Re-analysis of a meta-analysis about tryptophan and depression.

Martin Plöderl[1]

[1]Center for Inpatient Psychotherapy and Crisis Intervention, University Clinic for Psychiatry, Psychotherapy, and Psychosomatics, Paracelsus Medical University, Austria

This is a reanalysis of a meta-analysis about L-tryptophan blood levels and depression, which became part of the controversy around a recent umbrella review about the role of serotonin in depression. The reanalysis revealed major methodological limitations, raising doubts on the conclusions in the original publication that levels of tryptophan are lowered among depressed compared to non-depressed individuals. The data is also compatible with a null effect and no firm conclusion should be made.

Keywords: tryptophan, depression, meta-analysis

Background

This is a reanalysis of the meta-analysis by Pu et al. (2021) about L-tryptophan blood levels and depression, which became part of the controversy around a recent umbrella review about the role of serotonin in depression by Moncrieff et al. (2023). The authors of the review were reproached for not including the meta-analysis by Pu et al. (2021) by experts from the Science Media Centre and in a letter in response to the article signed by many leading psychiatrists (Jauhar et al., 2023; Science Media Center, 2022). For example: "The article shows that systematic umbrella reviews leave significant room for interpretation. Also, what you leave out can be as important as what you put in. The authors did not include a meta-analysis published in the same journal (Molecular Psychiatry) in 2021, the abstract of which concluded: '...our integrated results revealed that metabolic changes in the peripheral blood were associated with MDD, particularly decreased L-tryptophan...' (Science Media Center, 2022).

Despite the fact that the study by Pu et al. (2021) did not fulfill Moncrieff et al.'s (2023) pre-specified inclusion criteria, a closer look seems nonetheless important, given the role of the study in the controversy. In their meta-analysis, Pu et al. reviewed the evidence for blood metabolites, including L-tryptophan, and their association with depression. They included 26 studies in a random-effects meta-analysis, and it was found that the tryptophan levels were significantly lower among depressed individuals versus nondepressed controls, with a medium-sized difference of SMD = 0.46 (0.66 to 0.26). Pu et al. also reported that there were signs of publication bias, indicated by a highly significant Egger's test. After correcting for this bias with the trim-and-fill method, the effect even increased to SMD = -0.7. This is unusual, given that publication bias usually leads to an overestimation of true effects. This was one more reason for me to take a closer look at the study by Pu et al.

Methods

The data were kindly provided by the corresponding author Peng Xie via email on 2022-08-22. Replication of the meta-analysis and correction of publication bias was done with R, using the *meta* and *metafor* packages. The R code is available online (https://osf.io/he9x3/).
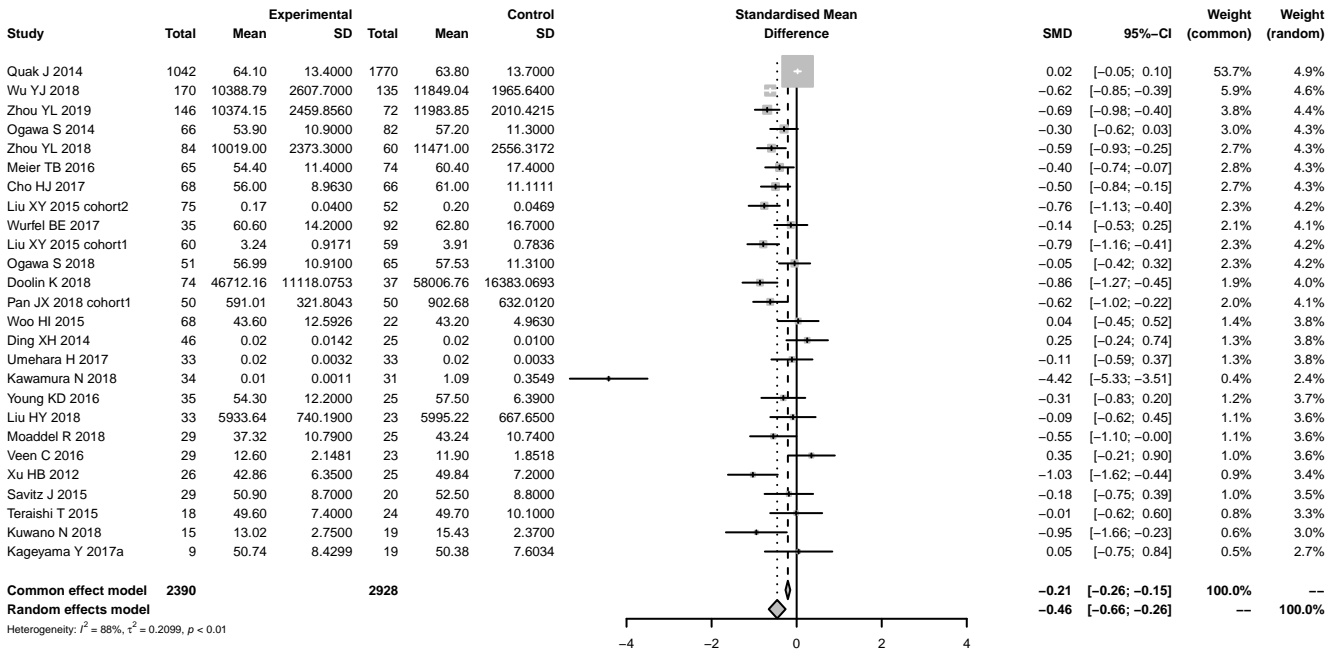
Results and Discussion

Replication of the meta-analyses

Using the *meta* package in R, my results from the random-effects meta-analysis were exactly the same as the ones reported by Pu et al. (2021) when using Cohen's d as measure of the standardized mean difference (SMD) and the DerSimonian-Laird estimator for τ^2: SMD = -0.46 (95% CI -0.66 – -0.26) (Figure 1). Heterogeneity is high: $\tau^2 = 0.21$ (95% CI 0.18 – 0.76), $I^2 = 87.7\%$ (95% CI 82.9 – 90.9]). Of note, the result from the fixed-effect model (not reported or discussed by Pu et al.) was also statistically significant but much smaller: SMD = -0.21 (95% CI -0.26 – -0.15).

When looking at the forest plot (Figure 1), there are three interesting features. First, the numerical values for means and standard deviations are strikingly different between the studies, likely because different measurements of tryptophan were used in different studies. It could be worth exploring with a meta-regression whether this can explain some of the large heterogeneity.

Figure 1

Forest plot of the meta-analysis for tryptophan studies. Studies are sorted by sample size.

Study	Experimental Total	Mean	SD	Control Total	Mean	SD	Standardised Mean Difference	SMD	95%–CI	Weight (common)	Weight (random)
Quak J 2014	1042	64.10	13.4000	1770	63.80	13.7000		0.02	[−0.05; 0.10]	53.7%	4.9%
Wu YJ 2018	170	10388.79	2607.7000	135	11849.04	1965.6400		−0.62	[−0.85; −0.39]	5.9%	4.6%
Zhou YL 2019	146	10374.15	2459.8560	72	11983.85	2010.4215		−0.69	[−0.98; −0.40]	3.8%	4.4%
Ogawa S 2014	66	53.90	10.9000	82	57.20	11.3000		−0.30	[−0.62; 0.03]	3.0%	4.3%
Zhou YL 2018	84	10019.00	2373.3000	60	11471.00	2556.3172		−0.59	[−0.93; −0.25]	2.7%	4.3%
Meier TB 2016	65	54.40	11.4000	74	60.40	17.4000		−0.40	[−0.74; −0.07]	2.8%	4.3%
Cho HJ 2017	68	56.00	8.9630	66	61.00	11.1111		−0.50	[−0.84; −0.15]	2.7%	4.3%
Liu XY 2015 cohort2	75	0.17	0.0400	52	0.20	0.0469		−0.76	[−1.13; −0.40]	2.3%	4.2%
Wurfel BE 2017	35	60.60	14.2000	92	62.80	16.7000		−0.14	[−0.53; 0.25]	2.1%	4.1%
Liu XY 2015 cohort1	60	3.24	0.9171	59	3.91	0.7836		−0.79	[−1.16; −0.41]	2.3%	4.2%
Ogawa S 2018	51	56.99	10.9100	65	57.53	11.3100		−0.05	[−0.42; 0.32]	2.3%	4.2%
Doolin K 2018	74	46712.16	11118.0753	37	58006.76	16383.0693		−0.86	[−1.27; −0.45]	1.9%	4.0%
Pan JX 2018 cohort1	50	591.01	321.8043	50	902.68	632.0120		−0.62	[−1.02; −0.22]	2.0%	4.1%
Woo HI 2015	68	43.60	12.5926	22	43.20	4.9630		0.04	[−0.45; 0.52]	1.4%	3.8%
Ding XH 2014	46	0.02	0.0142	25	0.02	0.0100		0.25	[−0.24; 0.74]	1.3%	3.8%
Umehara H 2017	33	0.02	0.0032	33	0.02	0.0033		−0.11	[−0.59; 0.37]	1.3%	3.8%
Kawamura N 2018	34	0.01	0.0011	31	1.09	0.3549		−4.42	[−5.33; −3.51]	0.4%	2.4%
Young KD 2016	35	54.30	12.2000	25	57.50	6.3900		−0.31	[−0.83; 0.20]	1.2%	3.7%
Liu HY 2018	33	5933.64	740.1900	23	5995.22	667.6500		−0.09	[−0.62; 0.45]	1.1%	3.6%
Moaddel R 2018	29	37.32	10.7900	25	43.24	10.7400		−0.55	[−1.10; −0.00]	1.1%	3.6%
Veen C 2016	29	12.60	2.1481	23	11.90	1.8518		0.35	[−0.21; 0.90]	1.0%	3.6%
Xu HB 2012	26	42.86	6.3500	25	49.84	7.2000		−1.03	[−1.62; −0.44]	0.9%	3.4%
Savitz J 2015	29	50.90	8.7000	20	52.50	8.8000		−0.18	[−0.75; 0.39]	1.0%	3.5%
Teraishi T 2015	18	49.60	7.4000	24	49.70	10.1000		−0.01	[−0.62; 0.60]	0.8%	3.3%
Kuwano N 2018	15	13.02	2.7500	19	15.43	2.3700		−0.95	[−1.66; −0.23]	0.6%	3.0%
Kageyama Y 2017a	9	50.74	8.4299	19	50.38	7.6034		0.05	[−0.75; 0.84]	0.5%	2.7%
Common effect model	**2390**			**2928**				**−0.21**	**[−0.26; −0.15]**	**100.0%**	--
Random effects model								**−0.46**	**[−0.66; −0.26]**	--	**100.0%**

Heterogeneity: $I^2 = 88\%$, $\tau^2 = 0.2099$, $p < 0.01$

Second, the study by Kawamura 2018 seems to be an outlier because the SMD is > 4. This was not discussed by Pu et al. A first guess would be that the effect size has been calculated incorrectly using standard errors instead of standard deviations. Excluding the Kawamura study had a larger impact on the result from the random-effects model, SMD = -0.36 (95% CI -0.52 – -0.20), than for the fixed-effect model, SMD = -0.19 (95% CI -0.25 – -0.13).

Third, and perhaps most importantly, there is one very large study (Quak 2014) with n = 2,812 participants with a near-perfect zero finding (SMD = 0.02) and a narrow confidence interval (95% CI -0.05 – 0.10). All the samples in the other studies are much smaller, ranging from n = 28 to n = 305. This likely explains the large difference between the random-effects and fixed-effect models. In random-effect models, the size of individual studies has less impact on the overall results, as their weights are much smaller compared to fixed-effect models (see the last two columns of Figure 1). Discrepancies between fixed-effect and random-effects models are related to the issue of reporting biases.
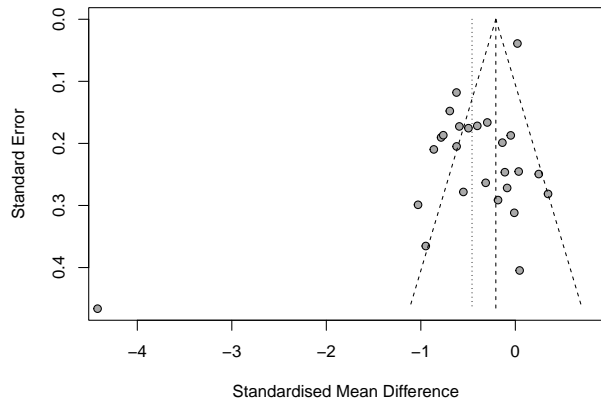
Reporting biases

Pu et al. analyzed and discussed publication bias, but they apparently failed to consider a recommendation in the Cochrane handbook (Sterne et al., 2008) which addresses the possible disagreements between the fixed-effect and random-effects models: "We recommend that when review authors are concerned about the influence of small-study effects on the results of a meta-analysis in which there is evidence of between-study heterogeneity ($I^2 > 0$), they compare the fixed-effect and random-effects estimates of the intervention effect. If the estimates are similar, then any small-study effects have little effect on the intervention effect estimate. If the random-effects estimate is more beneficial, review authors should consider whether it is reasonable to conclude that the intervention was more effective in the smaller studies." (p. 10:28).

The Cochrane Handbook also recommends: "If the larger studies tend to be those conducted with more methodological rigour, or conducted in circumstances more typical of the use of the intervention in practice, then review authors should consider reporting the results of meta-analyses restricted to the larger, more rigorous studies." (10:28)

In the worst case, the tryptophan studies resemble

Figure 2

Vertical lines correspond to the meta-analytical results (fixed-effect and random-effects models).


Figure 3

Funnel plot after correction for publication bias with the trim-and-fill method, using R's metabias function from the meta package. Gray dots are the existing studies, white dots are imputed studies.

the scenario from magnesium and myocardial infarction, as pointed out in the Cochrane Handbook (Sterne et al., 2008): "This is a well-known example in which beneficial effects of intervention were found in a meta-analysis of small studies, but were subsequently contradicted when the very large ISIS-4 study found no evidence that magnesium affected mortality (p. 10:27).

Publication bias seems to be an issue, as Pu et al. (2021) pointed out, based on the results of Egger's test for funnel-plot asymmetry. I was able to replicate this with the meta package (Egger's test: t = -3.22, df = 24, p = 0.004). These findings are clear evidence of asymmetry of the funnel plot (Figure 2).

When trying to apply the trim-and-fill method to correct for publication bias with the function *metabias* of the *meta*-package, the effect becomes very small and is no longer statistically significant: SMD = -0.08 (95% CI -0.29 – 0.13), p = 0.46. It was estimated that 10 studies were missing (Figure 3). This differs greatly from Pu et al.'s (2021) trim-and-fill analyses, where the magnitude of the effect increased to SMD = -0.70. The trim-and-fill method requires the analyst to make a number of decisions, and Pu et al. did not report all of their choices. I experimented with different options and found out that the discrepancy may have resulted from using different modeling approaches and choice of imputation parameters. It seemed that Pu et al. only used or reported the results for random-effects models both for imputing the missing studies and calculating the overall effect. In contrast, R's *metabias* function uses a fixed-effect model to estimate the missing studies and a random-effects model to calculate the overall effect. The reason for this is given in the documentation of the *metabias* function: "Simulation results (Peters et al. 2007) indicate that the fixed-random model, i.e. using a fixed-effect model to
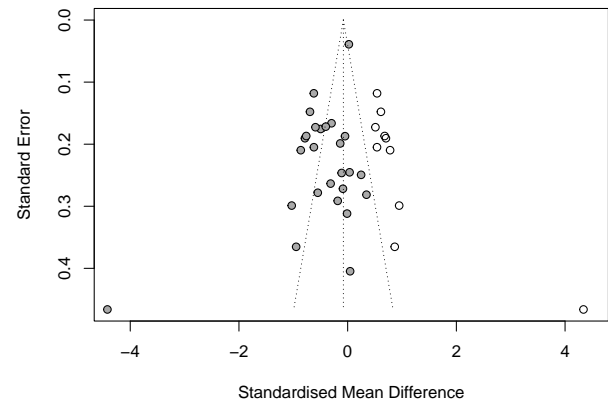
estimate the number of missing studies and a random-effects model to summaries results, (i) performs better than the fixed-fixed model, and (ii) performs no worse than and marginally better in certain situations than the random-random model. Accordingly, the fixed-random model is the default."

For imputing, it is also possible to decide on which side of the funnel plot studies should be imputed, or if this should automatically be decided by the results from Egger's test of asymmetry. As provided in the R code, Pu et al.'s results only appeared by imputing on the left-hand side of the funnel plot.

I informed the corresponding author of Pu et al. (2021), Dr. Peng Xie, about the discrepancies, and he kindly provided their STATA code, attempts to replicate my findings, and valuable arguments about how to best explain the differing results. He referenced to the warnings in the Cochrane handbook: "Therefore, 'corrected' intervention effect estimates from this method should be interpreted with great caution. The method is known to perform poorly in the presence of substantial between-study heterogeneity" (Sterne et al., 2008). I agree, but this raises several concerns for the study by Pu et al.

If estimations of publication bias depend strongly on the choice of a specific statistical method, it is not sufficient to only report results from one specific method, especially if the grossly different results from different methods were already known. Pu et al. (2021) also did not provide details about their statistical approach, hindering reproducibility. Furthermore, it seems that the fixed-random approach for the trim-and-fill method, as described in the documentation of the metabias func-

tion quoted above, would be a better choice than the one that Pu et al. made. Finally, as mentioned in the Cochrane Handbook and by the corresponding author, when there is large heterogeneity, the interpretation of bias-corrected results needs great caution. This was not mentioned as a limitation by Pu et al.

Conclusion

The review of L-tryptophan studies by Pu et al. (2021) was influential in the recent controversy surrounding the role of serotonin in depression (Jauhar et al., 2023; Moncrieff et al., 2023; Science Media Center, 2022). The re-analysis presented here revealed two major limitations of the study by Pu et al. First, by far the largest study had a near perfect null finding, leading to a much smaller effect in the (unreported) fixed-effect and (reported) random-effects models. This was not discussed, against recommendations. Second, the specific statistical method chosen for the estimation of publication bias was not justified and perhaps not optimal. The recommended method leads to a very small bias-corrected effect that is not statistically significant. Generally, in the presence of large heterogeneity, trim-and-fill methods are problematic, but this was again not discussed by Pu et al. The studies in the review are too heterogeneous to draw firm conclusions. If the one very large study is the most reliable, then this would be compatible with a zero difference in L-tryptophan level in depressed versus non-depressed people. Similarly, after correction for publication bias, the overall effect is also compatible with a null effect. Therefore, no firm conclusion should be drawn from Pu et al.'s study.

Author Contact

Martin Plöderl, Zentrum für stationäre Psychotherapie und Krisenintervention, Christian Doppler Klinik, Salzburg, Austria. m.ploederl@salk.at.
Orcid: https://orcid.org/0000-0003-4659-9314.

Conflict of Interest and Funding

Acknowledgments

Author Contributions

Martin Plöderl was the sole author of this study.

Open Science Practices

This article earned the Open Data, Open Materials, and Open Code badge for making the data, materials, and code openly available. It has been verified that the analysis reproduced the results presented in the article. The entire editorial process, including the open reviews, are published in the online supplement.

References

Jauhar, S., Arnone, D., Baldwin, D. S., Bloomfield, M., Browning, M., Cleare, A. J., Corlett, P., Deakin, J. W., Erritzoe, D., Fu, C., et al. (2023). A leaky umbrella has little value: Evidence clearly indicates the serotonin system is implicated in depression. *Molecular Psychiatry*, 1–4.

Moncrieff, J., Cooper, R. E., Stockmann, T., Amendola, S., Hengartner, M. P., & Horowitz, M. A. (2023). The serotonin theory of depression: A systematic umbrella review of the evidence. *Molecular psychiatry*, *28*(8), 3243–3256.

Pu, J., Liu, Y., Zhang, H., Tian, L., Gui, S., Yu, Y., Chen, X., Chen, Y., Yang, L., Ran, Y., et al. (2021). An integrated meta-analysis of peripheral blood metabolites and biological functions in major depressive disorder. *Molecular psychiatry*, *26*(8), 4265–4276.

Science Media Center. (2022). Expert reaction to a review paper on the 'serotonin theory of depression. https://www.sciencemediacentre.org/expert-reaction-to-a-review-paper-on-the-serotonin-theory-of-depression/

Sterne, J. A., Egger, M., & Moher, D. (2008). Addressing reporting biases. In *Cochrane handbook for systematic reviews of interventions*. Wiley Online Library.