The correlation between central and peripheral oxytocin concentrations: a systematic review and meta-analysis

Mathias Valstad, Cand. Psychol. [1, 2], Gail A. Alvares, Ph.D. [3, 4], Maiken Egknud, B.Sc. [1], Anna Maria Matziorinis, B.A. [1, 2], Ole A. Andreassen, MD., PhD. [1], Lars T. Westlye, Cand. Psychol., Ph.D. [1, 2], and Daniel S. Quintana, Ph.D. [1 *]

[1] NORMENT, KG Jebsen Centre for Psychosis Research, Division of Mental Health and Addiction, University of Oslo, and Oslo University Hospital, Oslo, Norway
[2] Department of Psychology, University of Oslo, Oslo, Norway
[3] Telethon Kids Institute, University of Western Australia, Australia
[4] Cooperative Research Centre for Living with Autism (Autism CRC), Long Pocket, Brisbane, Australia

* Corresponding author: Daniel S. Quintana, NORMENT, KG Jebsen Centre for Psychosis Research, Building 49, Oslo University Hospital, Ullevål, Kirkeveien 166, PO Box 4956 Nydalen, N- 0424 Oslo, Norway, Ph: +47 23 02 73 50, fax: +47 23 02 73 33, email: daniel.quintana@medisin.uio.no

Running title: Peripheral and central oxytocin concentrations

Abstract

There is growing interest in the role of the oxytocin system in social cognition and behavior. Peripheral oxytocin concentrations are regularly used to approximate central concentrations in psychiatric research, however, the validity of this approach is unclear. Here we conducted a pre-registered systematic search and meta-analysis of correlations between central and peripheral oxytocin concentrations. A search of databases yielded 17 eligible studies, resulting in a total sample size of 516 participants and subjects. Overall, a positive association between central and peripheral oxytocin concentrations was revealed [r=0.29, 95% CI (0.14, 0.42), p<0.0001]. This association was moderated by experimental context [Q_b(4), p=0.003]. While no association was observed under basal conditions (*r*=0.08, *p*=.31), significant associations were observed after intranasal oxytocin administration (*r*=0.66, *p*<.0001), and after experimentally induced stress (r=0.49, p=0.0011). These results indicate a coordination of central and peripheral oxytocin release after stress and after intranasal administration. Although popular, the approach of using peripheral oxytocin levels to approximate central levels under basal conditions is not supported by the present results.

Keywords: oxytocin; meta-analysis; blood plasma; central concentrations

1. Introduction

Oxytocin is a nine amino acid neuropeptide that acts on the widely distributed G-protein coupled oxytocin receptor in humans and almost all other vertebrate species (Horn and Swanson, 2013). Oxytocin is released both into the central nervous system (CNS) and peripheral circulation from neurosecretory cells in the paraventricular (PVN) and supraoptical (SON) nuclei of the hypothalamus, where most endogenous oxytocin is synthesized. Central and peripheral compartments of the oxytocin system are separated anatomically by the blood-brain barrier, that only in exceptional cases is appreciably permeated by oxytocin (Neumann and Landgraf, 2012).

Through central action, oxytocin is critically involved in a range of social behaviors and social cognitive functions (Guastella and MacLeod, 2012). Endogenous oxytocin levels appear to co-vary with social cognitive function at all levels of information processing in humans and other mammals, with similar observed effects after administration of exogenous oxytocin (Bartz et al., 2011). Growing clinical interest (Quintana et al., 2016a) has focused on neurodevelopmental and psychiatric conditions characterized by social cognition and behavioral impairments, such as autism spectrum disorder (ASD) (Alvares et al., 2016b; Guastella and Hickie, 2016) and schizophrenia (Shilling and Feifel, 2016), with the hope to explore the potential of oxytocin as a biomarker of these conditions, better understand their potential etiological pathways, and ultimately to ameliorate the associated social-cognitive and behavioral symptoms.

Several methodological approaches have been adopted to the study of oxytocin involvement in normal and impaired social behavior and cognition. These include the measurement of psychological or neurobiological outcomes after

administration of exogenous oxytocin, and the assessment of endogenous oxytocin concentration covariance with psychological phenotypes and psychiatric disorder status. While crucial to the latter, concentrations of oxytocin have been sampled within both of these research traditions. Although the social cognitive effects of oxytocin are attributed to central mechanisms, oxytocin concentrations have typically, but not universally, been sampled in peripheral fluids such as blood plasma, saliva, and urine (McCullough et al., 2013). Consequentially, that peripheral oxytocin concentrations approximate central bioavailability of the neuropeptide has been a crucial assumption in research where peripheral oxytocin concentrations are correlated with psychological phenotypes or psychiatric disorder status.

Although some animal research indicates that central release from the hypothalamus and peripheral release via the posterior pituitary is coordinated (Landgraf et al., 1988; Ross et al., 2009; Wotjak et al., 1998), other research does not support this (Amico et al., 1990; Robinson and Jones, 1982). Research is also mixed in humans, with some results consistent with related levels of central and peripheral endogenous oxytocin (Carson et al., 2014), while others report no significant associations (Kagerbauer et al., 2013). After exogenous oxytocin delivered via intranasal administration in humans, one study found a significant association between cerebrospinal fluid (CSF) and blood plasma concentrations of oxytocin (Wang et al., 2013), while another found no significant association (Striepens et al., 2013). Using peripheral oxytocin concentrations to index central concentrations is clearly appealing, given the more invasive procedures required to collect centrally circulating fluids in humans. However, it is currently unclear whether and when peripheral oxytocin measures can be used to index CNS concentrations and central oxytocin bioavailability.

The present systematic review and meta-analysis synthesized studies in which central and peripheral measures of oxytocin were simultaneously sampled into a summary effect size. The strength of the summary effect size is indicative of the plausibility of peripheral oxytocin as an index for central oxytocin concentrations. As eligible studies were likely to vary in a range of contextual specifications, several potential moderator variables were considered, including experimental paradigm, oxytocin sampling location, subject species, biochemical analysis methods, year of publication, and study quality. Such differences between contexts may contribute to variance in the correlations between central and peripheral oxytocin. Thus, it is possible that peripheral oxytocin can index central oxytocin concentrations in some contexts, but not others. Together, the purpose of this study was to examine whether, and under which circumstances, peripheral oxytocin is a correlate of central oxytocin concentrations.

2. Materials and Methods

The systematic search and meta-analysis was conducted in accordance with the PRISMA guidelines (Moher et al., 2009) (Supplementary table S1) and recent recommendations for conducting correlational meta-analyses (Quintana, 2015). Prior to the execution of the systematic search and meta-analysis, the protocol for this systematic review and meta-analysis was published (Valstad et al., 2016) and pre-registered on the PROSPERO registry (CRD42015027864).

2.1. Systematic literature search and inclusion of eligible studies

A systematic literature search was performed in two iterations to retrieve studies in which oxytocin had been simultaneously sampled in fluids or tissues located in

central (e.g., local extracellular fluid or CSF) or peripheral (e.g. blood plasma or saliva) regions of the body. In the first iteration, a search was performed, using Ovid, in Embase and Medline with the following combination of terms: (oxytocin) AND (concentration* OR level*) AND (plasma OR blood OR saliva* OR urin*) AND (central OR csf OR “cerebrospinal fluid”). The following constraints were applied to limit search results: the result should be (i) a full-text article or a conference abstract, (ii) written in English, that was (iii) published after 1971, when biochemical analysis of oxytocin content using enzyme immunoassay was made commercially available. Searches were conducted on April 1, 2016 and August 2, 2016, and resulted in a total of 572 studies. Out of these, 110 were relevant. A second iteration was performed in which citing articles and reference lists of included studies were examined for remaining relevant studies (Fig. 1). After retrieval, relevant studies were screened for inclusion based on the criterion that effect sizes for the correlation between central and peripheral concentrations of oxytocin must be obtainable. While 110 of the studies retrieved in the systematic search were relevant, only 17 of these satisfied this criterion.

2.2 Data extraction and management

Effect sizes and sample sizes were extracted from eligible studies. For some articles, effect sizes were stated explicitly, or directly obtainable through tables of individual values. In other articles, individual values were represented in graphs such as scatterplots, in which case a web plot digitizer (Rohatgi, 2015) was used for conversion of plots into numerical values. Since some articles contained both a scatterplot and a directly stated effect size, this plot digitizer was validated through comparing effect sizes provided by authors with plot digitizer outputs, revealing

almost perfect precision (Supplementary text S2). Some articles did not provide relevant effect sizes, individual values in tables, or scatterplots. Since 15 years is a common time frame for the retention of clinical data, authors of such articles published from 2001 were contacted and asked to provide effect sizes. Articles lacking this information that were published before 2001 (n = 68), and studies performed by authors that were not able to respond to the data request (n = 25), were not included in the meta-analysis. Data were extracted from all eligible studies using a custom data extraction form (Supplementary table S3).

2.3. Statistical analysis

Statistical analysis was performed with R statistical software version 3.2.4. (R Core Team, 2016), using the *MAc* (Del Re and Hoyt, 2012), *metafor* (Viechtbauer, 2010), and *multcomp* (Hothorn et al., 2008) R packages. The dataset and script to perform the analyses are available at https://osf.io/aj55y/

Prior to meta-analytic synthesis, raw effect sizes were transformed to Fischer's *z* for variance stabilization (Borenstein et al., 2009). Raw effect sizes given as Spearman's ρ were first transformed to Pearson's *r* according to Gilpin (1993), and then transformed to Fischer's *z* for meta-analysis. For studies reporting several effect sizes, or reporting one effect size based on repeated measures, within-study variance was estimated using a procedure described in the Supplementary text S4. A random effects model (DerSimonian and Kacker, 2007), where between-studies variance (τ^2) was estimated using a restricted maximum likelihood method, was used in the synthesis of individual effect sizes into a summary effect size. Outlier diagnostics were also performed to identify potential effect size outliers (Viechtbauer, 2010). Point estimates were converted back to Pearson's *r* for interpretive purposes.

The observed variance between studies may be due to heterogeneity (variance in the true effect sizes between studies) and within-study variance. Q, the significance of Q, and I^2 were computed in order to examine variance and heterogeneity among effect sizes of included studies. I^2 values of ~25%, ~50%, and ~75% were interpreted as low, moderate, and high, respectively (Higgins et al., 2003).

Potential moderator variables were defined *a priori* (Valstad et al., 2016). Some of the levels for moderator variables were also defined *a priori*, such as the levels *baseline condition* (lack of experimental intervention) and *intranasal administration* for the experimental paradigm moderator. Other levels of moderator variables were adjusted from pre-planned analyses *post hoc* based on the specific characteristics of included studies (for details, see Supplementary text S5). Due to the ambiguity of the concept “baseline”, an inclusive and a strict definition was adopted for sensitivity analysis, where the former was defined as lack of experimental manipulation, while the latter was defined as lack of experimental manipulation together with lack of specific context (e.g. lactation). For one of the studies (Striepens et al., 2013) effect sizes for the intranasal oxytocin (n = 11) and baseline (n = 4) conditions were not possible to disentangle, and the combined effect size was categorized in the intranasal subgroup. A sensitivity analysis for the moderator experimental paradigm was performed in which this study was removed. In all the included studies, peripheral oxytocin was sampled from blood, such that no moderator analysis for peripheral sample type was required. For most human participants ($n \geq 212$), central oxytocin was collected from the CSF by spinal puncture. A random effects model with separate estimates of between study variance was applied for all categorical moderator variables, yielding summary weighted mean effects and the significance of subgroup effects, which were calculated for each

subgroup. Although mammals share essential oxytocin system characteristics, such as production of oxytocin in the hypothalamus, peripheral and central release of oxytocin from hypothalamus, and a blood brain barrier that inhibits diffusion of oxytocin between the CNS and systemic circulation (Horn and Swanson, 2013), the between-species differences (Valstad et al., 2016) necessitated an additional analysis to examine the role of species in the different effects observed between experimental paradigms. When there were more than two subgroups, pairwise comparisons were performed between all moderator categories with Holm-adjusted p-values to control the family-wise error rate. Meta-regression models were fitted to account for heterogeneity of continuous moderator variables.

2.4. Data quality measures

Small study bias, which includes both publication bias and study quality bias (Egger et al., 1997; Schulz et al., 1995), was assessed by visually inspecting a funnel plot and performing Egger's regression test Egger et al., 1997). A significant test ($p < .05$) is indicative of small study bias. A contour enhanced funnel plot, which superimposes key areas of statistical significance ($p = .1$, $p = .05$, $p = .01$), was constructed to specifically assess for risk of publication bias (Peters et al., 2008). An over-representation of effect sizes in the key areas of significance is indicative of publication bias risk. Since the decision to report a specific effect size, in contrast to the decision to publish a study, is not directly dependent on sample size, the regression test for funnel plot asymmetry does not rule out the possibility that there could be a bias in the type of evidence that is *reported* in published studies. To examine whether this was a source of bias in the set of included studies, the included studies that explicitly stated effect sizes were compared to the studies where effect

sizes were obtained by other means, such as data scraping or author request. Furthermore, studies that reported the relevant effect size explicitly were separately examined for publication bias in order to test the possibility for publication bias among studies where the correlation between central and peripheral oxytocin was a research focus (i.e., focal studies).

There may also be issues with validity of the data that are internal to included studies. A custom risk of bias tool (Supplementary Table S6) was used (by ME and AMM) to systematically assess within-study risk of bias in included studies. This tool was developed by adapting the tool used in another meta-analysis (Alvares et al., 2016a) to the context of oxytocin research.

3. Results

17 studies yielding 32 effect sizes were included in the meta-analysis (Table 1; Fig. 1; Striepens et al., 2013, Carson et al., 2014, Martin et al., 2014, Kagerbauer et al., 2013, Neumann et al., 2013, Wang et al., 2013, Kojima et al., 2012, Williams et al., 2012, Sansone et al., 2002, Amico et al., 1990, Takeda et al., 1985, Takagi et al., 1985, Jokinen et al., 2012, Jin et al., 2007, Keverne and Kendrick, 1991, Engelmann et al., 2004, Kleindienst et al., 2004). The total number of participants/subjects across studies was 504. Among these, 256 were human, 237 were rodents, 7 were sheep, and 4 subjects were non-human primates.

3.1. Association between central and peripheral concentrations of oxytocin

There was a positive correlation between central and peripheral concentrations of oxytocin [r = 0.29, 95% CI (0.14, 0.42), $p < 0.0001$; Fig 2]. Egger's regression test revealed no overall evidence of small study bias ($p = .33$; Fig 3A), and no evidence

for small study bias among focal studies (p = .24). There was no significant difference between studies that reported [r = 0.35, 95% CI (0.18, 0.50)], and studies that did not explicitly report the relevant effect size [r = 0.15, 95% CI (-0.10, 0.38)]. An inspection of the contour enhanced funnel plot did not reveal an over-representation of effect sizes in the significance contours (Fig. 3B), indicating a low risk of publication bias. Furthermore, a meta-regression revealed that risk of bias did not influence effect sizes (p= 0.24; Fig 3C). Influence diagnostics identified one potential outlier (Wang et al., 2013). A sensitivity analysis, which involved re-analysis without the identified outlier, revealed a similar summary effect size as the original analysis that was also statistically significant [r = 0.23, 95% CI (0.11, 0.35), p = 0.0002]. As this sensitivity analysis suggested that this single effect size only had a modest effect on the overall meta-analysis, it was retained for the remainder of the analyses. In the total sample of included studies, there was a moderate-to-high level of heterogeneity [Q = 86.19, p < .0001, I^2 = 62.8% 95% CI (37%, 77%)]. Accordingly, moderator analyses were performed to identify sources of heterogeneity.

3.2. Impact of moderators on effect size

A moderator analysis revealed that part of the heterogeneity in the model was due to the type of experimental paradigm [$Q_b(4)$ = 16.03, p = 0.003; Fig 4A]. Across experimental paradigms, positive associations were observed for the intranasal oxytocin (IN-OT) condition (r = .66, p < .0001, k = 4) and after stress interventions (r = .49, p = 0.001, k = 5; Supplementary table S7). The IN-OT association was reproduced (r = 0.75, p < .0001, k = 3) in a sensitivity analysis where one effect size (Striepens et al., 2013) was removed due to containing some samples (n = 4) from a baseline condition. In contrast, no association was observed in the baseline condition

(r = 0.08, *p* = .31, k = 15). The subgroup effects for the peripheral oxytocin administration category (r = 0.29, *p* = .28, k = 3), as well as for the 'other' category (r = 0.30, *p* = .07, k = 5) were not significant. The results for the baseline condition were similar (*r* = 0.10, *p* = .27, k = 13) when applying a strict rather than inclusive extension of 'baseline'. A comparison of all possible pairwise comparisons with Holm corrected *p*-values revealed that the IN-OT point estimate was significantly greater than the baseline point estimate (*p* = .003; Fig. 4A). While there were no other significant pairwise comparisons, the increased stress point estimate compared to baseline point estimate was on the border of statistical significance (*p* = .14). When constrained to human studies, results for the levels of the experimental paradigm moderator were reproduced [$Q_b(2) = 7.65$, *p* = .02], with no significant correlation in the baseline condition [r = 0.05, 95% CI (-0.19, 0.30), *p* = .59, k = 7, I^2=0%], and a significant correlation in the intranasal condition [r = 0.71, 95% CI (0.34, 0.89), *p* = 0.013, k = 2 (Striepens et al., 2013; Wang et al., 2013), I^2 = 93%].

Analysis of the effect of central sampling location on effect sizes was on the border of significance [$Q_b(3) = 6.33$, *p* = .10; Fig. 4B], suggesting that specific brain sampling location differences may contribute to observed heterogeneity. Across levels of the central sampling location moderator, subgroup effects for hypothalamus (*r* = 0.42, *p* = 0.0004, k = 10), central amygdala (*r* = 0.52, *p* = 0.034, k = 3), and hippocampus (*r* = 0.50, *p* = 0.034, k = 3; Supplementary table S7) were significant. The subgroup effect for samples taken from CSF (*r* = 0.14, *p* = 0.14, k = 16) was not significant. Pairwise comparisons did not reveal any significant difference between any of these subgroups.

The moderator analysis for species was not significant [$Q_b(3) = 1.87$, *p* = 0.60; Fig. 4C], suggesting that species diversity might not contribute to heterogeneity

among effect sizes. Across levels of the species moderator, only the subgroup effect for rodents was significant (r = 0.35, *p* = .0004, k = 19; Supplementary table S7). The point estimates for human (r = 0.22, *p* = 0.081, k = 10), sheep (r=0.19, *p* = .55, k = 2), and non-human primate (r = -0.34, *p* = .53, k = 1) subgroups were not significant. There was no significant difference between any of the levels of this moderator variable. An exploratory mixed-effect meta-regression model was fitted to assess whether the influence of species type (human vs. rodent) on the correlation between central and peripheral levels varied between experimental paradigms (intranasal oxytocin vs. baseline vs. other). A Wald-type chi-square test did not reveal evidence for a significant interaction ($Q_m(2) = 0.48$, $p = .79$).

The biochemical analysis method moderator analysis was not significant [$Q_b(2) = 4.45$, $p = 0.11$; Fig. 4D], indicating that this moderator variable is not likely to contribute to heterogeneity among effect sizes. Across the levels of the biochemical analysis method moderator variable, subgroup effects for both RIA (*r* = 0.28, *p* = 0.0005, k = 24) and EIA (*r* = 0.42, *p* = 0.0035, k = 6) were significant. There was no significant effect for LC/MS (*r* = -0.2, *p* = 0.43, k = 2) and no significant differences between the levels of this moderator variable. The peptide extraction moderator analysis was not significant [$Q_b(1) = 0.06$, $p = 0.80$; Fig. 4E]. Both with (*r* = 0.31, *p* = 0.0002, k = 24) and without (*r* = 0.35, *p* = 0.0278, k = 6) extraction subgroup effects were significant (Supplementary table S7). Finally, the year of publication did not significantly moderate the relationship between central and peripheral oxytocin concentrations [Q(1) = 2.0, *p* = .15; Fig 3D].

4. Discussion

The present systematic meta-analysis revealed a positive correlation between concentrations of oxytocin in blood plasma and oxytocin concentrations in the CNS. However, the association was moderate and showed a high degree of heterogeneity, suggesting that the observed association might not be present across all contexts. Experimental paradigm was the moderator variable most likely to account for this heterogeneity. After IN-OT, as well as after an experimental stressor, there was a positive correlation between central and peripheral oxytocin concentrations. However, in the baseline condition, there was no evidence of correlation, neither for the entire sample of subjects, nor for any of the species analyzed separately. Notably, there was a statistically significant difference between the summary statistic for IN-OT studies and baseline studies. Given the lack of evidence for a correlation between peripheral and central oxytocin levels in the baseline condition, the data suggest blood plasma may not efficiently index central oxytocin concentrations under baseline conditions. Furthermore, this result provides additional indirect evidence for the effectiveness of the blood-brain barrier in restricting oxytocin diffusion between systemic circulation and the CNS (Neumann and Landgraf, 2012), as well supporting the hypothesis that under baseline conditions, hypothalamic oxytocin release into blood and into the CNS is uncoordinated (Amico et al., 1990).

There is a substantial body of research attempting to link peripheral oxytocin concentrations with psychological phenotypes or psychiatric disorder status. Since the social-cognitive effects of oxytocin have been assumed to arise from oxytocin action in the CNS, the assumption that peripheral and central oxytocin concentrations correlate in a baseline condition was crucial in the interpretation of the results from these two approaches (e.g. Hoge et al., 2008; Rubin et al., 2010). This assumption is called into question by the present data. These results may have two possible,

mutually exclusive, implications for the interpretation of studies within these research traditions: either the apparent social cognitive effects are type I errors produced by chance, or the demonstrated covariance between social cognition and endogenous oxytocin in systemic circulation arise from some phenomenon unrelated to central oxytocin levels. The former potential interpretation is consistent with the evidence of publication bias that has surfaced in the field of psychological and psychiatric oxytocin research (Lane et al., 2016; McCullough et al., 2013; Walum et al., 2016). Likewise, in a recent meta-analysis, basal levels of neither CSF nor plasma oxytocin were associated with psychiatric disorder status, with the exception of anorexia nervosa, which was associated with a reduction in plasma oxytocin levels (Rutigliano et al., 2016). The latter interpretation points to a potential peripheral mechanism for the observed social cognitive correlates of basal peripheral oxytocin concentrations. One potential causal mechanism is oxytocin action on peripheral tissues that provide afferent feedback to the CNS (Horn and Swanson, 2013).

In contrast to what was discovered under baseline conditions, this meta-analysis revealed a positive correlation between central and peripheral oxytocin after intranasal administration of oxytocin, both overall and when analysis was limited to studies in humans. Almost every study examining the effects of exogenous oxytocin on social cognition and behavior in normal and clinical populations have made use of the intranasal delivery route (Andari et al., 2010; Domes et al., 2007; Guastella et al., 2008; Kosfeld et al., 2005). The motivation behind administering oxytocin intranasally is to obtain non-invasive delivery of oxytocin into the brain. Although vasopressin, which is structurally similar to oxytocin, has been shown to enter the CSF after intranasal administration (Born et al., 2002), and intranasal oxytocin has been shown to enter the CSF in non-human primates (Lee et al., 2017) it is not entirely clear

where intranasally administered oxytocin travels in humans, or whether it actually reaches brain areas containing oxytocin receptors such as the hypothalamus or the amygdalae (Quintana et al., 2015a). However, recent work in humans comparing intranasal and intravenous oxytocin administration indicates that despite comparable peripheral oxytocin concentrations after both administration routes, social cognitive (Quintana et al., 2015b) and neural effects (Quintana et al., 2016b) were only observed after intranasal administration. Together, these results are consistent with a direct nose-to-brain transport of intranasally administered oxytocin via olfactory and trigeminal nerve fibers, although it is not clear whether the increased correlation after intranasal oxytocin administration stems from increased hypothalamic release of oxytocin, or simply from exogenous oxytocin reaching both CSF and peripheral circulation.

In this meta-analysis, a positive association was also found between central and peripheral concentrations of oxytocin after experimental stress induction. Stress induction involved either separation from a mother (Kojima et al., 2012), or a forced swim test (Williams et al., 2012). As the authors suggest (Williams et al., 2012), the hypothalamus-pituitary-adrenal (HPA) axis and related hormones such as corticosterone interact with the oxytocin system to regulate stress responses. Such interaction may occur through interneurons between magnocellular and parvocellular neurons in the PVN (Ferguson et al., 2008), from which oxytocin and corticotropin-releasing hormone are released. Furthermore, interaction may be mediated through corticosteroid effects on vasoconstriction and heart rate, which in turn could affect oxytocin release through baroreceptors and the vagal feedback system (Horn and Swanson, 2013; Quintana et al., 2015a).

There are some limitations to the study worth mentioning. First, although a total of 17 studies were included in the main analysis, some moderator analysis subgroups contained few studies, an extreme instance of which is the amygdala and hippocampus levels of the central sampling location moderator, where the data from only one study are included. Relatedly, several subgroups contained few or no human studies. In cases such as the stress subgroup, with no human studies, or the intranasal oxytocin subgroup, with two human studies (Striepens et al., 2013; Wang et al., 2013), generalization to humans is not straightforward and results should be considered preliminary. The effect size (Striepens et al., 2013) that was categorized to the intranasal subgroup despite representing a combination of intranasal (n =11) and baseline (n = 4) samples may reduce the reliability of the correlation estimate for intranasal oxytocin in humans. Second, to estimate variances for effect sizes from repeated measures, dependent samples variance estimation was used to control for dependency between samples. Since exact dependencies between repeated samples were unknown, there is a chance that variances for effect sizes obtained in repeated measures designs were slightly overestimated or underestimated, relative to variances for effect sizes obtained in single sample designs. A differential variance estimation would favor one of the two study types with respect to the relative weight they were afforded in the main analysis. However, since there is no a priori reason to believe that study type should impact upon the estimated effect sizes, it is unlikely that this potential bias had any considerable effects on the results. Third, even if there was no evidence for publication bias, or for bias in report of effect sizes, there may be some bias in the subjects sampled for studies where CSF was collected. Across included studies, some of the human participants had medical conditions (Carson et al., 2014; Striepens et al., 2013; Wang et al., 2013). Medical conditions

are often associated with pain, and pain may influence oxytocin release: in one study, chemical pain stimulation increased oxytocin release within the brain, but not in plasma (Yang et al., 2007). If pain leads to uncoordinated release, then this may bias the results of this meta-analysis in a negative direction. However, the strongest correlation between central and peripheral concentrations of oxytocin among included studies – which was also identified as a potential outlier – was observed in a sample of headache patients (Wang et al., 2013). This may point to the opposite possibility that pain could bias the effect sizes of this meta-analysis in a positive direction. To ensure that this study did not inflate the effect size for the IN-OT condition, a secondary analysis was performed with this study removed, yielding comparable results.

The collection of peripheral oxytocin measures to index central levels has obvious appeal given the difficulties surrounding central collection. However, research has yet to establish whether this is a valid experimental approach. The results of this meta-analysis indicate that there is a positive association between central and peripheral concentrations of oxytocin, but this association depends on experimental context. There was evidence for a positive association between central and peripheral concentrations of oxytocin after intranasal oxytocin administration and after experimental stress induction. However, as there was no evidence for an association between central and peripheral oxytocin concentrations under baseline conditions, future studies on the role of basal oxytocin in cognition or social behavior should avoid using peripheral oxytocin measures to make inferences on central oxytocin concentrations.

Acknowledgements

We thank Hege Kristin Ringnes (University of Oslo Library) for providing guidance on our systematic search strategy.

Funding sources

MV received salary support from the Research Council of Norway (RCN) via a grant for students in clinical psychology programmes. GAA is funded by the Cooperative Research Centre for Living with Autism (Autism CRC), established and supported under the Australian Government's Cooperative Research Centres Program. The Research Council of Norway (RCN) and OptiNose AS contributed to funding this review through a BIA project grant (219483) via salary support to DSQ and project support to OAA, LTW, and DSQ. LTW is supported by the South-Eastern Norway Regional Health Authority (2014097). DSQ is supported by an Excellence Grant from the Novo Nordisk Foundation (NNF16OC0019856). The funders they had no influence in the ideas contained in the manuscript and no role in the writing of the manuscript.

References

Alvares, G.A., Quintana, D.S., Hickie, I.B., Guastella, A.J., 2016a. Autonomic nervous system dysfunction in psychiatric disorders and the impact of psychotropic medications: a systematic review and meta-analysis. Journal of Psychiatry & Neuroscience 41, 89–104. doi:10.1503/jpn.140217

Alvares, G.A., Quintana, D.S., Whitehouse, A.J.O., 2016b. Beyond the hype and hope: Critical considerations for intranasal oxytocin research in autism spectrum disorder. Autism Research 1–7. doi:http://doi.org/10.1002/aur.1692

Amico, J.A., Challinor, S.M., Cameron, J.L., 1990. Pattern of Oxytocin Concentrations in the Plasma and Cerebrospinal Fluid of Lactating Rhesus Monkeys (Macaca mulatto,): Evidence for Functionally Independent Oxytocinergic Pathways in Primates*. The Journal of Clinical Endocrinology & Metabolism 71, 1531–1535. doi:10.1210/jcem-71-6-1531

Andari, E., Duhamel, J.-R., Zalla, T., Herbrecht, E., Leboyer, M., Sirigu, A., 2010. Promoting social behavior with oxytocin in high-functioning autism spectrum disorders. PNAS 107, 4389–4394. doi:10.1073/pnas.0910249107

Bartz, J.A., Zaki, J., Bolger, N., Ochsner, K.N., 2011. Social effects of oxytocin in humans: context and person matter. Trends in Cognitive Sciences 15, 301–309. doi:10.1016/j.tics.2011.05.002

Borenstein, M.H., Higgins, L.V., Rothstein, J., 2009. Introduction to meta-analysis. Wiley, Chichester, England

Born, J., Lange, T., Kern, W., McGregor, G.P., Bickel, U., Fehm, H.L., 2002. Sniffing neuropeptides: a transnasal approach to the human brain. Nat. Neurosci. 5, 514–516

Carson, D.S., Berquist, S.W., Trujillo, T.H., Garner, J.P., Hannah, S.L., Hyde, S.A., Sumiyoshi, R.D., Jackson, L.P., Moss, J.K., Strehlow, M.C., Cheshier, S.H., Partap, S., Hardan, A.Y., Parker, K.J., 2014. Cerebrospinal fluid and plasma oxytocin concentrations are positively correlated and negatively predict anxiety in children. Molecular Psychiatry 20, 1085–1090. doi:10.1038/mp.2014.132

Del Re, A.C., Hoyt, W.T., 2012. MAc: Meta-analysis with correlations. https://CRAN.R-project/package=MAc

DerSimonian, R., Kacker, R., 2007. Random-effects model for meta-analysis of clinical trials: An update. Contemporary Clinical Trials 28, 105–114. doi:10.1016/j.cct.2006.04.004

Domes, G., Heinrichs, M., Gläscher, J., Büchel, C., Braus, D.F., Herpertz, S.C., 2007. Oxytocin Attenuates Amygdala Responses to Emotional Faces Regardless of Valence. Biological Psychiatry 62, 1187–1190. doi:10.1016/j.biopsych.2007.03.025

Egger, M., Smith, G.D., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. BMJ 315, 629–634. doi:10.1136/bmj.315.7109.629

Ferguson, A.V., Latchford, K.J., Samson, W.K., 2008. The paraventricular nucleus of the hypothalamus – a potential target for integrative treatment of autonomic dysfunction. Expert Opinion on Therapeutic Targets 12, 717–727. doi:10.1517/14728222.12.6.717

Gilpin, A.R., 1993. Table for Conversion of Kendall“S Tau to Spearman”S Rho Within the Context of Measures of Magnitude of Effect for Meta-Analysis. Educational and Psychological Measurement 53, 87–92. doi:10.1177/0013164493053001007

Guastella, A.J., MacLeod, C., 2012. A critical review of the influence of oxytocin

nasal spray on social cognition in humans: Evidence and future directions. Hormones and Behavior 61, 410–418. doi:10.1016/j.yhbeh.2012.01.002
Guastella, A.J., Hickie, I.B., 2016. Oxytocin Treatment, Circuitry, and Autism: A Critical Review of the Literature Placing Oxytocin Into the Autism Context. Biological Psychiatry 79, 234–242. doi:10.1016/j.biopsych.2015.06.028
Guastella, A.J., Mitchell, P.B., Dadds, M.R., 2008. Oxytocin Increases Gaze to the Eye Region of Human Faces. Biological Psychiatry 63, 3–5. doi:10.1016/j.biopsych.2007.06.026
Higgins, J.P.T., Thompson, S.G., Deeks, J.J., Altman, D.G., 2003. Measuring inconsistency in meta-analyses. BMJ 327, 557–560. doi:10.1136/bmj.327.7414.557
Hoge, E.A., Pollack, M.H., Kaufman, R.E., Zak, P.J., Simon, N.M., 2008. Oxytocin Levels in Social Anxiety Disorder. CNS Neuroscience & Therapeutics 14, 165–170. doi:10.1111/j.1755-5949.2008.00051.x
Horn, J., Swanson, L., 2013. The autonomic motor system and the hypothalamus, in: Kandel, E.R., Schwartz, J.H., Jessel, T.M., Siegelbaum, S.A., Hudspeth, A.J. (Eds.), Principles of Neural Science. McGraw Hill Medical. pp 1056-1078
Hothorn, T., Bretz, F., Westfall, P., 2008. Simultaneous Inference in General Parametric Models. Biometrical Journal 50, 346–363. doi:10.1002/bimj.200810425
Kagerbauer, S.M., Martin, J., Schuster, T., Blobner, M., Kochs, E.F., Landgraf, R., 2013. Plasma Oxytocin and Vasopressin do not Predict Neuropeptide Concentrations in Human Cerebrospinal Fluid. Journal of Neuroendocrinology 25, 668–673. doi:10.1111/jne.12038
Kojima, S., Stewart, R.A., Demas, G.E., Alberts, J.R., 2012. Maternal Contact Differentially Modulates Central and Peripheral Oxytocin in Rat Pups During a Brief Regime of Mother–Pup Interaction that Induces a Filial Huddling Preference. Journal of Neuroendocrinology 24, 831–840. doi:10.1111/j.1365-2826.2012.02280.x
Kosfeld, M., Heinrichs, M., Zak, P.J., Fischbacher, U., Fehr, E., 2005. Oxytocin increases trust in humans. Nature 435, 673–676. doi:10.1038/nature03701
Landgraf, R., Neumann, I., Schwarzberg, H., 1988. Central and peripheral release of vasopressin and oxytocin in the conscious rat after osmotic stimulation. Brain Research 457, 219–225. doi:10.1016/0006-8993(88)90689-0
Lane, A., Luminet, O., Nave, G., Mikolajczak, M., 2016. Is there a publication bias in behavioral intranasal oxytocin research on humans? Opening the file drawer of one lab. Journal of Neuroendocrinology 28.
Lee, M.R., Scheidweiler, K.B., Diao, X.X., Akhlaghi, F., Cummins, A., Huestis, M.A., Leggio, L., Averbeck, B.B., 2017. Oxytocin by intranasal and intravenous routes reaches the cerebrospinal fluid in rhesus macaques: determination using a novel oxytocin assay. Molecular Psychiatry. Nature Publishing Group 56, 701.
McCullough, M.E., Churchland, P.S., Mendez, A.J., 2013. Problems with measuring peripheral oxytocin: Can the data on oxytocin and human behavior be trusted? Neuroscience & Biobehavioral Reviews 37, 1485–1492. doi:10.1016/j.neubiorev.2013.04.018
Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. Ann Intern Med 151, 264–269. doi:10.7326/0003-4819-151-4-200908180-00135
Neumann, I.D., Landgraf, R., 2012. Balance of brain oxytocin and vasopressin: implications for anxiety, depression, and social behaviors. Trends in

Neurosciences 35, 649–659. doi:10.1016/j.tins.2012.08.004
Peters, J.L., Sutton, A.J., Jones, D.R., Abrams, K.R., Rushton, L., 2008. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. Journal of Clinical Epidemiology 61, 991–996. doi:10.1016/j.jclinepi.2007.11.010
Quintana, D.S., 2015. From pre-registration to publication: a non-technical primer for conducting a meta-analysis to synthesize correlational data. Front. Psychol. 6, 839. doi:10.3389/fpsyg.2015.01549
Quintana, D.S., Alvares, G.A., Hickie, I.B., Guastella, A.J., 2015a. Do delivery routes of intranasally administered oxytocin account for observed effects on social cognition and behavior? A two-level model. Neuroscience & Biobehavioral Reviews 49, 182–192. doi:10.1016/j.neubiorev.2014.12.011
Quintana, D.S., Guastella, A.J., Westlye, L.T., Andreassen, O.A., 2016a. The promise and pitfalls of intranasally administering psychopharmacological agents for the treatment of psychiatric disorders. Molecular Psychiatry 21, 29–38. doi:10.1038/mp.2015.166
Quintana, D.S., Westlye, L.T., Alnæs, D., Rustan, Ø.G., Kaufmann, T., Smerud, K.T., Mahmoud, R.A., Djupesland, P.G., Andreassen, O.A., 2016b. Low dose intranasal oxytocin delivered with Breath Powered device dampens amygdala response to emotional stimuli: A peripheral effect-controlled within-subjects randomized dose-response fMRI trial. Psychoneuroendocrinology 69, 180–188. doi:10.1016/j.psyneuen.2016.04.010
Quintana, D.S., Westlye, L.T., Rustan, Ø.G., Tesli, N., Poppy, C.L., Smevik, H., Tesli, M., Røine, M., Mahmoud, R.A., Smerud, K.T., Djupesland, P.G., Andreassen, O.A., 2015b. Low-dose oxytocin delivered intranasally with Breath Powered device affects social-cognitive behavior: a randomized four-way crossover trial with nasal cavity dimension assessment. Transl Psychiatry 5, e602.
R Core Team, 2016 R: A language and environment for statistical computing. https://www.r-project.org/
Robinson, I.C.A.F., Jones, P.M., 1982. Oxytocin and Neurophysin in Plasma and CSF during Suckling in the Guinea-Pig. Neuroendocrinology 34, 59–63. doi:10.1159/000123278
Rohatgi, A., 2015. WebPlotDigitizer. http://arohatgi.info/WebPlotDigitizer/
Ross, H.E., Cole, C.D., Smith, Y., Neumann, I.D., Landgraf, R., Murphy, A.Z., Young, L.J., 2009. Characterization of the oxytocin system regulating affiliative behavior in female prairie voles. Neuroscience 162, 892–903. doi:10.1016/j.neuroscience.2009.05.055
Rubin, L.H., Carter, C.S., Drogos, L., Pournajafi-Nazarloo, H., Sweeney, J.A., Maki, P.M., 2010. Peripheral oxytocin is associated with reduced symptom severity in schizophrenia. Schizophrenia Research 124, 13–21. doi:10.1016/j.schres.2010.09.014
Rutigliano, G., Rocchetti, M., Paloyelis, Y., Gilleen, J., Sardella, A., Cappucciati, M., Palombini, E., Dell'Osso, L., Caverzasi, E., Politi, P., McGuire, P., Fusar-Poli, P., 2016. Peripheral oxytocin and vasopressin: Biomarkers of psychiatric disorders? A comprehensive systematic review and preliminary meta-analysis. Psychiatry Research 241, 207–20.
Schulz, K.F., Chalmers, I., Hayes, R.J., Altman, D.G., 1995. Empirical Evidence of Bias: Dimensions of Methodological Quality Associated With Estimates of Treatment Effects in Controlled Trials. JAMA 273, 408–412. doi:10.1001/jama.1995.03520290060030

Shilling, P.D., Feifel, D., 2016. Potential of Oxytocin in the Treatment of Schizophrenia. CNS Drugs 30, 193–208. doi:10.1007/s40263-016-0315-x
Striepens, N., Kendrick, K.M., Hanking, V., Landgraf, R., Wüllner, U., Maier, W., Hurlemann, R., 2013. Elevated cerebrospinal fluid and blood concentrations of oxytocin following its intranasal administration in humans. Scientific Reports 3, 3440. doi:10.1038/srep03440
Valstad, M., Alvares, G.A., Andreassen, O.A., Westlye, L.T., Quintana, D.S., 2016. The relationship between central and peripheral oxytocin concentrations: a systematic review and meta-analysis protocol. Systematic Reviews 2016 5:1 5, 1. doi:10.1186/s13643-016-0225-5
Viechtbauer, W., 2010. Conducting meta-analyses in R with the metafor package. J Stat Softw 36.
Walum, H., Waldman, I.D., Young, L.J., 2016. Statistical and Methodological Considerations for the Interpretation of Intranasal Oxytocin Studies. Biological Psychiatry 79, 251–257. doi:10.1016/j.biopsych.2015.06.016
Wang, Y.-L., Yuan, Y., Yang, J., Wang, C.-H., Pan, Y.-J., Lu, L., Wu, Y.-Q., Wang, D.-X., Lv, L.-X., Li, R.-R., Xue, L., Wang, X.-H., Bi, J.-W., Liu, X.-F., Qian, Y.-N., Deng, Z.-K., Zhang, Z.-J., Zhai, X.-H., Zhou, X.-J., Wang, G.-L., Zhai, J.-X., Liu, W.-Y., 2013. The interaction between the oxytocin and pain modulation in headache patients. Neuropeptides 47, 93–97. doi:10.1016/j.npep.2012.12.003
Williams, S.K., Barber, J.S., Jamieson Drake, A.W., Enns, J.A., Townsend, L.B., Walker, C.H., Johns, J.M., 2012. Chronic Cocaine Exposure During Pregnancy Increases Postpartum Neuroendocrine Stress Responses. Journal of Neuroendocrinology 24, 701–711. doi:10.1111/j.1365-2826.2012.02291.x
Wotjak, C.T., Ganster, J., Kohl, G., Holsboer, F., Landgraf, R., Engelmann, M., 1998. Dissociated central and peripheral release of vasopressin, but not oxytocin, in response to repeated swim stress: New insights into the secretory capacities of peptidergic neurons. Neuroscience 85, 1209–1222. doi:10.1016/S0306-4522(97)00683-0
Yang, J., Yang, Y., Chen, J.-M., Liu, W.-Y., Wang, C.-H., Lin, B.-C., 2007. Central oxytocin enhances antinociception in the rat. Peptides 28, 1113–1119. doi:10.1016/j.peptides.2007.03.003

Figure Legends

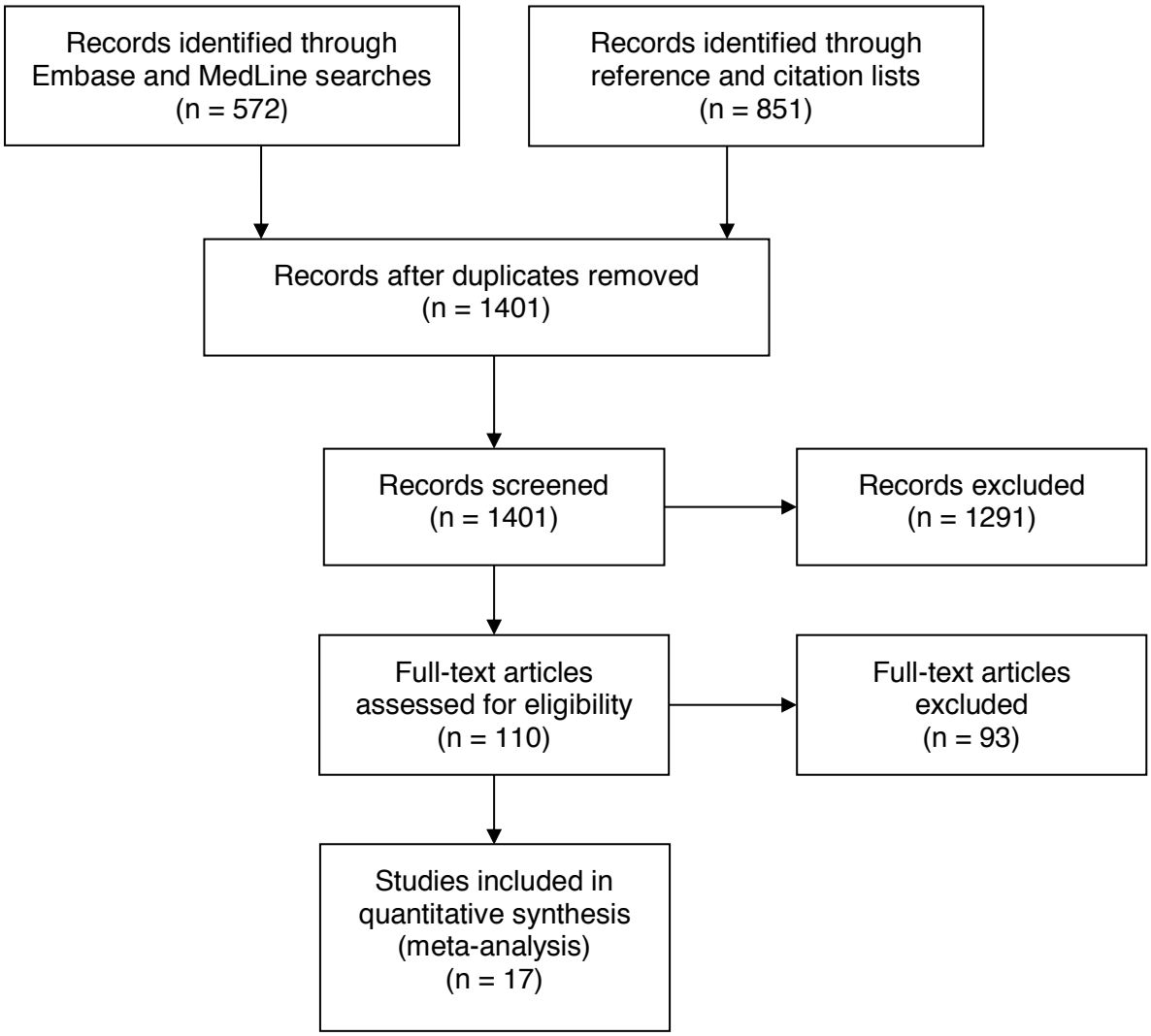
Figure 1. Flowchart of inclusion process. **n**: number of records.

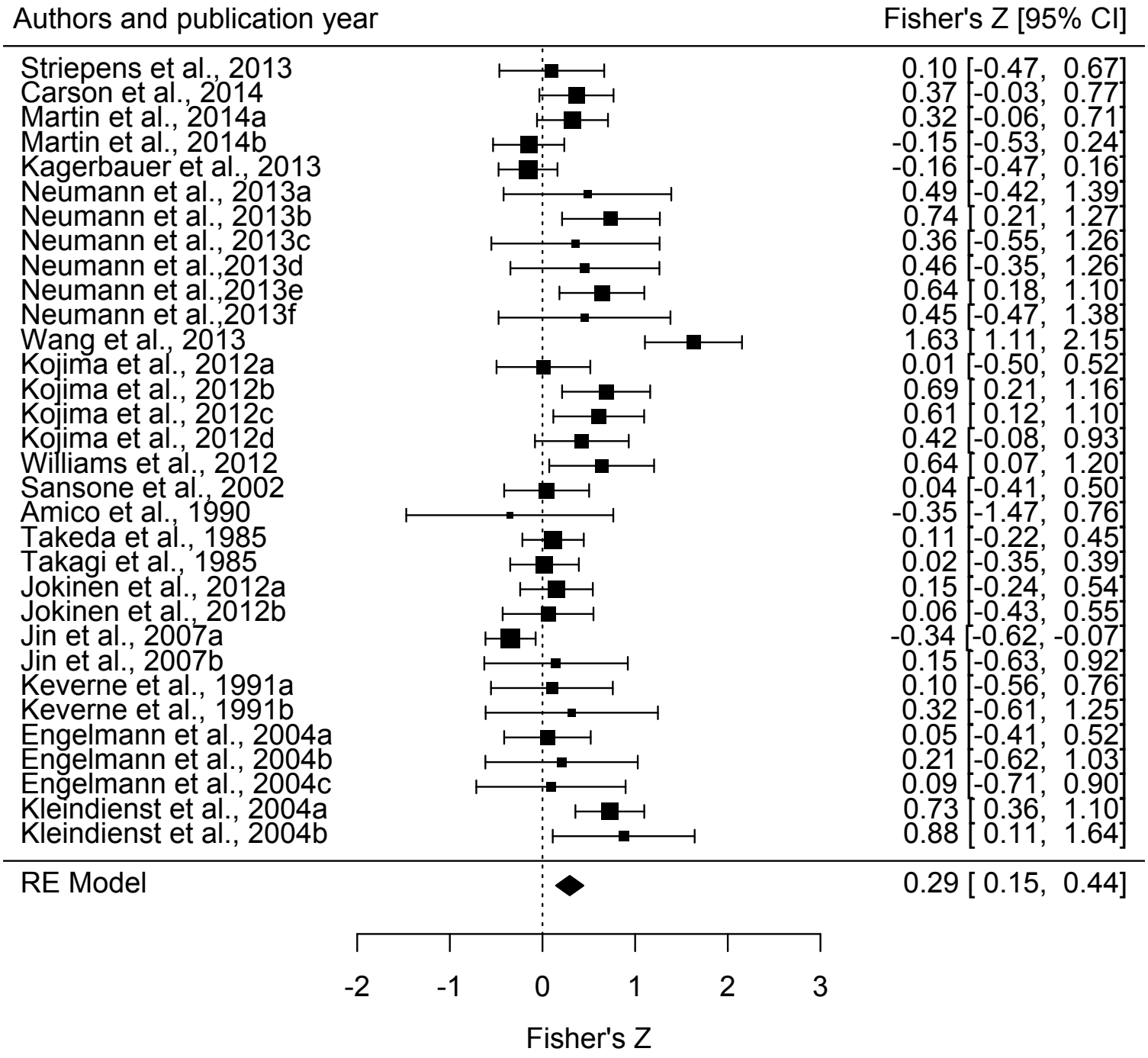
Figure 2. The relationship between peripheral and central oxytocin concentrations. Fisher's *Z* point estimates are depicted by filled squares, with square sizes indicating the relative weight of each study's effect size estimate in the analysis. The filled diamond reflects the overall summary effect size [Fisher's $Z = 0.30$, 95% CI (0.15 to 0.44), $p < 0.0001$]. Error bars and diamond width indicate 95% CIs. Note that the Fisher's Z summary point estimate slightly differs from the transformed Pearson's *r* point estimate, which is reported in text. **RE** = Random effects model.

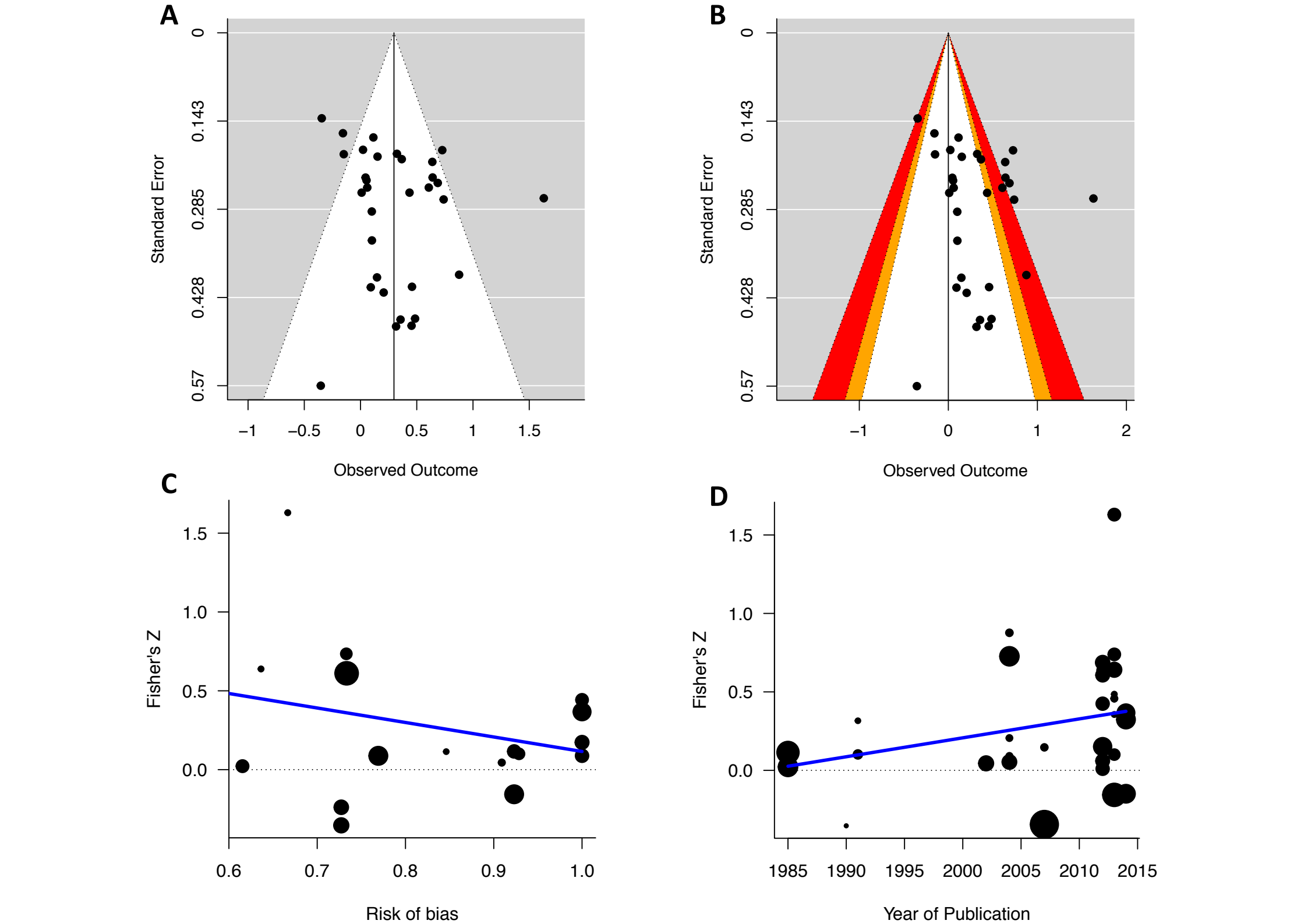
Figure 3. Tests of bias and the impact of moderator variables. Plots A and B illustrate the individual effect sizes on the horizontal axis and corresponding standard errors on the vertical axis. Visual inspection of plots does not reveal evidence of publication bias, as there does not appear to be any asymmetry (A) or an overrepresentation of effect sizes in the orange and red significance contours (B). Meta-regression models revealed no evidence that risk of bias (C; p=0.2204) or publication year (D; $p = 0.14$) had a significant influence on effect size. The solid blue lines in plots C and D represent the respective predicted Fisher's Z scores based on a mixed-effects models.

Figure 4. Subgroup effects for the moderators: experimental paradigm (A), central sampling location (B), subject species (C), biochemical analysis method (D), and sample extraction (E). Group summary point estimates are expressed as Pearson's r with 95% confidence intervals. There was a significant difference in the summary point estimates of intranasal administration and baseline experimental paradigm studies (Holm corrected $p = .002$).

Figure 1: Flowchart of inclusion process




Authors and publication year
Fisher's Z [95% CI]
Striepens et al., 2013 0.10 [-0.47, 0.67]
Carson et al., 2014 0.37 [-0.03, 0.77]
Martin et al., 2014a 0.32 [-0.06, 0.71]
Martin et al., 2014b -0.15 [-0.53, 0.24]
Kagerbauer et al., 2013 -0.16 [-0.47, 0.16]
Neumann et al., 2013a 0.49 [-0.42, 1.39]
Neumann et al., 2013b 0.74 [0.21, 1.27]
Neumann et al., 2013c 0.36 [-0.55, 1.26]
Neumann et al.,2013d 0.46 [-0.35, 1.26]
Neumann et al.,2013e 0.64 [0.18, 1.10]
Neumann et al.,2013f 0.45 [-0.47, 1.38]
Wang et al., 2013 1.63 [1.11, 2.15]
Kojima et al., 2012a 0.01 [-0.50, 0.52]
Kojima et al., 2012b 0.69 [0.21, 1.16]
Kojima et al., 2012c 0.61 [0.12, 1.10]
Kojima et al., 2012d 0.42 [-0.08, 0.93]
Williams et al., 2012 0.64 [0.07, 1.20]
Sansone et al., 2002 0.04 [-0.41, 0.50]
Amico et al., 1990 -0.35 [-1.47, 0.76]
Takeda et al., 1985 0.11 [-0.22, 0.45]
Takagi et al., 1985 0.02 [-0.35, 0.39]
Jokinen et al., 2012a 0.15 [-0.24, 0.54]
Jokinen et al., 2012b 0.06 [-0.43, 0.55]
Jin et al., 2007a -0.34 [-0.62, -0.07]
Jin et al., 2007b 0.15 [-0.63, 0.92]
Keverne et al., 1991a 0.10 [-0.56, 0.76]
Keverne et al., 1991b 0.32 [-0.61, 1.25]
Engelmann et al., 2004a 0.05 [-0.41, 0.52]
Engelmann et al., 2004b 0.21 [-0.62, 1.03]
Engelmann et al., 2004c 0.09 [-0.71, 0.90]
Kleindienst et al., 2004a 0.73 [0.36, 1.10]
Kleindienst et al., 2004b 0.88 [0.11, 1.64]
RE Model 0.29 [0.15, 0.44]
-2
-1
0
1
2
3
Fisher's Z


A
Standard Error
0
0.143
0.285
0.428
0.57
−1
−0.5
0
0.5
1
1.5
Observed Outcome
B
Standard Error
0
0.143
0.285
0.428
0.57
−1
0
1
2
Observed Outcome
C
Fisher's Z
1.5
1.0
0.5
0.0
0.6
0.7
0.8
0.9
1.0
Risk of bias
D
Fisher's Z
1.5
1.0
0.5
0.0
1985
1990
1995
2000
2005
2010
2015
Year of Publication

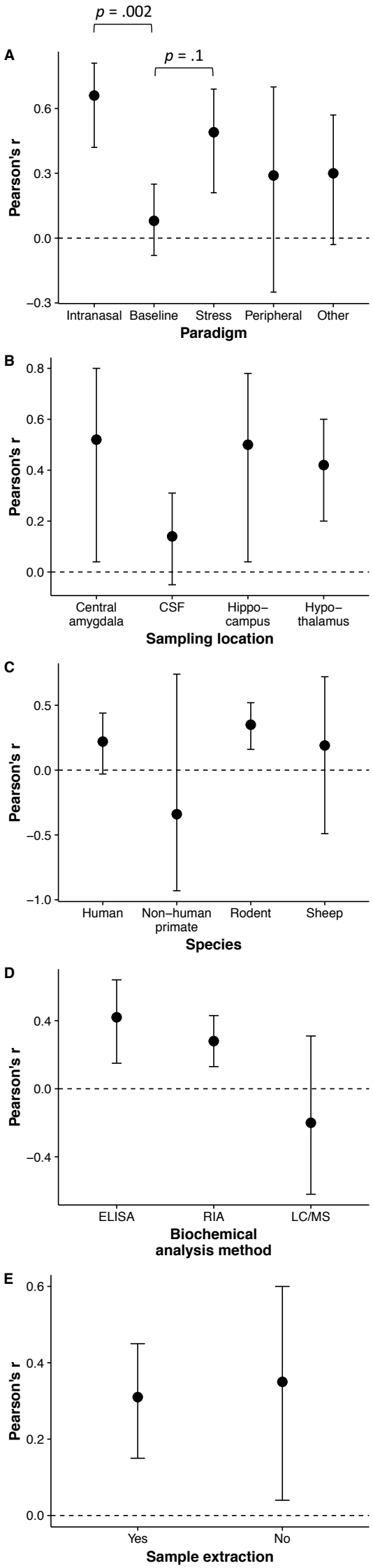

p = .002
p = .1
A
Pearson's r
0.6
0.3
0.0
−0.3
Intranasal
Baseline
Stress
Peripheral
Other
Paradigm
B
Pearson's r
0.8
0.6
0.4
0.2
0.0
Central amygdala
CSF
Hippo–campus
Hypo–thalamus
Sampling location
C
Pearson's r
0.5
0.0
−0.5
−1.0
Human
Non–human primate
Rodent
Sheep
Species
D
Pearson's r
0.4
0.0
−0.4
ELISA
RIA
LC/MS
Biochemical analysis method
E
Pearson's r
0.6
0.4
0.2
0.0
Yes
No
Sample extraction

Table 1: Overview of included studies

Study	*n*	*Species*	*Sex*	*Age*	*Status*	*csl*	*Conditions*	*rb*	*k*	*m*
(Striepens et al, 2013)	15	Human	M	19-64	Various	CSF	IN-OT, Baseline	0.93	1	15
(Carson et al, 2014)	27	Human	MF	4-64	Various	CSF	Baseline	1	1	27
(Martin et al, 2014)	41	Human	MF	19-81	Subarachnoid haemorrhage, Various	CSF	Baseline, Other	0.77	2	126
(Kagerbauer, 2013)	41	Human	MF	19-81	Various	CSF	Baseline	0.92	1	41
(Neumann et al, 2013)	27	Rats, mice	M	N.A.	N.A.	CeA, Hc	Baseline, IN-OT, IP-OT	0.73	6	232
(Wang et al, 2013)	17	Human	MF	19-64	Headache, Healthy	CSF	IN-OT	0.66	1	17
(Kojima et al, 2012)	74	Rats	MF	2w		Hyp	Baseline, Stress	1	4	74
(Williams et al, 2012)	15	Rats	F	N.A.	Postpartum	Hyp	Stress	0.64	1	26
(Sansone et al, 2002)	11	Rats	F	N.A.	Hypophys-ectomy, ovarectomy	CSF	Other	0.91	1	45
(Amico et al, 1990)	4	Monkeys	F	N.A.	Lactating	CSF	Baseline	0.73	1	40
(Takeda et al, 1985)	38	Human	MF	17-52	Pregnancy, Various	CSF	Baseline	0.85	1	38
(Takagi et al, 1985)	31	Human	F	19-45	Pregnancy, Various	CSF	Baseline	0.62	1	31
(Jokinen et al, 2012)	47	Human	MF	23-66	Suicidal, Healthy	CSF	Baseline	0.92	1	47
(Jin et al, 2007)	72	Mice	MF	2-5m	N.A.	CSF	Baseline, SC-OT	0.73	2	110
(Keverne and Kendrick, 1991)	7	Sheep	F	N.A.	Ovarectomy	CSF	Baseline, Other	1	2	105
(Engelmann et al, 2004)	18	Rats	M	N.A.	N.A.	Hyp	Baseline, Stress, Other	1	3	89
(Kleindienst et al, 2004)	20	Rats	M	~17w	N.A.	Hyp	Baseline, Other	0.73	2	84

Note. **n**: sample size, **csl**: central sampling location, **rb**: risk of bias (inversed), **CSF**: cerebrospinal fluid, **CeA**: central amygdala, **Hc**: hippocampus, **Hyp**: hypothalamus, **IN-OT**: intranasal oxytocin, **IP-OT**: intraperitoneal oxytocin, **SC-OT**: subcutaneous oxytocin, **k**: number of separate effect sizes included in meta-analysis, **m**: number of pairs of samples, **N.A.:** Data not available.

Supplementary Table S1: PRISMA checklist

Section/topic	#	Checklist item	Reported on page
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3-4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	4-5
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	5
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5-6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5-6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	5
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	5-6
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	6-7
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	7

Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	9
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	7
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	7
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	8
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	7-8
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	9
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	9
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	25
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	27
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	9-10
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	9
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	10-12
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	12-13
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	15-16
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	13-15,

FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	17

Supplementary Text S2: Validation of the web plot digitizer

One method for extracting correlations from articles employed in this meta-analysis was data scraping, in which scatterplots provided in published articles were screenshot and uploaded to a web plot digitizer (Rohatgi, 2015). Three studies (Amico et al., 1990; Takagi et al., 1985; Takeda et al., 1985) were included because we could scrape data from provided scatterplots. To make sure that this is a precise method for data extraction, we looked at the four studies included in this meta-analysis in which both a correlation and a scatterplot were provided (Carson et al., 2014; Kagerbauer et al., 2013; Kojima et al., 2012; Wang et al., 2013), and compared the correlations obtained from data scraping with the actual correlations stated in the articles.

Study	*r (web plot digitizer)*	*r (article)*
Wang	0.922	0.926
Kojima	0.597	0.596
Kagerbauer	-0.121	-0.149
Carson	0.598	0.56

The precision of the web plot digitizer was very high [r = 0.99, 95% CI (0.97, 0.99)]. There were, however, some slight differences in precision, which may be attributable to differences in scatterplot styles.

S3: Data extraction form for individual studies included in meta-analysis
A1: Assign for each study included. **B1**: Specify number of reported correlations. **B2**

Categories	Values
A Identification of study	
1. Study ID	
2. Authors	
3. Title	
4. Year of publication	
B Primary values	
1. Number of correlations	
2. Sample size	
3. Effect size	
4. Number of samples	
C Population	
1. Species	
2. Gender	
3. Mental health status	
4. Somatic health status	
5. Age	
D Moderators	
1. Study type	
2. Sampling location	
3. Biochemical analysis	
4. Extraction	
5. Level of coordination	
E Other	
1. Risk of bias	

For each correlation, specify sample size. **B3**: For each correlation, specify effect size and entity of correlaton. **B4**: For each correlation, specify included number of time-points for coordinated samples. **C3**: Specify mental health status of participants. **C4**: Specify somatic health status of participants. **C5**: Specify age of participants in range and/or mean and standard deviation. **D1**: Specify whether concentrations are sampled at a) baseline level, b) after exogenous OT administration, or c) after other experimental manipulations. **D2:** Specify the locations for central and peripheral samples. **D3**: Specify method for biochemical analysis of OT concentration. **D4**: Specify whether OT was extracted from the sampled substance. **D5**: Specify whether central and peripheral concentrations were sampled a) simultaneously (time interval < 2 mins), b) non-simultaneously, or c) not reported. **E1**: Report risk of bias expressed as a ratio of raw score to obtainable raw score for the study (see risk of bias form).

Supplementary text S4: Dependent samples variance estimation

For some included studies, multiple effect sizes were reported (e.g. Kojima et al., 2012). If these effect sizes were independent and representative of different moderators, such as in a between subjects study with an experimental and a control group (e.g. Martin et al., 2014), the separate groups were treated as separate studies in the meta-analysis. If, on the other hand, these effect sizes were not independent, such as in studies with repeated measures, preparatory analysis was performed according to planned contingent procedures prior to inclusion of resulting effect sizes to the main meta-study. In the case of repeated measures for the same participants under similar conditions, such as at different time-points in a baseline condition or after exogenous OT administration, separate effect sizes for the different time-points were pooled using a fixed effects model. Variance for the intra-condition pooled effect sizes was estimated with dependent samples variance estimation:

$$V_Y = \left(\frac{1}{m}\right)^2 \left(\sum_{j=1}^{m} V_i + \sum_{j \neq k} r_{jk} \sqrt{V_j} \sqrt{V_k}\right),$$

according to the assumption that the intra-condition separate effect sizes were associated (Borenstein et al., 2009). In all the cases where exact dependencies were unknown, an r=0.5 association between dependent effect sizes) was assumed, on the premise that the lack of robust variance estimation would default the association on either r=0 or r=1 (Borenstein et al., 2009). In the case of repeated measures for the same participants under dissimilar conditions, such as under baseline conditions versus after exogenous OT administration, separate effect sizes expressed different levels of the experimental paradigm moderator variable, and accordingly were not pooled, but rather included in the main meta-analysis directly. However, as these separate effect sizes were not independent, the weight they were afforded in the main meta-analysis was given by dependent samples variance estimation, where a total weight for the pooled effect sizes was distributed to separate effect sizes according to the number of samples in each separate effect size. In other studies, reported single effect sizes were based on repeated measures. For all of these, within-study variance was estimated using dependent samples variance estimation to control for the between measures dependence.

Since exact dependencies between repeated samples were unknown, there is a chance that variances for effect sizes obtained in repeated measures designs were slightly overestimated or underestimated, relative to variances for effect sizes obtained in single sample designs. A differential variance estimation would favor one of the two study types with respect to the relative weight they were afforded in the main analysis. However, since there is no a priori reason to believe that study type should affect effect sizes, this potential bias is not very disquieting. The functions and procedures for variance estimates are available at http://osf.io/aj55y/

Supplementary text S5: Moderators and moderator levels in protocol and in meta-analysis.

In the published protocol (Valstad et al., 2016), some of the a priori defined moderator variables were described with other levels than the ones used in the final moderator analysis. Such discrepancies are due to contingent characteristics of included studies that could not be predicted at the stage of meta-analysis planning and protocol registration. For the experimental paradigm moderator, a level was added to accommodate the effect sizes for samples obtained from subjects under experimentally induced stress (Engelmann et al., 2004; Kojima et al., 2012; Williams et al., 2012). Age moderation was omitted from analysis due to difficulty of comparison across species, while diagnosis and sex was omitted due to heterogeneity within effect sizes. In all the included studies, peripheral oxytocin was sampled from blood, such that no levels for peripheral sample location were required. Likewise, sample coordination was high in all included studies, such that no moderator analysis was required for this a priori identified moderator variable.

Supplementary Table S6: Risk of bias tool.

Criteria	Score
Category 1: Sample characteristics	
Human population	
1. Presence (or absence) of study-relevant diagnosis established by trained assessor [1p], according to standardized international criteria [1p].	/2
2. Reported comorbid diseases [1p] and medication use [1p].	/2
Non-human population	
3. Specified species [1p], source of individual animals stated [1p].	/2
Category 2: Internal Validity	
4. Potential confounding variables controlled for [2p], or accounted for [1p].	/2
5. All [2p] or most [1p] outcome measures have been validated.	/2
Category 3: Methodology and reporting	
6. Method of analysis stated for central [1p] and peripheral [1p] measure.	/2
7. Outcome data is complete, or incompleteness is accounted for [2p]	/2
8. Method of extraction described [1p].	/1
Category 4: Paradigm-specific criteria	
9. IF IN-OT: Proper instructions to participants for administration of OT [1p], measure of nasal cavity health [1p].	/2

Note: for a single study, a raw score of 11-15 is obtainable in principle. For each study, the risk of bias is expressed as a ratio of the actual raw score to its obtainable raw score. Abbreviations: OT = oxytocin, IN-OT = intranasal oxytocin.

Supplementary table S7: Table of subgroup effects and summary effect size

paradigm	*k*	*r*	*p*	*ci.l*	*ci.u*	*Q*	*p.h*	I^2
Baseline	15	0.08	0.310	-0.08	0.25	28.81	0.01	51%
IN-OT	4	0.66	<0.0001	0.42	0.81	16.02	0.001	81%
Stress	5	0.49	0.001	0.21	0.69	1.94	0.746	~0%
P-OT	3	0.29	0.284	-0.25	0.70	0.27	0.87	~0%
Other	5	0.30	0.074	-0.03	0.57	3.43	0.49	~0%

species	*k*	*r*	*p*	*ci.l*	*ci.u*	*Q*	*p.h*	I^2
Humans	10	0.22	0.081	-0.03	0.44	39.05	<0.0001	77%
Rodents	19	0.35	0.0004	0.16	0.52	43.88	0.0006	59%
Monkeys	1	-0.34	0.534	-0.93	0.74	0	1	NA
Sheep	2	0.19	0.546	-0.49	0.72	0.14	0.711	0%

csl	*k*	*R*	*p*	*ci.l*	*ci.u*	*Q*	*p.h*	I^2
CSF	16	0.14	0.143	-0.05	0.31	51.23	<0.0001	71%
Hypoth.	10	0.42	0.0004	0.20	0.60	11.85	0.222	24%
Ce.amy.	3	0.52	0.034	0.04	0.80	0.60	0.74	~0%
Hippoc.	3	0.50	0.034	0.04	0.78	0.23	0.89	~0%

analysis	*k*	*r*	*p*	*ci.l*	*ci.u*	*Q*	*p.h*	I^2
RIA	24	0.28	0.0005	0.13	0.43	59.08	<0.0001	61%
EIA	6	0.42	0.0035	0.15	0.64	4.85	0.44	0%
LCMS	2	-0.20	0.43	-0.62	0.31	1.38	0.24	27%

extraction	*k*	*r*	*p*	*ci.l*	*ci.u*	*Q*	*p.h*	I^2
Yes	24	0.31	0.0002	0.15	0.45	38.17	0.024	40%
No	6	0.35	0.0278	0.04	0.60	28.71	<0.0001	83%

	k	*r*	*P*	*ci.l*	*ci.u*	*Q*	*p.h*	I^2
Sum	32	0.29	>0.0001	0.14	0.42	86.19	<0.0001	63%

Note: for some moderators, k does not add up to 32, since not all effect sizes could be assigned to a level on that particular moderator. **k**: number of effect sizes, **r**: correlation coefficient, **p**: probability of r estimate given null parameter, **ci.l**: lower bound of confidence interval, **ci.u**: upper bound of confidence interval, **p.h**: probability of Q given no true differences between effect sizes, **P-OT**: peripherally administered OT, **csl**: central sampling location.