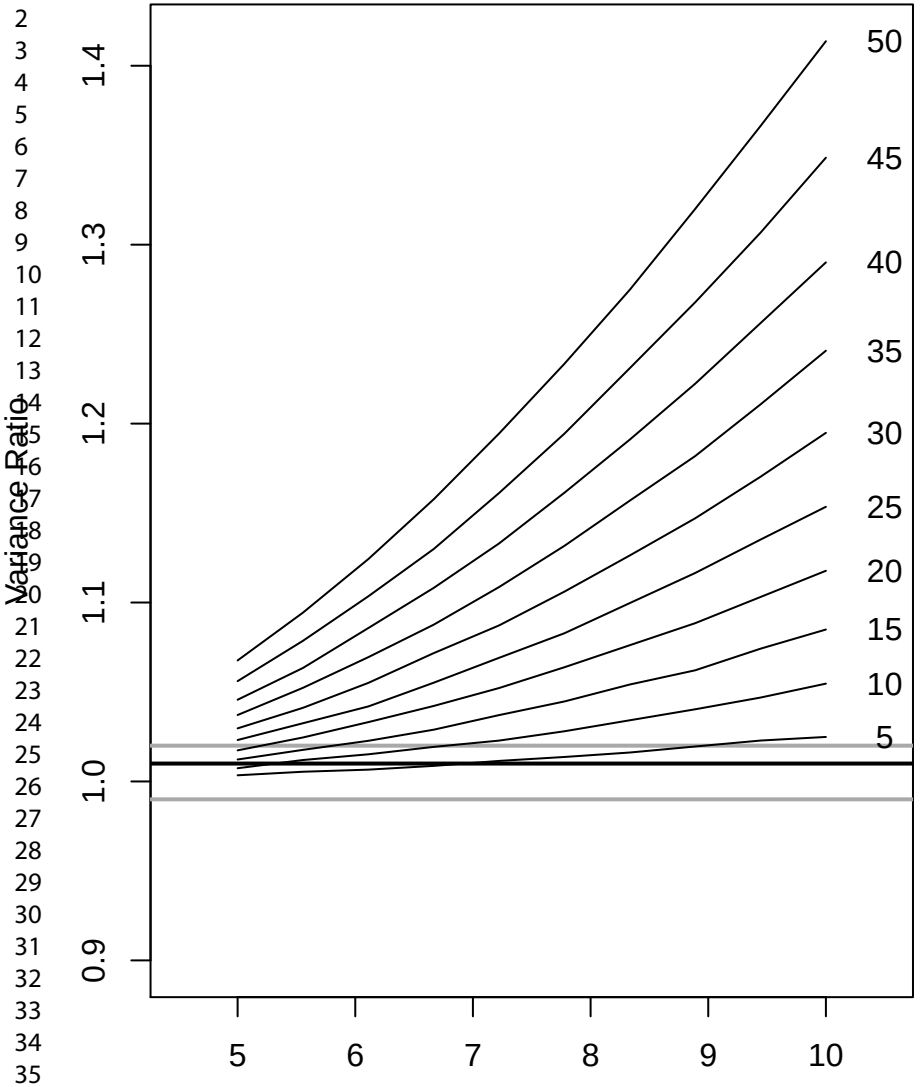
| | n | | | Variance Ratio (VR) Meta-Analysis | | | r (M, SD) | | Coefficient of Variance Ratio (CVR) Meta-Analysis | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Trials | AD | Placebo | VR (95% CI) | p | Q(df) | AD | Placebo | CVR (95% CI) | P | Q (df) |
| AD all | 169 | 32650 | 18746 | 1.01 (0.99-1.02) | .33 | 121.74 (168) | .55** | .52** | 0.82 (0.80-0.84) | .00 | 243.53 (168)** |
| SSRI | 89 | 13146 | 9024 | 1.01 (0.99-1.04) | .32 | 124.94 (88)**[a] | .56** | .45** | 0.83 (0.80-0.86) | .00 | 137.29 (88)** |
| SNRI | 53 | 9970 | 7226 | 1.00 (0.98-1.03) | .66 | 14.26 (52) | .53** | .56** | 0.81 (0.78-0.84) | .00 | 70.37 (52)** |
| Atypical | 54 | 8780 | 6293 | 1.00 (0.97-1.02) | .69 | 17.62 (53) | .71** | .74** | 0.83 (0.79-0.86) | .00 | 61.53(53) |
| Tricyclics | 11 | 754 | 771 | 1.04 (0.95-1.14) | .40 | 15.80 (10) | -.34 | .64* | 0.65 (0.58-0.72) | .00 | 6.96 (10) |

Notes:

p (M, SD) is the Pearson correlation of mean pre-post differences in depression (M) and the respective standard deviation (SD). Q(df) is the heterogeneity index for the meta-analysis. * $p < .05$, ** $p < .01$, [a]This result was caused by an outlier (Study Dube 2010, NCT00420004), and after removing this study, the heterogeneity index was Q(df=87)=54.54, p = 0.99.

**Legend for Figure 1.**

Figure 1. Results from simulation analyses (hypothetical variance ratios) for different definitions of benefiters (x-axis) and different percentages of benefiters (individual lines with the percentages on the right side of the line). The horizontal thick line is the result from our meta-analysis (VR = 1.01), the horizontal gray lines correspond with upper and lower limits of the confidence interval of the meta-analysis (0.99-1.02).



Supplementary Table 1: Results for trials where only end of treatment depression scores were available

| | n | | | Variance Ratio Meta Analysis | | | r (M, SD) | | Coefficient of Variance Ratio Meta-Analysis | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Trials | AD | Placebo | VR (95% CI) | p | Q(df) | AD | Placebo | CVR (95% CI) | P | Q (df) |
| AD all | 84 | 10879 | 7346 | 0.98 (0.96-1.00) | .07 | 52.53 (83) | .36** | .54** | 1.15 (1.11-1.18) | .00 | 92.67 (83) |
| SSRI | 40 | 5424 | 3763 | 0.99(0.96-1.03) | .71 | 28.60 (39) | .63** | .37* | 1.13 (1.08-1,17) | .00 | 35.23 (39)** |
| SNRI | 13 | 1440 | 1288 | 0.99 (0.94-1.05) | .73 | 4.50 (12) | -.49 | .62* | 1.15 (1.07-1.24) | .00 | 9.82 (12)** |
| Atypical | 31 | 4245 | 3092 | 0.97 (0.94-1.00) | .07 | 12.63 (30) | .43* | .14 | 1.12 (1.07-1.17) | .00 | 34.37 (30) |
| Tricyclics | 13 | 670 | 643 | 0.93 (0.86-1.01) | .07 | 9.18(10) | .09 | .72** | 1.37 (1.20-1.56) | .00 | 15.64 (12) |

Notes:

p (M, SD) is the pearson correlation of mean pre-post differences in depression (M) and the respective standard deviation (SD). Q(df) ist the heterogenity index for the meta-analysis. * p < .05, ** p < .01.

Supplementary Table 2: Meta-analytic results for variance ratios and coefficients of variance ratios, individual AD, only studies reporting pre-post-differences

| | n | | | Variance Ratio (VR) Meta-Analysis | | | r (M, SD) | | Coefficient of Variance Ratio (CVR) Meta-Analysis | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Trials | AD | Placebo | VR (95% CI) | p | Q(df) | AD | Placebo | CVR (95% CI) | P | Q (df) |
| Agomelatine | 0 | - | - | - | - | - | - | - | - | - | - |
| Amitriptyline | 11 | 754 | 771 | 1.04 (0.95-1.14) | .40 | 15.80 (10) | -34 | .64* | 0.65(0.58-0.72) | .00 | 6.96 (10) |
| Bupropion | 13 | 2095 | 1887 | 0.99 (0.94-1.03) | .99 | 2.57 (12) | .89** | .87** | 0.88 (0.82-0.94) | .00 | 7.70 (12) |
| Citalopram | 11 | 1819 | 1335 | 1.02 (0.97-1.07) | .50 | 4.35 (10) | .54** | .71* | 0.85 (0.79-0.92) | .00 | 4.83 (10) |
| Clomipramine | 0 | - | - | - | - | - | - | - | - | - | - |
| Desvenlafaxine | 7 | 2124 | 1241 | 1.00 (0.95-1.05) | .91 | 1.17 (6) | -.05 | -.03 | 0.83 (0.77-0.89) | .00 | 1.33 (6) |
| Duloxetine | 20 | 2953 | 2371 | 1.01 (0.97-1.05) | .66 | 1.81 (19) | .61** | .73** | 0.76 (0.71-0.81) | .00 | 25.63 (19) |
| Escitalopram | 16 | 2749 | 2114 | 1.06 (0.95-1.19) | .28 | 74.54 (15)**[a] | .65** | .05 | 0.90 (0.79-1.02) | .11 | 55.82 (15)** |
| Fluoxetine | 18 | 2765 | 1991 | 1.00 (0.96-1.05) | .95 | 6.96 (17) | .60** | .71** | 0.85 (0.78-0.92) | .00 | 20.58(17) |
| Fluvoxamine | 0 | - | - | - | - | - | - | - | - | - | - |
| Levomilnacipran | 5 | 1566 | 1032 | 1.03 (0.97-1.09) | .31 | 2.91 (4) | .20 | .45 | 0.83 (0.74-0.92) | .00 | 6.02 (4) |
| Milnacipran | 0 | - | - | - | - | - | - | - | - | - | - |
| Mirtazapine | 10 | 624 | 466 | 1.00 (0.91-1.09) | .96 | 8.74 (9) | .52 | .35 | 0.72 (0.63-0.83) | .00 | 10.91 (9) |
| Nefazodone | 9 | 767 | 528 | 1.01 (0.97-1.09) | .90 | 0.34 (8) | .27 | -.29 | 0.80 (0.71-0.91) | .00 | 5.09 (8) |
| Paroxetine | 37 | 4233 | 3333 | 1.00 (0.97-1.04) | .89 | 11.27 (36) | .28 | .26 | 0.82 (0.78-0.86) | .00 | 41.51 (36) |
| Reboxetine | 6 | 1035 | 944 | 0.99 (0.93-1.05) | .69 | 0.45 (5) | .51 | .66 | 0.92 (0.79-1.07) | .28 | 12.10 (5)* |
| Sertraline | 15 | 1534 | 1374 | 0.98 (0.90-1.06) | .57 | 27.03 (14) | .75** | .77** | 0.77 (0.71-0.83) | .00 | 11.68 (14) |
| Trazodone | 0 | - | - | - | - | - | - | - | - | - | - |

| | n | | | Variance Ratio (VR) Meta-Analysis | | | r (M, SD) | | Coefficient of Variance Ratio (CVR) Meta-Analysis | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Venlafaxine | 16 | 2292 | 1798 | 0.99 (0.95-1.04) | .82 | 6.64 (15) | .36 | .18 | 0.80 (0.75-0.85) | .00 | 11.86 (15) |
| Vilazodone | 5 | 1360 | 1064 | 0.99 (0.93-1.05) | .72 | 1.07 (4) | .55 | .28 | 0.81 (0.73-0.90) | .00 | 4.96 (4) |
| Vortioxetine | 14 | 3562 | 2098 | 1.00 (0.96-1.04) | .99 | 0.41 (13) | .32 | .79** | 0.81 (0.76-0.88) | .00 | 23.04 (13)* |

Notes:

r (M, SD) is the Pearson correlation coefficient of mean pre-post differences in depression (M) and the respective standard deviation (SD). Q(df) is the heterogeneity index of the meta-analysis. In case of AD without trials that reported pre-post differences, the results for the post-values of depression scores also resulted in VR ≈ 1.00 (Agomelatine: VR = 0.96, CI 0.92-1.01; Fluvoxamine VR = 0.99, CI 0.88-12; Trazodone: VR = 0.96, CI 0.88-1.06. For Clomipramine and Milnacipran, no placebo-controlled trials were available.

* p < .05, ** p < .01

[a]This result was caused by an outlier (Study Dube 2010, NCT00420004), and after removing this study, the heterogeneity index was Q(df=14)=7.36, p = 0.92.

Supplementary Table 3: Different estimations of heterogeneity

| | VR | CI-lower | CI-upper | $I^2$ (%) | p |
|---|---|---|---|---|---|
| **Estimator of heterogeneity** | | | | | |
| Restricted maximum-likelihood (same as in Winkelbaumer et al., 2019) | 1.01 | 0.99 | 1.02 | 0 | .24 |
| With Knapp and Hartung (KH) adjustment | 1.01 | 1.00 | 1.02 | 0 | .25 |
| Sidik-Jonkman (SJ) | 1.01 | 0.99 | 1.02 | 32 | .39 |
| With KH adjustment | 1.01 | 1.00 | 1.02 | 32 | .25 |
| Paule Mandel | 1.01 | 0.99 | 1.02 | 0 | .33 |
| With KH adjustment | 1.01 | 1.00 | 1.02 | 0 | .25 |
| Empirical Bayes | 1.01 | 0.99 | 1.02 | 0 | .33 |
| With KH adjustment | 1.01 | 1.00 | 1.02 | 0 | .25 |
| DerSimonian-Laird | 1.01 | 0.99 | 1.02 | 0 | .33 |
| With KH adjustment | 1.01 | 1.00 | 1.02 | 0 | .25 |
| | | | | | |
| **Unweighted results** (as recommended for $tau^2 = 0$ by Viechtbauer, 2010) | | | | | |
| Restricted maximum-likelihood (same as in Winkelbaumer et al. 2019) | 1.01 | 0.99 | 1.03 | 0 | .24 |
| Sidik-Jonkman (SJ) | 1.01 | 0.99 | 1.03 | 32 | .30 |
| | | | | | |
| **Using manual values for $tau^2$** | | | | | |
| $tau^2$=0.0108 (3 times the SJ) | 1.01 | 0.99 | 1.03 | 59 | .46 |
| $tau^2$=0.036 (10 times the SJ) | 1.01 | 0.98 | 1.04 | 83 | .59 |
| $tau^2$=0.36 (100 times the SJ) | 1.01 | 0.92 | 1.11 | 98 | .83 |
| | | | | | |