Coding Errors Lead to Unsupported Conclusions: A critique of Hofmann et al. (2015)

Donald R. Williams
University of California, Davis, USA

Paul-Christian Bürkner
Aalto University, Finland

Abstract

We have detected coding errors in the meta-analysis of Hofmann et al. (2015) who investigated the effect of intranasal oxytocin on psychiatric symptoms. We demonstrate that, after correcting these errors and reanalysing the data, the main conclusions of Hofmann et al. (2015) are no longer supported.

Keywords: meta-analysis, coding errors, intranasal oxytocin

Introduction

Due to converging evidence in animals and healthy human populations, oxytocin has been identified as potentially having therapeutic properties. As such, numerous randomized controlled trails have investigated the efficacy of intranasal oxytocin (IN-OT) on reducing psychiatric symptoms in clinical populations. As results have been mixed, meta-analytic reviews seeking to synthesize the extant literature have been published. One such review was published in Psychiatry Research (Hofmann et al., 2015). The authors concluded that IN-OT significantly improved psychiatric symptoms and found significant effects on depression, anxiety, psychotic symptoms, and general psychopathology. We found several errors in this paper and, when corrected, resulted in all null results (no significant effect of IN-OT) which suggests that the conclusions of Hofmann et al. (2015) are incorrect. The current letter therefore

has three aims: (1) we will outline several errors and raise questions regarding their analysis; and (2) we will perform a meta-analysis using the same primary studies and similar methods; and (3) we will conclude by stating the importance of issuing a correction.

Errors and questions

Effect size directions

While conducting a meta-analysis on a similar topic, we initially noticed Hofmann et al. (2015) misspecified the direction of one outcome. In other words, the primary study reported that the placebo group improved whereas the IN-OT group did not (Lee et al., 2013). However, Table 1 of Hofmann et al. (2015) reports a large effect of IN-OT (Hedges' $g = 1.07$). Furthermore, all of the outcomes reported in their Table 1 were positive which indicates IN-OT was superior to placebo in all instances. From the primary studies, however, we

extracted the relevant data and found that 6 out of the 16 outcomes used to compute the overall effect should have been negative. As seen in Figure 1 of this letter, all effects to the left of 0 had the wrong direction in Hofmann et al. (2015).

Possible Selection Bias

Decisions made during the research process can influence the presence of an effect (Gelman and Loken, 2014). This is true in meta-analyses, particularly when extracting only one outcome from several possibilities. Anagnostou et al. (2012) reported three outcomes on repetitive behavior. While two of the effects were either minimal (d = 0.13) or in the opposite direction (d = -0.22), Hofmann et al. (2015) selected the largest effect in support of IN-OTs efficacy (d = 0.64). While the Yale-Brown Obsessive Compulsive Scale produced the negative effect in Anagnostou et al. (2012) , the same scale produced a positive effect in Epperson et al. (1996) and was selected for inclusion by Hofmann et al. (2015). Dadds et al. (2014) reported two measures of repetitive behavior. For these outcomes, the placebo group showed improvement between pre- to post-test scores, whereas symptoms actually increased in the IN-OT group. From this study, Hofmann et al. (2015) again selected the outcome that was most favorable to IN-OT. However, this outcome (Child Autism Rating Scale) was not labeled as repetitive behavior in Dadds et al. (2014) while the outcomes that favored the placebo group were considered repetitive behavior. Finally, since multiple outcomes were extracted from some studies, the overall effect of IN-OT on psychiatric symptoms was computed on a subset of effects. Based on the effects reported in Table 1 of Hofmann et al. (2015), the average effect size was larger for the included outcomes (d = 0.83) than the excluded outcomes (d = 0.49).

Misspecified outcomes

We also believe several outcomes were not coded accurately in Hofmann et al. (2015). The majority of outcomes in the psychotic symptoms category were total scores from the Positive and Negative Symptoms Scale (PANSS). Total scores of the PANSS are a combination of negative symptoms, positive symptoms, and general psychopathology Kay et al. (1987). However, Hofmann et al. (2015) included two outcomes as psychotic symptoms that exclusively measured aspects of negative symptoms in schizophrenia. They also coded two Brief Psychiatry Reporting Scale (BPRS) outcomes as general psychopathology. Based on the contents of the scale and other meta-analyses on this topic (Oya et al., 2016), this should have either been coded as psychotic symptoms or they should have provided rationale for divergent

coding. All four of these outcomes were reported as positive which, in addition to the aforementioned errors, likely inflated their meta-analytic estimates.

Meta-analysis

Based on the methods provided in Hofmann et al. (2015), we attempted to replicate their procedures as closely as possible, including outcomes included, effect size calculation, and assessment of publication bias. We then analyzed the data in a manner that was consistent with the extant literature and previous meta-analyses on this topic. All computations were done in R and with the **metafor** package (Viechtbauer, 2010). We used the default settings in **metafor**, including REML for estimating the between-study variance and the Q-profile method for the corresponding confidence intervals. Two-tailed p-values are reported. Our fully reproducible analysis is available on OSF (https://osf.io/kd3en/).

Methods

The exact method used for effect size calculation is not entirely clear in Hofmann et al. (2015). Accordingly, we computed both Hedges' g (SMD) exclusively from the post-treatment scores and the standardized mean change with raw score standardization (SMCR), which is a measure of pre to post-treatment change (r = 0.7) compared between groups (IN-OT vs. Placebo). From their methods section, we think that an effect size similar to the SMCR was most likely used. In the present analysis, when a 95 % confidence interval (CI) excluded zero there was evidence for a significant effect at p-value < 0.05.

Replication Attempt. As seen in Figure 1, the overall estimates for psychiatric symptoms were not significant (SMD = 0.22, z = 1.67, p-value = 0.0953, 95 %-CI [-0.04, 0.47]; SMCR = 0.17, z = 1.23, p-value = 0.217, 95 % = CI [-0.10, 0.43]). Trim and Fill procedures indicated bias in SMD outcomes and, when corrected, the effect was reduced (SMD = 0.07, 95 % CI = [-0.18, 0.32]). There was significant between-study variance for the SMCRs (τ^2 = 0.15, p-value = 0.003), but not for the SMDs (τ^2 = 0.09, p-value = 0.1149). We then obtained estimates for specific symptoms (Table 1).The meta-analytic estimates for depression, anxiety, repetitive behaviors, and general psychopathology were all non-significant (CI's included zero). While using the outcomes reported in Hofmann et al. (2015) produced a significant SMD estimate for psychotic symptoms (Table 1), restricting the outcomes to total psychotic symptoms resulted in a loss of statistical significance.
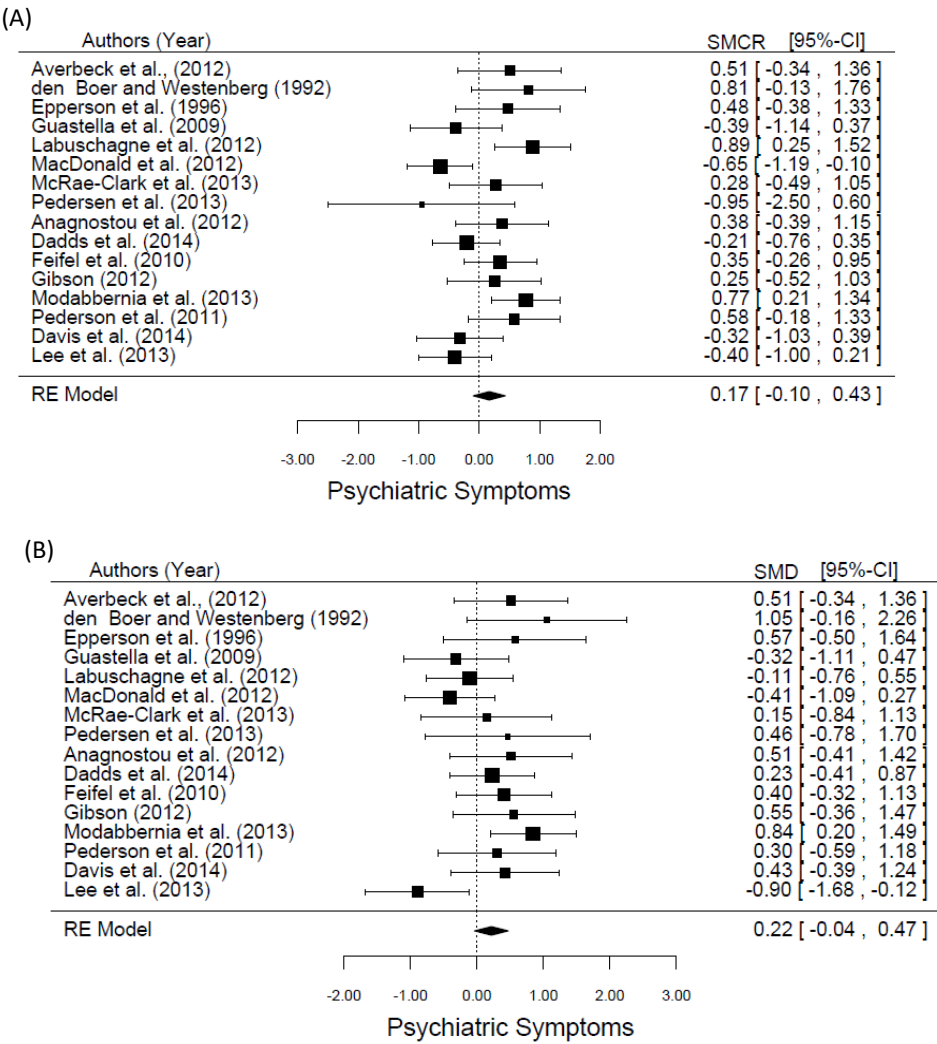
(A)

Authors (Year)	SMCR	[95%-CI]
Averbeck et al., (2012)	0.51	[-0.34 , 1.36]
den Boer and Westenberg (1992)	0.81	[-0.13 , 1.76]
Epperson et al. (1996)	0.48	[-0.38 , 1.33]
Guastella et al. (2009)	-0.39	[-1.14 , 0.37]
Labuschagne et al. (2012)	0.89	[0.25 , 1.52]
MacDonald et al. (2012)	-0.65	[-1.19 , -0.10]
McRae-Clark et al. (2013)	0.28	[-0.49 , 1.05]
Pedersen et al. (2013)	-0.95	[-2.50 , 0.60]
Anagnostou et al. (2012)	0.38	[-0.39 , 1.15]
Dadds et al. (2014)	-0.21	[-0.76 , 0.35]
Feifel et al. (2010)	0.35	[-0.26 , 0.95]
Gibson (2012)	0.25	[-0.52 , 1.03]
Modabbernia et al. (2013)	0.77	[0.21 , 1.34]
Pederson et al. (2011)	0.58	[-0.18 , 1.33]
Davis et al. (2014)	-0.32	[-1.03 , 0.39]
Lee et al. (2013)	-0.40	[-1.00 , 0.21]
RE Model	0.17	[-0.10 , 0.43]

-3.00 -2.00 -1.00 0.00 1.00 2.00

Psychiatric Symptoms

(B)

Authors (Year)	SMD	[95%-CI]
Averbeck et al., (2012)	0.51	[-0.34 , 1.36]
den Boer and Westenberg (1992)	1.05	[-0.16 , 2.26]
Epperson et al. (1996)	0.57	[-0.50 , 1.64]
Guastella et al. (2009)	-0.32	[-1.11 , 0.47]
Labuschagne et al. (2012)	-0.11	[-0.76 , 0.55]
MacDonald et al. (2012)	-0.41	[-1.09 , 0.27]
McRae-Clark et al. (2013)	0.15	[-0.84 , 1.13]
Pedersen et al. (2013)	0.46	[-0.78 , 1.70]
Anagnostou et al. (2012)	0.51	[-0.41 , 1.42]
Dadds et al. (2014)	0.23	[-0.41 , 0.87]
Feifel et al. (2010)	0.40	[-0.32 , 1.13]
Gibson (2012)	0.55	[-0.36 , 1.47]
Modabbernia et al. (2013)	0.84	[0.20 , 1.49]
Pederson et al. (2011)	0.30	[-0.59 , 1.18]
Davis et al. (2014)	0.43	[-0.39 , 1.24]
Lee et al. (2013)	-0.90	[-1.68 , -0.12]
RE Model	0.22	[-0.04 , 0.47]

-2.00 -1.00 0.00 1.00 2.00 3.00

Psychiatric Symptoms

Figure 1. (A) SMCR estimates. (B) SMD estimates. The effect from Averbeck et al. (2012) was computed from a *t*-statistic on post-treatment scores. Outcomes from MacDonald et al. (2013) were obtained from a figure using web plot digitizer Rohatgi, 2017. Pedersen et al. (2013) did not report pre- scores. Through email, Dr. Pederson confirmed that the authors of Hofmann et al. (2015) did not contact then in regards to pre-scores. As such, we used change scores (SMCR) from day 2 to day 3, while SMD was calculated from day 3. We used the same outcome for Dadds et al. (2014) as Hofmann et al. (2015) It should be noted, however, this was pre-treatment and follow-up scores (3 months later). Dr. Dadds confirmed that they did not collect post- data. Standard deviations (SD) for Modabbernia et al. (2013)

Table 1

Meta-analytic estimates for specific symptoms

	SMCR				
	ES	SE	z	*p*-value	95 % CI
Anxiety	0.09	0.17	0.5147	0.6067	[-0.24, 0.41]
Depression	0.29	0.27	1.0843	0.2782	[-0.23, 0.81]
Psychopathology	0.10	0.18	0.5664	0.5711	[-0.25, 0.46]
Psychotic	0.31	0.18	1.6814	0.0927	[-0.05, 0.66]
Repetitive	-0.06	0.21	-0.2889	0.7726	[-0.46, 0.35]
τ^2	0.09	0.06		0.0140	[0.01, 0.29]
	SMD				
Anxiety	0.08	0.18	0.4695	0.6387	[-0.26, 0.43]
Depression	0.28	0.27	1.0300	0.3030	[-0.25, 0.81]
Psychopathology	0.14	0.19	0.7301	0.4653	[-0.24, 0.52]
Psychotic	0.41	0.19	2.1278	0.0334	[0.03, 0.80]
Repetitive	0.15	0.22	0.6876	0.4917	[-0.28, 0.58]
τ^2	0.069	0.069		0.0561	[0.00, 0.40]

Note: SMCR: Standardized mean change with raw score standardization. SMD: Standardized mean difference (Hedges' *g*). ES: Effect size. τ^2: Residual between-study variance after accounting for symptom type.

Conclusion

Although Hofmann et al. (2015) is not a new article, and was recently retracted (Hofmann et al., 2016), there are several reasons this letter deserves attention. First, while they concluded that IN-OT had robust effects on several psychiatric symptoms, our analysis suggests that all effects were non-significant. Second, IN-OT research has become a very active field and ensuring correctness in the publish literature is a mental health priority. Third, there is overwhelming evidence from animal studies supporting the role of oxytocin in psychiatric disorders, especially those comprised of social dysfunction (Lim et al., 2005). By ensuring null results are represented in the literature, researchers might be compelled to improve current methods of delivery or dedicate more resources into developing pharmaceutical drugs that directly activate oxytocin receptors. Accordingly, we hope this letter not only results in a correction but also moves the field forward which is especially important because of the lack of effective treatments for certain aspects of these disorders.

Author Contact

Corresponding author: Donald R. Williams (email: drwwilliams@ucdavis.edu). Paul-Christian Bürkner (email: paul.buerkner@gmail.com).

Conflict of Interest and Funding

Author Contributions

DRW conducted the analysis and wrote the initial draft of the paper. PCB originally spotted the inconsistencies in Hofmann et al. (2015) and provided proofreading for the analysis and the paper. Author names are ranked in order of contribution.

Open Science Practices

This article earned the Open Data and the Open Materials badge for making the data and materials 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

Anagnostou, E., Soorya, L., Chaplin, W., Bartz, J., Halpern, D., Wasserman, S., Wang, A. T., Pepa, L., Tanel, N., Kushki, A., & Hollander, E. (2012). Intranasal oxytocin versus placebo in the treatment of adults with autism spectrum disorders: a randomized controlled trial. *Molecular autism*, *3*(1), 16. https://doi.org/10.1186/2040-2392-3-16

Averbeck, B. B., Bobin, T., Evans, S., & Shergill, S. S. (2012). Emotion recognition and oxytocin in patients with schizophrenia. *Psychological Medicine*, *42*, 259–266. https://doi.org/10.1017/S0033291711001413

Dadds, M. R., MacDonald, E., Cauchi, A., Williams, K., Levy, F., & Brennan, J. (2014). Nasal oxytocin for social deficits in childhood autism: A randomized controlled trial. *Journal of Autism and Developmental Disorders*, *44*(3), 521–531. https://doi.org/10.1007/s10803-013-1899-3

Epperson, C. N., McDougle, C. J., & Price, L. H. (1996). Intranasal oxytocin in obsessive-compulsive disorder. *Biological Psychiatry*, *40*(6), 547–549. https://doi.org/10.1016/0006-3223(96)00120-5

Gelman, A., & Loken, E. (2014). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. *Psychological bulletin*, *140*(5), 1272–1280. https://doi.org/dx.doi.org/10.1037/a0037714

Hofmann, S. G., Fang, A., & Brager, D. N. (2015). Effect of intranasal oxytocin administration on psychiatric symptoms: A meta-analysis of placebo-controlled studies. *Psychiatry Research*, *228*(3), 708–714. https://doi.org/10.1016/j.psychres.2015.05.039

Hofmann, S. G., Fang, A., & Brager, D. N. (2016). Notice of Retraction and Replacement: Hofmann et al. Effect of intranasal oxytocin administration on psychiatric symptoms: A meta-analysis of placebo-controlled studies. Psychiatry Research. 2015;228:708-714. *Psychiatry Research*. https://doi.org/10.1016/j.psychres.2016.10.055

Kay, S. R., Fiszbein, A., & Opler, L. a. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia bulletin*, *13*(2), 261–76.

Lee, M. R., Wehring, H. J., McMahon, R. P., Linthicum, J., Cascella, N., Liu, F., Bellack, A., Buchanan, R. W., Strauss, G. P., Contoreggi, C., & Kelly, D. L. (2013). Effects of adjunctive intranasal oxytocin on olfactory identification and clinical symptoms in schizophrenia: Results from a randomized double blind placebo controlled pilot study. *Schizophrenia Research*, *145*(1-3), 110–115. https://doi.org/10.1016/j.schres.2013.01.001

Lim, M. M., Bielsky, I. F., & Young, L. J. (2005). Neuropeptides and the social brain: Potential rodent models of autism. *International Journal of Developmental Neuroscience*, *23*(2-3 SPEC. ISS.), 235–243. https://doi.org/10.1016/j.ijdevneu.2004.05.006

MacDonald, K., MacDonald, T. M., Brüne, M., Lamb, K., Wilson, M. P., Golshan, S., & Feifel, D. (2013). Oxytocin and psychotherapy: A pilot study of its physiological, behavioral and subjective effects in males with depression. *Psychoneuroendocrinology*, *38*(12), 2831–2843. https://doi.org/10.1016/j.psyneuen.2013.05.014

Modabbernia, A., Rezaei, F., Salehi, B., Jafarinia, M., Ashrafi, M., Tabrizi, M., Hosseini, S. M. R., Tajdini, M., Ghaleiha, A., & Akhondzadeh, S. (2013). Intranasal oxytocin as an adjunct to risperidone in patients with schizophrenia: An 8-week, randomized, double-blind, placebo-controlled study. *CNS Drugs*, *27*(1), 57–65. https://doi.org/10.1007/s40263-012-0022-1

Oya, K., Matsuda, Y., Matsunaga, S., Kishi, T., & Iwata, N. (2016). Efficacy and safety of oxytocin augmentation therapy for schizophrenia: an updated systematic review and meta-analysis of randomized, placebo-controlled trials. *European Archives of Psychiatry and Clinical Neuroscience*, *266*(5), 439–450. https://doi.org/10.1007/s00406-015-0634-9

Pedersen, C. A., Smedley, K. L., Leserman, J., Jarskog, L. F., Rau, S. W., Kampov-Polevoi, A., Casey, R. L., Fender, T., & Garbutt, J. C. (2013). Intranasal Oxytocin Blocks Alcohol Withdrawal in Human Subjects. *Alcoholism: Clinical and Experimental Research*, *37*(3), 484–489. https://doi.org/10.1111/j.1530-0277.2012.01958.x

Rohatgi, A. (2017). WebPlotDigitizer. http://arohatgi.info/WebPlotDigitizer

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03