**Title:** Coding errors lead to unsupported conclusions: a critique of Hofmann et al. (2015)

To the Editor:

1. **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 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.

1. **Errors and questions**
   1. *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).

* 1. *Possible selection bias*

Decisions made during the research process can influence the presence of an effect (Gelman and Loken, 2013). 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).

* 1. *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., 2015), 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.

1. **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.

* 1. Methods

The exact method used for effect size calculation was unclear in Hofmann et al. (2015). While they discussed assuming a correlation (*r* = 0.7) between pre and post-treatment scores, the formula they provided did not contain a correlation. 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.

* 1. *Replication attempt*

As seen in Figure 1, the overall estimates for psychiatric symptoms were not significant (SMD = 0.22, 95 %-*CI* [-0.04, 0.47]; SMCR = 0.17, *CI* [-0.10, 0.43]). Trim and Fill procedures indicated bias in SMD outcomes and, when corrected, the effect was reduced (SMD = 0.07, *CI* = [-0.18, 0.32]). We then obtained estimates for specific symptoms. The meta-analytic estimates for depression (SMD = 0.28, 95 % *CI* [-0.25, 0.81]; SMCR = 0.29, *CI* [-0.23, 0.81]), anxiety (SMD = 0.08, *CI* = [-0.26, 0.43]; SMCR = 0.09, *CI* = [-0.24, 0.41]), repetitive behaviors (SMD = 0.15, *CI* = [-0.27, 0.57]; SMCR = -0.06, *CI* = [-0.46, 0.35]), and general psychopathology (SMD = 0.14, *CI* = [-0.24, 0.52]; SMCR = 0.10, *CI* = [-0.25, 0.46]) 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 (SMD = 0.41, CI = [0.03, 0.80]; SMCR = 0.31, CI = [-0.05, 0.66]), restricting the outcomes to total psychotic symptoms resulted in a loss of significance (SMD = 0.28, CI = [-0.22, 0.78]; SMCR = 0.21, CI = [-0.18, 0.61]).

1. **Conclusion**

Although Hofmann et al. (2015) is not a new article, 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.

Donald R. Williams1\*

Paul-Christian Bürkner2

1 Animal Behavior Graduate Group,

University of California, Davis

Davis, USA

2 Institute of Psychology, University of Muenster

Muenster, Germany

\*Corresponding Author

drwwilliams@ucdavis.edu

**Supplementary materials**

Data and R code for analyses presented in this letter are publicly available at Donald R. Williams’ Open Science Framework account (https://osf.io/mzcbr/)

**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. Mol. Autism 3, 16. doi:10.1186/2040-2392-3-16

Averbeck, B.B., Bobin, T., Evans, S., Shergill, S.S., 2011. Emotion recognition and oxytocin in patients with schizophrenia. Psychol. Med. 1–8. doi: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. J. Autism Dev. Disord. 44, 521–531. doi:10.1007/s10803-013-1899-3

Gelman, A., Loken, E., 2013. 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. Dep. Stat. Columbia.

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 Res. 228, 708–714. doi:10.1016/j.psychres.2015.05.039

Kay, S.R., Flszbein, A., Opfer, L.A., 1987. The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia. Schizophr. Bull. 13, 261.

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. Schizophr. Res. 145, 110–115. doi: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. Int. J. Dev. Neurosci. 23, 235–243. doi: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, 2831–2843. doi: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, 57–65. doi:10.1007/s40263-012-0022-1

Oya, K., Matsuda, Y., Matsunaga, S., Kishi, T., Iwata, N., 2015. Efficacy and safety of oxytocin augmentation therapy for schizophrenia: an updated systematic review and meta-analysis of randomized, placebo-controlled trials. Eur. Arch. Psychiatry Clin. Neurosci. 1–12. doi: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. Alcohol. Clin. Exp. Res. 37, 484–489. doi:10.1111/j.1530-0277.2012.01958.x

Rohatgi, A., 2011 WebPlotDigitizer Available from <http://arohatgi.info/WebPlotDigitizer/>. Ref Type: Online Source.

**Figure 1**

**

*Note*: (A) SMCR estimates. (B) SMD estimates. The effect from Averbeck et al. (2011) 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, 2011). 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) were attained from a figure using web plot digitizer.