



# Safety and tolerability of chronic intranasal oxytocin in older men: results from a randomized controlled trial

Jillian M. Rung<sup>1,2</sup> · Marilyn Horta<sup>1</sup> · Erin M. Tammi<sup>1</sup> · Eliany Perez<sup>1</sup> · Marite C. Ojeda<sup>1</sup> · Tian Lin<sup>1</sup> · Griffin Harris<sup>1</sup> · Jessie Somerville<sup>1</sup> · Dinia Salmeron<sup>1</sup> · Susan E. Beltz<sup>3</sup> · Bhanuprasad Sandesara<sup>4</sup> · David Feifel<sup>5</sup> · Natalie C. Ebner<sup>1,4,6</sup>

Received: 2 July 2020 / Accepted: 29 April 2021 / Published online: 12 May 2021  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

## Abstract

**Rationale** Most studies evaluating the safety and tolerability of intranasal oxytocin (OT) have not reported consistent adverse events (AEs), but they have largely focused on young men and single-dose administration. Thus, it is unclear whether these findings translate to older individuals and with longer administration periods.

**Objective** Extending previous work, this study investigated the safety and tolerability of chronic intranasal OT in generally healthy older men.

**Methods** Data were from a randomized, placebo (P)-controlled, double-blind clinical trial evaluating the effects of 4 weeks of self-administered intranasal OT (24 IU twice daily) in older adults with no major physical or cognitive impairments. Heart rate, blood pressure, urine osmolality, and serum metabolic biomarkers were obtained before and at the end of the intervention. AEs were collected during the first 3 weeks and 1 week after cessation of treatment.

**Results** Of 103 participants recruited, 95 were randomized and received the intervention (OT = 49, P = 46). OT had no significant impact on cardiovascular, urine, or serum measures. The AEs reported for both treatments were generally mild and few in number, though one participant assigned to OT and two assigned to P dropped out due to AEs. Relative to P, OT did not significantly increase the likelihood of reporting AEs, nor the number or severity of AEs reported.

**Conclusion** Chronic intranasal OT appears safe and well-tolerated in generally healthy older men. These findings provide support for continued human research on potential benefits of chronic OT in older adult populations.

**Keywords** Oxytocin · Aging · Safety · Adverse events · Intranasal administration · Male

## Introduction

Oxytocin (OT) is a naturally occurring nine-amino acid neuropeptide that is implicated in a variety of physiological processes including parturition and lactation (Russell and Leng 1998), appetite (Leng et al. 2008), sex (Gimpl and Fahrenholz 2001), and stress regulation (Neumann 2007). OT is also involved in social and cognitive domains (Guastella et al. 2012; Di Simplicio and Harmer 2016) across development (Carter 2003; Lefevre and Sirigu 2016; Horta et al. 2020b).

The discovery that OT can be non-invasively delivered to peripheral circulation and the central nervous system via nasal spray has prompted empirical investigations into the effects of OT administration across a variety of functional domains (Born et al. 2002; Quintana et al. 2015). In particular, various pathways to central and peripheral targets via intranasal OT administration have been proposed for

✉ Jillian M. Rung  
rung.jillian@gmail.com; jrung@ufl.edu

✉ Natalie C. Ebner  
natalie.ebner@ufl.edu

<sup>1</sup> Department of Psychology, University of Florida, P.O. Box 112250, Gainesville, FL 32611, USA

<sup>2</sup> Department of Epidemiology, University of Florida, Gainesville, FL 32611, USA

<sup>3</sup> Investigational Drug Service, University of Florida, Gainesville, FL, USA

<sup>4</sup> Department of Aging and Geriatric Research, Institute on Aging, University of Florida, Gainesville, FL, USA

<sup>5</sup> Department of Psychiatry, University of California, San Diego, USA

<sup>6</sup> Center for Cognitive Aging and Memory, Department of Clinical and Health Psychology, University of Florida, Gainesville, FL, USA

the modulation of social cognition (Quintana et al. 2015; Martins et al. 2020). OT has also been considered for the treatment of conditions such as autism spectrum disorder (ASD) (Yamasue 2016), pain (Rash et al. 2014; Lussier et al. 2019), schizophrenia (MacDonald and Feifel 2012), and drug addiction (Sarnyai and Kovács 1994). Intranasal OT has also recently received attention as a candidate for improving function across social and emotional domains in late adulthood (Huffmeijer et al. 2013; Ebner et al. 2014). In fact, such investigations probing the effects of intranasal OT among older individuals have generated promising findings pointing to a modulatory role of OT on brain function and socioemotional processing (Campbell et al. 2014; Ebner et al. 2015; Horta et al. 2019; cf. Grainger et al. 2019, 2018). This emerging evidence signals the importance of studying the OT system and determining the safety and tolerability of intranasal OT administration among older adults.

Safety information on intravenous OT administration in clinical contexts has been reported, including cardiovascular changes, nausea, vomiting, and headaches (MacDonald et al. 2011). In a review of 38 controlled human studies administering OT intranasally, MacDonald et al. concluded that acute (i.e., one-time, single dose) administration of intranasal OT (18–40 IU) in healthy and clinical populations (e.g., ASD, post-traumatic stress disorder, social anxiety disorder) composed of mostly young men, is well-tolerated, and not strongly associated with any adverse effects. Adverse events reported by individuals receiving OT included light-headedness, drowsiness, headache, nasal irritation, and dry mouth but a causal connection has not been established. Reports of the safety and tolerability of intranasal OT have been supported in subsequent reviews among pediatric populations (DeMayo et al. 2017) and individuals of varying age with ASD (Cai et al. 2018).

Despite the growing number of studies involving intranasal OT administration in older adults, there is limited information regarding its safety in that population. Several single-dose studies using 24 IU in older adults have reported no or inconsistent adverse effects (Grainger et al. 2018; Horta et al. 2019), while other studies did not specifically mention adverse effects (Campbell et al. 2014; Grainger et al. 2019). Two studies have evaluated adverse effects of OT as a function of repeated administration. In one of these studies, healthy older adults self-administered intranasal OT (40 IU) daily for 10 days, which was generally well-tolerated with only minor drowsiness reported in the OT group (Barraza et al. 2013). Comparable results were found in older individuals (mean age 66 years) with frontotemporal dementia at varying OT doses (24, 48, or 72 IU twice daily, or BID) administered for 1 week (Finger et al. 2015), which were corroborated by no negative impact on physiologic measures (e.g., blood pressure, sodium levels). Thus, at least in these smaller samples ( $N=23$ –39 participants; treatment groups

ranging from  $n=4$ –21), it appears that intranasal OT administration has no significant adverse effects in aging, including when administered repeatedly. The results from these trials are informative, but this knowledge is not comprehensive and may not be representative of potential adverse events with a dose administered in a more prolonged context (e.g., over several weeks).

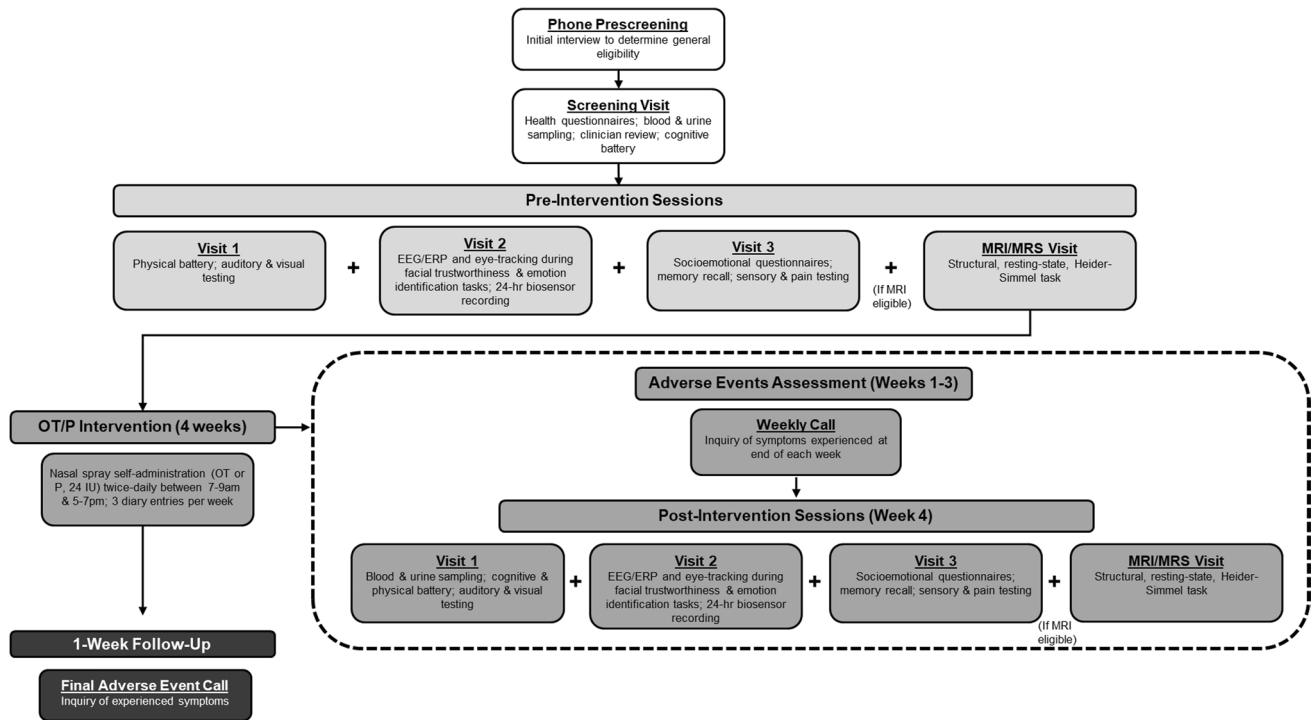
In this paper, we assessed the safety and tolerability of chronic (i.e., repeated, daily) intranasal OT over 4 weeks in generally healthy older men in the context of a randomized, placebo-controlled, double-blind clinical trial that used a dosage level (24 IU, BID) in line with previous experimental research among clinical populations (MacDonald and Feifel 2013). Specifically, we evaluated whether chronic intranasal OT was associated with compromised objective or self-reported safety, as indicated through measures of blood pressure and heart rate, sodium/hydration levels (osmolality), and/or changes in kidney and liver function, as well as self-reported adverse events.

## Methods

### Participants

One hundred and three older male volunteers (mean age:  $71.30 \pm 7.71$  years; range 55–95 years) enrolled in a clinical trial (clinicaltrials.gov, NCT02069431 Oxytocin and Aging Study) to examine the effects of chronic intranasal OT administration on (i) inflammatory biomarkers, (ii) socioemotional and physical functioning, and (iii) auditory and cognitive functioning in healthy aging. There was no indication for the sample other than aging. A CONSORT-format diagram (Online Resource 1) shows the flow of participants in the present report, including all participants who passed an initial phone pre-screening and were subsequently enrolled in the clinical trial. Participants were recruited through mail-outs and fliers in the community and on the university campus, retirement communities, newspaper ads, University of Florida (UF)-based participant registries, UF's community-recruitment service HealthStreet, and word of mouth. Recruitment began in February of 2016 and continued, on a rolling basis, through February 2020. Recruitment activities described above were conducted in Gainesville, FL, and the surrounding communities. All study procedures were approved by the University of Florida Institutional Review Board (UF IRB; Protocol #201300801), monitored by the UF IRB, the Data Safety Monitoring Board at the Institute on Aging, and the FDA (IND 100,860).

Inclusion criteria were 55 years of age and older, generally healthy and eligible to administer OT (i.e., no nasal obstruction), English-speaking, able and willing to provide consent, and a pre-treatment blood pressure less than



**Fig. 1** Diagram depicting the different phases of the experiment, the timing/order of experimental sessions, and the primary measures in each session/phase. Events encapsulated by the dashed outline occurred within the 4-week oxytocin (OT)/placebo (P) intervention,

with specific events taking place in the respective intervention weeks as noted. EEG/ERP, electroencephalogram/event-related potential; DTI, diffusion tensor imaging; MRI, magnetic resonance imaging; MRS, magnetic resonance spectroscopy

180/100 mm Hg. Participants had to score  $\geq 30$  on the Telephone Interview for Cognitive Status (Brandt, Spencer, & Folstein, 1988) for study eligibility. Primary exclusion criteria were as follows: history of hyponatremia or syndrome of inappropriate antidiuretic hormone secretion; use of vasoconstrictors (e.g., desmopressin, pseudoephedrine, or antidiuretic medication); low sodium ( $< 134$  mEq/L) coupled with high urine osmolality ( $> 1200$  L); psychogenic polydipsia; and heavy cigarette smoking or alcohol consumption. Study inclusion/exclusion criteria underwent changes throughout the course of the project, which included (1) increasing the maximum allowable blood pressure at pre-treatment to better accommodate the targeted age group (from an original 160/90), and (2) opening enrollment to females after securing additional study funds.

Participants were evaluated for eye-tracking and magnetic resonance imaging/spectroscopy (MRI/MRS) eligibility. Exclusion criteria for eye-tracking procedures were history of severe forms of glaucoma, macular degeneration, and cataract(s). Exclusion criteria for MRI/MRS procedures were history of brain surgery, serious brain damage or disease (e.g., aneurysm, stroke, or seizures), and any contraindication to MRI (e.g., severe claustrophobia, major surgery in the past two months, certain large pieces of metal in the body or any metal in the face or neck). Participants who

were eligible for MRI/MRS were invited to complete additional sessions involving these procedures.

At the end of each phase (pre-intervention, intervention, and post-intervention), participants were compensated with \$100. If participants completed MRI/MRS sessions, they received an additional \$50 after the second MRI/MRS visit. Thus, participants were compensated \$300 (non-MRI/MRS) or \$350 (MRI/MRS-eligible) for their participation in the entire study.

### Study design and procedure

A diagram of the study design and measures is presented in Fig. 1. From screening to follow-up, primary study activities spanned approximately 7 weeks. Sometimes, participants were contacted beyond this period when adverse event reports were not obtained as scheduled.

After consenting and all screening procedures, eligible participants completed three pre-intervention sessions for collection of cognitive, physiological, biological, behavioral, and self-report measures. Eligible participants also completed an optional pre-intervention MRI/MRS session. The specific measures included in this report and their timing are discussed in the following section “Measures.” Pre- and post-intervention sessions were conducted at the Institute on

Aging and at the McKnight Brain Institute at the University of Florida.

Participants were block randomized using a double-blind procedure to self-administer either 24 IU of intranasal OT or P twice daily (once each between 7 and 9AM, and between 5 and 7PM) for 4 weeks. Compounding, randomization, and distribution for the OT and P nasal spray were conducted by the Investigational Drug Service (IDS) at UF, with additional details on compounding and dosing provided in section “[Study drug, placebo, and administration adherence](#).” The IDS executed the randomization schedule, which was generated from the website randomization.com and used a block size of 4. On presentation for randomization, the study coordinator (who enrolled participants) provided a sequential randomization number and the IDS staff matched this number to the next sequential randomization number with the treatment (OT or P) in the randomization schedule. All nasal spray bottles were dispensed with a blinded label that indicated it contained either “Oxytocin or Placebo.”

During the first 3 weeks of the intervention, participants received weekly follow-up phone calls to monitor and record any potential adverse events associated with the spray administration. During the last week of treatment, participants returned for three end-of-intervention visits (hereafter referred to as “post-intervention”) that mirrored the pre-intervention visits. Participants did not administer the spray the morning of post-intervention visits to avoid measurement of acute effects, but resumed the treatment in the evening until completion of the 4-week intervention period. Post-intervention visits were conducted during the final week of administration, as opposed to after, to avoid the potential for withdrawal effects and their influence on measures. A follow-up phone call approximately 1 week after the last day of the intervention was conducted to evaluate any residual adverse effects.

## Measures

Only a subset of collected measures from this trial is presented here to depict the study sample and to evaluate safety and tolerability associated with chronic OT administration. These comprised cardiovascular measures, biomarkers indicating physiological functioning, and self-reported adverse events.

## Demographics and baseline characteristics

At pre-intervention, participants provided standard demographic information such as age, race, ethnicity, and years of education. All participants also provided their self-rated general physical and mental health/mood on a scale ranging from poor (1) to excellent (10); and all current medications including over-the-counter medications and vitamins/

supplements. The NIH Cognitive Toolbox was used to quantify participants’ fluid and crystallized cognition (Weintraub et al. 2013), represented by unadjusted mean scores.

## Safety-related measures

In the first pre- and post-intervention visits, research staff obtained participants’ heart rate (HR) and systolic and diastolic blood pressure (BP) at rest. At screening and post-intervention visit 1, urine and blood samples were collected from participants to assess safety biomarkers, which included urine osmolality (for kidney function) and key (primary) indices from a complete metabolic panel. The primary measures of analysis from the metabolic panel were blood urea nitrogen (BUN), creatinine, the ratio of the aforementioned (BUN:Creatinine), estimated glomerular filtration rate (eGFR), glucose, potassium, and sodium. Participants were instructed to not eat or engage in sexual activity or strenuous exercise in the 2 h before their appointment for the blood and urine collection, and to refrain from smoking, drinking alcohol, using any recreational drugs, or consuming any caffeine in the 24 h before coming in. Urine samples were collected in sterile plastic collection cups to test osmolality. Both urine osmolality and indices for the complete metabolic panel were measured according to standard procedures by Quest Diagnostics®.

## Adverse events

At the end of each of the first 3 weeks of nasal spray administration, and 1 week after the cessation of the administration period, participants were called and asked to report whether they had experienced 17 different symptoms in the last 7 days (by responding “Yes” or “No” after being read each one). When conducting these assessments, the following terms were used interchangeably to refer to adverse events: symptoms, physical reactions, and side effects. The checklist used to assess adverse events was based on MacDonald et al. (2011) and is presented in Online Resource 2. If a participant reported experiencing an adverse event, they were also asked to rate its severity (mild, moderate, or severe).

If participants were unavailable at the time of the call for a given week, a staff member called again the following day (unless requested otherwise by the participant) and research staff repeated this process as feasible or until the phone call was completed.<sup>1</sup> Due to difficulties contacting

<sup>1</sup> Due to an inability to reach participants (not answering phone calls, disconnected phone lines) or scheduling errors, some reports were missing ( $n=30$ ). We ultimately obtained 346 reports, or 92% of a total possible 376 reports (94 participants who received the intervention for at least 1 week, multiplied by four adverse event checklist queries).

some participants or scheduling errors, 128 adverse event reports (37% of total obtained) were obtained retrospectively, which was defined as reports obtained seven or more days late. These data were coded as such and late reporting was accounted for in the analyses. There were several off-schedule symptom reports that were recoded such that they represented the closest study week. For example, one participant who picked up their spray 5 days late was called on a weekly basis from the originally scheduled pick-up day, which resulted in all reports being 5 days early. This resulted in all of this participant's reports (with the exception of the first) better temporally representing the prior study week (e.g., week 2 symptoms were recoded as week 1, and week 3 as week 2); and the data were recoded as such.

In other instances, participants were called early (after two or fewer days of administration in a given intervention week); these early reports ( $n = 3$ ) were excluded from inferential analysis. Two of these early reports occurred during the first week of administration ( $OT = 1$ ,  $P = 1$ ), and one during the third week ( $P = 1$ ). Early reports were excluded from analyses for several reasons. First, there were too few instances of early reports to meaningfully control for their timing statistically. Second, depending on their week of occurrence, they either (i) did not follow a similar amount of time to gauge adverse events (e.g., allow similar time for irritation to begin or for participants to habituate to acute effects), or (ii) were too temporally proximal to on-time reports for other weeks.

As part of the adverse events assessment, participants were also asked to report any "other" symptoms they experienced. Severity ratings were not prompted for these "other" symptoms. Therefore, we did not include these additional symptoms in our primary analyses. Instead, we categorized them by likeness and scored them for valence (adverse/non-adverse).

### Study drug, placebo, and administration adherence

The synthetic OT and P (i.e., containing all ingredients that the active contained except for the synthetic OT powder) were compounded and dispensed at UF's IDS under IND 100,860 (sponsor: Dr. Feifel, University of California, San Diego). The IDS Pharmacy compounded a 120 unit/mL solution. Participants achieved the dose of 24 IU twice daily by administering one spray or 0.1 mL (12 IU) into each nostril twice daily for 28 days. All compounded OT and P batches had concentrations validated by an independent lab.

Adherence to the self-administration protocol was evaluated in two ways: first, from calculating an approximate proportion of the OT/P spray administered throughout the administration period, and second, from self-reported adherence logged in experimenter-provided paper logs. The proportion of spray used was calculated by taking the

difference between the weight of a full bottle of OT/P spray and the weight of the bottle upon return from participants, divided by the total amount used if participants had perfect adherence throughout the administration period. For simplification, the proportion and percentage-based measures of adherence were calculated without regard to the instruction for participants to skip the morning dose of spray during post-intervention visits. Thus, perfect adherence is represented at values of approximately 0.97/97.3%.

Participants were also provided a paper log to fill out during the intervention to track nasal spray adherence. Participants were instructed to write "Yes" or "No" for whether they administered the doses as instructed in the morning and evening for each study day, and to note the time the spray was administered. The percentage of doses taken was calculated as the total number of doses self-reported as administered throughout the intervention divided by the total number of reports provided; the number of reports provided was used to calculate percentages because reporting across study days/times was high (see "Results" section).

Degree of adherence is reported for both treatments, but these data are not included in primary analyses of safety or adverse events due to missing data for some participants on either or both bottle weights or nasal spray logs. This analytic choice was made because outcome measures (safety, adverse events) cannot have missing values on covariates/predictors in models; as such, including these adherence-related measures as covariates would mean excluding valid observations of the key measures of interest.

### Sample size determination

The enrollment target for the clinical trial was  $N = 300$  (150 males and 150 females), which was based on evaluating potential benefits of OT for reducing inflammation and preservation of physical, cognition, and socioemotional function in aging; and potential sex differences in any such effects. The full analytic sample for the present report comprised males who were randomized to receive either OT ( $n = 49$ ) or P ( $n = 46$ ) and received the intervention. The study was originally designed to investigate the potential benefits of OT in a single-sex sample (males;  $N = 150$ ) and as such recruitment was initially confined to males. When the COVID-19 pandemic began, the study was placed on hold. As it became evident that study activities would need to be halted for a significant duration to ensure participant safety, we decided to drop the aim of investigating potential sex differences and end the study given that we had achieved 70% of the original single-sex recruitment goal (103/150 males). The achieved  $N$  provided more than adequate power for adverse outcome measures. Due to the early termination and ultimately unbalanced proportion of males to females (see Online Resource

1 for details on recruitment numbers), we focused only on data from male participants for the present report.

## Data analysis

Before analysis of treatment effects, participants assigned to the OT and P groups were compared to assess any baseline (pre-intervention) differences in age, education, and cognition, as well as physical and mental health using Wilcoxon rank-sum tests. These same types of tests were used to assess differences in intervention adherence across groups (objective and self-reported). Wilcoxon rank-sum tests were used because the distribution of most of these variables violated assumptions of normality as determined by Shapiro-Wilks tests ( $p < 0.05$ ); as such, medians and first and third quartiles are reported for these measures instead of means and standard deviations. All inferential comparisons of frequencies of participants were conducted using Fisher's exact tests.

All safety outcomes (cardiovascular measures and biomarkers) were analyzed according to the following steps. First, the interaction between Treatment (OT/P) and Time (pre-/post-intervention) and their main effects were evaluated as predictors in robust linear mixed-models as fixed effects with a random intercept of Participant. The method used for conducting the robust mixed models does not include significance tests; as such, the  $p$ -values for the main effects and interactions from these models were manually calculated using the  $t$ -values corresponding to the estimated regression coefficients and Satterthwaite-approximated degrees of freedom (details provided below). Effect sizes for predictors ( $\delta$ ) were estimated as the model-estimated mean differences (i.e., the coefficients) divided by the square root of the summed variance components (participant-level and residual variances), which can be interpreted similarly to Cohen's  $d$  (see Brysbaert and Stevens, 2018 for example).

Analysis of adverse event data differed slightly to accommodate the nature of the outcome variables and the multiple ways of conceptualizing the reported adverse events. First, a generalized linear mixed-effects model assuming a binomial error distribution was used to analyze whether there were differences in the proportions of participants reporting any adverse events (Yes/No) across the study period (Time) as a function of Treatment (OT/P) (including both main effects and their interaction) while controlling for Report Timing. Time was represented as a categorical variable with four levels: the first 3 weeks of the intervention and the 1-week follow-up post-intervention. In this model, Report Timing was also included as a categorical variable (two levels: on-time and late), and only as a fixed effect. The designation of other variables as fixed and random was identical to that for the linear mixed models described above.

We also assessed differences in the number of reported adverse events and severity-weighted number of adverse events using generalized linear mixed models with a negative binomial error distribution (i.e., a discrete probability distribution with less restrictive assumptions on the mean and variance than a Poisson). In these analyses, participants who did not report any adverse events throughout the study were excluded (see section "Adverse events" below). The weighted number of adverse events was calculated as the number of mild, moderate, and severe adverse events, multiplied by 1, 2, and 3, respectively. For example, a participant reporting one mild and two moderate adverse events would receive a score of 5 ( $[1 \text{ mild} * 1] + [2 \text{ moderate} * 2]$ ). Model fit for negative binomial models was evaluated using simulation techniques (see Hartig 2019 for details). Either odds ratios ( $OR$ ) or rate ratios ( $RR$ ) were provided as effect size estimates from these models.

Finally, to ensure that aggregating symptom data did not mask the occurrence of specific symptoms due to OT, we also conducted exploratory analyses on a symptom-level basis. For each of the 17 symptoms, the number of participants who did versus did not experience the symptom was compared using Fisher's exact tests. These tests were conducted for each symptom, for each administration week, and the follow-up report. These analyses are considered exploratory due to the fact that the tests themselves are lower powered (given the lack of incorporating repeated measurements), and multiplicity concerns are high (17 symptoms  $\times$  4 weeks = 68 tests). They are conducted and included herein largely for illustrative purposes.

All analyses were conducted using  $R$  statistical software (R Core Team 2020). Descriptive statistics were conducted using functions in base  $R$ . Robust mixed models were conducted using the `robustlmm` package (Koller 2016) and Satterthwaite degrees of freedom using the `lmerTest` package (Kuznetsova et al. 2017). Generalized linear models were conducted using the `lme4` (for binomial outcome; Bates et al. 2015) and `glmmTMB` packages (for count-related data; Brooks et al. 2017). Model diagnostics for generalized linear models were conducted using the `DHARMA` package (Hartig 2019). For generalized linear models, test statistics for categorical predictors and their interactions were obtained using the `car` package (Fox and Weisberg 2019); and effect sizes ( $OR$ s and  $RR$ s) as well as model predictions from the `emmeans` package (Lenth 2020). The package `ggplot2` was used to create figures (Wickham 2016). Additional packages used are noted in the analysis code (see below). An alpha of 0.05 was used for all statistical tests; we did not correct for multiple comparisons/tests, which is more conservative to identify potential negative effects of the OT treatment. Study data and analysis code are available via the Open Science Framework (Rung et al. 2021).

## Results

Of the 103 male participants enrolled in the study (i.e., signed informed consent), 13 withdrew or were removed from the study (see Online Resource 1). Of these thirteen, seven participants ceased participation before randomization, and the other six dropped out after randomization: one before receiving OT, four during the intervention (OT,  $n = 1$ ; P,  $n = 3$ ), and one after completion of the intervention (assigned to OT). The OT-assigned participant who ceased participation after the intervention did so due to scheduling difficulties.

Of those who dropped during the intervention, the participant assigned to the OT group dropped due to experiencing intermittent nausea and upset stomach after beginning the nasal spray administration. One of the participants assigned to the P group dropped out for medical reasons unrelated to the study (had surgery after completing baseline sessions). The other two participants assigned to the P group who dropped during the intervention period did so due to adverse events. One of them ceased using the spray during treatment after going to the emergency room, which follow-up suggested was unlikely due to the spray. The other P participant who dropped out had an adverse reaction to a recreational drug during the intervention and subsequently withdrew.

Throughout the course of the study, there was one unanticipated adverse event that required immediate reporting to the IRB and eight reported at the time of annual continuing review. These nine unanticipated adverse events include the three described above that led to attrition. The unanticipated adverse event that required immediate reporting occurred with a participant who received P, and thus the adverse event was not related to OT. The only other unanticipated adverse events not yet discussed that were potentially related to OT administration involved feelings of malaise, headaches, and/or nausea, and these symptoms eventually subsided. There were roughly equal numbers of IRB-reported adverse events in both groups ( $n = 4$  and 5 in OT and P, respectively). None of these unanticipated adverse events was categorized as serious according to standard definitions. Symptoms associated with all unanticipated adverse events (immediately reported or reported during annual continuing review) were included in their respective reporting period for the analysis.

The two treatment groups, composed of individuals who were randomized and received the intervention ( $N = 95$ ), did not significantly differ in age, race or ethnicity, cognitive function (crystallized and fluid composite scores), measures of self-reported physical and mental health, or types of medications taken ( $p = 0.99$ ; see Table S1 in Online Resource 3). However, individuals in the OT group had obtained significantly more years of education

**Table 1** Descriptive statistics (median and Q1–Q3) and statistical results of treatment group comparisons of sample characteristics at pre-intervention

Characteristic	Treatment		<i>p</i> -value
	Oxytocin <i>n</i> = 49	Placebo <i>n</i> = 46	
Age (years)	73 (65–78)	68 (65–72)	.09
Race			.78
American Indian/Alaska Native	1 (2.04%)	2 (4.35%)	
Asian	1 (2.04%)	1 (2.17%)	
Black or African American	1 (2.04%)	2 (4.35%)	
White	46 (93.88%)	41 (89.13%)	
Ethnicity			.50
Hispanic or Latinx	2 (4.26%)	0 (0%)	
Not Hispanic or Latinx	47 (95.74%)	46 (100%)	
Education (years)	17 (16–19)	15 (14–18)	.01
Cognition			
Crystallized composite	121 (113–128)	119 (112–126)	.35
Fluid composite	91 (85–96)	92 (85–96)	.91
Health			
Physical	8 (7–9)	8 (7–9)	.52
Mental	9 (8–10)	9 (8–9)	.22

All values for descriptive statistics are rounded to the nearest full unit due to the majority involving integer or near-integer values. All *p*-values are from Wilcoxon signed-rank tests using Treatment as the independent variable. Years of education are missing for four participants (two in each treatment) and cognition measures for two participants (both in the OT group)

( $p = 0.01$ ). Descriptive statistics for these variables by Treatment are shown in Table 1 (OT vs. P).

Based on the available data, adherence to the self-administration protocol was overall high. Among those returning all dispensed spray bottles (OT = 24, 49.0%; P = 25, 54.3%), participants in each group self-administered a median of 90% of the total doses dispensed for home administration (first and third quartiles for OT [74%, 97%] and P [70%, 97%], respectively), which did not significantly differ across groups ( $W = 420$ ,  $p = 0.34$ ). This objective measure of adherence was paralleled by high self-reported self-administration adherence among those who provided nasal spray log information (OT = 28, 57.1%; P = 31, 67.4%), with participants in each treatment group reporting a median of 96.0% of doses administered (first and third quartiles for OT [88%, 100%] and P [93%, 98%], respectively). The proportion of participants who returned all provided spray bottles and completed at least partial nasal spray log reports did not differ across treatment groups ( $ps \geq 0.40$ ).

While all participants who were randomized and received the intervention were included in the analytic sample ( $N = 95$ ), not all participants contributed to each primary analysis below. Inclusion of a participant in each analysis was a function of whether the data were obtained for at least one time point on a given measure. While the specific number of participants contributing to each analysis varied across analyses (noted where relevant), all participants from the analytic sample ( $N = 95$ ) contributed to analyses of objective safety measures; and all those

who completed at least one weekly adverse event report ( $n = 93$ ) across the four reporting periods were included in analyses of symptoms/adverse events. In other words, there were no data exclusions in any analysis, with the exception of the three adverse event reports that were obtained “early.”

## Objective safety measures

Descriptive statistics for cardiovascular measures and for urine and serum biomarkers are shown in Table 2. The only measure that significantly changed (see Table S2 in Online Resource 3) was alkaline phosphatase (a significant Treatment  $\times$  Time interaction;  $t[87.3] = -2.11$ ,  $p = 0.04$ ). Relative to participants in the OT group who showed a slight decrease in alkaline phosphatase from pre- to post-intervention, those in the P group showed a slight increase (P group coded as the intercept;  $B = -4.05$ ;  $\delta = -0.24$ ). Regardless, values for this measure in both treatment groups were in the normal range at both time points. The full model results for all safety measures are provided in Table S2 (in Online Resource 3), and descriptive statistics based on the nominal results (in-range vs. out of range values) of urine osmolality and metabolic measures are provided in Table S3 (i.e., percentage of samples within-range by treatment group, including number of valid samples obtained; also in Online Resource 3). Overall, 4-week intranasal administration of OT did not produce any significant adverse changes in safety-related measures.

**Table 2** Mean (standard deviation) of cardiovascular, urine osmolality, and metabolic measures by Treatment at pre- and post-intervention

Measure	Pre-intervention		Post-intervention	
	Oxytocin	Placebo	Oxytocin	Placebo
<b>Cardiovascular</b>				
Heart rate (bpm)	68.33 (11.26)	67.26 (9.11)	65.76 (10.69)	63.76 (11.62)
Systolic BP (mmHg)	134.00 (16.91)	130.87 (18.43)	134.24 (16.22)	129.02 (18.72)
Diastolic BP (mmHg)	75.73 (9.95)	78.76 (10.43)	77.35 (11.04)	78.90 (12.72)
<b>Urine-based</b>				
Osmolality (mOsm/kg)	591.04 (222.08)	632.11 (255.51)	568.55 (221.02)	599.40 (288.48)
<b>Metabolic</b>				
BUN (mg/dL)	18.96 (5.80)	18.41 (7.17)	18.68 (5.24)	17.72 (5.35)
BUN:Creatinine ratio	17.96 (4.77)	18.44 (5.05)	17.43 (3.78)	17.63 (5.49)
Creatinine (mg/dL)	1.07 (0.22)	1.00 (0.28)	1.07 (0.23)	1.03 (0.26)
eGFR (mL/min/1.73 m <sup>2</sup> )	71.77 (16.69)	78.37 (16.65)	70.91 (16.17)	75.88 (16.41)
Glucose (mg/dL)	105.57 (29.09)	109.60 (42.56)	108.43 (46.36)	108.86 (33.02)
Potassium (mmol/L)	4.46 (0.41)	4.48 (0.49)	4.53 (0.58)	4.52 (0.42)
Sodium (mmol/L)	139.70 (2.68)	140.33 (2.35)	139.98 (2.50)	140.98 (2.37)

BP, blood pressure; BUN, blood urea nitrogen; eGFR, estimated glomerular filtration rate. Due to equipment issues or errors in data recording, BP and heart rate were missing for three participants at post-intervention (two in the oxytocin and one in the placebo group). There were no significant differences in any outcomes shown here from pre- to post-intervention by group. See information in Table S3 (in Online Resource 3) for numbers of participants represented in each measure from the complete metabolic panel



## Adverse events

Across the study period, most participants reported at least one adverse event: 40 (81.6%) of participants in the OT and 32 (72.7%) of participants in the P group. These proportions for study-wide adverse events were not significantly different across treatments (Fisher's exact  $p=0.33$ ). The frequencies and percentages of participants reporting any and each type of adverse event by treatment and week are shown in Table 3. The most frequently reported symptoms in both treatment groups were nasal-related (e.g., runny nose, stuffed nose, sneezing). This was also true for the most common "Other" types of adverse events (see Table 4), although there was variability in the types reported. Of the total number of participants who were reached for adverse events reporting ( $n=93$ ), 25 participants (27%) indicated various types of "other" adverse events, ten of whom were in the OT group (40%). Reporting "other" adverse events did not vary by Treatment (Fisher's exact test,  $p=0.16$ ).

As can be seen in the top row of Table 3, more than a third of participants in each treatment reported experiencing an adverse reaction in any given study week (38–66%).

The analysis of adverse event reporting (yes/no for each study week) revealed that participants in the OT group were not significantly more likely to report experiencing adverse events than participants in the P group, as reflected by a non-significant Treatment main effect ( $OR=2.64$ ,  $SE=1.66$ ). Table S4 (in Online Resource 3) contains test statistics and  $p$ -values for the findings on self-reported adverse events (any adverse event, total number of adverse events, and severity-weighted adverse events). There was also no significant main effect of Time, nor Treatment  $\times$  Time interaction, indicating that the likelihood of reporting adverse events across study weeks was consistent for both treatments. Report Timing, however, was a significant predictor of reporting adverse events: regardless of treatment or time in the intervention, individuals were more likely to report adverse events when phone calls were on-time (estimated probability,  $p=0.71$  [ $SE=0.07$ ]) as opposed to late ( $p=0.31$  [ $SE=0.08$ ];  $OR=5.37$  [ $SE=2.11$ ]).

While reporting adverse events was not uncommon, the number and severity of adverse events reported were relatively low in both treatment groups across the study. Figure 2 shows the median number of adverse events reported by each

**Table 3** Number of participants in each treatment for whom adverse event reports were obtained and number (and percentage) of participants in each treatment reporting any adverse event and specific types of adverse events during each study period

Adverse event/symptom	Study period							
	Week 1		Week 2		Week 3		Follow-up	
	Oxytocin $n=47$	Placebo $n=43$	Oxytocin $n=47$	Placebo $n=41$	Oxytocin $n=44$	Placebo $n=40$	Oxytocin $n=44$	Placebo $n=37$
Any	31 (66.0%)	21 (48.8%)	27 (57.4%)	20 (48.8%)	24 (54.5%)	21 (52.5%)	23 (52.3%)	14 (37.8%)
Abdominal stomach pain	2 (4.3%)	0 (0.0%)	2 (4.3%)	0 (0.0%)	2 (4.5%)	0 (0.0%)	2 (4.5%)	0 (0.0%)
Cough with phlegm	7 (14.9%)	6 (14%)	4 (8.5%)	5 (12.2%)	5 (11.4%)	5 (12.5%)	7 (15.9%)	4 (10.8%)
Drowsiness	7 (14.9%)	7 (16.3%)	4 (8.5%)	5 (12.2%)	5 (11.4%)	6 (15%)	5 (11.4%)	2 (5.4%)
Dry throat	8 (17%)	5 (11.6%)	6 (12.8%)	2 (4.9%)	7 (15.9%)	2 (5.0%)	6 (13.6%)	2 (5.4%)
Fainting	1 (2.1%)	1 (2.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (2.5%)	1 (2.3%)	0 (0.0%)
Fever	1 (2.1%)	2 (4.7%)	0 (0.0%)	1 (2.4%)	0 (0.0%)	2 (5.0%)	0 (0.0%)	0 (0.0%)
Heart rate change or palpitations	1 (2.1%)	2 (4.7%)	2 (4.3%)	2 (4.9%)	2 (4.5%)	2 (5.0%)	1 (2.3%)	0 (0.0%)
Hoarseness	6 (12.8%)	2 (4.7%)	3 (6.4%)	2 (4.9%)	7 (15.9%)	1 (2.5%)	5 (11.4%)	1 (2.7%)
Light-headedness	4 (8.5%)	5 (11.6%)	3 (6.4%)	4 (9.8%)	3 (6.8%)	3 (7.5%)	3 (6.8%)	1 (2.7%)
Nasal irritation	3 (6.4%)	2 (4.7%)	9 (19.1%)	2 (4.9%)	4 (9.1%)	3 (7.5%)	5 (11.4%)	0 (0.0%)
Nausea	1 (2.1%)	0 (0.0%)	3 (6.4%)	0 (0.0%)	1 (2.3%)	2 (5%)	0 (0.0%)	1 (2.7%)
Runny nose	13 (27.7%)	7 (16.3%)	9 (19.1%)	12 (29.3%)	9 (20.5%)	11 (27.5%)	12 (27.3%)	7 (18.9%)
shortness of breath	2 (4.3%)	1 (2.3%)	3 (6.4%)	1 (2.4%)	1 (2.3%)	0 (0.0%)	6 (13.6%)	1 (2.7%)
Sneezing	10 (21.3%)	9 (20.9%)	10 (21.3%)	9 (22%)	6 (13.6%)	7 (17.5%)	6 (13.6%)	5 (13.5%)
Sore throat	2 (4.3%)	2 (4.7%)	2 (4.3%)	3 (7.3%)	4 (9.1%)	3 (7.5%)	3 (6.8%)	2 (5.4%)
Stuffed-up nose	13 (27.7%)	5 (11.6%)	10 (21.3%)	8 (19.5%)	8 (18.2%)	11 (27.5%)	13 (29.5%)	6 (16.2%)
Tearing of the eyes	6 (12.8%)	0 (0.0%)	8 (17.0%)	6 (14.6%)	5 (11.4%)	4 (10.0%)	7 (15.9%)	1 (2.7%)

There were no significant differences between groups in the number of any symptoms reported (top row) across study weeks; the only significant difference between groups by symptom type was in "Tearing of the eyes" during week 1. Percentages are always calculated based on the number of reports obtained in each time period, per group (i.e., not based on the total possible sample). Symptoms or the lack thereof from reports coded as "early" ( $n=3$ ) are not included in the frequencies/percentages

**Table 4** Type of “other” adverse events reported, number of times each was reported, and number of participants reporting each adverse event per treatment across the study period

Adverse event/symptom	No. of times reported		No. of participants reporting	
	Oxytocin	Placebo	Oxytocin	Placebo
Affect/well-being	<b>3</b>	<b>3</b>	<b>3</b>	<b>3</b>
Depression	0	2	0	2
Irritability—decreased	1	0	1	0
Irritability—increased	1	0	1	0
Nervousness	1	0	1	0
Increased well-being	0	1	0	1
Appetite	<b>2</b>	<b>1</b>	<b>2</b>	<b>1</b>
Decreased	2	1	2	1
Blood pressure	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>
High	1	0	1	0
Low	0	1	0	1
Cognitive	<b>1</b>	<b>4</b>	<b>1</b>	<b>2</b>
Alertness—improved	0	1	0	1
Cognition—improved	0	3	0	1
Memory—improved	1	0	1	0
Dry mouth	<b>0</b>	<b>2</b>	<b>0</b>	<b>1</b>
Energy and weakness	<b>3</b>	<b>4</b>	<b>3</b>	<b>4</b>
General fatigue	1	1	1	1
More energy	2	2	2	2
Muscle weakness	0	1	0	1
Gastrointestinal	<b>0</b>	<b>5</b>	<b>0</b>	<b>4</b>
Constipation	0	3	0	2
Diarrhea/loose stool	0	2	0	2
Headache	<b>2</b>	<b>2</b>	<b>2</b>	<b>1</b>
Lightheadedness/dizziness	<b>0</b>	<b>3</b>	<b>0</b>	<b>2</b>
Lightheadedness	0	1	0	1
Dizziness	0	2	0	1
Nasal irritation	<b>2</b>	<b>1</b>	<b>2</b>	<b>1</b>
Nostril sensitivity	0	1	0	1
Bloody nose	1	0	1	0
Sniffles	1	0	1	0
Respiratory	<b>1</b>	<b>2</b>	<b>1</b>	<b>2</b>
Chest sensitivity	0	1	0	1
Cough	1	1	1	1
Sleep	<b>3</b>	<b>4</b>	<b>2</b>	<b>3</b>
Increased sleeping	0	1	0	1
Difficulty sleeping	1	2	1	1
Dream changes	2	1	1	1
Other	<b>2</b>	<b>2</b>	<b>1</b>	<b>2</b>
Cold symptoms	2	0	1	0
Sinus infection	0	1	0	1
Swelling of toes	0	1	0	1

Includes symptoms that were coded as coming from reports obtained early. Only one of the early reports contained any “other” symptoms (headache). Bolded numbers denote sums by category

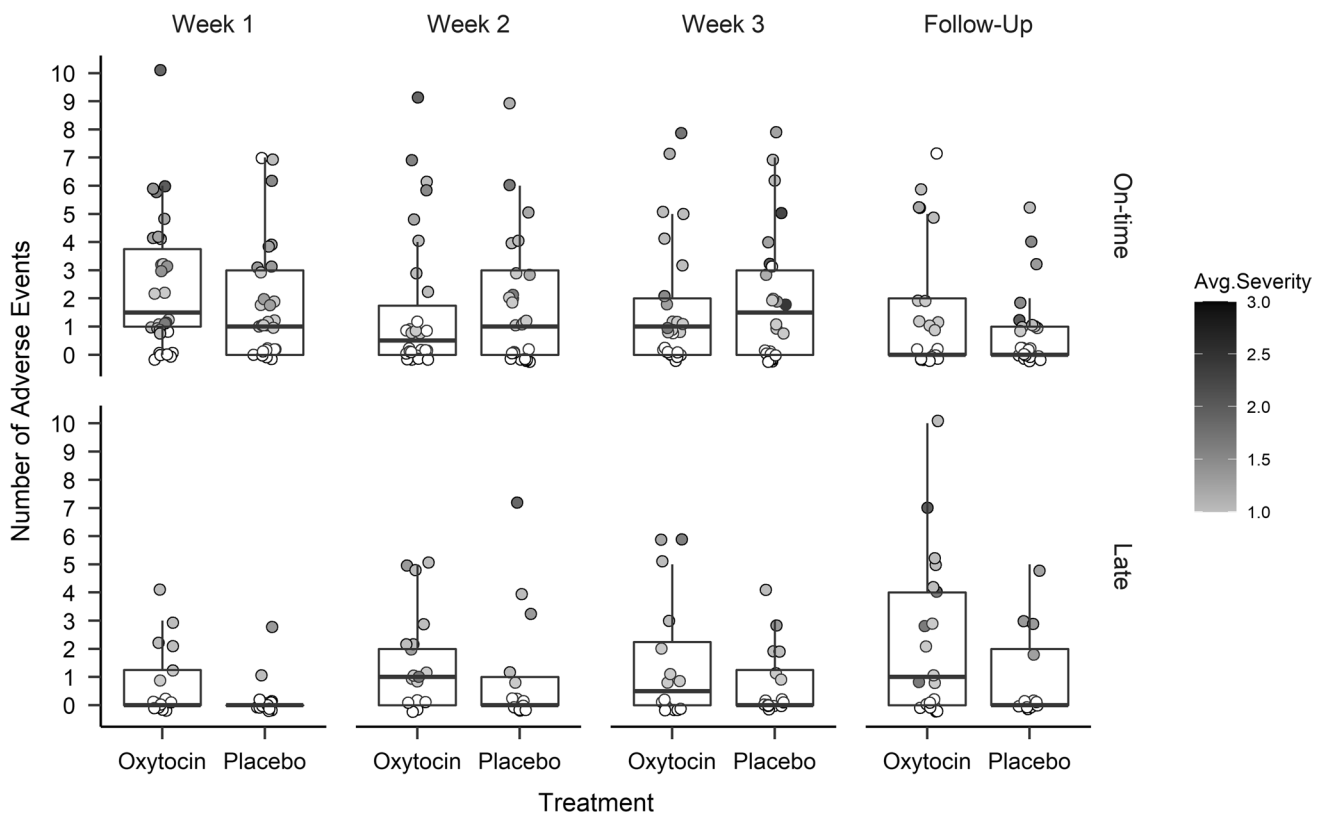
participant as a function of Treatment and week in the study, shaded according to the average severity of the reported adverse events. The median number of adverse events experienced was typically one or fewer, with an average severity of mild to moderate, indicated by the predominantly gray data points. For the analyses of both the number of adverse events and the severity-weighted adverse events, there were no significant treatment effects (see Table S4 in Online Resource 3 for test statistics and *p*-values). The numbers and severity of adverse events reported in the OT group were only nominally higher than those reported for the P group (*RRs* = 1.20 [SE = 0.25] and 1.24 [SE = 0.26], respectively). In both models, only Report Timing was a significant predictor (*ps* < 0.05; *RRs* = 1.46 [SE = 0.20] and 1.57 [SE = 0.24], for number of and severity-weighted number of adverse events, respectively). Consistent with the finding pertaining to reporting of adverse events above, reports completed on-time in both groups yielded more/more severe adverse events (1.77 [SE = 0.21]; 2.16 [SE = 0.26]) than those reported late (1.21 [SE = 0.18]; 1.38 [SE = 0.22]), across the two models.

Finally, the symptom-level analyses revealed no consistent effects of Treatment. The *p*-values for all tests are shown in Online Resource 3, Table S5. In brief, only one of the comparisons revealed a significant difference: individuals self-administering OT experienced “Tearing of the Eyes” (12.8%) more frequently than participants in the P group (0%), but only during week 1.

## Discussion

This study examined the safety and tolerability of chronically administered (24 IUs BID for 4 weeks) intranasal OT (vs. P) among generally healthy older men. Our findings corroborate and extend those of previous research indicating that intranasal OT administration is safe and not significantly associated with adverse events including in aging (MacDonald et al. 2011; Barraza et al. 2013; Finger et al. 2015). Evidenced by the low attrition during the intervention, OT was well-tolerated; and it did not increase the likelihood of reporting adverse events, nor produce significantly more adverse events in number or severity. Together, the lack of apparent differences in adverse events and objective safety measures across the two treatments broadly supports OT as safe and well-tolerated in older adults over 4 weeks of twice-daily (24 IUs) intranasal self-administration.

This research has several strengths in providing important information on the safety and tolerability of intranasal OT. First, safety and tolerability of intranasal OT were assessed over a longer period of administration and at a higher dose than in most previous studies (i.e., 4 weeks, 48 IUs daily), which is necessary for determining safe, therapeutically relevant applications of OT (Horta et al. 2020a). This study



**Fig. 2** Boxplots showing the median and interquartile range of the number of reported adverse events (circles are individual participant values) across the study by treatment. Adverse event reporting is paneled by whether the report was completed on-time (close to the end of the indicated week) or late (see “Methods” for details). The shade

of the circle indicates the average severity of the reported adverse event, with light gray corresponding to mild (value = 1.0) and black corresponding to severe (value = 3.0); white indicates no severity data were obtained (either because no adverse events were reported, or this information was missing). Circles are jittered to reduce overlap

is also the largest examination ( $N=95$ ) of safety and tolerability of repeated intranasal OT administration among older adults to date. Before this investigation, there was one study examining safety and adverse events of chronically administered intranasal OT in a sample of older individuals as a primary outcome ( $N=23$ ; Finger et al. 2015); and another trial that provided a brief overview of safety and side effects in generally healthy older adults ( $N=39$ ; Barraza et al. 2013). In both of these prior reports, intranasal OT was generally well-tolerated and posed no significant risk to safety. The results of the present study support generalizability of Finger and colleagues’ findings to healthy older male participants; and extend the conclusions of Barraza et al. (2013) to longer durations (from 10 to 28 days) and with a higher dose. The present study benefitted from a mixed design, which yielded additional experimental control in the form of repeated assessment of symptoms/adverse events, as well as a variety of objective safety measures.

While the present study was the largest examination of safety and tolerability of repeated intranasal OT administration in an older adult population yet, future extensions of this work will benefit from even larger samples to allow the study

of relevant moderators. This research is especially pertinent considering interest in social-cognitive applications for OT (e.g., treatment of depression and anxiety disorders) (Macdonald and Feifel, 2013). For example, intranasal OT administration can impact social-cognitive processes differently by sex (Bartz et al. 2019). To comprehensively conclude that intranasal OT is a safe and viable intervention in aging, it is necessary to conduct more chronic administration studies that consider moderators (e.g., sex) and their interplay with conditions of clinical interest, including in aging. Investigation of phenomena in both sexes is undeniably important, but inclusion of both men and women in relatively small- $N$  studies could inadvertently mask important findings should they be sex-dimorphic.

Future extensions of this work will also benefit from more standardized approaches of assessing adverse events and objective safety of intranasal OT (see DeMayo et al. 2017 for guidelines in pediatric populations that can be carried over to other populations of interest). Namely, obtaining adverse events reports from participants in a consistent, timely manner, and preferably at the originally scheduled times, is critical as delayed follow-ups

may introduce bias. Coding the timeliness of adverse event reports and including their timeliness in analysis for the present study revealed this to be a key factor—and the most robust predictor of adverse event occurrence and frequency. Gauging adverse effects through text messaging systems or other forms of momentary assessment may help ameliorate delays in collecting reports, although the appropriateness of these methods based on the population under investigation must be considered.

While our study has several strengths and provides crucial data on the safety of intranasal OT, some limitations should be noted. For example, the present study did not systematically assess participants' tendency for reporting adverse events at baseline nor their perceived drug allocation. This information would have been beneficial to control for and assess its influence on adverse event reporting. While no significant treatment differences in adverse events or their reporting were found herein, these could be important factors that moderate individuals' experience of the drug and should be accounted for in future studies; this is especially relevant given that OT can increase focus on feelings (Ebner et al. 2015). Unfortunately, we were unable to address the role of reporting tendency and/or perceived drug allocation on participants' adverse event reporting in the present data. Future research should also consider including measures of trust in study staff. Although the robustness of such effects has been questioned (e.g., Nave et al. 2015; cf. Van IJzendoorn and Bakermans-Kranenburg 2012), OT could influence interpersonal trust, which may impact the willingness to report adverse events. This seems like an unlikely possibility in the present study given the lack of differentiation in prevalence of reporting across OT and P groups and because rapport was equally established between participants and study staff over multiple sessions of assessment. Regardless, such influences would be worthwhile for direct investigation in the future.

Furthermore, obtaining consistent, objective measures of self-administration adherence would have been ideal to include as a covariate in the analyses. While we were able to obtain multiple measures of adherence, our data for these were incomplete and therefore could not be considered in the mixed-effects models in the present study (as observations cannot be missing on predictor variables). What we were able to obtain for intervention adherence suggested that adherence was high and did not differ across treatment groups; as such, these data suggest that our conclusions about safety and tolerability are reflective of the dosing protocol as it was designed; and that our conclusions are unlikely to be confounded by differential adherence across treatments. However, because a sizeable amount of these data were missing, we could not evaluate other possibilities that would be relevant to consider, such as symptoms

being more likely among those with better self-administration adherence (or among those who slightly exceeded the amounts they were instructed to administer). For such purposes, adherence could be included as a time-varying (e.g., if amounts of spray used per week were recorded) or time-invariant (e.g., total amount of spray used) covariate in models. Future work should ensure these measures are consistently obtained for all study participants. Incentivizing spray bottle return and nasal spray log completion may assist in such efforts. Perhaps in the future there may also be equipment to facilitate real-time data collection on nasal spray usage (e.g., time and dosage of administration), similar to the Medication Event Monitoring System (MEMS®) devices. Unfortunately, at present, MEMS® devices are only available for drugs administered in pill and injectable forms.

Another limitation of the present research was that the adverse event checklist primarily focused on potential physical symptoms (e.g., headache, nasal irritation). While a sizeable proportion of participants reported "other" symptoms (and with great variability), inclusion of an array of adverse events and symptom types is important. Current work in our lab (clinical trial #NCT03878589; Understanding Cognition, Oxytocin, and Pain in Elders) now uses a more comprehensive checklist that includes potential physical, social (e.g., sexual behavior), and emotional (e.g., emotional experiences, mood) symptoms. DeMayo et al. (2017) also make suggestions on materials for standardizing adverse event assessments in young populations, including using the Safety Monitoring Uniform Reporting Form by Greenhill et al. (2004). We propose that similar measures should be adapted for assessing adverse events in older populations with special consideration of the greater vulnerability imposed by poly-pharmacological use in this age group (Brahma et al. 2013).

Finally, given that some potential clinical applications of OT have been demonstrated in preclinical models (e.g., reducing abuse liability and relapse of opioid self-administration in non-humans; Zanos et al. 2014; Fan et al. 2019), it will be important to consider exploring self-reported and objective safety and tolerability measures with doses comparable to those used in non-humans and to test the dosing limits of chronic intranasal OT. Overall, prior research indicates that OT has limited abuse potential, though a small percentage of participants in prior research have reported that OT produces feelings of calmness and euphoria (MacDonald et al. 2011) and reinforcing effects (Dolan et al. 2020). While positive reactions were not explicitly included in our checklist assessment, very few "other" effects reported pertained to subjective effects that might be of concern from an abuse liability standpoint (e.g., no

reports of “feeling high”; see Table 4), further supporting that OT is unlikely addictive.

## Conclusions and implications

The findings from the present study demonstrate the safety and tolerability of chronically administered intranasal OT in generally healthy older men. Participants who self-administered chronic OT for 4 weeks were no more likely to report adverse events, nor did they experience a greater number (or severity) of adverse events than those self-administering placebo. Considering the duration and number of participants enrolled in this clinical trial, there were few IRB-reportable adverse events, none was considered serious, and few led to study withdrawal—supporting that chronic OT at the dose administered herein, the frequency, and duration is well-tolerated in generally healthy older men. Importantly, those receiving chronic OT also did not exhibit clinically significant differences in any objective measures of safety (e.g., cardiovascular and physiological functioning). Thus, our findings support pursuit of an emerging line of research on OT’s potential beneficial effects in aging. More research is needed on how chronic intranasal OT may impact individuals in aging. Therefore, future studies should extend this research on the safety and tolerability of chronic intranasal OT to females, older adult clinical populations, and other age cohorts under systematic variation of treatment schemes.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00213-021-05862-3>.

**Acknowledgements** The authors thank Devon Weir and the Social-Cognitive and Affective Development lab and the Institute on Aging research teams at the University of Florida for help with study implementation, data collection, and management.

**Funding** While working on this manuscript, NCE was in part supported by the National Institutes of Health (NIH) funded Claude D. Pepper Older Americans Independence Center (P30AG028740); MH was supported by the National Institute on Aging (NIA, T32AG020499); and JR and MH were supported by the UF Substance Abuse Training Center in Public Health (National Institute on Drug Abuse, T32DA035167). Data collection was supported by the Departments of Psychology (College of Liberal Arts and Sciences) and Aging & Geriatric Research (College of Medicine); the Claude D. Pepper Older Americans Independence Center (P30AG028740); the Center for Cognitive Aging & Memory, PRICE-CTSI-IOA ARG DTD 03–26–2008; the Evelyn F. & William L. McKnight Brain Institute; NIA R01AG059809; Aptar Pharma (for spray pumps); and SGD Pharma (for spray bottles). The content of this paper is solely the responsibility of the author(s) and does not necessarily represent the official views of the NIH.

## Declarations

**Conflict of interest** The authors declare no competing interests.

## References

- Barraza JA, Grewal NS, Ropacki S et al (2013) Effects of a 10-day oxytocin trial in older adults on health and well-being. *Exp Clin Psychopharmacol* 21:85–92. <https://doi.org/10.1037/a0031581>
- Bates D, Mächler M, Bolker BM, Walker S (2015) Fitting linear mixed-effects models using lme4. *J Stat Softw* 67:1–48. <https://doi.org/10.18637/jss.v067.i01>
- Born J, Lange T, Kern W et al (2002) Sniffing neuropeptides: a transnasal approach to the human brain. *Nat Neurosci* 5:514–516. <https://doi.org/10.1038/nn0602-849>
- Brahma DK, Wahlang JB, Marak MD, Sangma MC (2013) Adverse drug reactions in the elderly. *J Pharmacol Pharmacother* 4:91–94
- Brandt J, Spencer M, Folstein M (1988) The telephone interview for cognitive status. *Neuropsychiatry, Neuropsychol Behav Neurol* 1:111–117
- Brooks ME, Kristensen K, van Benthem KJ, et al (2017) glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R J* 9:378–400. <https://doi.org/https://doi.org/10.32614/rj-2017-066>
- Brybaert M, Stevens M (2018) Power analysis and effect size in mixed effects models: a tutorial. *J Cogn*. <https://doi.org/10.5334/joc.10>
- Cai Q, Feng L, Yap KZ (2018) Systematic review and meta-analysis of reported adverse events of long-term intranasal oxytocin treatment for autism spectrum disorder. *Psychiatry Clin Neurosci* 72:140–151. <https://doi.org/10.1111/pcn.12627>
- Campbell A, Ruffman T, Murray JE, Glue P (2014) Oxytocin improves emotion recognition for older males. *Neurobiol Aging* 35:2246–2248. <https://doi.org/10.1016/j.neurobiolaging.2014.04.021>
- Carter CS (2003) Developmental consequences of oxytocin. *Physiol Behav* 79:383–397. [https://doi.org/10.1016/S0031-9384\(03\)00151-3](https://doi.org/10.1016/S0031-9384(03)00151-3)
- DeMayo MM, Song YJC, Hickie IB, Guastella AJ (2017) A review of the safety, efficacy and mechanisms of delivery of nasal oxytocin in children: therapeutic potential for autism and Prader-Willi syndrome, and recommendations for future research. *Pediatr Drugs* 19:391–410. <https://doi.org/10.1007/s40272-017-0248-y>
- Di Simplicio M, Harmer CJ (2016) Oxytocin and emotion processing. *J Psychopharmacol* 30:1156–1159. <https://doi.org/10.1177/026981116641872>
- Dolan SB, Berry MS, Johnson PS, Johnson MW (2020) Potential for limited reinforcing and abuse-related subjective effects of intranasal oxytocin. *J Psychopharmacol*. <https://doi.org/10.1177/026981119867607>
- Ebner NC, Horta M, Lin T et al (2015) Oxytocin modulates meta-mood as a function of age and sex. *Front Aging Neurosci* 7:175. <https://doi.org/10.3389/fnagi.2015.00175>
- Ebner NC, Kamin H, Diaz V et al (2014) Hormones as “difference makers” in cognitive and socioemotional aging processes. *Front Psychol* 5:1595
- Fan XY, Shi G, Zhao P (2019) Reversal of oxycodone conditioned place preference by oxytocin: promoting global DNA methylation in the hippocampus. *Neuropharmacology* 160:107778. <https://doi.org/10.1016/j.neuropharm.2019.107778>
- Finger EC, MacKinley J, Blair M et al (2015) Oxytocin for frontotemporal dementia: a randomized dose-finding study of safety and tolerability. *Neurology* 84:174–181. <https://doi.org/10.1212/WNL.0000000000001133>
- Fox J, Weisberg S (2019) An {R} Companion to Applied Regression, Third. Sage, Thousand Oaks, CA
- Gimpl G, Fahrenholz F (2001) The oxytocin receptor system: structure, function, and regulation. *Physiol Rev* 81:629–683

- Grainger SA, Henry JD, Steinvik HR et al (2018) Intranasal oxytocin does not reduce age-related difficulties in social cognition. *Horm Behav* 99:25–34. <https://doi.org/10.1016/j.yhbeh.2018.01.009>
- Grainger SA, Henry JD, Steinvik HR, Vanman EJ (2019) Intranasal oxytocin does not alter initial perceptions of facial trustworthiness in younger or older adults. *J Psychopharmacol* 33:250–254. <https://doi.org/10.1177/0269881118806303>
- Greenhill LL, Vitiello B, Fisher P et al (2004) Comparison of increasingly detailed elicitation methods for the assessment of adverse events in pediatric psychopharmacology. *J Am Acad Child Adolesc Psychiatry* 43:1488–1496. <https://doi.org/10.1097/01.chi.0000142668.29191.13>
- Guastella AJ, MacLeod C, Graustella AJ, MacLeod C (2012) A critical review of the influence of oxytocin nasal spray on social cognition in humans: evidence and future directions. *Horm Behav* 61:410–418. <https://doi.org/10.1016/j.yhbeh.2012.01.002>
- Hartig F (2020) DHARMA: Residual diagnostics for hierarchical (multi-level/mixed) regression models. <https://CRAN.R-project.org/package=DHARMA>
- Horta M, Kaylor K, Feifel D, Ebner NC (2020a) Chronic oxytocin administration as a tool for investigation and treatment: a cross-disciplinary, systematic review. *Neurosci Biobehav Rev* 108:1–23. <https://doi.org/10.1016/j.neubiorev.2019.10.012>
- Horta M, Pehlivanoglu D, Ebner NC (2020b) The role of intranasal oxytocin on social cognition: an integrative human lifespan approach. *Curr Behav Neurosci Reports* 7:175–192. <https://doi.org/10.1007/s40473-020-00214-5>
- Horta M, Ziaei M, Lin T et al (2019) Oxytocin alters patterns of brain activity and amygdalar connectivity by age during dynamic facial emotion identification. *Neurobiol Aging* 78:42–51. <https://doi.org/10.1016/j.neurobiolaging.2019.01.016>
- Huffmeijer R, van IJzendoorn MH, Bakermans-Kranenburg MJ (2013) Ageing and oxytocin: a call for extending human oxytocin research to ageing populations—a mini-review. *Gerontology* 59:32–39. <https://doi.org/10.1159/000341333>
- Koller M (2016) Robustlmm: an R package for robust estimation of linear mixed-effects models. *J Stat Softw* 75:1–24. <https://doi.org/10.18637/jss.v075.i06>
- Kuznetsova A, Brockhoff PB, Christensen RHB (2017) lmerTest Package: tests in linear mixed effects models. *J Stat Softw* 82:1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lefevre A, Sirigu A (2016) The two fold role of oxytocin in social developmental disorders: a cause and a remedy? *Neurosci Biobehav Rev* 63:168–176. <https://doi.org/10.1016/j.neubiorev.2016.01.011>
- Leng G, Onaka T, Caqueneau C et al (2008) Oxytocin and appetite. *Prog Brain Res* 170:137–151. [https://doi.org/10.1016/S0079-6123\(08\)00413-5](https://doi.org/10.1016/S0079-6123(08)00413-5)
- Lenth R (2020) emmeans: estimated marginal means, aka least-squares means. <https://CRAN.R-project.org/package=emmeans>
- Lussier D, Cruz-Almeida Y, Ebner NC (2019) Musculoskeletal pain and brain morphology: oxytocin's potential as a treatment for chronic pain in aging. *Front Aging Neurosci* 11:338
- MacDonald E, Dadds MR, Brennan JL et al (2011) A review of safety, side-effects and subjective reactions to intranasal oxytocin in human research. *Psychoneuroendocrinology* 36:1114–1126
- MacDonald K, Feifel D (2013) Helping oxytocin deliver: considerations in the development of oxytocin-based therapeutics for brain disorders. *Front Neurosci* 7:35. <https://doi.org/10.3389/fnins.2013.00035>
- MacDonald K, Feifel D (2012) Oxytocin in schizophrenia: a review of evidence for its therapeutic effects. *Acta Neuropsychiatr* 24:130–146. <https://doi.org/10.1111/j.1601-5215.2011.00634.x>
- Martins DA, Mazibuko N, Zelaya F et al (2020) Effects of route of administration on oxytocin-induced changes in regional cerebral blood flow in humans. *Nat Commun* 11:1–16. <https://doi.org/10.1038/s41467-020-14845-5>
- Nave G, Camerer C, McCullough M (2015) Does oxytocin increase trust in humans? A critical review of research. *Perspect Psychol Sci* 10:772–789. <https://doi.org/10.1177/1745691615600138>
- Neumann ID (2007) Stimuli and consequences of dendritic release of oxytocin within the brain. *Biochem Soc Trans* 35:1252–1257. <https://doi.org/10.1042/BST0351252>
- Quintana DS, Alvares GA, Hickie IB, Guastella AJ (2015) Do delivery routes of intranasally administered oxytocin account for observed effects on social cognition and behavior? A two-level model. *Neurosci Biobehav Rev* 49:182–192
- R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Rash JA, Aguirre-Camacho A, Campbell TS (2014) Oxytocin and pain: a systematic review and synthesis of findings. *Clin J Pain* 00:1–10. <https://doi.org/10.1097/AJP.0b013e31829f57df>
- Rung JM, Horta M, Ebner NC (2021) Side effects of chronic OT in older men: data and analyses. <https://doi.org/10.17605/OSF.IO/PNX9U>
- Russell JA, Leng G (1998) Sex, parturition and motherhood without oxytocin? *J Endocrinol* 157:343–359
- Sarnyai Z, Kovács GL (1994) Role of oxytocin in the neuroadaptation to drugs of abuse. *Psychoneuroendocrinology* 19:85–117
- Van IJzendoorn MH, Bakermans-Kranenburg MJ (2012) A sniff of trust: meta-analysis of the effects of intranasal oxytocin administration on face recognition, trust to in-group, and trust to out-group. *Psychoneuroendocrinology* 37:438–443. <https://doi.org/10.1016/j.psyneuen.2011.07.008>
- Weintraub S, Dikmen SS, Heaton RK et al (2013) Cognition assessment using the NIH Toolbox. *Neurology* 80:S54–S64. <https://doi.org/10.1212/WNL.0b013e3182872ded>
- Wickham H (2016) ggplot2: elegant graphics for data analysis. Springer-Verlag, New York
- Yamasue H (2016) Promising evidence and remaining issues regarding the clinical application of oxytocin in autism spectrum disorders. *Psychiatry Clin Neurosci* 70:89–99. <https://doi.org/10.1111/pcn.12364>
- Zanos P, Georgiou P, Wright SR et al (2014) The oxytocin analogue carbetocin prevents emotional impairment and stress-induced reinstatement of opioid-seeking in morphine-abstinent mice. *Neuropsychopharmacology* 39:855–865. <https://doi.org/10.1038/npp.2013.285>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.