PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	What are the chances for personalized treatment with antidepressants? Detection of patient by treatment interaction with a variance ratio meta-analysis
AUTHORS	Plöderl, Martin; Hengartner, Michael

VERSION 1 – REVIEW

REVIEWER	Stephen Senn Consultant Statistician, UK.
REVIEW RETURNED	20-Oct-2019

GENERAL COMMENTS	This is an interesting paper providing potentially useful evidence regarding the scope (or not) for personalising treatment in depression. I have one concern regarding results that the authors should address. Otherwise I have only minor comments regarding language etc.<P> <B>Major comment</B> I worry about the Q statistics in Table 1. Three seem to be very much on the low side given the degrees of freedom (see attached page of critical values – Please contact the plublisher for this file). In other words the variation is much less than one would expect by chance. I see two possible explanations 1) The trials are not completely randomised (this will almost certainly be the case) and common blocking factors are contributing to the variance of both arms 2) The method of analysis is incorrect a) the transformation of the variance ratio is not Normal or b) the variance that is assigned to the transformation is incorrect. Please check this point carefully. You can perhaps use your simulation to do so. You might be interested to have a look at (Senn, S.J., et al., Random main effects of treatment : A case study with a network met a-analysis. Biometrical Journal, 2019. 61.) where we tried to apply a meta-analysis to a similar problem. (However, there is no guarantee that this is appropriate to your problem or even correct! You should investigate first the method you have been using.)<P> <B>Minor comments</B> 1) P7 L20 'limited to demonstrating' not 'limited to demonstrate' 2) P9 L57 'as many' not 'as much'. 3) General use of 'in case'. Your use is best avoided. This is because, in everyday English when one says something like ?I shall take my umbrella in case it rains' this means that the umbrella will be taken because if it rains, it will be useful to have one. It does not mean that the umbrella will be taken if and only if it rains. So, for example, you could replace P8 L20 etc with something like 'Where trials compared different...' AD..

REVIEWER	Giovanni Ostuzzi

	University of Verona, Italy
REVIEW RETURNED	05-Nov-2019

GENERAL COMMENTS	This is a very interesting and innovative topic, still overlooked in psychiatry as compared to other fields. The paper provides an original contribution to the rapidly growing debate around personalized interventions in psychiatry. The methodology is accurate and in line with widely recognized quality standards. The rationale is well-explained. The discussion is not free from some degree of technicality, possibly challenging for the lay reader, however this is probably unavoidable given the sophistication of the study design. Furthermore, authors provide a clear synthesis of results. Overall, results are critical about the possible role of precision medicine for people suffering from depression, and this might generate an interesting debate.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Stephen Senn

Institution and Country: Consultant Statistician, UK.

Please state any competing interests or state 'None declared': None of which I am aware. I

keep a full declaration of interest online here

http://www.senns.demon.co.uk/Declaration_Interest.htm

Please leave your comments for the authors below

This is an interesting paper providing potentially useful evidence regarding the scope (or not) for personalising treatment in depression. I have one concern regarding results that the authors should address. Otherwise I have only minor comments regarding language etc

Major comment

I worry about the Q statistics in Table 1. Three seem to be very much on the low side given the degrees of freedom (see attached page of critical values). In other words the variation is much less than one would expect by chance. I see two possible explanations 1) The trials are not completely randomised (this will almost certainly be the case) and common blocking factors are contributing to the variance of both arms 2) The method of analysis is incorrect a) the transformation of the variance ratio is not Normal or b) the variance that is assigned to the transformation is incorrect. Please check this point carefully. You can perhaps use your simulation to do so. You might be interested to have a look at (Senn, S.J., et al., Random main effects of treatment : A case study with a network met a-analysis. Biometrical Journal, 2019. 61.) where we tried to apply a meta-analysis to a similar problem. (However, there is no guarantee that this is appropriate to your problem or even correct! You should investigate first the method you have been using.)

Reply: We thank the reviewer for detecting this potential problem, possible explanations, providing the critical chi-square values, and for pointing out the hierarchical Bayesian meta-analytic approach where the treatment effect is modelled as being random, too.

We guess that the reason for lacking variance across trials, indicated by the Q-Values but also by the low I^2 values (not presented in the manuscript but below in the table) is likely caused by incomplete randomization of trials. Furthermore, it is well known that patients are selected according to several characteristics (baseline depression severity, no suicide ideation, no comorbidity, or by previous response to medication). This perhaps narrows down the variation between trials.

The reviewer also requested to check if the analysis is wrong or inappropriate due to lacking normality assumptions. We tried to clarify these issues with the following steps.

First, we inspected the log-transformed variance ratios, and it seems that these are normally distributed (see attachment).

Second, we manually calculated the variance ratios and their variances using equations Nr. 9 and Nr. 10 provided by Nakagawa et al. (2015) and cross-checked them with the results provided by R's metafor package. The results were identical, so the calculations seem to be correct.

$$\ln \mathrm{VR} = \ln\left(\frac{s_E}{s_C}\right) + \frac{1}{2(n_E - 1)} - \frac{1}{2(n_C - 1)}, \qquad \text{eqn 9}$$

$$s^2_{\ln \mathrm{VR}} = \frac{1}{2(n_C - 1)} + \frac{1}{2(n_E - 1)}. \qquad \text{eqn 10}$$

Third, we checked if there are convergence problems within R's metafor package, i.e., if there are sufficient numbers of iterations to estimate tau^2. It turned out that tau^2 quickly converged within a few iterations for different estimators.

Fourth, we used different estimators of heterogeneity in the meta-analysis, but the results were roughly the same (see attached table). Only for the Sidik Jonkman estimator the heterogeneity increased to I^2 of 32%, but the result for the VR and the CI remained the same (only the p-value increased slightly).

Furthermore, as recommended by Viechtbauer (2010), when $tau^2 = 0$, we also checked the unweighted results and results with the Knapp and Hartung adjustment. Again, this did not change the results, except that some CI-Values increased by 0.01.

Fifth, we run sensitivity analyses by increasing/varying the heterogeneity with the parameter tau2 in the metafor-package. The results remained quite robust. The confidence intervals for the meta-analytic results widened for large values of tau.

Estimator of heterogeneity	VR	CI	Cu	I^2 (%)	Q (df=169)	p
Restricted maximum-	1.01	0.99	1.02	0	121.74	.24
likelihood (same as in						
Winkelbaumer et al., 2019)						
With Knapp and Hartung	1.01	1.00	1.02	0	121.74	.25
(KH) adjustment						
Sidik-Jonkman (SJ)	1.01	0.99	1.02	32	121.74	.39
With KH adjustment	1.01	1.00	1.02	32	121.74	.25
Paule Mandel	1.01	0.99	1.02	0	121.74	.33
With KH adjustment	1.01	1.00	1.02	0	121.74	.25
Empirical Bayes	1.01	0.99	1.02	0	121.74	.33
With KH adjustment	1.01	1.00	1.02	0	121.74	.25
DerSimonian-Laird	1.01	0.99	1.02	0	121.74	.33
With KH adjustment	1.01	1.00	1.02	0	121.74	.25
Unweighted results (in case of						
tau2 = 0, as recommended by						
Viechtbauer (2010)						
Restricted maximum-	1.01	0.99	1.03	0	121.74	.24
likelihood (same as in						
Winkelbaumer et al. 2019)						
Sidik-Jonkman (SJ)	1.01	0.99	1.03	32	121.74	.30
Using manual values for tau2						
tau2=0.0108 (3 times the SJ)	1.01	0.99	1.03	59	121.74	.46
tau2=0.036 (10 times the SJ)	1.01	0.98	1.04	83	121.74	.59
tau2=0.36 (100 times the SJ)	1.01	0.92	1.11	98	121.74	.83

We did not apply a hierarchical meta- analysis as suggested in Senn et al. (2019), i.e., an analysis with an extra level assuming that the main effect of a treatment is also random. We think that this deserves an extra paper with both efficacy measures (standardized mean placebo-drug differences and a variance ratio meta-analysis).

To account for the reviewer's points, we added statement in the discussion section:

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 supplement). 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 (Zimmerman et al., 2019).

References

Nakagawa, Shinichi, et al. "Meta-analysis of variation: ecological and evolutionary

applications and beyond." *Methods in Ecology and Evolution* 6.2 (2015): 143-152.

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. Journal of statistical Software, 36 (3).

Zimmerman, Mark, et al. "Have Treatment Studies of Depression Become Even Less Generalizable? Applying the Inclusion and Exclusion Criteria in Placebo-Controlled Antidepressant Efficacy Trials Published over 20 Years to a Clinical Sample." *Psychotherapy and Psychosomatics* 88.3 (2019): 165-170.

Minor comments

1) P7 L20 'limited to demonstrating' not 'limited to demonstrate'
2) P9 L57 'as many' not 'as much'.
3) General use of 'in case'. Your use is best avoided. This is because, in everyday English when one says something like ?I shall take my umbrella in case it rains' this means that the umbrella will be taken because if it rains, it will be useful to have one. It does not mean that the umbrella will be taken if and only if it rains. So, for example, you could replace P8 L20 etc with something like 'Where trials compared different...' AD..

Reply: Thank you for pointing out these language issues! We corrected that.

Reviewer: 2

Reviewer Name: Giovanni Ostuzzi

Institution and Country: University of Verona, Italy

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

This is a very interesting and innovative topic, still overlooked in psychiatry as compared to other fields.

The paper provides an original contribution to the rapidly growing debate around personalized interventions in psychiatry.

The methodology is accurate and in line with widely recognized quality standards.

The rationale is well-explained. The discussion is not free from some degree of technicality, possibly challenging for the lay reader, however this is probably unavoidable given the sophistication of the study design. Furthermore, authors provide a clear synthesis of results. Overall, results are critical about the possible role of precision medicine for people suffering from depression, and this might generate an interesting debate.

Reply: Thank you very much for this encouraging feedback. We fully agree and with this reviewer and also hope that our paper will stipulate a critical debate about precision psychiatry.

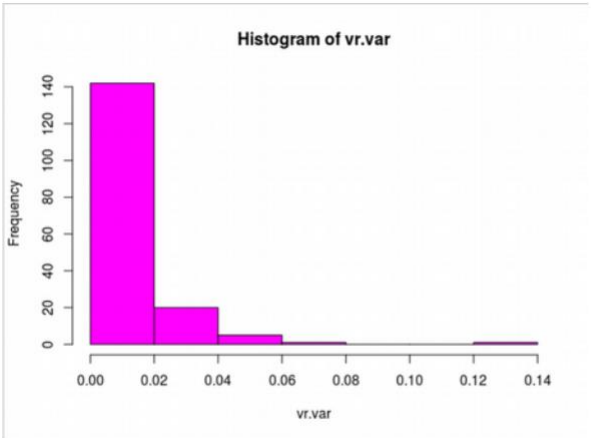
Additional Changes:

We detected a typo in the original data-set by Cipriani et al. (Placebo instead of placebo). This lead to the inclusion of one additional SNRI-trial, but the results were essentially unchanged (the main results in the abstract were similar to the second decimal).

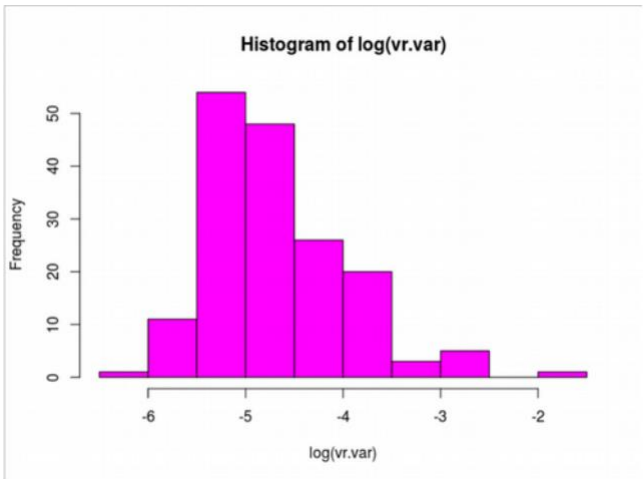
Variance of the Variance-Ratio – Effect of log-transformation

All Antidepressants

Distribution of the variance of the VR

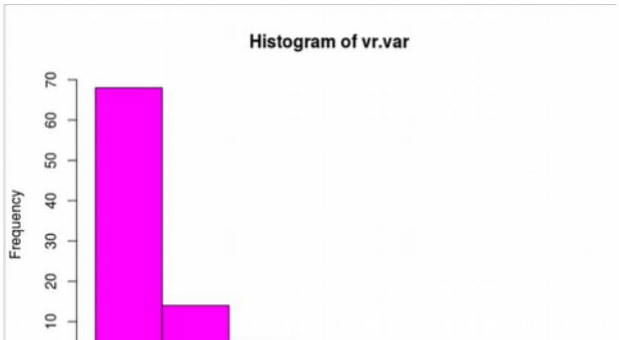


After log-transformation

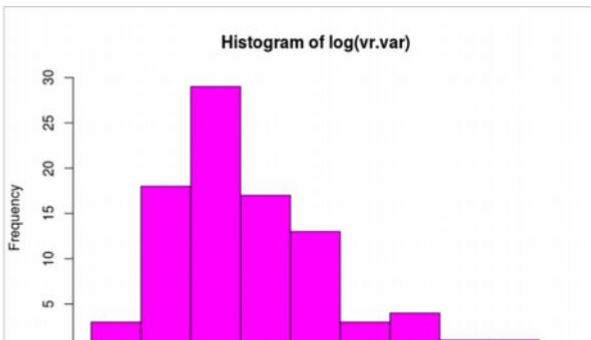


SSRI

Distribution of the variance of the VR

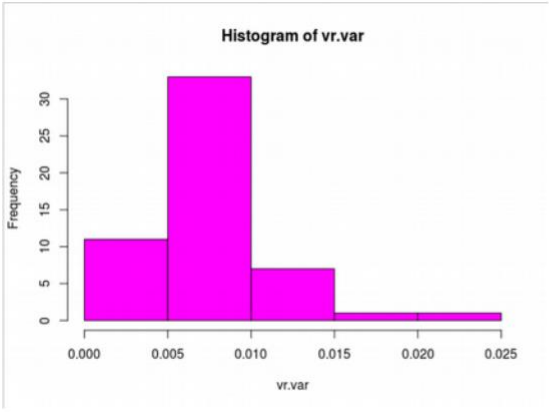


After log-transformation

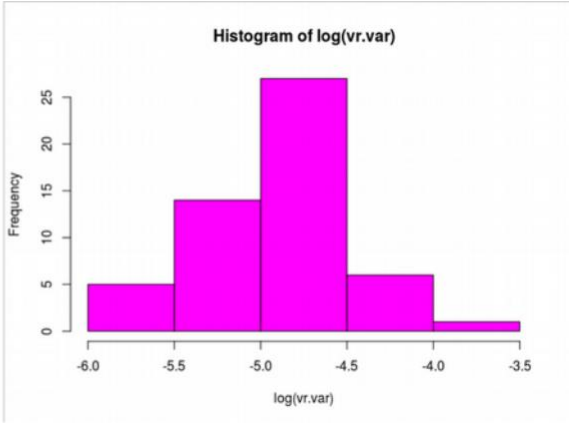


SNRI

Distribution of the variance of the VR

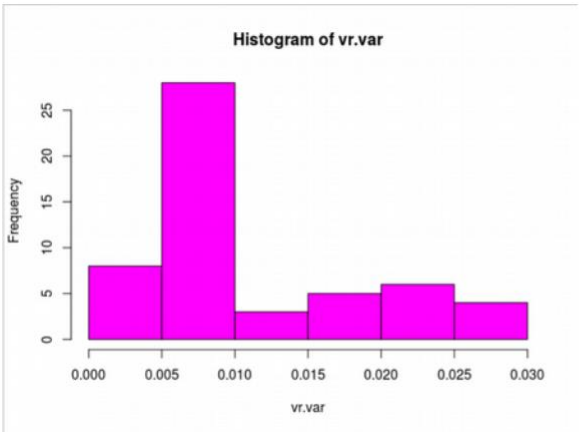


After log-transform

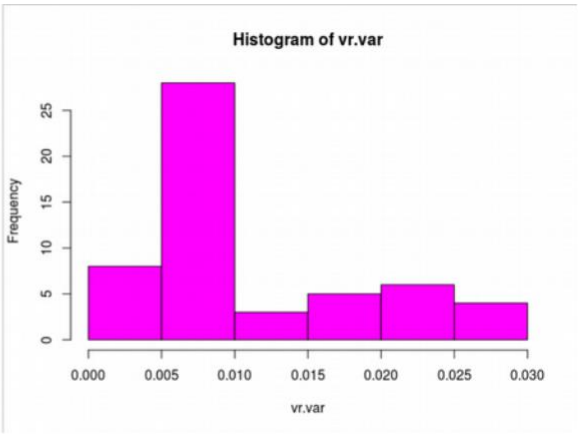


Atypicals

Distribution of the variance of the VR

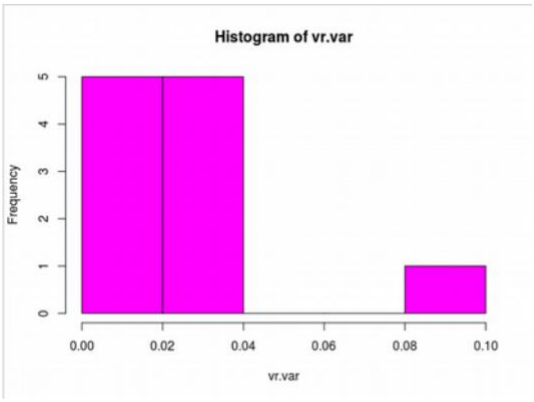


After log-transformation

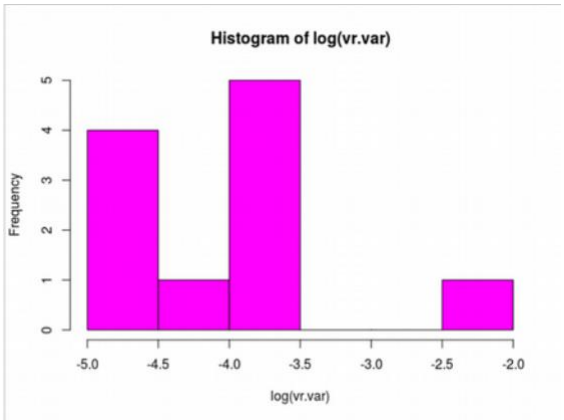


Tricyclics

Distribution of the variance of the VR



After log-transformation



VERSION 2 – REVIEW

REVIEWER	Stephen Senn Consultant statistician, UK
REVIEW RETURNED	05-Dec-2019

GENERAL COMMENTS	I still worry about the Q Statistics.(Thank you for having carried out extensive investigations including some along the lines I proposed.) You write ',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,' . However, In fact, my worry was that the VRs did not vary <I>enough, even assuming pure random variation. <P> However, I accept that you have thoroughly investigated the problem. It may be that your paper will inspire others to investigate further and this worry should not hold up publication.