COMMENT OPEN

Check for updates

There was no call for immediate implementation of "Tetris" in clinical practice: Response to the commentary by Halvorsen et al. (2024)

Camille Deforges[1], Yvonnick Noël[2], Susan Ayers[3], Emily A. Holmes[4], Vania Sandoz[5], Valérie Avignon[1,5], David Desseauve[6], Julie Bourdin[7], Manuella Epiney[8] and Antje Horsch[1,5]✉

Molecular Psychiatry; https://doi.org/10.1038/s41380-024-02766-4

We thank Halvorsen and colleagues for their commentary [1] on the Swiss TrAumatic biRth Trial (START) [2], a multicenter, double-blind, randomized controlled trial (RCT) with an active control group. START tested the efficacy of a single-session intervention, carried out within six hours following an unplanned cesarean section, to prevent maternal symptoms of childbirth-related posttraumatic stress disorder (CB-PTSD). It was a next step in our research program, in which a previous proof-of-principle RCT showed that women receiving this intervention had fewer intrusive childbirth-related memories during the first postpartum week than those receiving routine care [3]. At the end of our publication [2], we concluded that the next step would be an implementation study. However, in stark contrast to what Halvorsen et al. claim, this was not a call for immediate implementation of the intervention in clinical practice. Halvorsen et al.'s title claim is thus incorrect and misleading.

Halvorsen et al. rightly note that we did not operationalize the primary outcomes, as preregistered, by analyzing means, but rather by counting the symptoms present (scored ≥2, in accordance with validated ratings [4, 5]). Primary outcomes were group differences in the presence and severity of maternal CB-PTSD symptoms at six weeks postpartum on PCL-5[1] and CAPS-5[2] subscale and total scores. These primary outcome measures remain the same in the trial pre-registration (NCT03576586), study protocol [6] and the paper [2]. The way we analyzed them as symptoms counts is detailed in the paper, and we agree we should also have explicitly written this differed from the pre-registration. There was a strong statistical argument for using this different type of analysis. We analyzed PCL-5 and CAPS-5 (ordinal scales) using dedicated categorical models from Item Response Theory (IRT) [7]. Among IRT models, only Rasch models, when validated using suitable fit statistics [8], result in a global measure of severity from nominal presence of symptoms, or an ordinal assessment of their severity [e.g., 9, 10], that is reducible to a summed score. In START, routine IRT analyses showed that none of these models had an acceptable fit to the data (the RMSEA fit statistic ranged from 0.07 to 0.0943), and that a zero-inflation effect was present in the score data. The first result disqualified classical summed scores as a valid participant descriptor, while the second implied the rejection of means and standard deviations as valid group summaries, hence the deviation from the pre-registration. We therefore switched to symptom counts as a more appropriate measure of presence and severity of CB-PTSD symptoms. It would have been misleading to use invalid indicators, simply because they had been preregistered. Furthermore, we agree with Halvorsen et al. on the value of the recent CONSORT Outcome Extension [11], but it was not published at the time we finalized our manuscript. Halvorsen et al. incorrectly state that we did not respond to their request for information, but we replied to them well before we saw their commentary and our offer to meet went unanswered.

Halvorsen et al. describe our primary outcome results as "non-significant" and criticize our conclusion that the intervention had been beneficial. However, they seem to have misunderstood how a statistical approach based on Akaike Information Criterion (AIC) [12] allows to conclude in favor of an intervention effect. For each outcome variable, presented as a symptom count, we needed to infer i) a proper distributional model, ii) the presence of a potential zero-inflation effect, and iii) the existence of a group effect - either on the zero or non-zero part of the distribution. Given that distribution comparisons cannot be obtained from standard tests, we used a model comparison approach using information criteria, allowing us to test for all three aspects in a unique decisional procedure. The inferential decision criterion was the AIC: an intervention effect is statistically validated if the inclusion of the group variable, be it on the zero-inflation or non-zero part of the model, translated into a *diminished* AIC[3]. But while a reduced AIC indicates an intervention effect, it does not indicate the direction of the effect (i.e., if

[1]PTSD Checklist for DSM-5
[2]Clinician-Administered PTSD Scale for DSM-5

[3]Importantly, as a measure of the expected distance (in the Kullback-Leibler sense) to the true structure of the data [12], the computation of this distance for one model does not impact the same computation for another model on the same data. Thus, multiplicity corrections are inappropriate here.

[1]Institute of Higher Education and Research in Healthcare, University of Lausanne, Lausanne, Vaud, Switzerland. [2]Department of Psychology, Rennes 2 University, Rennes, France. [3]Centre for Maternal and Child Health Research, City, University of London, London, UK. [4]Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden. [5]Department Woman-Mother-Child, Lausanne University Hospital, Lausanne, Vaud, Switzerland. [6]Department of Obstetrics, Grenoble Alpes University Hospital, Grenoble, France. [7]Institute of Pedagogy and Applied Research, Limésy, France. [8]Department of Woman, Child and Teenager, Geneva University Hospitals, Geneva, Switzerland.
✉email: antje.horsch@chuv.ch

Received: 17 August 2024 Revised: 19 September 2024 Accepted: 23 September 2024
Published online: 04 October 2024

beneficial or negative). In the second stage, we therefore examined the sign of each coefficient individually (a negative β corresponds to *reduced* symptoms). While we reported their *p*-values, this was not necessary: as long as the AIC is lower, one can conclude a group effect. Note that, we chose to only comment on those effects for which *both* a lower AIC *and* a significant effect *p*-value were found, which constitutes an unusually conservative approach, contrary to what Halvorsen et al. suggest.

Thus, based on AIC and model coefficients, we reiterate that the intervention reduced the number of self-reported symptoms of CB-PTSD at both six weeks (total number of symptoms; intrusions and arousal subscales but not avoidance or negative alteration in cognitions and mood subscales) and six months postpartum (total number of symptoms; negative alterations in cognition and mood scale, as well as arousal subscales but not intrusions and avoidance subscales).

Halvorsen et al. expressed concern over missing outcome cases. Moving to symptom counts requires discarding participants with missing responses for fair between-participants and between-group comparisons. We note that >95% of cases were complete (only 6/128 cases at six weeks, and 5/113 at six months were discarded) as indicated in the paper [2]. Halvorsen et al. also wished for more information on participants who dropped out. This would indeed have been useful, but our ethical approval did not allow us to contact participants who dropped out. Nevertheless, the drop-out rate of 7.5% between randomization and primary outcome is well below the expected 20% [6].

Our paper sought to report results transparently, including null findings. For example, we stated that null findings on CAPS-5 and the intrusion diary were contrary to expectations (p.3847). We refrained from drawing mechanistic conclusions, which this type of clinical trial does not allow. In the discussion, we twice urged caution against overinterpreting results, e.g., "*the absence of group differences in clinical interviews warrants caution in interpreting the effects of the intervention*" and "*our results cautiously confirm its efficacy in the secondary prevention of CB-PTSD symptom development*".

We are pleased Halvorsen et al. acknowledge the strengths of this study: adequate randomization and efforts to maintain blinding. START was designed with a multidisciplinary steering committee of international experts, who oversaw data collection and approved the independent statisticians' advice to conduct primary outcome analyses by symptom count (7/6/2022). The intervention effect was assessed using both self- and clinician-reported validated measures up to six months. START had regular oversight by an independent trial monitor. Analyses were conducted by an independent statistician blinded to group allocation. The sample size was calculated based on the effect sizes found in our previous RCT [3], with power calculations carried out by an independent statistician and approved by the ethics committee[4].

In summary, we sought to conduct START with care and rigor from planning to publication. Participants who received the intervention developed fewer self-reported CB-PTSD symptoms than those in the active control group, for up to six months post-intervention, in accordance with AIC statistics. This pattern of results constitutes a meaningful step forward, given the unmet need for interventions for mothers after traumatic childbirth and that the intervention is acceptable, requiring few resources. We readily acknowledge that START has several limitations, as highlighted [2], but we do not see that these justify the criticism of being "*actively misleading*". Halvorsen et al.s' claim that we made a "*premature call for implementation of Tetris in clinical practice*" is incorrect, since nowhere in our paper did we call for the intervention to be immediately used in clinical practice. Rather, we wrote that "*Future research may thus evaluate its implementation*". Implementation research is a form of research by definition done before actual clinical practice.

[4]It is well-known from simulation studies [13] that applying sample size calculations based on the Gaussian distribution to non-symmetric count distributions, leads to conservative results, inflating the required sample size by as much as 50%, depending on the effect size. To take but one example, replicating on the total PCL-5 symptom count at six months of the simulation by Cundill and Alexander [13], and using the AIC, a total sample size of $N = 86$ participants would have been sufficient to detect an average difference of one unit count (from 1.55 to 2.55, which we observe), in the presence of overdispersion (we observe a dispersion parameter of $k = 0.53$), with a power of 0.80.

REFERENCES

1. Halvorsen JO, Wessel I, Cristea IA Premature call for implementation of Tetris in clinical practice: a commentary on Deforges et al. (2023). Mol Psychiatry. 1–2 2024 https://doi.org/10.1038/s41380-024-02642-1.
2. Deforges C, Sandoz V, Noel Y, Avignon V, Desseauve D, Bourdin J, et al. Single-session visuospatial task procedure to prevent childbirth-related posttraumatic stress disorder: a multicentre double-blind randomised controlled trial. Mol Psychiatry. 2023;28:3842–50.
3. Horsch A, Vial Y, Favrod C, Harari MM, Blackwell SE, Watson P, et al. Reducing intrusive traumatic memories after emergency caesarean section: A proof-of-principle randomized controlled study. Behav Res Ther. 2017;94:36–47.
4. Blevins CA, Weathers FW, Davis MT, Witte TK, Domino JL. The posttraumatic stress disorder checklist for DSM-5 (PCL-5): Development and initial psychometric evaluation. Journal of Traumatic Stress. 2015;28:489–98.
5. Weathers FW, Bovin MJ, Lee DJ, Sloan DM, Schnurr PP, Kaloupek DG, et al. The clinician-administered PTSD scale for DSM-5 (CAPS-5): Development and initial psychometric evaluation in military veterans. Psychol Assess. 2018;30:383–95.
6. Sandoz V, Deforges C, Stuijfzand S, Epiney M, Vial Y, Sekarski N, et al. Improving mental health and physiological stress responses in mothers following traumatic childbirth and in their infants: study protocol for the Swiss TrAumatic biRth Trial (START). BMJ Open. 2019;9:e032469.
7. van der Linden WJ Handbook of Item Response Theory: Volume 3: Applications. New York, NY: Chapman and Hall/CRC; 2017.
8. Maydeu-Olivares A Evaluating the fit of IRT models. Handbook of item response theory modeling: Applications to typical performance assessment. Multivariate applications series. New York, NY: Routledge/Taylor & Francis Group; 2015. 111-27.
9. Deforges C, Stuijfzand S, Noël Y, Robertson M, Breines Simonsen T, Eberhard-Gran M, et al. The relationship between early administration of morphine or nitrous oxide gas and PTSD symptom development. J Affect Disorders. 2021;281:557–66.
10. Deforges C, Noel Y, Eberhard-Gran M, Garthus-Niegel S, Horsch A. Prenatal insomnia and childbirth-related PTSD symptoms: A prospective population-based cohort study. J Affect Disord. 2021;295:305–15.
11. Butcher NJ, Monsour A, Mew EJ, Chan AW, Moher D, Mayo-Wilson E, et al. Guidelines for reporting outcomes in trial reports: The CONSORT-Outcomes 2022 Extension. JAMA. 2022;328:2252–64.
12. Akaike H A New Look at the Statistical Model Identification. In: Parzen E, Tanabe K, Kitagawa G, editors. Selected Papers of Hirotugu Akaike. Springer Series in Statistics. New York, NY: Springer New York; 1974. 215-22.
13. Cundill B, Alexander ND. Sample size calculations for skewed distributions. BMC Med Res Methodol. 2015;15:28.

FUNDING

Open access funding provided by University of Lausanne.

COMPETING INTERESTS

Antje Horsch is a management committee member of COST Action CA22114. Emily Holmes is on the Board of Trustees of the MQ Foundation. She has developed an imagery-competing task intervention for intrusive memories, and training in using it (ANEMONE™). She receives book royalties from Guildford Press and Oxford University Press and receives occasional honoraria for conference keynotes and clinical workshops. She receives funding from The Wellcome Trust (223016/Z/21/Z), the Swedish Research Council (2020–00873), OAK Foundation (OCAY-18-442), and Rannís

The Icelandic Research Fund. Other authors declare no competing interests. The funder of the START study, the Swiss National Science Foundation, had no role in the study design, data collection, data analysis, data interpretation, report writing, or the decision to submit the paper.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Antje Horsch.