

Mechanism-Focused Randomized Controlled Trials in Youths: Another Step Uphill

Argyris Stringaris, M.D., Ph.D., Jamilah Silver, M.A.

In this issue, Hwang et al. (1) report on a randomized controlled trial (RCT) of intranasal oxytocin for young people (mean age around 14 years) with severe irritability. For clinicians, RCTs like that of Hwang et al. hold the important promise of direct clinical translation. RCTs are the best way to estimate the causal effect of a treatment on a given outcome. In the best-case scenario, randomization balances (and therefore neutralizes) confounders, and the presence of an adequate and blinded control condition allows one to estimate the effect of the intervention net of biases such as expectation. The Hwang et al. study is appealing for an additional reason in that it tests a fairly clearly articulated mechanistic hypothesis. The authors' idea is that oxytocin may be beneficial to children with severe irritability by changing the way they process emotions, as measured using functional MRI (fMRI). Combining the rigor of RCTs with a mechanistic hypothesis could catalyze the much-needed change in our field and facilitate early, rational interventions.

Oxytocin is a peptide, a chain of nine amino acids, produced in the hypothalamus and stored in the posterior pituitary; it acts as a neuropeptide on other neurons and as a hormone on peripheral organs, such as the uterus and the mammary glands. While much talked about in neuroscience, oxytocin is mostly used in obstetrics and gynecology, at least since Blair-Bell reported in the early 20th century that "an extract of the pituitary, pituitrin" could be used to induce labor by enhancing uterine contractions (2). In recent decades, oxytocin has acquired a special status in public perception, being described as the "love hormone," as a result of its reported positive effects on affiliative behaviors ranging from parent-child bonding to sexual intercourse and overall social interactions (3).

Inevitably, this has attracted interest from psychiatrists: If oxytocin has salutary effects on social interactions, could it be used to treat disorders in which such interactions are impaired? Answers to this question are rather mixed. Let's take autism spectrum disorder (ASD) as an example, in which social interactions are aberrant. There is meta-analytic evidence of lower endogenous oxytocin levels in children (but not adults) with ASD compared with those without (4). However, when oxytocin was administered intranasally to young people with ASD (ages 3–17 years) in a reasonably powered randomized controlled trial (5), it did not differentiate from placebo on a range of outcomes, including core

ASD symptoms and social and cognitive functioning. Similarly mixed are the findings in antisocial personality disorder, where a review of (mainly smaller) studies (6) showed that oxytocin may enhance or decrease prosocial behaviors, partly depending on context.

Given all this, was it a good idea for the authors to try to assess the role of oxytocin in irritability? The answer is a resounding yes. Never mind the mixed results so far; oxytocin has plausible actions on systems that are very relevant to health and well-being, most notably empathy (7), and Hwang et al. decided to test whether it might affect irritability, a phenotype for which aberrant social-emotion processing may play a central role.

Irritability, defined as interindividual differences in proneness to anger, has received increasing attention from neuroscience and psychopathologists in recent years. A review article, also in this issue of the *Journal* (on which we are among the coauthors) (8), documents its role as a precursor (and genetic relative) of depression and anxiety, its role in adverse social outcomes, and its neurobiology, which is partly rooted in frustrative non-reward. Yet, despite its public health implications, irritability has received little attention (9) as a treatment target (outside the use of antipsychotics in ASD).

Hwang et al. report on a parallel-group double-blind RCT of youths with irritability (as defined by a score ≥ 4 on the self-report version of the Affective Reactivity Index [ARI]) and diagnoses of disruptive behavior disorders or disruptive mood dysregulation disorder who were randomized to daily treatment with either intranasal oxytocin or placebo for 3 weeks. The authors report on the clinical outcomes (irritability) and imaging outcomes.

There were two preregistered main outcomes: the Clinical Global Impressions severity scale (CGI-S) score and change in the ARI score. The authors found statistical evidence for improvement in CGI-S score, but there was no statistically significant change in ARI score. In terms of secondary outcomes, the preregistered outcome was less

Bringing to bear an experimental medicine approach to pressing issues in mental health is important, particularly when testing well-circumscribed mechanistic hypotheses, as the authors do here.

specific, a change in fMRI brain signal. The authors analyzed their data using region-of-interest approaches and found that neural responses to emotional stimuli were reduced in areas of the posterior cingulate and dorsolateral prefrontal cortex in the oxytocin group compared with the placebo group. They also found that this change in fMRI signal was associated with changes in clinical measures.

There are several admirable aspects to this study. First, as noted above, bringing to bear an experimental medicine approach to pressing issues in mental health is important, particularly when testing well-circumscribed mechanistic hypotheses, as the authors do here. Second, anyone who has conducted an RCT knows how difficult it is to recruit and complete such studies, particularly when they involve young people. This difficulty, combined with the chronic underfunding of youth mental health, compounds the scarce evidence base for treatments in youth irritability. The authors should be commended for their efforts. Third, the authors combine clinical outcomes (their primary endpoints) with neurobiological measurement, a practice that can potentially advance both clinical and mechanistic knowledge.

What did they find and how confident can we be in the study results? The authors wisely state that their results are “initial and preliminary,” and this is indeed how they should be understood. This is first and foremost because of the sample size ($N=25$ and $N=27$ for oxytocin and placebo groups, respectively), which is relatively small (as noted above, recruitment always poses a problem, and the authors also report that several participants in each group withdrew from the study prematurely). Small sample sizes can lead to both over- and underestimation of the magnitude of effects, termed type I and type II errors, respectively, and can even mislead in terms of sign (direction of effect) (10). The dissociation between the ARI and CGI results (and the null results in a range of secondary outcomes) are yet another important reason to seek replication. The study is particularly underpowered for higher-order analyses, such as those involving interactions in the fMRI data. The next reason to be cautious in the interpretation of the results is that the number of possible imaging analyses is vast and that it was not constrained a priori (at the point of preregistration) by the authors. Their results build on plausible previous results by their research group, but the possibility of a chance finding and/or overfitting should not be underestimated. Third, the cutoff of an ARI score ≥ 4 for study entry is quite high and restricts inferences to a group of quite irritable children.

The study results provide important lessons for all of us in terms of choosing mechanistic targets and testing them in rigorous designs, such as the RCT. What should be the next steps following this study? In our opinion, there are a number of systemic barriers that need to be overcome to improve rigorous mechanism-focused research in young populations. The first is the perennial problem of recruitment into studies. What changes need to be made in the ways we encourage and support the recruitment of young people into research while

upholding the highest standards of safety and ethics? Responding to this problem will require us to revisit questions about how society and lawmakers view young people's capacity to consent to interventional studies. It will also require that researchers and their institutions help build the necessary public trust for people to take part in medical research—trust being a central issue in public health, as was demonstrated during the COVID pandemic (11). Understanding the barriers for youths in research participation, including matters of inequitable distribution of information or resources, will be paramount.

Second, multisite collaborations are likely needed for intervention research, yet most of them are formed ad hoc (i.e., in response to a particular grant call). Instead, we should consider building consortia for intervention research in the way we have for observational research (e.g., the Enhancing Neuroimaging Genetics Through Meta-Analysis [ENIGMA] Consortium); these can be response ready and provide high-quality training and quality assurance (the latter being an issue that has often been a topic of discussion for large multisite pharmacological trials, particularly in youths). Third, issues of measurement should be given more attention and funding. Methods such as fMRI, used in this study, are expensive and hard to scale; would the cingulate signals in this study be detectable by EEG, which is cheaper and more scalable? Moreover, we have learned that cognitive tasks that show outstanding replicability can be problematic when it comes to detecting interindividual differences (12); this too is a matter that needs further research effort. Finally, intervention research should be additionally incentivized. Current research metrics (e.g., citation rates) used for promotion and scientific recognition do not weigh the effort and persistence required to carry out interventional research in young people; it should therefore not come as a surprise if early-career researchers choose to avoid such a tough field of study.

AUTHOR AND ARTICLE INFORMATION

Division of Psychiatry (Stringaris) and Division Psychology and Language Sciences (Stringaris, Silver), Faculty of Brain Sciences, University College London; First Department of Psychiatry, National and Kapodistrian University of Athens, Athens, Greece (Stringaris); Department of Psychology, Stony Brook University, Stony Brook, N.Y. (Silver).

Send correspondence to Dr. Stringaris (a.stringaris@ucl.ac.uk).

The authors report no financial relationships with commercial interests.

Accepted February 9, 2024.

Am J Psychiatry 2024; 181:259–261; doi: 10.1176/appi.ajp.20240123

REFERENCES

1. Hwang S, Suk J-W, Meffert H, et al: Neural responses to intranasal oxytocin in youths with severe irritability. *Am J Psychiatry* 2024; 181:291–298
2. “Pituitrin”: landmarks in its history. *Br Med J* 1925; 3365:1
3. Froemke RC, Young LJ: Oxytocin, neural plasticity, and social behavior. *Annu Rev Neurosci* 2021; 44:359–381
4. John S, Jaeggi AV: Oxytocin levels tend to be lower in autistic children: a meta-analysis of 31 studies. *Autism* 2021; 25:2152–2161

5. Sikich L, Kolevzon A, King BH, et al: Intranasal oxytocin in children and adolescents with autism spectrum disorder. *N Engl J Med* 2021; 385:1462–1473
6. Gedeon T, Parry J, Völlm B: The role of oxytocin in antisocial personality disorders: a systematic review of the literature. *Front Psychiatry* 2019; 10:76
7. Fragkaki I, Cima M: The effect of oxytocin administration on empathy and emotion recognition in residential youth: a randomized, within-subjects trial. *Horm Behav* 2019; 114:104561
8. Leibenluft E, Allen LE, Althoff RR, et al: Irritability in youths: a critical integrative review. *Am J Psychiatry* 2024; 181: 275–290
9. Towbin K, Vidal-Ribas P, Brotman MA, et al: A double-blind randomized placebo-controlled trial of citalopram adjunctive to stimulant medication in youth with chronic severe irritability. *J Am Acad Child Adolesc Psychiatry* 2020; 59:350–361
10. Gelman A, Carlin J: Beyond power calculations: assessing type S (sign) and type M (magnitude) errors. *Perspect Psychol Sci* 2014; 9:641–651
11. Algan Y, Cohen D, Davoine E, et al: Trust in scientists in times of pandemic: panel evidence from 12 countries. *Proc Natl Acad Sci U S A* 2021; 118:e2108576118
12. Hedge C, Powell G, Sumner P: The reliability paradox: why robust cognitive tasks do not produce reliable individual differences. *Behav Res Methods* 2018; 50:1166–1186