

# Commentary: The Signal and the Noise—questioning the benefits of puberty blockers for youth with gender dysphoria—a commentary on Rew et al. (2021)

Alison Clayton<sup>1</sup> , William J. Malone<sup>2</sup>, Patrick Clarke<sup>3</sup>, Julia Mason<sup>4</sup> & Roberto D'Angelo<sup>5</sup>

<sup>1</sup>University of Melbourne, Melbourne, Vic, Australia

<sup>2</sup>Department of Medicine, Idaho College of Osteopathic Medicine, Boise, ID, USA

<sup>3</sup>University of Adelaide, Adelaide, SA, Australia

<sup>4</sup>Calcagno Pediatrics, Gresham, OR, USA

<sup>5</sup>Institute of Contemporary Psychoanalysis, Los Angeles, CA, USA

In less than a decade, there has been a sharp rise in the numbers of young people presenting with gender dysphoria (GD). Today, the majority are adolescents, many with post-puberty adolescent-onset transgender histories, and suffering from mental health and neurodevelopmental comorbidities (De Vries, 2020; Zucker, 2019). Furthermore, there is controversy and heated debate in the literature on this topic (Dubicka, 2021). This lack of scientific consensus highlights the need for any published literature on the topic of GD to be carefully evaluated.

In this commentary, we critically examine a systematic review of the evidence for puberty blockers for GD youth that was recently published in this journal (Rew, Young, Monge, & Bogucka, 2021). Our aim is to highlight problems with this review that compromise its findings and conclusions.

## Brief description of Rew et al.'s (2021) study

Rew et al. described undertaking a “critical” and “systematic” literature review on the topic of puberty blockers for GD youth. They identified nine studies for review and, on the basis of these, concluded that puberty blockers have “few serious adverse outcomes,” and “several potential positive ones.” Rew et al.'s abstract highlighted two key conclusions: the “potentially life-saving benefits” of puberty blockers; and a need for rigorous research. Their “implications,” “conclusion,” and “key practitioner message” sections appeared to claim that the literature supports the use of puberty blockers for the early puberty subgroup of GD youth.

## Overview of our concerns

We agree with Rew et al.'s conclusion that more rigorous research is required in the area of management of GD in youth. However, in our view, their review suffers from methodological oversights, including the omission of relevant studies and suboptimal analysis of the quality of the included studies. As a result, the authors overstate the certainty of the potential positive outcomes and minimize the potential adverse outcomes of puberty blockers. Importantly, their statement, that a “positive

outcome” of puberty blockers is “decreased suicidality in adulthood,” is a misinterpretation of a single cross-sectional study. This study's design was incapable of determining causation, and adult suicidality was not one of the measured outcomes (Turban, King, Carswell, & Keuroghlian, 2020).

Contrast Rew et al.'s (2021) conclusions with another recently completed systematic review of puberty blockers for GD youth, commissioned by England's NHS and conducted by The National Institute for Health and Care Excellence (NICE) (2020). The NICE review concluded that studies investigating the benefits or adverse effects of GnRH analogs (puberty blockers) were of “very low certainty using modified GRADE.” They noted that any outcome differences that were found could have represented changes of “questionable clinical value,” or, as the studies themselves were “not reliable,” could have been “due to confounding, bias or chance.” They suggest that if controlled studies are not possible, then reliable comparative studies are required.

These findings came just after NHS England suspended the use of puberty blockers for new patients under the age of 16, following the High Court's judgment that children so young could not consent to the unknown risks of these drugs. The Karolinska Institute in Sweden suspended the use of puberty blockers as treatment for GD youth outside of clinical trials following this review, citing multiple physical risks, including to bone development (Nainggolan, 2021). Finland also sharply curtailed the use of these drugs after their systematic review arrived at similar conclusions about the uncertain risk/benefit profile (COHERE, 2020).

We are concerned that Rew et al.'s review will mislead clinicians unfamiliar with the literature into prescribing puberty blockers to GD youth with confidence, when the only clinical stance supported by the evidence is that of extreme caution. This is also underscored by the fact that the research literature in this field is rapidly evolving. For example, a recently published study, that attempted to demonstrate the benefits of the Dutch puberty suppression protocol in the UK setting, failed to show any psychological benefit (Carmichael et al., 2021).

### *Limitations in study selection strategy*

The review published by Rew et al. has important limitations that compromise its usefulness for clinical decision-making. Rew et al. identified only 151 potentially eligible studies, while the NICE review found 525 studies. One possible explanation for this could be their limited study search strategy. Another possible explanation is that Rew et al. did not conduct a comprehensive search so that, in omitting one of the largest electronic databases—EMBASE, they may have overlooked relevant evidence.

Notably, the final set of nine studies reviewed by Rew et al. is missing at least one key study on puberty blockers and psychosocial functioning (Costa et al., 2015), and two other studies examining the risks of puberty blockers on bone density (Joseph, Ting, & Butler, 2019; Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015). It is unclear to us whether these studies were omitted due to the limited database search or whether the evaluators decided to exclude these studies, and if so for what reason. These three studies were all included in the NICE (2020) review. Although it has to be kept in mind that all the NICE reviewed studies' findings were assessed as "very low certainty," the Costa et al. study provided comparative evidence and found no significant difference in psychosocial functioning between a group of adolescents receiving puberty blockers plus psychosocial support, and a group receiving only psychosocial support, at eighteen months (the study end period) (Biggs, 2019). In addition, the Costa study was cited by the Finnish gender identity services in their policy change, which now recommends psychotherapy alone as first-line treatment.

### *Failure to adequately assess certainty of the study findings*

It is our contention that the reviewers did not adequately assess the certainty of the reviewed studies' findings. For example, they used the Joanna Briggs Institute checklist to assess Turban et al. (2020), the study from which their message that puberty blockers reduce adult suicidality and have "potentially life-saving benefits" derives. This checklist can overemphasize whether studies report information and underemphasize the assessment of study validity. Below, we show how Rew et al. applied this tool to Turban et al. (2020), and the important study limitations it overlooked.

*Was the exposure measured in a valid and reliable way? (Q3)* Rew et al. answered "yes" to this question. We believe it should be "no." The exposure to puberty blockers was based on a self-report, with 73% of those respondents, who answered yes, claiming they began to use puberty blockers after the age of 18. It was noted that the respondents likely confused puberty blockers with other hormonal interventions (Biggs, 2020; D'Angelo et al., 2020). Although Turban et al. attempted to reduce the effects of this confusion by excluding certain participants from the sample, no adequate correction was possible. This introduced a significant risk of bias.

*Were confounding factors identified and strategies to deal with them stated? (Q5, Q6)* Rew et al. answered "yes" to both questions. We believe the answer to the latter question should be "no." For example, while one key confounding factor—prior mental health status—was indeed correctly identified by Turban et al., no strategy

was articulated to deal with it. When discussing their finding that puberty suppression is associated with lower lifetime suicidality, they acknowledged that "reverse causation cannot be ruled out: it is plausible that those without suicidal ideation had better mental health when seeking care and thus were more likely to be considered eligible for pubertal suppression" (Turban et al., 2020). This is one of the most serious limitations of the study, introducing a high risk of bias, and reducing the certainty of the findings.

In addition, while two questions ask about the subject selection criteria and whether the subjects and the setting were described in detail (Q1, Q2), these questions do not attempt to assess the impact of the sample composition. Affirmative ("yes") and "not applicable" answers to these questions, respectively, masked the fact that the study participants were not required to have a diagnosis of GD, and that the participant demographics were markedly different from the US population of transgender adults (D'Angelo et al., 2020), which negatively impacts the study's applicability/generalizability.

Rew et al. aggregated the answers to the checklist questions, with the Turban et al.'s study earning an 86% mark and a "good quality" rating. Even if we sideline the issue of any scoring inaccuracy, using such a simplistic scoring category is misleading since it implies that all questions are equally important, which is clearly not the case.

We also note, what appears to be, at least one error in Rew et al.'s assessment and reporting of study outcomes. In Table 2, they reported that Turban et al.'s positive outcome findings included decreased past-month psychological distress, past-month binge drinking, and lifetime illicit drug use. However, Turban et al.'s univariate analysis showed only one of these three outcomes, past-month psychological distress, showed any significant difference, and this significance disappeared once demographic variables were controlled for in the multivariable analysis.

A more rigorous tool to assess Turban et al.'s study would be ROBINS-I (The Risk of Bias of Non-randomized Studies of Interventions) (Sterne et al., 2016). This tool focuses on confounding, selection bias, classification and deviations from intervention, measurement of outcome, missing data, and selective reporting, and the extent to which the study design minimized biases and yielded trustworthy results. Given this, applying the ROBINS-I tool would find that the Turban et al.'s study is at a critical risk of bias.

### *Misleading statements regarding puberty blockers and suicidality*

We are concerned that Rew et al.'s discussion of evidence about suicidality is unbalanced and misleading. Reading that puberty blockers had "positive outcomes [of] decreased suicidality in adulthood" will likely be understood as indicating causation. However, Turban et al. (2020), where this claim originates, noted that their study design did not allow for determination of causation, and "reverse causation" (individuals without suicidal ideation had better mental health and were more likely to be considered eligible for puberty blockers) was a plausible alternative explanation.

Further, there is a critical difference in meaning between "lifetime," and "adulthood." Not only does the

latter erroneously imply a pre–post effect (i.e., access to puberty blockers in childhood reduces suicidality in adults), which was not detectable in the study, but a measure of “adulthood suicidality,” which Rew et al. claim was impacted, was never included in the original study (Turban et al., 2020).

There is also unclear use of the term suicidality, which exaggerates the implication of Turban et al.’s findings. Suicidality is a broad term, which is comprised of suicide attempts, plans, and ideation, and indeed this was the manner it was used by Turban et al. It is also important to note that Turban et al. made no assessment of completed suicides. Turban et al. assessed six areas of suicidality (including recent and lifetime suicide attempts, recent ideation with plans, recent and lifetime ideation) and found no association between puberty blockers and suicidality measures on five of the six areas. The only association was with “lifetime suicidal ideation.” Of course, any suicidal ideation is concerning, but suicide attempts are generally considered of higher concern, in terms of suicide risk assessment, than suicidal ideation (Ryan & Oquendo, 2020).

Rew et al.’s inaccurate language further intensifies in the final sentence of their abstract, which described puberty blockers as “potentially life-saving.” This exaggerated claim is misleading, since there is no evidence to support it.

#### *Absence of an appropriate process for making clinical recommendations*

Finally, the authors appear to recommend the use of puberty blockers in the “key practitioner messages” box and in the “implications” section of their paper. Making recommendations requires not only evidence about benefits and harms on all health outcomes that are important for decision-making (which this review provides in a suboptimal way), but also considerations about patients values and preferences, ethics, acceptability, resources, costs, etc. (Andrews et al., 2013). All these considerations are balanced by making value judgments, which should be documented and reported explicitly and transparently. Rew et al. failed to do this, which, in our view, further undermines the credibility of their clinical practice recommendations.

#### *Clinician reflections on the state of the GD literature*

Rew et al.’s review illustrates a concerning trend, that we have observed in the GD literature, to overstate the evidence underpinning clinical practice recommendations for youth with GD. New publications reference prior ones with increasing and unwarranted confidence, and with the risk of misleading clinicians regarding the state of evidence. There is also a marked asymmetry in outcomes reporting: findings of positive outcomes of medical interventions are trumpeted in abstracts, while their profound limitations remain behind the paywall, thus, below the radar of busy clinicians.

Rew et al.’s paper demonstrates these types of issues. To start, the Turban et al.’s paper described a noncausal association between puberty blockers and “lifetime suicidal ideation,” carefully avoiding making a causal claim (although, arguably, implying it). Then, Rew et al., whose findings on suicidality are based solely on this Turban et al.’ study, rewrite this finding to create the strong

impression of causality—that puberty blockers reduce adult suicidality and are “potentially life-saving.” Subsequently, a recent Commentary and Editorial in the *Lancet* both directly state that puberty blockers reduce suicidality, and the latter adds the extraordinary claim that “removing these treatments is to deny life.” The only reference provided for these claims is the Rew et al. (2021) paper (Baams, 2021; *Lancet* editorial, 2021).

This resembles the game of “Telephone,” in which a message is whispered from person to person distorting the original meaning of the message. However, this is not a game, and these types of errors can cause harm. Clinicians relying on Rew et al.’s review are likely to misinform patients and families about the risk/benefit profile of puberty blockers. Can such patients really be considered as giving informed consent?

The clear signals emerging from the various reviews of the available evidence of the use of puberty blockers for GD youth are that there is very low certainty of the benefits of puberty blockers, an unknown risk of harm and there is need for more rigorous research. The clinically prudent thing to do, if we aim to “first, do no harm,” is proceed with extreme caution, especially given the rapidly rising case numbers and novel GD presentations. We must also, collectively, raise the bar on the quality of publications, in order to accurately educate clinicians and help patients make truly informed decisions that may impact for the rest of their lives.

#### **Acknowledgements**

The study received no external funding. Open Access fees were provided by the Society for Evidence-Based Gender Medicine. We would also like to thank the Society for Evidence-based Gender Medicine (SEGM) for providing access to several experts who helped shape this commentary and ensure its accuracy. Specifically, we would like to thank Dr. Romina Brignardello Petersen for contributing her methodological expertise; Dr. Michael Biggs for reviewing the accuracy of the claims relating to puberty blockers and suicidality made in this review, as well as relating to the developments in the United Kingdom; and to Ema Syrulnik for her help with the preparation of this manuscript. The authors have declared that they have no competing or potential conflicts of interest.

#### **Ethical information**

No ethical approval was required for this commentary.

#### **Correspondence**

Alison Clayton, University of Melbourne, Melbourne, Vic, Australia; Email: [alclayton@student.unimelb.edu.au](mailto:alclayton@student.unimelb.edu.au)

#### **References**

- Andrews, J.C., Schünemann, H.J., Oxman, A.D., Pottie, K., Meerpohl, J.J., Coello, P.A., . . . & Guyatt, G. (2013). GRADE guidelines: 15. Going from evidence to recommendation—Determinants of a recommendation’s direction and strength. *Journal of Clinical Epidemiology*, *66*, 726–735.
- Baams, L. (2021). Equity in paediatric care for sexual and gender minority adolescents. *The Lancet Child & Adolescent Health*, *5*, 389–391.
- Biggs, M. (2019). A letter to the editor regarding the original article by Costa et al: Psychological support, puberty suppression, and psychosocial functioning in adolescents with gender dysphoria. *Journal of Sexual Medicine*, *16*, 2043.

- Biggs, M. (2020). Puberty blockers and suicidality in adolescents suffering from gender dysphoria. *Archives of Sexual Behavior*, 49, 2227–2229.
- Carmichael, P., Butler, G., Masic, U., Cole, T.J., De Stavola, B.L., Davidson, S., ... & Viner, R.M. (2021). Short-term outcomes of pubertal suppression in a selected cohort of 12 to 15 year old young people with persistent gender dysphoria in the UK. *PLoS One*, 16, e0243894.
- COHERE. (2020). Medical treatment methods for dysphoria associated with variations in gender identity in minors – recommendation. Palveluvalikoimaneuvosto. Available from: <https://palveluvalikoima.fi/en/recommendations> [last accessed 23 May 2021].
- Costa, R., Dunsford, M., Skagerberg, E., Holt, V., Carmichael, P., & Colizzi, M. (2015). Psychological support, puberty suppression, and psychosocial functioning in adolescents with gender dysphoria. *The Journal of Sexual Medicine*, 12, 2206–2214.
- D'Angelo, R., Syrulnik, E., Ayad, S., Marchiano, L., Kenny, D.T., & Clarke, P. (2020). One size does not fit all: In support of psychotherapy for gender dysphoria. *Archives of Sexual Behavior*, 50, 7–16.
- de Vries, A.L.C. (2020). Challenges in timing puberty suppression for gender-nonconforming adolescents. *Pediatrics*, 146, e2020010611.
- Dubicka, B. (2021). Editorial: Evidence, policy and practice – gold standard, good enough or doing it differently? *Child and Adolescent Mental Health*, 26, 1–2.
- Joseph, T., Ting, J., & Butler, G. (2019). The effect of GnRH analogue treatment on bone mineral density in young adolescents with gender dysphoria: Findings from a large national cohort. *Journal of Pediatric Endocrinology and Metabolism*, 32, 1077–1081.
- Klink, D., Caris, M., Heijboer, A., van Trotsenburg, M., & Rotteveel, J. (2015). Bone mass in young adulthood following gonadotropin-releasing hormone analog treatment and cross-sex hormone treatment in adolescents with gender dysphoria. *The Journal of Clinical Endocrinology & Metabolism*, 100, E270–E275.
- Nainggolan, L. (2021). Hormonal Tx of youth with gender dysphoria stops in Sweden. *Medscape*. Available from: <https://medscape.com>
- National Institute for Health and Care Excellence (NICE). (2020). Evidence Review: Gonadotrophin releasing hormone analogues for children and adolescents with gender dysphoria. Available from: <http://evidence.nhs.uk>
- Rew, L., Young, C., Monge, M., & Bogucka, R. (2021). Review: Puberty blockers for transgender and gender diverse youth – a critical review of the literature. *Child and Adolescent Mental Health*, 26, 3–14.
- Ryan, E.P., & Oquendo, M.A. (2020). Suicide risk assessment and prevention: Challenges and opportunities. *Focus*, 18, 88–99.
- Sterne, J.A.C., Hernán, M.A., Reeves, B.C., Savović, J., Berkman, N.D., Viswanathan, M., ... & Higgins, J.P.T. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*, 355, i4919.
- The Lancet Child Adolescent Health. (2021). A flawed agenda for trans youth. *The Lancet Child & Adolescent Health*, 5, 385.
- Turban, J.L., King, D., Carswell, J.M., & Keuroghlian, A.S. (2020). Pubertal suppression for transgender youth and risk of suicidal ideation. *Pediatrics*, 145, e20191725.
- Zucker, K.J. (2019). Adolescents with gender dysphoria: Reflections on some contemporary clinical and research issues. *Archives of Sexual Behavior*, 48, 1983–1992.

Accepted for publication: 10 September 2021